UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper A08/05

Applications and Policy

Five Reasons For Measuring Period Fertility

Máire Ní Bhrolcháin

Abstract

Five reasons for measuring period fertility are distinguished: to describe fertility time trends, to explain these, to anticipate future population prospects, to provide input parameters for formal models, and to communicate with non-specialist audiences. The paper argues that not all measures are suitable for each purpose, and that tempo adjustment may be appropriate for some objectives but not others. In particular, it is argued that genuine timing effects do not bias or distort measures of period fertility as dependent variable. Several different concepts of bias or distortion are identified in relation to period fertility measures. Synthetic cohort indicators are a source of confusion since they conflate measurement and forecasting. Anticipating future fertility is more akin to forecasting than to measurement. Greater clarity about concepts and measures in the fertility arena could be achieved by a stronger emphasis on validation. Period incidence and occurrence-exposure rates have a straightforward interpretation. More complex period fertility measures are meaningful only if a direct or indirect criterion can be specified against which to evaluate them. Their performance against that criterion is what establishes them as valid or useful. Discussion of tempo adjustment and allied issues in demographic measurement might profit from the development of a theory of measurement in demography, comparable to the axiomatic systems devised in e.g. physics, psychology and some areas of economics such as price index theory.

Five reasons for measuring period fertility[1]

Máire Ní Bhrolcháin
Division of Social Statistics
University of Southampton
Southampton SO17 1BJ
UK

Email: mnb2@soton.ac.uk

November 2007

Acknowledgments: I thank David Freedman, Laurent Toulemon and Mike Murphy for helpful comments on previous drafts of the paper. Any errors that remain are mine.

[1] This is a revised and extended version of M. Ní Bhrolcháin "Why we measure period fertility," Southampton Social Statistics Research Institute Working Paper A07/04, June 2007. Online at: http://www.s3ri.soton.ac.uk/publications/applications.php?s=10 and http://eprints.soton.ac.uk/46437/

Abstract

Five reasons for measuring period fertility are distinguished: to describe fertility time trends, to explain these, to anticipate future population prospects, to provide input parameters for formal models, and to communicate with non-specialist audiences. The paper argues that not all measures are suitable for each purpose, and that tempo adjustment may be appropriate for some objectives but not others. In particular, it is argued that genuine timing effects do not bias or distort measures of period fertility as dependent variable. Several different concepts of bias or distortion are identified in relation to period fertility measures. Synthetic cohort indicators are a source of confusion since they conflate measurement and forecasting. Anticipating future fertility is more akin to forecasting than to measurement. Greater clarity about concepts and measures in the fertility arena could be achieved by a stronger emphasis on validation. Period incidence and occurrence-exposure rates have a straightforward interpretation. More complex period fertility measures are meaningful only if a direct or indirect criterion can be specified against which to evaluate them. Their performance against that criterion is what establishes them as valid or useful. Discussion of tempo adjustment and allied issues in demographic measurement might profit from the development of a theory of measurement in demography, comparable to the axiomatic systems devised in e.g. physics, psychology and some areas of economics such as price index theory.

Keywords

Fertility; measurement; period fertility; tempo adjustment; validation; forecasting; projection; fertility timing; fertility quantum; total fertility rate; TFR; representational measures; pragmatic measures; axiomatic measurement theory.

1. Introduction

This paper investigates whether it is reasonable to adjust measures of period fertility for tempo effects. The subject has been much discussed in recent years, stimulated by the elegant and sophisticated adjustment to the total period fertility rate proposed by Bongaarts and Feeney (1998). Tempo adjustment of the total fertility rate (TFR) was proposed in the context of unexpectedly low levels of fertility in developed countries in recent decades. These have provoked much discussion of fertility prospects, with debate being concerned with timing effects and the likelihood of future recuperation in fertility, as well as with measurement issues *per se* (Golini 1998; Lesthaeghe and Willems 1999; van Imhoff and Keilman 2000; Kim and Schoen 2000; Zeng and Land 2001, 2002; Kohler and Philipov 2001; Frejka and Calot 2001; Frejka and Ross 2001; Kohler and Ortega 2002a,b; Bongaarts 2002; Smallwood 2002; Lutz et al 2003; Morgan 2003; Schoen 2004; Sobotka 2004a,b). The paper aims to make a contribution to recent debate primarily by structuring discussion around the variety of uses to which period fertility measures are put. Several key points are emphasised—the need to validate fertility measures empirically, the difficulties posed by single figure summary indicators and synthetic cohort measures, the distinction between measuring fertility and forecasting it, and the obscurity of the period quantum concept.

The paper addresses period fertility measurement in low fertility populations. The focus is on period indicators because it is these that have recently been the subject of debate (Schoen 2004: 807). However, the period vs. cohort issue is not discussed, the relative merits of the two perspectives being incidental to the paper's concerns. To seek clarification of period measurement is not to imply that period take precedence over cohort fertility. The assumption is only that the attempt to measure fertility in period mode is a valid research objective. Indicators of period fertility are widely used in academic and applied demography, by scholars on each side of the period/cohort debate, as well as by those who are agnostic on the subject. The need for period measures is independent of whether we view period trends as essentially a reflection of cohort forces or vice versa. Taking the utility of period measures as given, the objectives of the paper are to examine whether and when tempo adjustment is appropriate, in principle, and to suggest some ways in which we can advance our understanding of time trends in fertility, from both a substantive and a measurement perspective.

2. Approaches to measurement

Demography is probably the social science discipline most closely associated with measurement. Population science has accumulated considerable expertise in the production of accurate population statistics, in devising measures of static and dynamic aspects of population processes from data in a variety of forms, as well as in assessing data quality (Crimmins 1993; Caldwell 1996). Demography has an extensive range of techniques for estimating population parameters from incomplete and defective data, with e.g. Manual X (United Nations 1983) testifying to this capacity for demographic alchemy. Nevertheless, while having a substantial body of measurement procedures, skill in estimation, and a fund of knowledge concerning these, demography has not engaged in the kind of systematic theorising about measurement long established in physics and adapted further in psychometrics and psychophysics (Luce 1996, Hand 2004). Thus far, population science has not developed an axiomatic approach to demographic measurement comparable to price index theory (Fisher 1922, cited in Hand 2004; Balk 1996), an area similarly rooted in the concerns of official statistical agencies. This may be because the rate of technical change is slower in demography, due to the small size of the field, than in larger disciplines such as statistics and econometrics (Preston 1993: 596). Whatever the reasons, the absence of a theory of measurement means that we lack a set of disciplinary principles by which to evaluate competing measures, or even a conceptual framework within which to situate such a discussion, beyond the basics of rate construction and standardisation. As a result, underlying the practical questions that arise in the issue of tempo adjustment are some foundational questions that are as yet unacknowledged and unexplored in demographic thinking.

A key underlying issue is the precise nature of demographic measurement. Hand (2004) distinguishes two principal approaches to, or aspects of, measurement: representational and pragmatic. In representational measurement, the relationships between the attributes studied are reflected in, or modelled by, the relationships between the numbers representing them. In pragmatic measurement by contrast the numbers are chosen by the investigator to some extent arbitrarily, by a set of conventions that do not mimic the real world phenomenon's attributes or behaviour. Operationalism, in which the concept in question is defined by the measurement procedure itself, is an extreme form of pragmatic measurement. Perhaps because of its strong practical roots and focus, demography as a discipline has not yet addressed the question inherent in this distinction. What is the formal status of the measures that we routinely use? To what extent do they represent in a realistic way the phenomena they are intended to measure and how far do they incorporate elements decided on and agreed by convention? A considered approach to the question of tempo adjustment seems to lead us into that terrain.

On the principle that "(t)he use to which an index will be put will be the determining factor in its construction" (Hand 2004: 267), the paper approaches the issues by considering the purposes for which period fertility is measured, and then discusses how far tempo adjustment is appropriate to each of the objectives discussed. The main reasons for which period indicators of fertility are employed appear to be as follows: to describe time trends in fertility, to explain these, to anticipate future population trends, to provide input parameters for formal population models, and to convey information on fertility trends to non-specialist audiences[2]. In my view, the measures most suitable for these objectives differ. Fertility indices that are adjusted for period change in the timing of childbearing—tempo adjusted measures—may be appropriate for some purposes, but not for others. No one fertility index or set of indices is best suited to all purposes.

However, without agreed principles of measurement, how can we assess whether any proposed indicator is suitable for a given purpose? In the present perspective, the suitability of a period fertility indicator is assessed by validation, having first specified clearly what the indicator is intended to measure. Validation is central to much applied demography—e.g. the validation of census counts by means of coverage surveys,

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

demographic analysis, administrative checks and the like, or in evaluating the completeness of vital registration or of demographic surveillance systems, or in the accuracy of survey reports (Caldwell 1996; Preston et al. 2001; Siegel and Swanson 2004). Much validation in demography concerns the accuracy of straightforward quantities that present no conceptual difficulties of mensuration, though plenty of practical ones—e.g. counts of persons or vital events, or reports of age or birth weight and other personal characteristics. Validation is also carried out in relation to more complex concepts such as unmet need, fertility intentions, family size preferences, and the planning status of births—for recent studies of this kind, see Bankole and Westoff (1998), Casterline et al (2003), Dixon-Mueller and Germain (2007), Curtis and Westoff (1996), Marston and Cleland (2003), Joyce et al (2002) and Koenig et al (2006). The evaluation of projections for forecast accuracy might also be considered a form of validation, though as much of methods as of measures (Keilman 1997; Ahlburg et al 1998; Wilson and Rees 2005).

Aggregate indicators of fertility in period mode are not usually explicitly validated, their relative merits normally being assessed on practical (and relevant) criteria such as how specific or well-standardised they are. There is an extensive literature on validating measures in a wide range of disciplines[3], and the subject has been discussed in particular depth in relation to psychological testing. Validity theory includes three general types of

[3] A search on the terms “valid* AND measur*” in the science and social sciences databases of the ISI Web of Knowledge brings up over 100,000 citations. Of the first 100,000 of these, the top 10 subject areas are, in order: public, environmental and occupational health, psychiatry, electrical and electronic engineering, clinical psychology, clinical neurology, analytical chemistry, radiology, meteorology, and environmental sciences.

validation (Rupp and Pant 2006). Criterion validity can be assessed where an independent measure of the concept in question is available. Content validity is based on expert agreement about the content of a proposed measure. Construct validity involves embedding a measurement procedure within a network of concepts and hypotheses concerning the construct to be measured. In considering the question of tempo-adjustment, these concepts of validation will be useful. Indeed, some of the criticisms levelled thus far at tempo-adjusted measures arise from an implicit validation perspective. While validation appears to be a productive way forward where controversy arises about approaches to period fertility measurement, this paper does not attempt to discuss in depth the validation of demographic measures.

If a phenomenon is not well understood, and that procedures for measuring and describing it are still evolving, attempts to explore appropriate forms of measurement may effectively also become an investigation of the phenomenon itself. The history of the measurement of temperature provides an example. Establishing a temperature scale was interdependent with developing an understanding of the phenomenon of heat itself. That is, poorly understood aspects of heat—supercooling and superheating, for example—cut across attempts to arrive at fixed points (freezing and boiling) by which to anchor a coherent, stable scale of temperature. This in turn provoked investigations of the behaviour of the process, whose results fed back into the construction of a scale of temperature (Chang 2004). In the same way, the attempt to find a sensible way of measuring period fertility trends—at any rate for explanatory purposes—almost certainly requires us to associate measurement effort with substantive investigation of how and why fertility changes through time. Progress in measuring and in understanding period fertility are likely to be interdependent.

3. Objectives of fertility measurement

3.1 Describing fertility trends

Demography is regarded by many social scientists as primarily a descriptive discipline (e.g. Moffitt 2003) though it also has strong theoretical roots and a long-standing focus on explanation (Crimmins 1993; Caldwell 1996). A natural starting point is, therefore, the measurement of period fertility for descriptive purposes[4].

Much of the output of national statistical offices and international statistical bodies such as the UN and its various regional agencies, WHO and the Council of Europe, is highly descriptive. But annual series of birth counts and fertility measures are compiled for purposes beyond simple description. Collecting and processing data on annual births is costly, as is the production of accurate fertility series. It is neither for popular dissemination, nor to supply academic demographers with the raw material for their research, that official statistical agencies compile such figures. Data collection and publication are, on the contrary, justified on policy grounds—to provide an empirical base for population projections, for other forms of local, and national, and international administration and planning. Studies reporting new approaches to measurement often present much descriptive material. But they can have an analytical focus too—not only in contrasting alternative descriptions of fertility trends but also in measuring the components (through e.g. disaggregation by order) of an aggregate measure, and thus identifying its demographic determinants (e.g. Whelpton 1946; Ryder 1980; Feeney and Yu 1987; Ní Bhrolcháin 1987; Hoem 1993; Rallu and Toulemon 1994).

[4] 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.

It is not altogether clear whether the use of period fertility measures for descriptive purposes can help to clarify the arguments for and against tempo adjustment. We could take the view that “anything goes”: in descriptive investigations the measures used are either the only feasible ones or are chosen according to the investigator’s preference. There are two senses in which this is obviously true. First, data for areas or time-periods of considerable interest are often incomplete or defective, or less than comprehensive, and so the choice of measures can be severely constrained. Second, in the fertility arena there are many measures to choose from, and as yet no one measurement approach has gained widespread agreement as the method of choice. Nevertheless, an idiosyncratic preference will not necessarily guarantee an author publication, or an audience, or the acceptance by the demographic community of either measurement approach or associated findings. To assess the appropriateness of a measurement approach, we need to know what the aim of the description is. For example, in describing regional differentials in fertility, is the objective to identify historical continuities in regional profiles, or similarities in time-trends between specific types of region? In the first case, either period or cohort measures might be appropriate; in the second, the indicators selected might differ according to whether the underlying question concerned regional differences in the causes of fertility change, or regional variation in fertility prospects for projection purposes.

Where different measures give a different account of time trends, we will want to know both why this is and which gives the more accurate picture. Such a stance is not purely intellectual. The question as to which measure gives a more accurate reading can matter significantly for policy purposes, especially in evaluating the success of a population policy or intervention. Feeney and Yu (1987: 78) comment in connection with fertility in China in 1980-81, when two versions of the TFR gave discrepant results, that “it will not suffice to say that the answer depends on what measure of fertility we use. We must ask why the two measures differ, and which provides, in this particular case, the better representation of 'the level of fertility.'” Feeney and Yu go on to argue for the superiority of the parity-progression based measures on the grounds that they are less subject to distortion as a result of changes in tempo—distortion meaning in this case bias resulting from a changing composition by parity and duration of women at risk. They also argued for multiplicative parity progression indicators on the grounds that they were more stable than the additive, age-based equivalent. Both of these arguments appear to reflect a concern with the stability of period measures, and so to reflect a concern with future prospects for fertility.

Leaving aside instances where incomplete sources limit the choice of indicator, the selection of period fertility measures for descriptive purposes seems to depend on the utility of those indicators for the other purposes discussed in this paper—explanation, prediction and so on[5]. An author whose primary purpose is description will want to present indicators that reflect those aspects of fertility trends that matter, whether for scientific or practical purposes. Descriptive accounts of fertility trends can stand alone, but they are most informative when based on measures of clear significance to the

[5] Caldwell (1996: 309) has remarked that “some theory, perhaps barely articulated, must underlie all analysis, especially in a discipline so enamoured with the detection and explanation of change.”

context. The link between the selection of measures for description and the other purposes discussed in this paper is most apparent when it comes to validation. What does validation mean in a descriptive context? How can we validate a descriptive account of fertility trends? In my view, the answer depends on the underlying reason why a description of the trends is of interest in any particular context, on the scientific or practical purposes for which such information is useful.

3.2 Explaining fertility trends

This section now considers period fertility measured for the purpose of explaining time-trends. In the present perspective, the ultimate explanation for fertility time trends would take the form of a substantive or behavioural model as distinct from an empirical or descriptive model (Freedman 1985; Cox 1990; Hand 2004). A substantive model, and the theoretical explanation it embodies, would be as close as possible to representing the real-world processes giving rise to the fertility rates analysed. The fertility measures embedded in such a model would be as close to representational in type as they can be. Like many indicators in social science, fertility indices are hybrid in form. Counts of births and populations at risk are the basis for the measures used, and so there is a strong representational element. But standardising rates, or synthesising or adjusting them, introduces a conventional aspect to the mix, and so dilutes the representational component. An alternative to substantive or behavioural modelling is explanation via an empirical or descriptive model, one that aims to account for as much as possible of the variance in a dependent variable, but without seeking to represent the modus operandi of the phenomena generating the indicators. Most of my argument assumes that it is ultimately a substantive model of fertility change that we seek. Nevertheless, empirical modelling is also useful, and probably more so where the dependent variables—measures of annual fertility—are chosen for their closeness to the underlying behavioural phenomena rather than on conventional grounds.

To understand the causes of change in period fertility, we require a form of measurement that reflects period fertility in its role as explanandum. The properties desirable in a measure of fertility as dependent variable may be quite different from the desiderata in period fertility indices that serve other functions. In particular, there is no *a priori* reason why a measure of period fertility as explanandum should be expected to predict future or longer-run trends. Ideally, specification of fertility as dependent variable should flow from an explicit behavioural model or theory, and in any case choice of measure usually embodies some assumptions about underlying processes, whether acknowledged or not. Leridon (2006) notes, however, that there is no generally accepted and successful theory of fertility change, and that this view is shared by several other leading commentators in the field (Lesthaeghe and Surkyn 1988; van de Kaa 1996; Kirk 1996). Hence, even though we have many candidate theoretical perspectives to explore (Crimmins 1993; Morgan and Taylor 2006), we have as yet little by way of well-established theory to guide us in specifying suitable measures. Nevertheless, some practical guidelines can be suggested.

In a theoretical explanatory context, a measure of period fertility as dependent variable needs to meet both substantive and demographic-statistical criteria. Two substantive issues will be discussed here: the treatment of personal and historical time, and the issue of scale. Personal time and historical contingency are central behavioural issues in the context of fertility measurement . How far are childbearing decisions influenced by (a) current state (e.g. parity, age, duration) vs. previous fertility history (beyond that represented in current state) vs. intentions or expectations regarding either future or overall lifetime fertility and (b) in a historical context, by current, past or expected future social and economic circumstances? A pure period-based behavioural model such as Ryder's (1973: 504) suggestion that couples may make childbearing decisions sequentially—month by month or year by year—will lead us to choose fertility indicators that are strictly period-based, and to look for determinants of this kind also (see also Ryder 1980).[6] An alternative is the cohort approach explicitly adopted by Butz and Ward (1979) whose indices of period quantum[7] (or level) and timing, in both their ex-post and ex-ante forms, reflect the supposition that annual childbearing decisions are taken relative to an overall desired or expected family size. Lee (1980: 208) is also clear about his behavioural assumptions—that the ultimate desired family size of any cohort is not fixed but may change over time, and that at each age "the annual birth rate is a fixed proportion of additional desired fertility." Both of these see period rates as essentially driven by a cohort target, fixed in the case of Butz and Ward but, for Lee, subject to alteration through the life course. In recent years, research on the behavioural underpinnings of fertility change has, however, not advanced greatly, and fewer linkages have been made

[6] Ryder has been a strong advocate of the cohort approach to fertility but confessed to doubts on the question when confronted with aggregate change in US fertility combined with the findings of the US fertility surveys (Ryder 1973).
[7] The quantum idea essentially refers to the level of fertility, but with any timing component removed. It will be argued, however, that the concept is ill-defined in a period context.

between behavioural assumptions, on the one hand, and either measurement schemes or aggregate trends, on the other.

Scale is another central issue, having both substantive and statistical aspects: detail in measurement will vary with time-scale and the level of generality sought. For broad brush treatment or long range perspectives such as those of e.g. Frejka and Ross (2001) or Caldwell and Schindlmayr (2003)—where gross change or differentials are the focus of interest—explanatory ideas will probably be general and systemic, and a total fertility rate (TFR) of some kind may well be adequate. When the TFR is effective in this context, it is not because it represents “family size” in successive calendar years, in any real sense, but because it summarizes annual age-specific fertility rates, and so is a general indicator of level. In fact, when used in this way the period TFR could be replaced by e.g. TFR/35—the average period fertility rate at single years of age—with no loss. In either case, only variation in level is represented. This may be adequate if attempting to account for a major shift in the level of fertility—e.g. from a TFR in the region of 5 to one of around 2. But for short to medium run series, in low fertility settings, a greater degree of resolution is to be expected both in fertility as the object of explanation and in the explanatory factors proposed. In this context, a single figure summary—whether a conventional or adjusted TFR, or an average annual age specific fertility rate—is a coarse dependent variable, and more refined measurement is called for. The principal refinement is that rates or probabilities are used and that these are specific by parity, both because parity dependence is a key feature of fertility behaviour in contracepting populations and because the time-path of fertility rates can vary substantially by parity, particularly low vs. higher order births (Ryder 1980, 1986; Isaac et al. 1982; Namboodiri 1981; Ní Bhrolcháin 1987; Feeney and Yu 1987; Feeney and Lutz 1991; Toulemon and Rallu 1994; Kohler and Ortega 2002b; Sobotka 2004b). In addition, progressions of differing orders appear to be subject to different influences (Bulatao 1981; Namboodiri 1981; Isaac et al 1982; Yang 1994; Andersson 2000). Measures specific by parity have the important substantive property that they reflect directly the sequential nature of family building, and thus of the life-course, and hence are close to the behavioural processes giving rise to aggregate change.

The principal demographic-statistical requirement for measures of period fertility as dependent variable is that they give a fair representation of temporal change. There is, however, no independently ascertainable, true value of period fertility against which to evaluate a proposed period fertility indicator. An assessment of the validity of the time trends depicted cannot therefore be made directly. We can, however, specify some minimal conditions (though a rigorous approach to the subject is a matter for future research). Where used as a dependent variable, an indicator of period fertility should be uninfluenced by nuisance factors—that is, any substantial demographic influences on fertility rates that we acknowledge as occurring but do not wish to explain. The effect of such factors can be removed in several ways—by increased specificity, by standardizing, or by some form of regression. Thus, age specific rates have the rationale or removing age structure effects—rates vary substantially by age, and age structure is accepted for many analytical purposes as a given rather than as something to be explained. Beyond age, the distribution of women by parity is an important potential nuisance factor, in that, as noted earlier, birth rates vary by parity, and parity specific rates often follow different trends. The biasing effect of the parity distribution can be removed by parity specific rates, thus reinforcing the substantive case for parity specific indicators. Statistical considerations would also lead us to doubt the utility as a dependent variable of a single-figure summary indicator, such as the TFR, when time-trends in age- or parity-specific indicators differ.

Further disaggregation or standardization by age and/or duration may also be required, for analogous reasons. Age-parity or parity and duration-specific rates, or rates specific for both age and parity, as well as period parity progression ratios synthesising these, can be used for the purpose. Arguments can be advanced in favour of age and/or duration as controls, but the relative merits of age and/or duration controls in attempts to explain time trends in parity specific fertility remain to be established. Note that in removing the effects of age and/or duration, we implicitly accept as given the variation in rates by age and/or duration and so decline to account for these.

Should we adjust period measures to remove tempo effects when fertility is a dependent variable? The argument for such correction is that the TFR is biased or distorted by timing change. A first point to note is that even when timing is stable the conventional TFR, considered as a dependent variable, may misrepresent period change, because it is standardized only for age and so is influenced by any changes that occur in the parity distribution. We can remove this nuisance factor, as noted above, by introducing parity-specificity in our rates, or by standardizing for parity in some way. However, what if the timing of fertility is changing—do we then need to adjust our measures to remove the tempo influence, when fertility is the explanandum? On the present view, the answer is no, provided that our measures are appropriately specified. This is because genuine timing influences are an integral part of fertility as dependent variable. The argument is illustrated via the start of childbearing, which is subject to the largest timing effects[8].

A progressive delay in first births will result in an increase in the proportion of women at younger ages who are childless, and so at risk of a first birth. In such conditions, time-trends either in age specific rates or in unconditional first birth rates by age (incidence rates) have two components: changing proportions childless at each age, due to later timing—a compositional effect—and change in the birth propensities of childless women—the true period trend of interest. The timing-related compositional effect can be eliminated by making rates parity specific. That is, age specific birth rates to women of parity zero remove the compositional effects that result in spurious timing related influences. More generally, methods have been available for some time that remove this compositional timing-related effect from period measures—age-parity specific rates for first birth together with parity- and duration-specific rates for later births (or parity-, age- and duration-specificity), and period life tables that synthesize these (Henry 1953; Feeney

[8] In England and Wales, for example, the mean age at all live births rose by three years (from 26.5 to 29.5) between 1977 and 2005, and the mean age at (true) first birth by 3.2 years (from 24.4 to 27.6) during the same period (ONS *Birth Statistics*. *Series FM1*, various years). These figures suggest that rises in the age at second and later births over the period are largely due to the delayed timing of first birth. See also United Nations (2003: Chapter III.C).

and Yu 1987; Ní Bhrolcháin 1987; Murphy and Berrington 1993; Rallu and Toulemon 1994; Kohler and Ortega 2002a).

When the timing of first births is changing, genuine tempo effects will be present in age-parity specific rates, manifest as a shift along the age axis in the birth rates of childless women. *These are not distorting to measures of period fertility as dependent variable, because real tempo change is part of what we are, or should be, trying to explain.* For example, the baby boom of the late 1950s and 1960s was partly due to accelerated childbearing (Butz and Ward 1979; Ryder 1980). If part of the explanation is that post-war prosperity, full employment and high wages induced couples to marry earlier and have children sooner than in preceding periods, that faster pace of family formation needs to be represented on the left hand side of the equation. Similarly, later childbearing is one aspect of what we need to explain in relation to current fertility trends in developed societies. To this end, the full effect of genuine timing change should be retained in measures of fertility as dependent variable, rather than adjusted away. 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.

An analogy may help to illustrate the argument. 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 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.

There appears to be little agreement about whether the quantum idea is applicable in a period context and if so what it means (see e.g. the varying definitions and views of Butz and Ward 1979; Ryder 1980; Pressat and Wilson 1988; Brass 1990; Murphy 1993; Bongaarts and Feeney 1998; van Imhoff and Keilman 2000; van Imhoff 2001; Bongaarts 2002; Koehler and Ortega 2002a; Schoen 2004). Note in particular Brass's (1990: 472) comment that "(t)he concept of the 'quantum' is so bound up with the family size achieved over the reproductive period that the cohort total fertility must serve as a focus." And while several specifications of period quantum and tempo measures have been suggested (e.g. Butz and Ward 1979; Ryder 1980; Bongaarts and Feeney 1998; Kohler and Ortega 2002a), these may be no more than mathematical abstractions, and may have no real-world reference. If we regard period tempo and quantum as representing the behaviour of individuals or aggregates, we would need to ground these constructs in empirical reality. To do so, evidence would be required that, at the individual level, decisions about timing and quantity are made independently rather than being a joint process. Ryder (1980) puts this argument with great clarity. Ryder's paper is best known for having made estimates of the quantum and tempo of cohort fertility, and analysed them into their components. However, he concludes with some highly sceptical comments about whether quantum and tempo are behaviourally distinct. In deliberating on the process giving rise to changing quantum and tempo he suggests that reproductive decisions may incorporate both quantum and tempo elements simultaneously, and proposes that quantum and tempo "are to some degree manifestations of the same underlying behaviour" (ibid.: 44). He concludes 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: 45). The argument presented here is very much in tune with this two-fold view of Ryder's—that tempo and quantum are interwoven in real-world decisions, and that specifying appropriate measures, and the justification for separating tempo and quantum in an explanatory context, depends on the nature of the real-world decision processes that give rise to fertility events.

Tempo adjustment of period fertility as dependent variable could potentially be argued for if several conditions were to hold: that in period mode, the quantum of fertility and its timing are separable in a quantitative sense, that they reflect distinct aspects of the underlying behavioural process (though how distinct is an open question), and that they respond differently to change in social, economic and other determining factors. 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. On the question of response, 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.[9] Empirical

[9] For example, the duration-specific indicators for birth orders 2 and above of Hoem (1990) and Andersson et al (2006) give much clearer evidence of the timing shifts associated with the Swedish speed premium effect than do the adjusted mean ages of Figure 4b of Kohler and Philipov (2001). The latter would not, alone, be convincing evidence of a speed premium effect whereas the analyses by duration are very compelling

evidence demonstrating that quantum and tempo are genuinely distinct aspects of the behavioural processes underlying period fertility movements would be required to clarify the matter further. Annual data of both immediate and longer-term fertility intentions and on the decision-making process would be required for the purpose.

If we take a representational approach to measurement, establishing a link between behavioural processes and demographic indices—whether of quantum, tempo or the two combined—is an essential step in arguing for the relevance of the indices concerned for explanatory purposes. If the two cannot be firmly linked, then either the indicator(s) proposed, or the explicit or implicit behavioural concepts are of unproven value in explaining time trends, even though the indices may have instrumental value as predictors. On the other hand, we can choose to consider period fertility measurement to be primarily pragmatic in form. If so, we nevertheless need to be explicit about what period indices are intended to measure, and, unless an operationalist approach is adopted, to suggest how they can be validated.

How can we tell whether an indicator is suitable as a dependent variable—that is, how can we validate the measure? No independent criterion is available by which to assess how well a period measure, or a set of such measures, reflects temporal change in fertility. But an indirect check on the validity of a measure of period fertility as dependent variable *is* available—viz. explanatory success. An indicator or set of indicators of period

---

indeed. Of course, the Swedish maternity pay regulations are framed in terms of time since previous birth, and so measures that are specific by duration since previous birth naturally fit the structure of the maternity pay incentive.

fertility as explicandum can be considered useful or valid to the extent that it is 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. An effective explanation of post-war fertility trends in developed countries would be based, at least in part, on some form of quantitative model of time series data, since an array of ever changing influences are almost certainly at work. However, systematic attempts at explaining aggregate fertility trends via statistical or econometric models are and have been rare. The Easterlin approach has not been successful in meeting empirical tests (Waldorf and Byun 2005) and there has been little recent work on aggregate movements in fertility, grounded in behavioural theoretical assumptions, such as that of Butz and Ward (1979) and Lee (1980). Hence we currently have no well-formulated explanatory framework, backed up by solid evidence, to underwrite particular approaches to period fertility measurement, even though good substantive grounds exist for e.g. advocating parity specific measures. Nevertheless, the key point here is that a convincing explanation, and not a check against cohort values, is the appropriate criterion for evaluating an indicator of period fertility as dependent variable.

3.3 Anticipating the future

A further purpose for constructing indicators based on period fertility is to anticipate future population parameters—whether prospects for fertility or for population growth, size and composition. This is often the explicit or implicit rationale for studying current or recent trends. The outlook for fertility is a natural preoccupation in a profession whose most sought-after applied function is providing information on future population trends.

It is also a natural focus of interest when fertility levels are causing policy concern, as is currently the case in many low fertility societies (Lutz and Skirbekk 2005; McDonald 2006). Many of the difficulties that beset period fertility measurement arise from attempting to expand its time reference into the future. Indicators that reflect some aspect of the (prospective) longer run are clearly more difficult to create from the data of a single period, and inherently less satisfactory, than are measures intended to reflect purely current reproductive performance. If an indicator is sought purely for prediction purposes, it does not matter what form it takes as long as it performs well as a predictor. Successful prediction does not require theory, even though theory may be helpful in this regard.[10] Importantly, however, even a very well-established theory may give little or no predictive power—Lieberson and Lynn (2002) use the example of evolutionary theory to highlight this point.

Period-based measurement can be geared to reflecting population prospects in several ways. One objective is to estimate the fertility of cohorts—discussed here in terms of birth cohorts, though marriage cohorts and parity cohorts may also be of analytical interest. Second, period fertility may be examined for indications about future fertility trends, in a broader and less specific sense. Forecasting or projection is a third approach to the fertility of the future. Finally, period fertility may be measured with a view to evaluating future population growth prospects.

[10] Keyfitz (1982) argued that causal knowledge of past population change had not yet proved useful for forecasting purposes. Sanderson (1998) argues that that judgment needs to be revisited.

3.3.1 Cohort fertility

Where data are available on completed childbearing, measuring cohort fertility is straightforward. A variety of indices of both level and timing can be specified, up to the limits of data quality and sample size. Practical problems may arise regarding e.g. whether immigrants should be included, but the indices themselves are straightforward to define. However, where the childbearing of the cohorts in question is incomplete, period rates are used in the estimation process and it is then that difficulties arise.[11] Leaving aside projection, which is discussed in a later section, conversion of period fertility into cohort terms takes two forms: calculating synthetic cohort measures and demographic translation.

3.3.1.1 Synthetic cohort indicators

The simplest and most common demographic device for bridging the period-cohort gap is the synthetic cohort indicator. Such measures convert the rates of a single period into an estimate of experience that in reality extends over many years, and in some cases over a lifetime. To fulfil this role, the assumption is required that the age and/or duration-specific rates of a given period obtain at successive ages/durations. The best known hypothetical cohort fertility index is the conventional TFR, normally presented as an estimate of the mean family size of an imaginary cohort experiencing a particular period's age-specific fertility rates throughout their childbearing years.

That period synthetic cohort indicators are often thought of effectively as cohort estimators is evident from the long-standing criticism that those constructed on an

---

[11] 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…"

additive basis can produce on occasion results that are impossible in a real cohort. Ever since Whelpton (1946, 1949)[12], the sum of period age specific first birth rates has been considered a defective indicator because it can reach values greater than 1 (see, for example, Park 1976, Ryder 1990, Bongaarts and Feeney 1998; van Imhoff and Keilman 2000; Bongaarts 2002). The period TFR is also regarded as faulty on that argument as it is the sum of birth order specific components, each of which is potentially subject to this apparent inconsistency. The explanation for the feature is that it is due to timing effects. However, if the TFR were regarded as *sui generis* or as a statistical summary, such a criticism would not make sense.

In the literature on tempo adjustment, the charge is levelled at the TFR that it is biased or distorted whenever the timing of fertility is changing. However, terminology varies, with some sources using the concept of bias, either alone or interchangeably with the idea of distortion (Bongaarts and Feeney 1998; Bongaarts 2002; Bongaarts and Feeney 2000; van Imhoff and Keilman 2000; Kohler and Philipov 2001; Zeng and Land 2001, 2002) and others referring exclusively to the notion of distortion (Bongaarts 1999; Bongaarts and Feeney 2000; Frejka and Ross 2001; Kohler et al. 2002; Schoen 2004; Sobotka 2004a). The terms will be used interchangeably throughout the paper. The bias in question is not statistical, in that there is no question of a probability distribution for the

[12] The original logic of this criticism is that e.g. a sum of first birth rates greater than unity implied that the conditions represented by the age specific rates were temporary could not continue indefinitely into the future. And while indicators based on a multiplicative basis are less volatile than those constructed additively (e.g. Feeney and Yu 1987; Rallu and Toulemon 1994), the lack of realism in the assumption of fixed rates remains unaltered. Another way of construing the criticism that a sum of first birth rates exceeding 1 is absurd is that it reflects the outcome under a stable population scenario of "current conditions."

TFR (see also Zeng and Land 2002, note 1). The most reasonable way of construing discussion in this area is that it is measurement bias that is at issue. If the distortion or bias generally referred to in the tempo adjustment literature is indeed measurement bias, then there must be something which is measured in a biased way by the TFR. What is that something?[13] Recent commentary is not clear on what it is that is held to be misrepresented by the TFR, but two interpretations are possible.

A first possibility is that it is a construct "period fertility" that is wrongly measured by the TFR and that any period measure of fertility that is influenced by timing effects is distorted by definition. This is the kind of proposition that might be a candidate principle in a theory of, or a set of axioms relating to, demographic measurement. The position would have to be proposed and supported explicitly, however, since it is not self-evidently true that a change in tempo necessarily distorts a measure of period fertility. Some scholars (Bongaarts and Feeney 1998, 2000; Zeng and Land 2002) appear to espouse a view somewhere close to this. For the point to be sustained, however, the alternative view, viz. that genuine timing effects are integral to, and do not distort, a period fertility indicator, at least when considered as explanandum, would have to be shown to be faulty.

A second way of construing distortion as measurement bias is to recall Ryder's (1964) use of the term distortion—what he meant was that the period TFR was a distorted

[13] See also van Imhoff (2001: 27) who asked the question: "what does a PTFR stand for?" His answer was that it is really attempting either to measure current births, in which case we should count these directly, or to get an idea of generational replacement, in which case we should be attempting to estimate cohort TFRs.

version of cohort values. This appears to be what many commentators have in mind in referring to tempo distortions. However, the assertion that bias is present only when timing is changing reflects a theoretical rather that an empirical construal of the TFR. As a measure of a real, as distinct from a theoretical, cohort quantity, the TFR is, in fact, *always* mistaken and not solely when fertility timing is changing. The conventional TFR is equivalent to cohort fertility only where age-specific rates are either fixed or randomly distributed around a given period's values. Such stability is rare. Hence, the conventional period TFR is almost always biased as a real-world cohort estimator since it corresponds only by chance to the mean family size of any cohort at risk during the period.

Thus, bias in the TFR as an indicator of real cohort values is present whether or not the timing of fertility is changing. Shifts in fertility tempo may of course increase the bias—that is, create a greater discrepancy between period and corresponding cohort TFRs, particularly if only a timing change occurs with no change in cohort quantum. There are, therefore, two sources of bias in the TFR as an measure of real cohort fertility—one due to the non-fixity of age specific rates and the other due to tempo change.

Is adjustment of the period TFR for tempo effects warranted, in principle, in this context? The answer is probably yes, *if* the objective is to get closer to cohort mean family size than the figure given by the period TFR, and *if* a case can be made that the timing change is a short-run phenomenon. The same goes for other synthetic indicators, such as period parity progression ratios, *if* the objective is to get closer to the cohort equivalent. The case for adjustment is that tempo effects are often (but not always) offset in later periods to some degree—though the extent to which compensating movements in rates occurs is unpredictable (Lesthaeghe 2001; Frejka and Calot 2001; Billari and Kohler 2004). While tempo adjustment can do nothing to correct for the bias due to non-fixity of the rates, it can sometimes be expected to reduce the bias in period synthetic indicators, considered as estimates of cohort quantities. The argument for tempo adjustment is empirical—based on the behaviour of fertility series in the past, as well as judgment about the likelihood that a current shift is mainly a timing phenomenon—rather than theoretical, though currently available adjustment procedures are derived on theoretical grounds. The case against is that we have no way at present of distinguishing a short term timing shift from a long-run change in level. Whatever method is used, biases will remain in tempo adjusted period measures considered as estimates of real cohort quantities, both because rates schedules rarely if ever remain fixed and because timing changes may be neither as systematic nor as persistent as recently proposed adjustment procedures assume.

3.3.1.2 Demographic translation

A further type of conversion, demographic translation, goes somewhat beyond the synthetic cohort principle and treats the transfer from period to cohort as a systematic problem to be solved. Translation seeks a determinate, formal mathematical relationship between the level and timing of period and cohort fertility (Ryder 1964; Foster 1990) Translation can proceed from period to cohort or vice versa. In extending translation to non- repeatable events, Keilman (1994: 343) stated the problem as “how to translate the time-dependent indices of the quantum and tempo of the process obtained from a period perspective into quantum and tempo indices for cohorts.” For accuracy, demographic translation requires smooth patterns of change much simpler than those found in observed fertility series. It fulfils a useful function in providing an account of overall fertility trends that dampens the more volatile period picture. On one view, the task of translation is impossible since "in real life the factors involved in explaining the link between period and cohort quantum are so complex and subtle… that we will never be able to describe it completely" (van Imhoff 2001: 25). If this is so, then translation in either direction, though an interesting theoretical problem, and offering a useful tool for projection and modelling purposes, cannot be empirically successful since period influences are irregular, unpredictable, and appear not to have cyclical features.

3.3.1.3 Validation

Criterion validity can clearly be applied where a period indicator is intended to estimate cohort fertility, with the criterion being the fertility of real cohorts. The conventional period TFR performs badly on this test, having long been known to be a poor estimator of cohort mean family size. The discrepancy is often illustrated graphically by the much larger swings in the period TFR than in cohort total fertility (for recent examples see van Imhoff and Keilman 2000; Frejka and Calot 2001; Smallwood 2002; Schoen 2004). It is this shortcoming that has, in part, motivated the search for improved ways of converting between period and cohort, among them tempo adjustment. Adjustment methods vary in how well they estimate cohort quantities. Bongaarts and Feeney do not see their adjusted TFR as an estimator of cohort fertility but several commentators have construed it in this way (van Imhoff and Keilman 2000; van Imhoff 2001; Kohler and Ortega 2002a; Smallwood 2002; Schoen 2004; Sobotka 2003). Evidence suggests the adjusted TFR has limited accuracy on an annual basis (van Imhoff and Keilman 2000; van Imhoff 2001; Smallwood 2002; Schoen 2004) but Bongaarts and Feeney (1998) show that an average of the annual adjusted TFR values tracks cohort fertility well. Neither the Kohler and Philipov (2001) nor the Kohler and Ortega (2002a) measures appear to have been evaluated against the cohort equivalents. Butz and Ward's (1979) Average Completed Fertility index, advocated by Schoen (2004) as a measure of period quantum, performs better than the alternatives considered by Schoen in approaching cohort values. But it has several acknowledged limitations as a period measure, requiring in particular that cohort fertility be already known, and thus conditioning on the future. It can only be used prospectively as an estimator of cohort fertility by substituting annual intentions data for completed cohort fertility. Its performance in that mode has not been evaluated.

Demographic translation has the unambiguous objective of converting between cohort and period formats, and so the criterion of success in each case is perfectly clear: when translating from cohort to period, observed period fertility is the yardstick, and real cohort outcomes are the criterion when translating in the other direction. Ryder's classic translation procedure approximates more closely than the period TFR to cohort values over the medium term, but is nevertheless quite inaccurate. Indeed, Ryder (1964) acknowledged that its empirical performance is severely limited by the mathematical approximations required (see also Murphy 1993; van Imhoff 2001).

3.3.2 Future fertility

Beyond explicit attempts to estimate cohort levels of reproduction, period fertility may be the vehicle for a more general discussion of fertility prospects—either indirectly or, if directly, with a fairly inexact time reference. Such discussion can involve measurement to varying degrees. At one extreme, analyses, arguments and views may be offered on possible futures, without any special attention to the choice of indicator. At the other, the subject may be addressed via an approach to measurement that implicitly but clearly reflects a reference to future reproductive performance. A broad reference to the future is found in the discourse of fertility definitions and measurement in a variety of ways. It appears to be what e.g. van Imhoff (2001) 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 expressions that refer to the completed fertility "implied by" current rates. The concepts of "true" or "underlying" level of fertility could also be argued to refer to longer-run fertility levels in some non-specific sense (at least on one reading—another will be considered in a later section). These terms are never applied to simple incidence or occurrence-exposure rates, but carry the connotation of a mean family size, or parity progression ratio, or other synthetic quantity. In contrasting a measured with a true or underlying quantity, they implicitly draw a distinction between the temporary fertility conditions of the current period and longer term fertility levels. Also the real as opposed to synthetic realization of such quantum indicators occurs over long stretches of time rather than in individual periods.

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. Bongaarts and Feeney suggest, for example, that the adjusted TFR gives "a better answer to the question of how many births women will have if current childbearing behaviour continues into the future" (Bongaarts and Feeney 1998: 285). While certainly based on the rates of a single period, and in that sense a pure period measure, it can be argued that assumptions about future fertility are inherent to the index, as of all synthetic cohort measures that are interpreted as such. This also appears to be a reasonable inference from the fact that tempo adjustment has been understood, in practice, as carrying implications for long-run trends in fertility (Lesthaeghe and Willems 1999; Morgan and King 2001; Kohler and Ortega 2002a; Bongaarts 2002). The upward correction to recent period TFRs resulting from tempo adjustment has been interpreted as implying that period fertility rates are temporarily low and hence as implicitly predicting a future recovery in fertility, though doubts have been expressed as to the likelihood or extent of such recuperation (Lesthaeghe and Willems 1999; Frejka and Calot 2001; Sobotka 2004a).

If the objective is to get some idea of longer-run mean family size, adjusting for timing change can be justified on the same pragmatic grounds as apply to estimating cohort fertility. The record suggests that declines or rises in period rates associated with changes in the age at childbearing are often though not always compensated for in subsequent years. If grounds can be found for interpreting a short-run trend as primarily a timing change, adjusting for timing effects may be, in principle, a reasonable procedure to adopt in attempting to get closer to longer-run quantum estimates. However, as in the cohort case, such adjustment can be considered a type of forecasting and methods need to be evaluated for their success in this respect.

While the appropriate criterion for evaluating period estimators of cohort parameters is clear-cut, the task of evaluating period measures with an extended time reference but conceived in other terms—measures of longer run fertility levels, of underlying completed fertility levels, or of period quantum—is more complex. An index intended to reflect longer-run fertility levels in some sense can be evaluated by some kind of longer-run averaging procedure (see e.g. the criterion used by Bongaarts and Feeney 1998). Both the ideas of underlying fertility and of fertility quantum appear to be ill-defined as empirical entities in a period context—as noted earlier, there seems to be little agreement as to what quantum means in relation to a period though the idea is perfectly clear when applied to a cohort. As a result, it is not obvious how we could assess whether an indicator proposed as a measure either of underlying fertility, or of period quantum, is valid, and successful in fulfilling its intended purpose. The meaning of such concepts could be considerably clarified by specifying empirical criteria against which to evaluate them.

3.3.3 Projection

Projection or forecasting is the final way of anticipating the future. As a source of information on either a generalized notion of current and future fertility levels or cohort fertility in particular, projection or forecasting has several merits. 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 must be made explicit. In addition, where cohort fertility is the target, the accumulated fertility experience of incomplete cohorts up to the base year of the projection can be incorporated into estimates of their future completed fertility. Treating the anticipation of future fertility as a forecasting problem, rather than as something to be inferred from current period rates, appears both more transparent and empirically more realistic (on this point see also Lesthaeghe and Willems 1999 and Schoen 2004).

Improving the accuracy of fertility forecasts is a continuing concern for applied demography. A potentially powerful approach that has been pursued less actively than it might be is the projection of parity progression ratios on either a period or cohort basis. Feeney (1985) appears to have been the first to propose this strategy and some useful findings have been reported by Toulemon and Mazuy (2001) and by Sobotka (2004). Kohler and Ortega (2002a) combine this approach with tempo adjustment, using their adjusted parity progression measures as a basis for projection under several scenarios. Again, these need to be evaluated for forecast accuracy.

The criterion against which fertility projections should be evaluated is unambiguous—fertility out-turn, whether cohort or period (again a form of criterion validity). Nevertheless, assessing the comparative performance of fertility projections over time is potentially complex, requiring analysis not only by initial date, and duration of the projection, but also by period (see Keilman's (1990) analysis of forecast errors in a framework equivalent to an age-period-cohort analysis). Fertility projection has had limited success in low fertility societies and there is little evidence that fertility projections based on a cohort approach are any more accurate than those based on period lines (Booth 2006; Keilman 1990; Keilman and Pham 2004; Lee 2004). In addition to evaluation for forecast accuracy, fertility projections could usefully be compared with other methods—adjustment, translation—of anticipating future fertility.

3.3.4 Future population growth prospects

A further role for period fertility measures is to gain some idea of future population growth prospects. The customary figure of a TFR of 2.1 as reflecting replacement level fertility in low mortality populations reflects this usage. Reliance on the TFR for this kind of purpose is understandable in the context of intense concern in the post-war period initially on future prospects for high fertility populations, and more recently population prospects in low fertility societies. Nevertheless, it is well known that a single year's TFR is no indication of future growth prospects, and that the TFR can be below replacement for decades without resulting in population decline (Smallwood and Chamberlain 2005). The idea of a replacement level TFR assumes, of course, stable conditions just like its predecessor, the NRR. And despite the severe criticisms to which the NRR and other reproduction rates have been subject (Stolnitz and Ryder 1949, 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. But it is worth recalling Hajnal's (1947: 162) comment that to get an idea of long run population prospects "(t)here can be no mechanical formula which can be applied year in, year out." He was arguing against the NRR as an indicator, but the point applies equally to the period TFR in all its forms—additive, multiplicative, adjusted or unadjusted for tempo. Hajnal's solution is also still relevant—that 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..

Validation of the 2.1 figure in low mortality populations, or of analogous figures in higher mortality contexts, is in one sense superfluous—the figure is a theoretical one that follows necessarily from assuming a closed population subject to stable rates. However, as deployed empirically, validation could be said to be necessary: how often and for how long have real populations had fertility below replacement level while nevertheless continuing to grow, and how frequently has the reverse occurred? Again, criterion validity would be at issue, with population growth at selected time points from a base year being the outcome measure. A focus on how often the 2.1 and analogous figures have misrepresented population growth trends might stimulate a more sophisticated approach to population reproductivity.

3.4 Formal models

Beyond description, understanding, and anticipating the future, a further way in which indices of period fertility can be deployed is theoretical. This is really not a measurement activity at all, but a form of population modelling. The representation of the TFR as the mean family size of a hypothetical cohort is an instance. In a theoretical population with fixed age specific rates, it estimates cohort fertility without bias. Under changing tempo, however, the conventional TFR is biased as an measure of cohort fertility, in the sense that it is not equivalent to the cohort mean family size in a theoretical population subject indefinitely to the rates, and to the timing shift, of a given period. In a theoretical context, procedures to adjust for tempo change can be designed around the particular type of tempo shift assumed to operate. These are clearly appropriate if the objective is to estimate theoretical cohort fertility. This could be an alternative way of construing the Bongaarts and Feeney adjusted TFR—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—and is the preferred interpretation of Zeng and Land (2001) and of Rodriguez (2006).

It can be argued that the concept widely used in demography of the mean family size "implied by" the rates of a given period is essentially of this type: a theoretical construct rather than an empirical measure. Clearly nothing in the real world is literally implied, in a logical sense, by the rates of a single period, since real world populations are neither stationary nor, for the most part, stable. However, we might concede that orders of magnitude are probably implicit in period rates—women who are of childbearing age in a period when the TFR is, say, 6 are extremely unlikely to have a mean family size of 2, for example, and vice versa. But estimates at the degree of resolution that is usually sought in attempts to refine period fertility measures cannot be considered to be logically implied by the rates of a single period, except in a theoretical context.

Evaluating indices of fertility that are defined within a theoretical population model is essentially a matter of checking mathematical derivations, and possibly also the theoretical coherence and utility of a particular hypothetical measure. No empirical criterion is relevant for evaluating a measure construed as reflecting a hypothetical entity within a population model. An empirical yardstick becomes necessary only when a theoretical specification is regarded as measuring a real world process.

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. The present paper would be incomplete without a discussion of the role of such communication vis a vis measurement issues. This is because academic demography has strong ties with the official agencies that produce the population statistics that are the raw material of population science. Such links are mutually beneficial, but they may place an unwelcome, if mostly unnoticed, constraint on how academic demographers think about measurement issues. In Brass's (1990: 455) view, the practical need to communicate population trends to non-specialists has resulted in a preference in demography for single figure summary indices of fertility, such as the TFR[14]. He thought that meeting this practical demand had resulted in a neglect of more complex approaches using multiple indicators that are potentially technically superior.

The way we justify our measures may also be influenced by this need to communicate to others outside the profession. To evaluate fertility indicators for communication purposes, the natural criteria are how easy they are to produce and how accurately the indicators in question can be interpreted for and by non-specialist audiences. The TFR has clear advantages in that respect. And while intelligibility is a virtue in a measure intended for popular dissemination, it should not influence our choices for scientific purposes. We can be fairly sure, for example, that physicists and astronomers do not

[14] Brass is worth quoting at length on this subject. Controversy regarding cohort and period measures of fertility were, he suggested "a consequence of the desire to describe a complex, multiple component phenomenon by a single index which expresses the significant features, particularly of the most recent tendencies. 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: 455).

justify their measurement procedures by whether lay people can understand them. Nor should we. In all, user-friendliness is an essential feature in a measure intended for a general audience, but is a low priority when deciding on measures for technical and scientific analysis.

The period TFR is the leading fertility indicator published by statistical agencies in most developed countries and, as a single figure summary, appears reasonably adequate for the purpose[15]. The crude birth rate or the general fertility rate or any of a number of other indices could serve the same purpose and do so in more restricted vital registration systems. Are tempo adjusted measures needed in this more popular context? The view taken here is that they are unnecessary, since changes in timing can be conveyed by reporting time trends in the mean or median age at birth or, for preference, at first birth. One argument for doing so is that an unadjusted period synthetic indicator misleads the public about their likely future lifetime experience. That argument, however, relies on presenting and interpreting period synthetic measures as implicit forecasts, and on the assumption that adjusting them for timing effects would improve the forecast. But period measures should not be interpreted or presented as a prediction of actual or likely future experience. If this is what a statistical office wishes to convey, explicit forecasts would be a more appropriate vehicle.

In a policy setting, non-specialist users may often not be primarily interested in an annual fertility index *per se*. Rather, medium to long-run broad population prospects may be of

[15] Brass (1990: 456) suggests that the "widespread adoption" of the TFR was due to "[s]implicity, convenience and propaganda ."

greater pertinence for general policy and public information purposes and these depend not only on fertility but also on age structure, mortality and migration. National policy decisions on e.g. planning facilities and services, maternity provision, family policy, immigration, or pension provision require information going far beyond recent trends in a summary fertility index—projections of annual birth numbers, of cohort fertility, of population size and structure and so on. For this reason, an unrefined index of fertility may be quite adequate for most non-technical users' purposes, since it will usually be supplemented by other kinds of data relevant to future prospects, at least in a policy context.

4. Discussion

Several arguments relating to recent debates on fertility measurement are advanced in this paper. First, there are a variety of reasons for measuring period fertility, and not all measures are suited to each purpose. Period measures are relative to the purpose in hand. Recent debates on fertility measurement have tended to overlook the multiple objectives for which period fertility measures can be employed. Second, a major distinction is between period measures intended, on the one hand, to represent period fertility as dependent variable, and, on the other, to anticipate either longer run fertility levels, whether of cohorts or in a more diffuse sense, or population prospects themselves. Third, the validity of indicators of each type needs to be assessed, and clarity is required on the appropriate criterion in each case. Indicators based on concepts that lack a convincing direct or indirect criterion, or a systematic validation procedure, may be of limited value as measures or predictors of empirical conditions.

The monitoring and measurement of fertility trends mostly has the explicit or implicit purpose of attempting to divine the future, usually by approximating long run mean family size in some sense. This is a perfectly valid aim and is not surprising, given the applied orientation of much demographic activity. One consequence is, however, that measurement and forecasting are conflated in demographic thinking about fertility. What we think of as measurement is often really a form of forecasting and merging these distinct activities results in unnecessary confusion. Evaluating how best to measure period fertility as dependent variable, and how to analyze the factors associated with aggregate time trends in a statistically systematic way, is an important research objective. It, and the measurement effort it entails, is a distinct problem from that of assessing future fertility prospects, whether on a cohort or period basis, or future population growth. But period synthetic indicators effectively roll the two into one. Acknowledging the double duty that synthetic fertility indicators are asked to carry, however implicitly, could liberate ideas about period fertility measures. It could stimulate measurement approaches in which, on the one hand, period measures are expected to reflect no more than what occurs in a period and, on the other, indicators intended for forecasting need not be confined to a single period's fertility. A measure summarizing one period's rates, however transformed, is a slim basis for predicting the future. Fertility time trends do not follow any known deterministic law. If we want to estimate future completed fertility, it appears short-sighted to ignore the accumulated fertility of the recent past, whether in period or cohort form—but that is what synthetic cohort measures do, even when adjusted for shifting tempo. Attempts to anticipate future trends would be better served by an empirical search, by means of statistical or econometric models, for systematic relationships between the fertility performance of the past—whether in period or cohort format—and that of the future.

Another consequence of demography's future orientation is that although commentary on past fertility trends is extensive and although many explanatory schemes have been put forward (Morgan and Taylor 2006), a statistically systematic approach to explaining aggregate trends in low fertility societies has been neglected in demography, and with it the specification of measures of period fertility as dependent variable. Further possible reasons for this neglect include the limited success of attempts to model time series in the past (Lee 2004), the shortcomings of conventional micro-economic theory which tends to inform such work—inappropriately, since it attempts to explain an aggregate phenomenon in terms of micro-level theory (Murphy 1992)—the neglect of theory better suited to the realities of fertility variation, especially parity-specificity (Leibenstein 1974; Namboodiri 1981) and the recent dominance of micro-level analytical approaches.

The sense in which the terms bias and distortion are used in the literature on adjustment is not always clear. Two different concepts of bias or distortion have been distinguished in the present paper. Measurement bias is present when the indicator chosen gives a systematically mistaken reading of, or misrepresents, the phenomenon in question. The claim that the TFR is biased when timing is changing appears to be rooted in this concept, and can be interpreted in three possible ways. One is that a period measure influenced by timing effects is by definition biased. A second reading is that the TFR is biased relative to real cohort values, and a third, relative to theoretical cohort values. In the second of these, bias is present whether timing is changing or not, and in the third, only when timing is changing. Measures of period fertility as dependent variable are not designed as indicators of cohort values, but can be biased in a further sense, by the presence of nuisance factors, if inadequately standardized. The paper has argued that, provided indices are appropriately standardized, timing change is not a source of bias or distortion, but is integral to period fertility as dependent variable—part of what we should be trying to explain.

The ubiquity in demography of some version of the period TFR is, for all its faults, neither accidental nor arbitrary. It is, after all, the counterpart of cohort mean family size, which ultimately reflects population reproductivity. It is "good value" as an indicator, being standardized for *the* major demographic influence on vital rates, while requiring relatively simple inputs. It gives an excellent idea of crude orders of magnitude and so is useful for broad-based comparisons. It is unexceptional in a theoretical context. Finally, it has an attractively meaningful interpretation for individuals, and so is a successful device in communicating with general audiences. That it is easily understood by non-specialists is often cited as an advantage, but this has no bearing on its merits as a scientific tool. Where, as currently in developed societies, fertility response through time is distinctively articulated across age and life-stages, the crudity of the TFR—at least in its role as dependent variable—needs to give way to detailed measures that are more convincing, and more defensible, in statistical and behavioural terms. Besides, the TFR performs poorly as a forecaster in real-world situations where precision matters (period parity progression ratios may well do better, but have not been evaluated for forecasting purposes). If the question is whether long-term mean family size will be closer to e.g. 1.8 than to 2.1 or 1.6 vs. 1.3, we require a demographic technology that goes beyond estimating out-turn by the assumption that this year's rates are fixed. Redefining the question at issue as a forecasting rather than a measurement problem could be a productive way forward (see also Lesthaeghe and Willems 1999; Schoen 2004).

The synthetic cohort device is deeply engrained in demographic thinking. Ironically, it is called on precisely because the hypothesis on which it is based—stable rates—is rarely valid. Period fertility fluctuates, sometimes sharply, and the period TFR cannot therefore be relied on to reflect long run completed family size. This has given rise to concepts such as period quantum and "underlying" or "true" fertility, or the total fertility "implied by" current rates, that essentially assert a contrast between the apparent and the real. Such terms are widely employed in fertility analysis and are probably subject to a process of reification (Wilson and Oeppen 2003). But they are ill-defined and lack a clear empirical reference. The view taken here is that, if meaningful at all, these constructs refer to longer-run fertility in some sense. Such concepts can be useful only if there is clarity about their intended status as theoretical vs empirical entities and as representational or pragmatic in form. Where intended as empirical concepts, a method needs to be specified by which the proposed indices can be validated.

Finally, there has recently been a decided lack of attention to the substantive processes underlying aggregate fertility change—the question of how decisions are made in personal time, and how these are influenced by historical change, both short and long-term. For example, the idea of postponement has been adopted widely in demography in the last decade or so to describe recent trends in fertility in developed societies. No clear behavioural model has been put forward to give substance to the concept, and little or no attention has been given to testing the implied behavioural model against alternatives (Ní Bhrolcháin and Toulemon 2005). Behavioural mechanisms have been overlooked despite the massive scale of the move to later childbearing, and although the forecast implicit in the postponement idea of a recuperation in fertility has been challenged (Lesthaeghe and Willems 1999; Frejka and Calot 2001). The behavioural underpinnings of fertility change need to come back on the agenda of fertility studies and with it research on the formal modelling of aggregate change. Arriving at an agreed effective approach to measuring period fertility and at an understanding of how and why fertility changes may well be interdependent.

5. Summary

The central arguments of the present paper are as follows.

1. Period fertility is measured and analyzed for a number of objectives: for descriptive purposes, to explain past trends, to predict or anticipate future fertility, as input to theoretical models, and to communicate with non-specialist audiences. Any proposal for an index of period fertility should specify which of these objectives it is intended to serve.
2. The properties desirable in a period fertility indicator depend on its intended role. An index of period fertility as explanandum need not, and should not be expected to, function as a predictor of future fertility trends. An index of fertility intended for scientific analysis need not be readily interpretable by non-specialist audiences.

Indicators chosen for popular communication need not be those preferred for detailed scientific purposes.

3. Period fertility measures can be established as useful, ultimately, only if they are validated by an external criterion. Explanatory success is the criterion appropriate when measuring period fertility as dependent variable. By contrast, approximation to cohort values, or to longer-run fertility outcomes in a more diffuse sense, is the proper yardstick when attempting to anticipate future fertility. Validation is a more tangled issue in the case of measures intended exclusively for descriptive purposes.
4. Demographers often interpret the question of what current fertility "is" to mean "what fertility is and is likely to be in the future." An answer is usually expected in terms of the mean number of children per woman.
5. Synthetic cohort measures embody this conceptual elision and, as routinely interpreted, function as implicit forecasts. Their conventional use and interpretation confuses measurement and prediction.
6. When regarded simply as statistical summaries, synthetic measures such as the total fertility rate are uncomplicated in interpretation and can be evaluated purely on their statistical, as distinct from their demographic, attributes.
7. Tempo effects unquestionably occur. Genuine timing effects that are not due to compositional factors are an essential part of period fertility as dependent variable. There appears to be no justification at present for removing real timing effects from indices of period fertility in its role as explanandum.

8. Adjusting a synthetic measure for tempo effects amounts to claiming to improve on the forecast implicit in the routine interpretation of such measures. However, tempo-adjustment has not yet been shown to forecast successfully.
9. Unlike its cohort counterpart, the idea of fertility "quantum" is not clearly defined in a period context. Indices can be specified that are plausibly describable in this way, but we still lack evidence that these are anything more than mathematical abstractions. Their utility could be established if they were shown to represent a real-world aspect of period fertility or to predict successfully. However, if their role is to predict some future parameter, they are not simultaneously measures of the present.
10. If we wish to estimate future or longer-run levels of fertility, we should project or forecast explicitly.
11. If the objective is to predict either cohort fertility or longer run fertility levels in a more general sense, *any* transformation of annual birth numbers and rates that improves the forecast is justified. Such transformations should not be confused with measures of current fertility, and they need not have the properties that a period fertility measure might be expected to have.
12. The notion of the mean family size "implied by" current rates has a precise meaning only in a theoretical population model. In an empirical context, orders of magnitude could be said to be implied, in a loose sense, by the rates of a period. However, that is not by logical implication but by accumulated empirical observation.

6. References

Ahlburg, D.A., W. Lutz, and J. Vaupel. 1998. “Ways to improve population forecasting,” pp. 191-198 in W. Lutz, J.W. Vaupel and D.A. Ahlburg (eds). *The Frontiers of Population Forecasting* (Supplement to *Population and Development Review*), New York: Population Council.

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

Andersson, Gunnar. 2000. “The impact of labour-force participation on childbearing behaviour: Pro-cyclical fertility in Sweden during the 1980s and the 1990s,” *European Journal of Population* 16: 293-333.

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.

Balk, B.M. 1995. “Axiomatic price index theory: a survey,” *International Statistical Review* 63: 69-93.

Bankole, A. and C.F. Westoff. 1998. “The consistency and validity of reproductive attitudes: Evidence from Morocco,” *Journal of Biosocial Science* 30: 439-455.

Billari, Francesco and Hans-Peter Kohler. 2004. “Patterns of low and lowest-low fertility in Europe,” *Population Studies* 58: 161-176.

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.

Booth, Heather. 2006. “Demographic forecasting: 1980 to 2005 in review,” *International Journal of Forecasting* 22: 547-581.

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.

Bulatao, Rodolfo A. 1981. “Values and disvalues of children in successive childbearing decisions,” *Demography* 18: 1-25.

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

Caldwell, John C. 1996. “Demography and social science,” *Population Studies* 50: 305-333.
Caldwell, John C. and Thomas Schindlmayr. 2003. “Explanations of the fertility crisis in modern societies: a search for commonalities,” *Population Studies* 57: 241-263.

Casterline, J.B., F. El-Zanaty, and L.O. El-Zeini. 2003. “Unmet need and unintended fertility: Longitudinal evidence from Upper Egypt,” *International Family Planning Perspectives* : 158-166.

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

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

Crimmins, E. 1993. “Demography: the past 30 years, the present, and the future,” *Demography* 30: 579-591.

Curtis, S.L. and C. Westoff. 1996. Intention to use contraceptives and subsequent contraceptive behavior in Morocco,” *Studies in Family Planning* 27: 239-250.

Dixon-Mueller, R. and A. Germain. 2007. “Fertility regulation and reproductive health in the millennium development goals: The search for a perfect indicator,” *American Journal of Public Health* 97: 45-51.

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.

Fisher, I. 1922. *The Making of Index Numbers*. Boston: Houghton Mifflin.

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.

Golini, Antonio. 1998. “How low can fertility be? An empirical exploration,” *Population and Development Review* 24: 59-73.

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”. Stockholm Research Reports in Demography No. 75, University of Stockholm.

Hoem, Jan M., Alexia Prskawetz, and Gerda Neyer. 2001. “Autonomy or conservative adjustment? The effect of public policies and educational attainment on third births in Austria, 1975-1996,” *Population Studies* 55: 249-261.

Isaac, Larry, Philips Cutright, Elton Jackson and William R. Kelly. 1982. “Period effects on race- and parity-specific birth probabilities of American women, 1917-1976: A new measure of fertility,” *Social Science Research* 11: 176-200.

Joyce, R. Kaestner and S. Korenman. 2002. “On the validity of retrospective assessments of pregnancy intention,” *Demography* 39: 99-213.

Keilman, Nico. 1990. “Analyzing ex-post observed errors in a series of population forecasts,” *Zeitschrift für Bevölkerungswissenschaft* 17: 411-432.

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

Keilman, Nico. 1997. “Ex-post errors in official population forecasts in industrialised countries,” *Journal of Official Statistics* 13: 245-277.

Keilman, Nico and Dinh Quang Pham. 2004. "Time series based errors and empirical errors in fertility forecasts in the Nordic countries," *International Statistical Review* 72: 5-18.

Keyfitz, N. 1982. "Can knowledge improve forecasts?" *Population and Development Review* 8: 729-251.

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.

Kirk, Dudley. 1996. "Demographic transition theory," *Population Studies* 50: 361-387.

Koenig, M., R. Acharya, S. Singh and T.K. Roy. 2006. "Do current measurement approaches underestimate levels of unwanted childbearing? Evidence from rural India," *Population Studies* 60: 243-256.

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, Ronald. D. 1980. "Aiming at a moving target. Period fertility and changing reproductive goals," *Population Studies* 34: 205-226.

Lee, Ronald. 2004. "Quantifying our ignorance: stochastic forecasts of population and public budgets," in Linda J. Waite (ed.) *Aging, Health, and Public Policy: Demographic and Economic Perspectives*, Supplement to *Population and Development Review*.

Leibenstein, Harvey. 1974. "An interpretation of the economic theory of fertility: promising path or blind alley?" *Journal of Economic Literature* 12: 457-479.

Leridon, Henri. 2006. "Demographic effects of the introduction of steroid contraception in developed countries," *Human Reproduction Update* 12: 603-616.

Lesthaeghe, Ron. 2001. “Postponement and recuperation: Recent fertility trends and forecasts in six Western European countries,” Interface Demography, Vrije Universiteit Brussel. IPD-WP 2001-1.

Lesthaeghe, Ron and Johan Surkyn. 1988. “Cultural dynamics and economic theories of fertility change,” *Population and Development Review* 14: 1-45.

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

Lieberson, S. and F.B. Lynn. 2002. “Barking up the wrong branch: scientific alternatives to the current model of sociological science”, *Annual Review of Sociology*, 28: 1-19.

Luce, R.D. 1996. “The ongoing dialog between empirical science and measurement theory,” *Journal of Mathematical Psychology* 40: 78-98.

Lutz, Wolfgang, Brian C. O’Neill and Sergei Scherbov. 2003. “Europe’s population at a turning point,” *Science* 299: 1991-1992.

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

Marston, C. and J. Cleland. 2003. “Do unintended pregnancies carried to term lead to adverse outcomes for mother and child? An assessment in five developing countries,” *Population Studies* 57: 77-93.

McDonald, Peter. 2006. “Low fertility and the state: The efficacy of policy,” *Population and Development Review* 32: 485-510.

Moffitt, R. 2003. “Causal analysis in population research: An economist's perspective,” *Population and Development Review* 29: 448-458.

Morgan, S. Philip. 2003. “Is low fertility a twenty-first-century demographic crisis?” *Demography* 40: 589-603.

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.

Morgan, S. Philip and Miles G. Taylor. 2006. “Low fertility at the turn of the twenty-first century,” *Annual Review of Sociology* 32: 375-399.

Murphy, Mike. 1992. “Economic models of fertility in post-war Britain – A conceptual and statistical re-interpretation,” *Population Studies* 46:235-258.

Murphy, Mike. 1993. “Time-series approaches to the analysis of fertility change,” in M. Ní Bhrolcháin (ed.) *New Perspectives on Fertility in Britain*. London: HMSO, 51-66.

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.

Namboodiri, N. Krishnan. 1981. "On factors affecting fertility at different stages in the reproduction history: an exercise in cohort analysis," *Social Forces* 59: 1114-1129.

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 and Laurent Toulemon. 2005. "Does postponement explain the trend to later childbearing in France?" *Vienna Yearbook of Population Research* 2005: 83-107.

Ortega, J.A. and H-P. Kohler. 2002. "Measuring low fertility: Rethinking demographic methods. MPIDR Working Paper WP 2002-001. (Max Planck Institute for Demographic Research, Rostock). URL: http://www.demogr.mpg.de.

Park, C.B. 1976. "Lifetime probability of additional births by age and parity for American women, 1935-1968: A new measurement of period fertility," *Demography* 13: 1-17.

Pressat, R. and C. Wilson. 1988. *The Dictionary of Demography*. Oxford: Blackwell.

Preston, S. 1993. "The contours of demography: estimates and projections," *Demography* 30: 593-606.

Preston, S., P. Heuveline, and M. Guillot. 2001. *Demography. Measuring and Modeling Population Processes.* Oxford, Blackwell.

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.

Rupp, A.A. and H.A. Pant. 2006. "Validity theory," *Encyclopedia of Measurement and Statistics*. Sage Publications. URL: http://www.sage-reference.com/statistics/Article_n471.html. Accessed November 2007.

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

Ryder, Norman. 1973. “A critique of the National Fertility Study,” *Demography* 10: 495-506.

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.

Ryder, Norman. 1990. “What is going to happen to American fertility,” *Population and Development Review* 16: 433-454.

Sanderson, W. 1998. “Knowledge can improve forecasts: a review of selected socioeconomic population projection models,” pp. 88-117 in W. Lutz, J.W. Vaupel and D.A. Ahlburg (eds). *The Frontiers of Population Forecasting* (Supplement to *Population and Development Review*), New York: Population Council.

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

Siegel, J.S. and D.A. Swanson (eds). 2004. *The Methods and Materials of Demography. 2nd edition*. London: Elsevier.

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

Smallwood, Steve. 2002. “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, Tomas. 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, Tomas. 2004a. “Is lowest-low fertility in Europe explained by the postponement of childbearing?” *Population and Development Review* 30: 195-220.

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

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

United Nations. 1983. *Manual X. Indirect Techniques for Demographic Estimation*. New York: United Nations.

United Nations. 2003. *Partnership and Reproductive Behaviour in Low-Fertility Countries.* ESA/P/WP.177 Revised version for the web. URL: ww.un.org/esa/population/publications/reprobehavior/partrepro.pdf. Accessed March 2005.

Van de Kaa, Dirk. 1996. "Anchored narratives: the story and findings of half a century of research into the determinants of fertility," *Population Studies* 50: 389-432.

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.

Waldorf, Brigitte and Pillsung Byun. 2005. "Meta-analysis of the effect of age structure on fertility," *Journal of Population Economics* 18: 15-40.

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.

Wilson, T. and P. Rees. 2005. "Recent developments in population projection methodology: a review," *Population, Space and Place* 11: 337-360.

Yang, Quanhe. 1994. "Determinants of the decline in parity progression ratios in China, 1979-1984," *International Family Planning Perspectives* 20: 100-107.

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.