UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper A08/10

Applications and Policy

Tempo And The TFR

Máire Ní Bhrolcháin

Abstract

The paper proposes that whether period indicators are biased by timing effects depends on the objective of measurement. Several kinds of bias in the TFR are identified. Five reasons for measuring period fertility are distinguished: to explain fertility time trends, to anticipate future prospects, to describe trends, to provide input parameters for formal models, and to communicate with non-specialist audiences. Genuine timing effects are not biasing where period fertility is the explanandum, but are distorting where the aim is to estimate cohort fertility. Synthetic measures such as the TFR have a number of known defects. Alternatives to the TFR are available, and seem a more defensible solution to current problems than tempo adjustment. Tempo adjustment could be more fruitfully considered a form of modelling rather than empirical measurement. The measurement of period fertility could benefit by a more statistical approach and less reliance on indicators requiring stable assumptions.

Tempo and the TFR

Máire Ní Bhrolcháin

Division of Social Statistics
University of Southampton
Southampton SO17 1BJ
UK

August 2008
revised September and December 2008

Acknowledgments: I thank David Freedman, Mike Murphy, Laurent Toulemon, and Chris Wilson for helpful comments on earlier versions of this paper. Any errors are mine.

Abstract

The paper proposes that whether period indicators are biased by timing effects depends on the objective of measurement. Several kinds of bias in the TFR are identified. Five reasons for measuring period fertility are distinguished: to explain fertility time trends, to anticipate future prospects, to describe trends, to provide input parameters for formal models, and to communicate with non-specialist audiences. Genuine timing effects are not biasing where period fertility is the explanandum, but are distorting where the aim is to estimate cohort fertility. Synthetic measures such as the TFR have a number of known defects. Alternatives to the TFR are available, and seem a more defensible solution to current problems than tempo adjustment. Tempo adjustment could be more fruitfully considered a form of modelling rather than empirical measurement. The measurement of period fertility could benefit by a more statistical approach and less reliance on indicators requiring stable assumptions.

Keywords

Fertility; measurement; period fertility; total fertility rate; TFR; synthetic cohort measures; tempo adjustment; fertility timing; fertility quantum; projection

1. Introduction

Are conventional measures of fertility in period mode biased or distorted, and if so what should we do about it? The question was first raised by Whelpton (1946) in relation to the then leading period fertility indicators, gross and net reproduction rates (GRR and NRR). During the 1930s and 1940s, the lack of realism in the stable population assumptions underpinning these synthetic indicators had become particularly apparent. Scholars realised that the GRR and NRR could give a misleading account of levels and trends in fertility, particularly when the timing of fertility is changing (Whelpton 1946, Hajnal 1947). In the 1940s and 1950s, several solutions were proposed to overcome the deficiencies in period fertility indices. However, reproduction rates were eclipsed only briefly, having reappeared in the guise of the total fertility rate (TFR), a closely related index that has dominated the field ever since, but has all the faults that led to the demise of its predecessors as period measures.

The question of tempo effects in period fertility measures, while an ever present concern, has a renewed practical importance currently because of the steady move to later childbearing seen in most developed countries in recent decades. Bongaarts and Feeney (1998, 2000, 2006) have responded to the clear evidence of tempo effects in period measures of recent fertility with an innovative methodological approach. They suggest that tempo-related bias in period indicators of both fertility and mortality can and should be removed by means of an adjustment to the TFR that improves on the conventional index. This proposal has been the subject of much debate and has also been extended in various ways (Lesthaeghe and Willems 1999; van Imhoff and Keilman 2000; Kim and

Schoen 2000; Bongaarts and Feeney 2000, 2006; Zeng and Land 2001, 2002; van Imhoff 2001; Kohler and Philipov 2001; Kohler and Ortega 2002a,b; Smallwood 2002b; Schoen 2004; Sobotka 2004a; Barbi et al. 2008). Discussion thus far has centred mainly on the properties and performance of the Bongaarts-Feeney measure and associated indices. The present paper takes a somewhat different approach and examines the rationale for tempo adjustment *per se*.

The paper examines first what is meant by the terms “bias” and “distortion” in period measures, and identifies several distinct types of bias to which the TFR may be subject. On the principle that the purpose of analysis determines whether a tempo effect is biasing, the paper distinguishes five reasons for measuring period fertility (c.f. Ní Bhrolcháin 1992, 1994). It then examines whether tempo effects are a source of bias in measures deployed for each purpose, and, if so, whether adjustment is a good solution to the problem. The concluding section summarises the paper’s findings in tabular form, and discusses some strategic aspects of fertility measurement. Several conclusions emerge. First, tempo adjusted measures are inappropriate for some purposes, and where appropriate, of uncertain validity. Second, adjustment is predicated on the use of synthetic cohort indicators. This approach to period measurement is paradigmatic in demography, but its utility and role need fundamental reassessment. Third, alternatives to tempo adjustment are available and offer a more realistic and statistically defensible way of handling timing effects and the anticipation of the future. Finally, the tempo adjustment approach can be more fruitfully considered as a theoretical model of fertility change rather than as a method of empirical measurement.

2 Bias or distortion in period measures

The recent literature on tempo adjustment ascribes bias or distortion to the TFR whenever the timing of fertility is changing. Terminology varies, some sources using the concept of bias, either alone or interchangeably with the idea of distortion (Bongaarts and Feeney 1998; Bongaarts 2002; van Imhoff and Keilman 2000; Kohler and Philipov 2001; Zeng and Land 2001, 2002; Kohler and Ortega 2002b) and others referring exclusively to the notion of distortion (Bongaarts 1999; Bongaarts and Feeney 2000; Frejka and Ross 2001; Kohler and Ortega 2002a; Kohler et al. 2002; Schoen 2004; Sobotka 2004a). The two terms are considered to be synonymous in the present paper. That period fertility measures are influenced by timing effects is generally agreed. But are tempo effects biasing or distorting? The answer depends on clarifying what exactly we mean by 'bias' or 'distortion', and on the purpose of analysis. The present section is given to clarifying the nature of bias, and the sections that follow discuss measures for different analytical objectives.

In the context of period fertility measurement, what is meant by bias? Statistical bias is not the issue, since there is no question of a probability distribution for the TFR (see e.g. Zeng and Land 2002, note 1). Two meanings of bias or distortion in the TFR appear to be implicit in recent discussion: that the TFR is subject (a) to confounding or (b) to measurement bias, or to both of these. Confounding is at issue where the TFR is used to assess period trends in fertility, and measurement bias where it is seen as representing the level of fertility.

2.1 Bias as confounding

The TFR can be considered biased as a measure of period change. That is, ignoring level, the difference between the TFR in successive years/periods need not give an accurate account of the change in fertility from year to year. Being standardised only for age, the TFR can be influenced by year on year change in the distribution of women by parity. Change in fertility as measured by the TFR may therefore be confounded with change in the parity distribution, since birth rates vary by parity (Whelpton 1954, Henry 1953; Ryder 1986; Ní Bhrolcháin 1987; Feeney and Yu 1987; Rallu and Toulemon 1994; Andersson 2004; Sobotka 2004; Prskawetz et al. 2008; Toulemon et al. 2008). Confounding may also result from change in the distribution by age or duration within parity, since rates vary by these factors also. Such confounding, referred to as bias of type A here, is exacerbated by tempo shifts but occurs also in its absence.[1] Standard methods are available to remove the bias arising from compositional change. Measures specific for age and parity, or for duration and parity for orders two and above, remove the spurious, i.e. compositional, effect of timing change on period measures, and so deal with the bias due to confounding between time and the exposure distribution. Summary measures derived from such rates—period parity progression ratios, regression-standardised indices—are also free of a spurious timing influence. Such measures, however, retain the effect of real timing change—i.e. change along the age or duration axis in the schedule of rates—and are therefore suitable where it is appropriate that indicators reflect genuine

[1] Timing has this effect because advance or delay in births of a given order alters the composition by parity of the population at risk, and also influences the distribution by age or duration since previous birth of women of a given parity. The effect is particularly marked when the timing of first birth changes. However, a pure change in level can also impact on the parity distribution and hence result in potential confounding.

tempo change, as will be discussed in later sections.

2.2 Measurement bias

Measurement bias occurs where there is a systematic difference between a measure of something and its true value.[2] If the TFR is distorted in this sense, what real world phenomenon or theoretical construct does it measure in a biased way? The recent adjustment literature is less than clear on the subject. But, if mis-measurement is in question, it must be one of the three possibilities that follow, labelled for convenience B1 to B3.

B1. A first reading is that any measure of period fertility that is influenced by timing effects is considered to be distorted by definition, and that it is a construct "period fertility" that is wrongly measured by the TFR. Some scholars appear, implicitly, to espouse a view somewhere close to this, in seeing tempo effects as necessarily distorting to period measures (see e.g. Bongaarts and Feeney 1998, 2000; Kohler and Philippov 2001; Zeng and Land 2002; Bongaarts 2002). This is a potentially legitimate position, for which specific arguments could be envisaged. However, one can equally argue that genuine tempo effects are an essential component of the fertility of a period (Ní Bhrolcháin 1992, p.615; 1994). The case for each approach would have to be argued explicitly, in the context of any particular objective.

B2. A second way of understanding distortion in the TFR as measurement bias is to recall Ryder's use of the term. What he meant was that period parameters are "distorted reflections of cohort behaviour", and that the time series of period and cohort

---

[2] Note that throughout the paper it is assumed that fertility rates and indices are measured accurately in the sense that numerators and denominators are measured without error.

values do not coincide when fertility timing is changing (Ryder 1964, p.79; see also Ryder 1980). This seems what many commentators have in mind in asserting that period measures are biased or distorted, i.e. that period measures are erroneous indicators of real cohort values (see van Imhoff and Keilman 2000; van Imhoff 2001; Kohler and Ortega 2002a; Smallwood 2002b; Schoen 2004; Sobotka 2003). Thus, on this second interpretation tempo change introduces bias into the TFR considered as a measure of (real) cohort fertility. The TFR, and analogous synthetic indicators, are certainly biased in this sense. The conventional period TFR measures (real) cohort fertility accurately only if age-specific rates are either fixed or randomly distributed around a given period's values. Such stability is rare. In the fluctuating conditions obtaining in most empirical populations, the bias is exacerbated by timing change but is present even when timing is constant. Hence the period TFR is a poor indicator of associated cohort values for two reasons: because rates are not fixed and because fertility timing changes.

B3. A third way of construing distortion as measurement bias is entirely within a theoretical framework—that, in some theoretical population scenarios, period measures are biased indicators of cohort parameters. In most discussion of tempo adjustment, tempo change is treated as the exclusive cause of bias in period indices. But as we saw in the preceding paragraph, the TFR in empirical populations is a biased indicator of cohort total fertility whether or not timing is changing. It is only in theoretical populations that bias in the period TFR can be due exclusively to tempo change. Constant age specific rates give a constant TFR that is equivalent to the cohort TFR. In a theoretical scenario where the schedule of rates is fixed but moving along the age-axis—i.e. that timing is changing—the period TFR is a biased measure of cohort total fertility. A view of the TFR as biased only when tempo alters therefore makes sense only if it refers to the relationship between period and cohort TFR in theoretical rather than in empirical populations. Hence, the notion of bias informing the tempo adjustment debate appears to apply primarily to model populations.

2.3 Overview of bias

As a measure of period change, the period TFR is biased due to confounding between the exposure distribution and calendar time, a feature which is exacerbated by tempo change but can be resolved by the use of more specific rates/probabilities or indicators based on these. It suffers from measurement bias in three possible senses. Those for whom period timing is a nuisance factor will consider it definitionally biased as a measure of the level of period fertility. It is also biased as a real-world cohort estimator, timing effects contributing to but not being the sole cause of this type of bias. Finally, it is biased as a cohort estimator in theoretically specifiable scenarios; it is only in such model populations that tempo change is the exclusive source of measurement bias.

3. Objectives of fertility measurement

By and large, the merits and demerits of fertility measures are discussed in the population literature without reference to the ways in which period indicators are used. As a result, the attributes desirable in period measures tend to be thought of in an undifferentiated way. However, Hand (2004, p.267) notes that “(t)he use to which an index will be put will be the determining factor in its construction.” On that principle, the present section attempts to identify the leading purposes for measuring fertility, and examines the rationale for tempo adjustment in each case. The main reasons for measuring period fertility appear to be as follows: to explain time-trends in fertility, to anticipate future trends, to describe fertility, to provide input parameters for formal population models, and to convey information on fertility trends to non-specialist audiences.[3] Are timing effects biasing to measures of period fertility for all of these purposes? In the view presented here, it is essential to retain tempo effects in period fertility indicators used for some, though not all objectives.

3.1 Explaining fertility trends

We can set about explaining fertility time-trends in two ways. The strongest form of explanation would consist of a theory entailing a substantive model, i.e. a model that represents as far as possible the real-world processes giving rise to the fertility rates analysed (Freedman 1985; Cox 1990; Hand 2004). An alternative, less demanding form of explanation, and much more common in social science, is via an empirical or descriptive model. Such a model would aim to account for as much as possible of the variance in a dependent variable, but is not designed to represent the *modus operandi* of the underlying phenomena. The arguments that follow apply broadly to both types of approach, but some distinctions are drawn in discussion.

3.1.1 Are timing effects a source of bias? Of the two kinds of bias discussed in Section 2 above, it is confounding that is relevant to time trends in the TFR as dependent variable. Year on year movements in the TFR can be biased by compositional factors, due to confounding between the exposure distribution and calendar time. Tempo change

---

[3] These objectives may not be exhaustive. For example, period fertility indicators may also be used to evaluate interventions—e.g. in assessing the need for, and in monitoring, family planning programs. Other purposes might include the monitoring of fertility trends among specific sub-groups, such as migrants.

is a contributory factor, but not the sole cause of this bias. Is tempo adjustment an appropriate way to tackle the problem? The answer is no, on the present view, for two reasons.

The first reason why an adjusted TFR is not a good solution to the confounding problem is that total fertility measures of any kind are in principle unsuitable as dependent variables in explaining period trends. The TFR and analogous synthetic indicators are inappropriate in two ways:

(a) Period change is, by and large, multi- rather than uni-dimensional, and so cannot be represented by a single figure summary index. A summary indicator may, on the other hand, be adequate for analyses of the long run, intended to account for gross changes in level.

(b) Synthetic cohort indicators are in an inappropriate metric to represent the phenomena of a calendar period because they refer to the cumulative experience of many years (Ní Bhrolcháin 1992). As a result, they are unsuitable as dependent variables in any substantive model of the underlying process in its period aspect. Synthetic indicators can, on the other hand, be treated as simple statistical summaries, and in that guise they may be a suitable dependent variable for empirical as distinct from behavioural models. Nevertheless, point (a) above still applies, that no single figure summary measure can accurately represent fertility series over time when trends are heterogeneous across exposure categories.

A second reason for rejecting tempo adjusted measures as dependent variables is that genuine tempo effects are not a source of bias in measures of period fertility as dependent variable. This is because real timing shifts are part and parcel of period fertility change. Where period indicators are the dependent variable, they are estimates of a period phenomenon, not cohort estimators, and so are not subject to bias of types B2 and B3 of Section 2.2 above. As measures of change, they are potentially subject to confounding (bias of type A above), but provided this bias has been removed by selecting appropriate measures (specific for parity etc., as detailed in Section 2.1 above), the (genuine) tempo effects that remain are intrinsic to the dependent variable. True tempo effects consist of the shifting in personal time of births of each order rather than change in composition with respect to such factors. They are an integral part of what we have to explain. Our task as investigators is to account for the entirety of change in period fertility, rather than just a subset of it. Hence, the full effect of real timing change should be included in the dependent variable, rather than removed by adjustment.[4] In sum, removing the tempo component from period fertility as explanandum denudes it of an essential and often substantial component of change (Ní Bhrolcháin 1992, 1994). The greater the timing shift in a period, the worse the impact of adjustment on our measures, because a larger part of what is happening in a period, from an explanatory angle, is removed. These arguments are implicitly counter to the idea that period measures influenced by true timing effects are distorted by definition (bias of type B1 above).

An extreme example, the Year of the Fire Horse in Japan, helps to illustrate how the removal of timing effects would be misconceived when explaining period trends. In order to explain the fertility dip in Japan in 1966, it would be absurd to use as dependent

[4] The same applies to changes in variance discussed by Kohler and Philipov (2001). They too need to be accounted for in substantive terms rather than removed as a nuisance factor, when period fertility is a dependent variable.

variable a time series of tempo adjusted TFRs. Any timing element in that year's fertility is integral to the impact of folk belief in that year, as it is to the effect of other prominent period events.[5] Another example is the speed premium in Sweden: in that case, timing is not only a large part of the period effect, its precise detail allows a strong case to be made for a causal influence of maternity pay provision on fertility (Hoem 1990; Andersson 1999). The same applies to less pronounced period fluctuations. For example, accelerated childbearing was a sizeable ingredient of the baby boom of the late 1950s and 1960s (Butz and Ward 1979; Ryder 1980). If part of the explanation of the baby boom is that post-war prosperity, full employment, and high wages gave rise to accelerated marriage and childbearing relative to preceding periods, that faster pace of family formation must be represented on the left hand side of the equation. Similarly, the later childbearing prominent in developed societies in recent decades would have to be represented in the dependent variable in any attempt to account for current period trends.

An analogy may be useful. Consider a car travelling for a fixed duration of time. Its speed varies during the journey—rounding a sharp bend or going uphill, it slows down, while on the straight or downhill it travels faster. Speed may vary also depending on terrain, traffic, the driver's inclinations and so on. Saying that a well-standardized period fertility indicator is distorted is like saying that a measure of the car's speed at an arbitrarily chosen point in the journey, or when the car is changing speed, is mistaken. It may well give a biased estimate of average speed over the journey as a whole, but it gives an accurate account of the car's speed at the point at which this was measured. If we

[5] Of course, if we want to estimate the effect of the folk belief on cohort fertility, we would have to look explicitly at cohort outcomes.

think in terms of “underlying” speed or average speed during a journey, and whether and how it can be inferred from speed at a given stage along the way, we are measuring something other than speed at a particular time-point. We would, in addition, either have to construct models and make assumptions for the purpose, or investigate the properties of a large number of such journeys to generate an empirical basis for the estimate. The analogy is not perfect, but may be sufficient to highlight several key points. Our task in explaining period fertility trends is analogous to accounting for the speed of the car at successive points during its journey. Explaining episodes of acceleration and deceleration is comparable to explaining tempo effects on period fertility. These two together are, however, a different problem from either measuring or explaining average speed or distance travelled during the journey—a task analogous to estimating cohort or longer run fertility levels. Schoen (2004) has used the car analogy for a different purpose—to argue for the importance of cohort fertility—and assumes that the driver has an intended destination, though one that may alter during the journey. In the present case the analogy is between the car’s trajectory and aggregate fertility movements, and no assumption is needed about intentions regarding either destination, speed, or duration of the journey.

3.1.2 Period quantum and tempo The argument thus far is that timing effects are central to period fertility as dependent variable. However, that does not rule out the disaggregation of period fertility into tempo and quantum components, each to be the focus of explanation. Making estimates of period level and timing effects separately for explanatory purposes could potentially be argued for if several conditions were to hold. These are that in period mode:

1) the quantum of fertility and its timing are separable in a quantitative sense;

2) quantum and tempo measures reflect distinct aspects of the underlying behavioural process, where we are seeking a substantive/behavioural explanation rather than an empirical one;

3) that quantum and tempo respond differently to change in social, economic, and other determining factors.

Are tempo and quantum separable in a quantitative sense? Indices can be specified that, on their face, represent the quantum and tempo aspects of period fertility (e.g. Butz and Ward 1979; Ryder 1980; Foster 1990; Bongaarts and Feeney 1998; Kohler and Ortega 2002a).[6] However, quantum and tempo, though separable in theory, tend to covary in practice. The most familiar evidence of this is in the near universal tendency of cohort fertility series to reflect the fluctuations in corresponding period series, but with a lesser amplitude.[7] Time series presented by Schoen (2004) illustrate the point further. Schoen's Table 1 gives a number of quantum indicators for the US 1917-2001: the conventional TFR, the Bongaarts and Feeney adjusted TFR (TFR*, as originally specified and TFR** in a slightly revised version), the Butz and Ward Average Completed Fertility (ACF) measure,[8] together with the mean age at childbearing, and corresponding (true)

---

[6] In period parity progression life tables, the period parity progression ratios may also be considered measures of level and the synthetic birth intervals measures of timing (cf Ní Bhrolcháin 1987).

[7] This phenomenon has in recent years been referred to as a tempo-quantum interaction; see Kohler and Ortega (2002a,b), Kohler et al. (2002), Lutz and Skirbekk (2005), Goldstein et al. (2003) and Sobotka (2004a).

[8] Butz and Ward's ACF, Schoen's preferred measure of period quantum, is obtained as the period TFR inflated/deflated by an index representing the pace of childbearing in the period; the timing

cohort total fertility for part of the series. From these the period timing indices associated with each quantum measure have been obtained. Corresponding to the ACF we have the timing measure TFR/ACF (Butz and Ward 1979; Schoen 2004). Two specifications of a Bongaarts and Feeney timing index are calculated for each of the two versions of the B-F TFR: additive (BF timing =TFR-TFR*) and ratio versions (BF timing = TFR/TFR*). For each pair of indices, the timing and quantum indicators are positively associated—the correlation between ACF and its associated timing index is 0.83 (1917-97), and for the four versions of the Bongaarts and Feeney timing and quantum indices the correlations are between 0.32 and 0.45 (1918-1997). On the ACF evidence, periods of faster timing are also periods of higher levels of fertility. The same is true of the Bongaarts and Feeney indices, though the association is not as strong, possibly because the TFR_{adj} indicator can be erratic. Overall, these figures demonstrate that period quantum and tempo indices vary jointly (on this point, see also van Imhoff and Keilman 2000).

A second question is whether period quantum and tempo are behaviourally distinct. Estimates of tempo and quantum effects such as those of Butz and Ward (1979), Ryder (1980) and Foster (1990) are fascinating, but may be no more than mathematical abstractions with little or no representational content. While informative and suggestive they can however be considered exploratory, intermediate quantities, rather than definitive accounts of the real-world processes depicted. Are decisions about timing and quantity at the individual level made independently rather than as an integral whole? The question matters if we want to understand how and why fertility changes, and would be

---

index measures the extent to which an above or below average proportion of cohorts' overall fertility occurs in the period in question.

central to constructing a substantive model. Behavioural issues of this kind have been overlooked in recent debate on measurement, but Ryder (1980) has an excellent discussion of the subject. This classic paper is best known for having made estimates of the quantum and tempo of cohort fertility, and analysed them into their components. Ryder concludes, nevertheless, with some highly sceptical comments about how distinct these are, suggesting that quantum and tempo “are to some degree manifestations of the same underlying behaviour” (ibid., p.44). He is of the opinion that “we cannot, in principle, make a statistical separation of the tempo and quantum facets of fertility” and that fertility time series data, in themselves, will not allow appropriate measurements to be specified “in the absence of behavioural surveys designed to explore the structure of intentions and the use of means to fulfil those intentions” (ibid., p.45). Thus, this pioneering authority on estimating fertility quantum and tempo had serious doubts as to whether these were distinct aspects of real world decisions. In all, establishing how far quantum and tempo are separate aspects of the behavioural processes underlying period fertility movements requires detailed annual data of both immediate and longer-term fertility intentions and on the decision-making process, in addition to standard fertility series.

On the third point considered above, if period quantum and tempo are influenced either by different factors, or differentially by the same factors, then they may reflect genuinely distinct processes; otherwise, they are a single, undifferentiated entity. Instances can be found of changes in timing in reaction to socio-economic determinants—the Swedish speed premium effect being a very clear-cut case (Hoem 1990; Andersson 1999; Andersson et al. 2006). But it is not obvious that such instances are exclusively due to timing effects, nor that currently available indices of timing would represent them accurately.

3.1.3 Alternatives to tempo adjustment Alternatives to tempo adjustment are available for explanatory purposes. The natural way to address confounding bias is to use parity specific rates, specific also for age and/or duration. In the present view, the best choice in contemporary developed societies is a combination of age-parity specific rates/probabilities for women of parity 0, and parity and duration specific rates/probabilities for women of parity 1 and above, as in Murphy and Berrington (1993). The reason for the latter choice is that age-parity specific indicators for births of order 2+ are influenced by the timing in previous years of births of earlier orders, whereas parity-duration specificity starts the clock anew at each birth. Such measures fulfil two important requirements: they reflect both the heterogeneity in time trends by parity and genuine tempo effects, i.e. temporal shifts in the schedule of rates along the age or duration axis. Period parity progression ratios and synthetic timing indicators can be constructed from such rates, and may also be suitable as dependent variables. If a single figure summary is required, period parity progression ratios can be used to generate a total fertility figure that is free of confounding (bias A of Section 2.1) but will still include some timing effects. And while synthetic measures are, on the present view, inauthentic to period phenomena, such indicators may be usefully thought of as standardised summaries, rather than as hypothetical cohort indicators.

3.1.4 Validation Ultimately, measures to be used for any purpose need formal evaluation. No independent criterion is available against which to validate period measures in explanatory studies. But we do have an indirect check on the performance of a period measure as dependent variable: viz. explanatory success. Indicators of period fertility as explanandum can be considered valid to the extent that they are embedded in an empirically successful explanation of period trends—a form of construct validity. As Ryder has suggested, we will know we have the right measures when we have a good explanation of time trends. Thus far, we lack convincing, well-documented explanations for period time trends that could adjudicate between different measurement approaches. Nevertheless, success in explanation, rather than a check against cohort values, is the appropriate criterion by which to evaluate indicators of period fertility as dependent variable.

3.2 Anticipating the future

A second major objective for measuring fertility in period mode is to anticipate future population parameters. This is often the explicit or implicit rationale for studying recent trends, and entirely natural in a discipline whose most sought-after applied function is to inform about the future. Period based fertility measures are used in several ways to get a handle on population prospects. One is to estimate the fertility of cohorts. A second is to look for indications of future fertility trends in a less specific sense. A third is the use of period fertility measures to encapsulate population growth prospects in simple form. Each of these roles for period indicators is considered here in turn.

3.2.1 Cohort fertility Where data are available on completed childbearing, measuring cohort fertility is straightforward, subject only to the limits of data quality and sample size.[9] However, where cohort childbearing is incomplete, period rates are used in the estimation process and that presents some challenges.[10] Synthetic cohort indicators are the principal means by which period fertility is converted into cohort terms. Demographic translation is another approach but will not be discussed here (Ryder 1964; Keilman 1994).

The TFR re-expresses the age-specific rates of a single period as a mean family size, a quantity that can, in reality, only be arrived at over a lengthy time-span. The price paid for this transformation is the assumption that the age and/or duration-specific rates of a given period obtain at successive ages/durations, and so are fixed through time. The period TFR has long been known to be an unreliable indicator of the mean family size of associated cohorts.[11] The discrepancy is often illustrated graphically by the much larger

---

[9] Practical problems may arise regarding e.g. whether immigrants should be included, but the indices themselves are uncomplicated.

[10] In a paper that has influenced the present one, van Imhoff (2001: 24-5) expresses the matter thus: "A particularly important struggle faced by demographic analysts is how to arrive at statements about family formation processes from a cohort perspective…from data that are collected on an annual basis…".

[11] It can be said that the distinction between period and cohort is so well known in demography that the two types of index are thought of as quite distinct. Nevertheless the TFR and analogous synthetic indices are often treated effectively as cohort estimators. Evidence of this habitual conceptual double-think, paradigmatic in the subject, is found in the long-standing criticism that period indices constructed on an additive basis can occasionally produce results that are impossible in a real cohort (see, for example, Park 1976, Ryder 1990, Bongaarts and Feeney 1998, Bongaarts 2002). It has been known since the 1940s that variation in age-specific rates

swings in period than in cohort total fertility (for recent examples see van Imhoff and Keilman 2000; Frejka and Calot 2001; Smallwood 2002b; Schoen 2004; see also Sobotka 2003). Considered as a cohort estimator, the TFR suffers from measurement bias, as we saw in Section 2.2 above, and timing change contributes to the bias. There is no objection in principle to attempting to remove this bias by adjusting for tempo, unlike the case where period measures are explananda. If timing effects could be identified reliably and removed successfully, period indicators could certainly be improved on to approximate corresponding cohort quantities. Hence, it is perfectly reasonable to seek a means of doing this, if the objective is to get closer to cohort values. Does tempo adjustment achieve this in practice?

The evidence thus far is mixed. There are scattered data—proportions having a first birth in the US or proportions marrying in France—suggesting that adjusted period values can be closer to cohort values than are the period equivalent, though limited evidence is available thus far (Sobotka 2003; Bongaarts and Feeney 2006). But time series of annual adjusted TFRs do somewhat less well, not showing a reliable improvement over the conventional TFR in approximating to the total fertility of associated cohorts (van Imhoff and Keilman 2000; Smallwood 2002b; Sobotka 2003, 2004a; Schoen 2004). Empirical support at present for the use of tempo adjustment as a method of estimating cohort fertility is, therefore, equivocal. It could not, in any case, be a complete solution since it would do nothing about the bias resulting from the non-

---

though time is the cause of these apparent anomalies, and that these are exacerbated by timing change. If the period TFR were regarded as a pragmatically useful statistical summary rather than as a cohort estimator, no anomaly would be perceived.

constancy of age specific rates (see Section 2.2 above). For many demographic scholars getting better approximations to cohort values is the inherent demographic logic of tempo adjustment (van Imhoff and Keilman 2000; van Imhoff 2001; Kohler and Ortega 2002a; Smallwood 2002b; Schoen 2004; Sobotka 2003). Note, however, that this was not the original objective of tempo adjustment though in subsequent work Bongaarts and Feeney (2006, p.145) come closer to this perspective, regarding the adjusted period measures as "approximate measures of lagged cohorts" where "patterns of change … are close… to the translation assumptions."[12]

3.2.2 Future fertility Beyond attempts to estimate cohort levels, period fertility may be the vehicle for more general discussion of fertility prospects. A broad reference to the future is found in the discourse of fertility measurement in a variety of ways. It appears to be what e.g. van Imhoff (2001, p.24) has in mind when he says that by "level of fertility" we mean something like "how many children do people have, on average." It also seems essentially what is meant by widely-used references to "true" or "underlying" fertility, or the completed fertility "implied by" current rates (at least one reading—another interpretation will be considered in Section 3.4 below).

One interpretation of the Bongaarts and Feeney adjusted TFR, and related measures, is that they serve this function—i.e. are an attempt to infer longer run fertility in some non-specific sense, though not cohort fertility. This is one of the ways in which tempo adjustment has been understood in practice by demographic analysts—i.e. as carrying implications for long-run trends in fertility (Lesthaeghe and Willems 1999;

[12] See also Winkler Dworak and Engelhardt (2004) for comparison of adjusted period marriage rates against cohort equivalents.

Morgan and King 2001; Frejka and Ross 2001; Kohler and Ortega 2002a; Bongaarts 2002). The upward correction to recent period TFRs when adjusted for tempo has been interpreted as implying that period fertility rates are temporarily low and hence as implicitly predicting a future recovery in fertility. However, evidence supporting the interpretation of TFR_{adj} as predicting longer run fertility is as yet rather weak, as in the cohort case, and doubts have been expressed as to the likelihood or extent of such recuperation (Lesthaeghe and Willems 1999; Frejka and Calot 2001; Frejka and Ross 2001; Sobotka 2004a).

3.2.3 Future population growth prospects A further role for period fertility measures is to gain some idea of future population growth prospects, as reflected in regarding a TFR of 2.1 as constituting replacement fertility. The idea of a replacement level TFR assumes stable conditions just like its predecessor, the NRR. It has been known since the 1940s that a single year's TFR is no indication of future growth prospects. And despite the severe criticisms to which the NRR and other reproduction rates have been subject (Whelpton 1946, Hajnal 1947, Stolnitz and Ryder 1949, Dorn 1950, Hajnal 1959), we continue to use the TFR as a kind of reproduction rate—testimony perhaps to the recurring need to have a quick and easy way of conveying something about future growth prospects.

The concept of replacement level fertility is, in an empirical sense, known to be invalid in that the TFR can be below replacement for decades without resulting in population decline (Smallwood and Chamberlain 2005). The reason is, of course, that the long-run, stable assumptions are almost never valid. Timing effects are certainly a source of bias in the TFR treated as reproduction rate, since they reflect an additional departure from stability. But to use a tempo adjusted TFR in place of the conventional TFR in this context would be to take the notion of replacement level fertility more seriously than is warranted for practical purposes.

3.2.4 Alternatives to tempo adjustment Where attempting to estimate cohort or other future outcomes, there is no argument in principle against the removal of timing effects from period measures, unlike period fertility as explanandum. The only criterion relevant is how well an adjusted measure performs for the purpose. We have seen that tempo adjusted period measures have had some but limited success in approximating cohort quantities better than do unadjusted versions, and that the record in predicting future fertility movements is patchy also. As in the explanatory case, several alternatives are available.

Instead of relying on a period synthetic indicator to assess future prospects, we should undertake detailed analysis of trends in series of disaggregated fertility indicators to assess the exact nature of the trends (c.f. Hajnal 1947, Dorn 1950). This is the approach taken by several recent critics of tempo adjustment (Lesthaeghe and Willems 1999; van Imhoff 2001; Sobotka 2004b). This approach recognises the force of the admonition that “(t)here can be no mechanical formula which can be applied year in, year out” (Hajnal 1947, p.162) to give an accurate idea of long run population prospects. Hajnal’s argument concerned the NRR as a period indicator. But it applies to the period TFR in all its forms—additive, multiplicative, adjusted or unadjusted for tempo—since it is based in the inapplicability of stable assumptions as a basis for anticipating the future, as well as on the heterogeneity of disaggregated fertility series.

The second alternative is an extension of the first: that we treat the anticipation of future fertility explicitly as a forecasting problem, rather than as an inference to be made from current period rates. Projection or forecasting has several merits as a guide to either cohort fertility or a generalized notion of future fertility levels. The estimates produced are presented as projections rather than as measures, the inherent uncertainty of the estimates is acknowledged, and assumptions about future movements in rates are made explicit in the process. In addition, a forecasting approach is much less constrained than tempo adjustment. Tempo adjustment treats every individual period in isolation, with each point estimate of future fertility derived from a single period's rates together with the timing change since the previous period. A forecast, by contrast, can be based on an examination of the fine detail of period change, rather than on a routinized transformation of period rates, one year at a time. Forecasts of cohort fertility can, importantly, use the cumulated fertility of cohorts to the base period, potentially valuable information. No assumptions regarding stable conditions, or constant shape in fertility schedules, or long run change in timing are required, but such assumptions can be incorporated where appropriate, and adapted to apply for variable and selected spans of time. In all, if we want to get an idea of the ultimate mean family size of incomplete cohorts, or of future levels of period fertility, an explicit forecast is a more transparent, versatile, and potentially more powerful method of achieving this (cf. Lesthaeghe and Willems 1999, van Imhoff 2001, Kohler and Ortega 2002, and Schoen 2004).

3.3 Describing fertility trends

Demographers do a great deal of descriptive work, particularly directed to delineating "levels and trends". When the period TFR is used to measure trends in a descriptive context, it is certainly biased by confounding (Section 2.1 above), and so may give a mistaken account of both direction and degree of change in fertility. Specificity for parity and age/duration can deal with this, as discussed above. Whether the period TFR is biased as to level depends on what the investigator means by "level of fertility"—oddly, the concept in period mode has no generally agreed meaning. One way of construing it is simply that a rough and ready index of annual childbearing propensity is required, net of age structure. In this role, the TFR is not a synthetic cohort measure but a statistical summary of age specific rates. It could e.g. be replaced by TFR/35, representing the average age specific rate during the year, without any change in meaning. In either form, the principal deficiency of the TFR as summary indicator is that fertility change is multi-dimensional rather than uni-dimensional, and that a single figure indicator cannot represent this adequately. Where, on the other hand, an investigator thinks of "level" as a cohort mean family size, the TFR suffers from measurement bias and, as argued above, the utility of a tempo adjusted measure will depend on its performance in prediction.

In a descriptive context, indicators will often be chosen on conventional grounds or constrained by data availability. But we can assess descriptive measures only in a secondary way, relative to the uses to which the descriptive account is put. The underlying aim of a descriptive account may not be obvious, but any inferences drawn from it may be a guide. Often the subtext is a reflection on what is driving the trends or on future implications, and the arguments relating to choice of measures set out in Sections 3.1 and 3.2 above will apply.

3.4 Formal models

A further way in which indices of period fertility are deployed is theoretical. This is not measurement at all, but a form of population modelling. Wachter's (2005, p.202) comment that “(a)ny measure measures something” prompts an interesting question: *is* the TFR indeed a measure or is it, rather, a theoretical parameter? If we think of the TFR simply as a statistical summary of period rates, it is a workaday measure, with no statistical or demographic pretensions. However, the TFR as classically conceived—the mean family size of a hypothetical cohort subject to the rates of a particular period—is a theoretical parameter, not an empirical measure. A synthetic cohort is not an empirical entity, and its hypothetical mean family size is not a measure of anything in the real world.[13] The same is true of synthetic indicators in general, if interpreted as such. They are not at all comparable to apparently analogous measurements such as speed. The miles per hour expression, though stated in hypothetical terms (distance that would be travelled in an hour at a given speed), also allows us to state distance travelled during a very short interval at the point when the measurement is taken. That is a measure of something real. Nothing similar about real world births or birth rates can be derived from the period TFR, in its classic interpretation.

Such reasoning leads naturally to interpreting the TFR_{adj} and allied indicators as theoretical rather than empirical concepts. This is the approach of Zeng and Land (2001) and of Rodriguez (2006), who interpret the Bongaarts and Feeney TFR_{adj} as the cohort mean family size in a theoretical population with the age-order specific rates of a given period and subject continuously to the tempo change of that period. It also appears to be the logic implicit in the tempo adjustment literature. One reason for making the distinction is that scholars in this area (Bongaarts and Feeney 1998, 2006, Zeng and Land

[13] The constant use of these ideas has probably resulted in a process of reification (Wilson and Oeppen 2003): we come to believe in the existence of real counterparts to our routine measures.

2001, Kohler and Philippov 2001, Kohler and Ortega 2002a) view the presence of distortion or bias in the TFR as arising exclusively from tempo change. Such thinking relates to theoretical rather than empirical scenarios, since it is only in theoretical populations that bias in the TFR as a measure of cohort fertility is confined to periods when tempo is changing. A second reason for considering TFR_{adj} a theoretical rather than an empirical measure is in the rationale offered for it—that it provides “a better indication of the level of completed fertility implied by current fertility behaviour” (Bongaarts and Feeney (1998, p.285). But the mean family size “implied by” the rates of a given period, though a paradigmatic concept in demography, is an imaginary mathematical construct rather than an empirical quantity. Nothing in the real world is literally implied by the rates of a single period, since empirical populations are not stable, for the most part. Finally, Bongaarts and Feeney’s tempo adjustment ideas could potentially have a more fruitful role if regarded as a theoretical model of fertility change, with potential applications when the requisite assumptions are met. It is one among a large variety of potential models of aggregate change in fertility that can be envisaged, and could be seen as a stimulus to the development of such models.[14]

In a theoretical context, procedures to adjust for tempo change can be designed around the particular type of tempo shift assumed to operate. They can be applied without further comment if the objective is to estimate the cohort fertility of a theoretical population of the specified type. However, in order to argue for the use of a theoretically

[14] The tempo adjustment idea in the mortality arena has come to be regarded in this way, with differing specifications of period life expectancy indicators corresponding to different models of how the processes underlying mortality change operate in the real world (see Barbi et al. 2008).

specified adjustment procedure as an empirical measure, it would have to be shown first that the model conditions under which the adjusted indicator has a particular meaning hold empirically, either for each application, or widely enough for the procedure to be recommended for general purposes.

3.5 Communication and public information

A final reason for choosing an index of period fertility is to convey information on fertility trends to non-specialist audiences of various kinds. Such communication needs may place an unnoticed constraint on how academic demographers think about measurement issues, in view of the strong ties that academic demography has with official statistical agencies (Brass 1990).

Is a tempo adjusted version of the TFR useful in this more popular context? The view taken here is in the negative, regardless of the purpose of reporting on trends. Where tempo change is a significant aspect of trends, the classic TFR can be accompanied by information on timing, particularly on changing age at first birth. An unrefined index such as the TFR can satisfy most non-technical users' needs, if supplemented by a variety of other data relevant to future prospects, at least in a policy context. One argument for adjustment is that unadjusted period synthetic indicators mislead the public about their likely lifetime experience. But that assumes that period synthetic measures are presented to popular audiences as an implicit forecast, which is unwarranted, whether indicators are adjusted or not. If statistical offices wish to convey future likely experience, an explicit forecast is the appropriate vehicle, not a period synthetic index.

In commenting on the widespread adoption of the TFR, William Brass was of the view that the preference for single figure summary indicators arose from attempts to simplify communication. He thought this unfortunate since it had resulted in a neglect of approaches using multiple indicators that are potentially technically superior (Brass 1990, p.455). The TFR has the attraction of intelligibility to non-specialists, and it is natural that the indicators used by statistical agencies are chosen for ease of interpretation by users as well as ease of production.[15] But interpretative simplicity is irrelevant when choosing measures for technical purposes. Physicists and astronomers do not justify their measurement procedures by whether lay people can understand them: why should we? Besides, a projected long-term completed family size is just as user friendly as a classic or tempo-adjusted version. And users can also be educated to expect more refined and varied indicators, such as age-specific and parity specific rates. Finally, there is no need for annual press releases on the latest fertility trends, since one year's fertility never has long term significance. Publicity can be given rather to the more considered reflection on the long run that accompanies population projections (c.f. Hajnal ,1959).

4. Discussion

The paper has addressed the question of tempo-adjustment in principle by clarifying first how, in a statistical sense, the terms bias or distortion have been used in recent debate. Two distinct types of bias are implicit in the tempo adjustment literature: confounding and measurement bias. The paper then identifies the main purposes for which we measure period fertility and asks, in each case, whether timing effects are a source of bias, and, if so, of what kind, and whether adjustment is a sensible way to deal with it. The

[15] All the same, such transparency is unnecessary. Numerous arbitrary measurement scales are in use, and communicated meaningfully to popular audiences in the media, from temperature itself to Richter scale earthquakes.

conclusions reached are summarised in Table 1.

Genuine timing effects are not biasing or distorting, on the present argument, where period fertility is dependent variable, and so adjustment is inappropriate for this purpose. In non-specialist communication, tempo effects can be handled well by existing methods; in a descriptive context, adjustment will often be irrelevant. By contrast, genuine tempo effects *are* a source of bias when period synthetic measures are being used as a proxy for the cohort equivalents, or to predict longer term fertility, or in specific theoretical scenarios. Tempo adjustment is therefore a candidate technique in empirical situations when using a single year's rates to estimate completed cohort fertility or attempting to predict future fertility. Does it work well in that context? Thus far the evidence is mixed. Tempo adjustment has been evaluated by most commentators primarily against cohort estimates of e.g. proportions having a first birth or cohort total fertility. It has not yet been shown to improve reliably on period measures as estimators or predictors, though there are instances in which it appears to do well (Lesthaeghe and Willems 1999; van Imhoff and Keilman 2000; Smallwood 2002b; Sobotka 2003; Winkler-Dworak and Engelhardt 2004; Schoen 2004; Bongaarts and Feeney 2006, Figures 8-13). Alternatives discussed include the use of disaggregated measures, particularly parity specific ones, to replace the reliance on single figure summary indicators, with close analysis of recent trends of disaggregated indicators a prerequisite both for explanatory purposes and for assessing future prospects. We need in particular to avoid the conflation of measurement and forecasting to which synthetic cohort indicators give rise. Estimates of the completed fertility of incomplete cohorts, and of future likely period fertility, should be based routinely and explicitly on forecasts (cf. Lesthaeghe and

Willems 1999; Kohler and Ortega 2002a; Schoen 2004).

For Bongaarts and Feeney (1998, 2000) the TFR_{adj} measure is designed to give a better reading of current fertility rather than to estimate cohort or future fertility. What they have is mind here is the “level of completed fertility implied by current fertility behaviour” and an “answer to the question of how many births women will have if current childbearing behaviour continues into the future” (Bongaarts and Feeney 1998, p.285). This conceptualisation is conventional in the field but needs to be re-examined. A first difficulty is that there is no explicit specification how the TFR_{adj} is intended to be used, whether explanatory, predictive, descriptive, theoretical or communicative roles are envisaged. This is unsurprising since demographic measures are often proposed without specifying the objective they are to serve. As a result, it is not clear how TFR_{adj} should be evaluated. If we do not know the precise purpose of analysis, we cannot decide whether one measure of period fertility gives a better reading than another. A second difficulty is that, as discussed in Section 3.4, the “family size implied by current rates” concept refers to a hypothetical rather than an empirical entity. Hence, it does not quantify any real attribute of the calendar year in question. There do not appear to be grounds for preferring one hypothetical calculation to another as a means of quantifying empirically the fertility of a calendar period.

It is useful to consider the historical context of hypothetical cohort indicators. The defects of the GRR and NRR as measures of period fertility conditions were clearly recognised in the 1940s and 1950s. Reproduction rates had been routinely used as indices of time trends up to then, but rapid shifts in fertility in the 1930s and 1940s brought the realisation that the stable assumptions underwriting their quantitative relevance did not hold in empirical populations. In addition, fertility series were seen to be heterogeneous, dependent on parity and personal time (age or duration). A number of alternatives were suggested to replace them—parity and age- or duration-specific measures, cohort analysis and period parity progression ratios (Whelpton 1946, 1949, 1954; Hajnal 1947, 1959; Stolnitz and Ryder 1949; Henry 1953). Had these methods been adopted as standard, the TFR, a reproduction rate in all but name, would not have its current prominence. But the TFR has been the leading fertility indicator since the 1960s., its widespread use being in Brass's view attributable to "[s]implicity, convenience and propaganda" (Brass 1990, p.456).[16] Its adoption went counter to the lessons learnt from the GRR and NRR, though these were not forgotten entirely.

The parity specific measures devised in response to past concerns about period synthetic summary indicators remain the best current alternative to the biases present in the TFR as a measure of period change. They also have a potential role in projecting cohort outcomes and future fertility. These methods have undergone further development more recently (Ní Bhrolcháin 1987, Feeney and Yu 1987, Rallu and Toulemon 1994). And while period parity progression ratios are, on the present argument, problematic as

[16] Brass is worth quoting at length on this subject. The attractions of the TFR were, he thought, due to the misconceived desire for a single figure summary index. "If the demand for a simple index is relaxed there is no great problem in providing an array of measures which in combination show the characteristics and dynamics of a population's fertility. However this leaves the burden of the interpretation to the user. The search for the single index is [a part] of the process of simple presentation to non-demographers of the evidence on what is happening to fertility, and consequently, on what might be its path in the future" (Brass 1990, p.455).

synthetic cohort indicators, they may be useful as standardised summaries of parity specific rates. Tempo adjusted versions of period parity progression ratios have been proposed (Kohler and Ortega 2002a) and are in some respects preferable to a tempo adjusted TFR. Nevertheless, reservations still apply to such measures if in the role of dependent variable. Regression-standardised indicators devised by Hoem (1993) have been employed to report fertility trends in Sweden and elsewhere and are possibly the most under-used recent development in methods of fertility analysis (for applications see e.g. Andersson 1999, 2004).

These alternatives require more data and analytical resources than the computation of a tempo adjusted TFR. The need for more complex data has doubtless contributed to the neglect of parity specific indicators. But current data resources can be better exploited. Available data can be deployed to generate parity specific measures: for example, Luther et al. (1990) and Murphy and Berrington (1993) have shown that reasonable estimates of period parity progression ratios and, of course, of their constituent rates, may be obtained by the own children method from household surveys. Where either true birth order data or information on the distribution of women by parity are not available, age-parity specific rates can be reconstructed (Smallwood 2002a; Chamberlain and Smallwood 2004). Such compilations provide not only greater resolution in the depiction of current trends, but also facilitate alternative and potentially promising approaches to forecasting (see Feeney 1985, Toulemon and Mazuy 2001, Kohler and Ortega 2002a, Sobotka 2004b). Indirect methods may also help to expand our resources in this area (cf. Henry 1953, Feeney 1991). Finally, the parity specific regression standardised indicators devised by Hoem (1993) merit a great deal more attention than they have received thus far, particularly but not solely in an explanatory context.

We need to seek a more empirical and statistically defensible approach to demographic measurement than methods that assume a stable population scenario. The record of investigations conducted along these lines is a decidedly positive one. A case in point is the demonstration of the speed-premium effect in Sweden, well documented by a set of regression standardised indicators, specific by parity and duration (Hoem 1993; Andersson 1999), owing nothing at all to stable assumptions. Tempo-quantum decomposition is also illustrative. None of the three most informative analyses of this kind—those of Butz and Ward (1979), Ryder (1980) and Foster (1990)—required stable population assumptions. Nor did Lee's (1980) moving target theory. These are, rather, demographically aware mathematical and statistical analyses, and illustrate well the potential pay-off to a statistically less constrained, more open-minded, and sophisticated approach to demographic measurement and analysis.

The tempo issue has further implications of scientific and practical importance. The essence of the problem is that we cannot distinguish short- or medium-term period timing shifts from longer-term quantum changes. When we see a rise or decline in period fertility accompanied by a shift to younger or older ages at birth, the state of the art does not allow us to say whether it is merely a timing effect—and so will be fully compensated by altered rates at older ages—or a mixture of timing and level, or the start of a long-run change in the level of fertility. A key priority for fertility research is, therefore, to develop empirical methods of telling these processes apart, contemporaneously if possible, and retrospectively if not. Analysis of the rates alone is unlikely to solve the problem.

Identifying these components successfully almost certainly requires detailed knowledge of the factors driving fertility trends in any particular case. This was how a scale for temperature was finally developed successfully. The study of underlying processes—i.e. the behaviour of various substances across a range of temperatures—was essential to establishing a sound measuring instrument (Chang 2004). Similarly, technical and substantive analysis of fertility are unlikely to progress fruitfully in isolation from each other. Work on measurement needs to be integrated with investigation of the causes of fertility change at the aggregate level.

Table 1. Summary of conclusions on tempo adjustment of period fertility measures, together with alternatives.

Purpose	***Are genuine timing effects biasing?***	***Is there a case for tempo adjustment?***	***Has tempo adjustment been validated?***	***Alternative solutions***
Explanation	No	No, but separation into period quantum and tempo might be argued for.	No	Ideally, parity and age and/or duration specific measures, or indices standardised for these, but depends on time-scale
Anticipation				
Estimate cohort values	Yes, probably worsen fit between period and cohort or longer term indicators	Depends entirely on empirical success.	Limited. Results mixed thus far	Project/forecast cohort fertility explicitly
Predict longer run fertility			No	Detailed analysis of time series of recent trends. Project/forecast explicitly.
Evaluate growth prospects	Yes, stable assumptions even less valid under changing tempo		No	Mathematics of population reproductivity needs to be developed further
Description	Depends on purpose	Depends on objective of investigator	No	Depends on ultimate purpose
Population modelling	Yes, under classic assumptions but alternative models specifiable	Adjustment neither required nor ruled out by theoretical considerations	Only mathematical consistency required[17]	Perhaps greater variety of population models
Communication	Depends on purpose	Inadvisable; little justification	Not relevant	Present a greater range of indicators. Rely more on long-term projections of fertility

17 However, if a theoretical indicator is applied to an empirical situation and treated as a measure, it has to be shown that the assumed conditions hold sufficiently generally for the use of the measure to be defensible.

5. References

Andersson, Gunnar. 1999. Childbearing trends in Sweden 1961-1997, *European Journal of Population* 15: 1-24.

Andersson, Gunnar. 2004. Demographic trends in Sweden: an update of childbearing and nuptiality up to 2002, *Demographic Research* 11(4): 95-110.

Andersson, Gunnar, Jan M. Hoem and Ann-Zofie Duvander. 2006. Social differentials in speed-premium effects in childbearing in Sweden, *Demographic Research* 14: 51-70.

Barbi, E., J. Bongaarts and J.W. Vaupel (eds). 2008. *How Long Do We Live? Demographic Models and Reflections on Tempo Effects.* Berlin and Rostock: Springer and Max Planck Society.

Bongaarts, John. 1999. The fertility impact of changes in the timing of childbearing in the developing world, *Population Studies* 53: 277-289.

Bongaarts, John. 2002. The end of the fertility transition in the developed world, *Population and Development Review* 28: 419-443.

Bongaarts, John and Griffith Feeney. 1998. On the quantum and tempo of fertility, *Population and Development Review* 24: 271-291.

Bongaarts, John and Griffith Feeney. 2000. On the quantum and tempo of fertility: Reply, *Population and Development Review* 26: 560-564.

Bongaarts, John and Griffith Feeney. 2006. The quantum and tempo of life-cycle events, *Vienna Yearbook of Population Research*: 115-151.

Brass, W. 1990. Cohort and time period measures of quantum fertility: Concepts and methodology, pp. 455-476 in H.A. Becker (ed.) *Life Histories and Generations, Vol. 2.* Utrecht: ISOR.

Butz, William P. and Michael P. Ward. 1979. Will US fertility remain low? A new economic interpretation, *Population and Development Review* 5: 663-688.

Chamberlain, J. and S. Smallwood. 2004. Estimates of true birth order for Scotland, 1945-1999, *Population Trends* 117: 27-42.

Chang, H. 2004. *Inventing Temperature.* Oxford: Oxford University Press.

Cox, D. 1990. Role of models in statistical analysis, *Statistical Science* 5: 169-174.

Dorn, Harold. 1950. Pitfalls in population forecasts and projections, *Journal of the American Statistical Association* 45 (251): 311-334.

Feeney, Griffith and Jingyuan Yu. 1987. Period parity progression measures of fertility in China, *Population Studies* 41: 77- 102.

Feeney, Griffith. 1985. Parity progression projection, in: *International Population Conference, Florence 1985,* Vol. 4. Liège: International Union for the Scientific Study of Population: 125-136.

Feeney, Griffith. 1991. Fertility decline in Taiwan: a study using parity progression ratios, *Demography* 28(3): 467-479.

Foster, A. 1990. Cohort analysis and demographic translation, *Population Studies* 44: 287-315.

Freedman, D.A. 1985. Statistics and the scientific method, pp. 343-366 in W. Mason and S. Fienberg (eds) *Cohort Analysis in Social Research*. New York: Springer Verlag.

Frejka, Tomas and Gérard Calot. 2001. Cohort reproductive patterns in low fertility countries, *Population and Development Review* 27: 103-132.

Frejka, Tomas and John Ross. 2001. Paths to subreplacement fertility: The empirical evidence, *Population and Development Review* 27 (*Supplement: Global Fertility*

Transition): 213-254.

Goldstein, Joshua, Wolfgang Lutz and Sergei Scherbov. 2003. Long-term population decline in Europe: the relative importance of tempo effects and generational length, *Population and Development Review* 29: 699-707

Hajnal, John. 1947. The analysis of birth statistics in the light of the recent international recovery of the birth-rate, *Population Studies* 1: 137-164.

Hajnal, John. 1959. The study of fertility and reproduction: a survey of 30 years, in Milbank Memorial Fund, *Thirty Years of Research in Human Fertility*. New York: Milbank Memorial Fund, pp. 11-37.

Hand, D.J. 2004. *Measurement. Theory and Practice*. London: Arnold.

Henry, L. 1953. *Fertility of Marriage: A New Method of Measurement* (Paris, 1953; English translation, 1980, UN/ESCAP Population Studies Translation Series No. 3)

Hoem, J. 1990. Social policy and recent fertility change in Sweden, *Population and Development Review* 16: 735-747.

Hoem, J. 1993. Classical demographic methods of analysis and modern event-history analysis. *IUSSP International Population Conference Proceedings*, Vol. 3: 281-291.

Keilman, Nico. 1994. Translation formulae for non-repeatable events, *Population Studies* 48: 341-357.

Kim, Young J. and Robert Schoen. 2000. On the quantum and tempo of fertility: limits to the Bongaarts-Feeney adjustment, *Population and Development Review* 26: 554-559.

Kohler, Hans-Peter and Dimiter Philipov. 2001. Variance effects in the Bongaarts-Feeney formula, *Demography* 38: 1-16.

Kohler, Hans-Peter and Jose Antonio Ortega. 2002a. Tempo-adjusted period parity progression measures, fertility postponement and completed cohort fertility, *Demographic Research* 6: 92-144.

Kohler, Hans-Peter and Jose Antonio Ortega. 2002b. Tempo-adjusted period parity progression measures: assessing the implications of delayed childbearing for cohort fertility in Sweden, the Netherlands, and Spain, *Demographic Research* 6: 145-190.

Kohler, Hans-Peter, Francesco C. Billari and Jose Antonio Ortega. 2002. The emergence of lowest-low fertility in Europe during the 1990s, *Population and Development Review* 28: 641-680.

Lee, R.D. (1980) Aiming at a moving target: period fertility and changing reproductive goals, *Population Studies* 34: 205-226.

Lesthaeghe, Ron and Paul Willems. 1999. Is low fertility a temporary phenomenon in the European Union? *Population and Development Review* 25: 211-228.

Luther, N.Y., G. Feeney and W. Zhang, W. 1990. One-child families or a baby boom? Evidence from China's 1987 one-per-hundred survey, *Population Studies* 44: 341-357.

Lutz, Wolfgang and Vegard Skirbekk. 2005. Policies addressing the tempo effect, *Population and Development Review* 31: 699-720

Morgan, S. Philip and Rosalind Berkowitz King. 2001. Why have children in the 21st century? Biological predisposition, social coercion, rational choice, *European Journal of Population* 17: 3-20.

Murphy, Mike and Ann Berrington. 1993. Constructing period parity progression ratios from household survey data, in M. Ní Bhrolcháin (ed.) *New Perspectives on Fertility in Britain*. London: HMSO, 17-32.

Ní Bhrolcháin, Máire. 1987. Period parity progression ratios and birth intervals in

England and Wales 1941-1971, *Population Studies* 41: 103-125.

Ní Bhrolcháin, Máire. 1992. Period paramount? A critique of the cohort approach to fertility, *Population and Development Review* 18: 599-629.

Ní Bhrolcháin, Máire. 1994. Past history, synthetic indicators and period fertility, *Population (An English Selection)* 6: 116-120.

Prskawetz, Alexia, Tomáš Sobotka, Isabella Buber, Henriette Engelhardt and Richard Gisser. 2008. Austria: Persistent low fertility since the mid-1980s, *Demographic Research* 19 (12): 293-360.

Rallu, Jean-Louis and Laurent Toulemon. 1994. Period fertility measures: The construction of different indices and their application to France, 1946-89, *Population (An English Selection)* 6: 59-94.

Rodriguez, G. 2006. Demographic translation and tempo effects: An accelerated failure time perspective, *Demographic Research* 14 (6): 85-110.

Ryder, Norman. 1964. The process of demographic translation, *Demography* 1: 74-82.

Ryder, Norman. 1980. Components of temporal variations in American fertility, in R.W. Hiorns (ed.). *Demographic Patterns in Developed Societies*. London: Taylor and Francis, pp. 15-54.

Ryder, Norman. 1986. Observations on the history of cohort fertility in the United States, *Population and Development Review* 12: 617-643.

Schoen, Robert. 2004. Timing effects and the interpretation of period fertility, *Demography* 41: 801-819.

Smallwood, S. 2002a. New estimates of trends in births by birth order in England and Wales, *Population Trends* 108: 32-48.

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

Smallwood, Steve and Jessica Chamberlain. 2005. Replacement fertility, what has it been and what does it mean? *Population Trends* 119: 16-27.

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.

Sobotka, Tomáš. 2004a. Is lowest-low fertility in Europe explained by the postponement of childbearing? *Population and Development Review* 30: 195-220.

Sobotka, Tomáš. 2004b. *Postponement of Childbearing and Low Fertility in Europe*. Amsterdam: Dutch University Press.

Stolnitz, G.J. and N.B.Ryder. 1949. Recent discussion of the net reproduction rate, *Population Index* 15: 114-128.

Toulemon, Laurent and Magali Mazuy. 2001. Cinq projections de fécondité fondées sur une hypothèse de stabilité de comportements, *Population* 56: 647-656.

Toulemon, Laurent, Ariane Pailhé and Clémentine Rossier. 2008. France: High and stable fertility, *Demographic Research* 19 (16): 503-556.

van Imhoff, Evert 2001. On the impossibility of inferring cohort fertility measures from period fertility measures, *Demographic Research* 5(2): 22-64.

van Imhoff, Evert and Nico Keilman. 2000. On the quantum and tempo of fertility: comment, *Population and Development Review* 26: 549-553.

Wachter, Kenneth. 2005. Tempo and its tribulations, *Demographic Research* 13: 201-222.

Whelpton, Pascal K. 1946. Reproduction rates adjusted for age, parity, fecundity, and marriage, *Journal of the American Statistical Association* 41: 501-516.

Whelpton, Pascal K. 1949. Cohort analysis of fertility, *American Sociological Review* 14: 735-749.

Whelpton, Pascal K. 1954. *Cohort Fertility*. Princeton: Princeton University Press.

Wilson, Chris and Jim Oeppen. 2003. On reification in demography, in: J. Fleischhacker, H.A. de Gans and T.K. Burch (eds.) *Population, Projections and Politics*. Amsterdam: Rozenberg, 113-129.

Winkler-Dworak, Maria and Henrietta Engelhardt. 2004. On the tempo and quantum of first marriages in Austria, Germany, and Switzerland: Changes in mean age and variance, *Demographic Research* 10 (9): 231-264.

Zeng, Yi and Kenneth Land. 2001. A sensitivity analysis of the Bongaarts-Feeney method for adjusting bias in observed period total fertility rates, *Demography* 38: 17-28.

Zeng, Yi and Kenneth Land. 2002. Adjusting period tempo changes with an extension of Ryder’s basic translation equation, *Demography* 39: 269-285.