Reconstructing Past Fertility Schedules for Hard-to-Estimate Groups from CPS Data

by

Carl P. Schmertmann, *Florida State Univ.* [schmertmann@fsu.edu]

C. Gray Swicegood, *Univ. of Illinois* [swice@uiuc.edu]

Frank D. Bean, *Univ. of California – Irvine* [fbean@uci.edu]

ABSTRACT

We develop methods that use combined June CPS samples to estimate time series of age-specific fertility rate (ASFR) schedules for subpopulations not identified in US vital statistics reports. We use a new model for the shape of the ASFR schedule, and new statistical methods for extracting fertility information from CPS data. These innovations allow estimation of plausible and useful time series of ASFRs over 1960-2000, even for relatively small groups. Our main goal is to produce age- and time-specific estimates of the fertility of Mexican-born and other immigrant women while they reside in the US. These rates are important for understanding the expected and potential contribution of immigrant fertility to future US population change. The methods that we develop are also applicable to many other subpopulations of interest.

Submitted for presentation at the annual meeting of the Population Association of America, Boston MA, 1-3 April 2004. This research was supported by grant #R01 HD38791 from the US National Institutes of Health.

This is a draft. Please do not cite or quote without the authors' permission.

Introduction

Demographers studying vital events need both numerators (event counts) and denominators (person-years at risk). Census or vital registration data are often inadequate for sociologically-oriented studies, because they lack one (or both) of these prerequisites. In the case of fertility research, analysts need to use specialized surveys if they wish to study childbearing behavior in socioeconomic groups for which numerators are not available from birth certificates (e.g., women of Irish ancestry), or groups for which population estimates are unreliable or nonexistent (e.g., women 30-34 born in South America).

In the case of US fertility, the next-best data source in such cases is usually the June fertility supplement to the Current Population Survey (CPS). Rindfuss, Morgan, and Offutt (1996), for example, used June CPS data in a study of historical trends in race and education-related differentials in US fertility. Morgan et al. (1999) found CPS data superior to vital statistics for studying racial differences in first birth timing. Swicegood, Morgan, and Rindfuss (1984) concluded that pooling samples from June surveys in separate years was a useful (and safe) strategy for expanding sample sizes.

In this paper we use June CPS data for a similar purpose – namely, estimating late twentieth-century (1960-2000) fertility patterns among immigrants living in the US. In addressing this specific estimation problem, we also present a new and very general approach to using pooled CPS samples that allows maximum use of the available fertility information. Our primary goal is to establish the feasibility of using fertility histories collected in June CPS data, which are retrospective and often incomplete, to reconstruct historical fertility schedules for groups not included in vital statistics or census reports.

Data

June CPS fertility supplements, taken at irregular intervals over 1986 to 2000, collected retrospective birth histories for women at or above childbearing ages. The age range of these women varied slightly between surveys, as did the completeness of the birth histories. In all surveys, women were asked their age, number of children ever born, and the month and year of their most recent birth (if any). Some June CPS surveys also contain data on the month and year of woman's first birth. The June 1995 survey was the most complete, with month and year data on births 1…4 for each woman (if they occurred) in addition to the month and year of last birth (birth L). We restrict our samples to CPS surveys that *also* collected information on mother's nativity, immigration history, and parents' places of birth; these surveys are from June 1986, 1988, 1994, 1995, 1998 and 2000.

Table 1: Counts of Women and Births in June CPS Fertility Supplements

	1986	1988	1994	1995	1998	2000	TOTAL
# women 15-64	32,481	29,531	31,850	47,410	27,024	26,629	**194,925**
# foreign-born	3,557	3,188	3,437	5,162	3,228	3,426	**21,998**
Which births recorded?	1,L	1,L	L	1234,L	L	L	
# Children Ever Born	44,754	40,325	39,690	79,903	33,387	32,903	**270,962**
# Undated Births	9,926	8,652	21,315	2,542	17,790	17,601	**77,826**
# Dated Births	34,828	31,673	18,375	77,361	15,597	15,302	**193,136**
# Dated US Births	33,629	31,008	17,992	73,484	15,267	14,909	**186,289**
pre 1960	78	4	0	7,743	0	0	**7,825**
1960-1969	5,495	3,431	50	15,749	2	0	**24,727**
1970-1979	14,180	12,069	2,391	17,876	780	348	**47,644**
1980-1989	13,876	15,504	7,643	19,891	4,589	3,511	**65,014**
1990-2000	0	0	7,908	12,225	9,896	11,050	**41,079**

Table 1 provides a summary of the six surveys used in our analysis. In the table births are identified as "dated" if the month and year of birth are known, and as "undated" otherwise. Dated births are further subdivided by whether or not they occurred in the US, and (if they occurred in the US) by decade. Together the June CPS fertility supplements contain fertility histories for nearly 200 thousand women aged 15-64. These women reported a total of approximately 271 thousand children ever born, with timing information available for over 193 thousand births (71%), but unavailable for another 78 thousand (29%).

Nonhomogeneous Poisson Model of Birth Timing

The primary difficulty in using CPS birth data for historical fertility estimation is that timing information is unavailable for some births. If a CPS survey collected dates for at most n births, then some of the births to women with parity $B > n$ will have missing dates. We propose to overcome this difficulty by constructing a continuous-time event history model in which even partial birth histories have well-defined likelihoods. The model allows us to use both parity and timing data when fitting the fertility model, and to combine information from multiple (June CPS) surveys that cover similar historical periods with different levels of detail about birth timing.

We assume that waiting times to births are drawn from a nonhomogeneous Poisson process, with age-specific hazard rate $f(a)$ and cumulative hazard $F(a)$. The term "nonhomogeneous" refers to the variability of rates across ages; all women at a given age are assumed to face an identical hazard. In such a model, $f(a)$ also equals the density of expected events at age a, $F(a)$ equals expected parity at age a, the number of births to a woman in any age interval (a,b) has a Poisson distribution with parameter $F(b)$-$F(a)$, and numbers of births in any pair of non-overlapping intervals are independent (Ross 1983:46-47). These are strong mathematical assumptions. The assumption that age *alone* determines fertility hazards is clearly simplistic. We relax this in the model below by also adding period effects, but we have omitted parity, immigration history, and many other factors that may belong in a model of fertility risks. We do not argue that our model is complete; we do assert that it is a useful device for summarizing the recent history of age-specific fertility patterns.

In the age-specific model both complete and partial timing histories have well defined likelihoods. Suppose, for example, that a CPS survey collects data on ages at first and last births only. Consider two women with these histories:

#1. Age 41. Two children, born at exact maternal ages 21 and 29
#2 Age 41. Four children, with the first born at age 21 and the last at age 29

The likelihood of the first woman's (complete) history is

$$e^{-F(21)} \cdot f(21) \cdot e^{-(F(29)-F(21))} \cdot f(29) \cdot e^{-(F(41)-F(29))}$$

where the two $f(\ )$ terms represent births at known times and the exponential terms represent the probabilities of observing zero births in each of the intervals (0,21), (21,29), and (29,41). The likelihood of the second woman's (incomplete) history is

$$e^{-F(21)} \cdot f(21) \cdot e^{-(F(29)-F(21))} \cdot \left[F(29) - F(21)\right]^2 / 2! \cdot f(29) \cdot e^{-(F(41)-F(29))}$$

where the multiplicative terms have the same meanings, except that the middle term now represents the probability of observing two (rather than zero) births over age interval (21,29). The log likelihoods for these women have intuitive forms:

(#1) $\quad - F(41) + \ln f(21) + \ln f(29)$

(#2) $\quad - F(41) + \ln f(21) + \ln f(29) + 2\ln\left[F(29) - F(21)\right] - \ln 2!$

The likelihood of a birth history is based on the density at any known ages of birth *f*, and on the fact that births over an interval (*a*,*b*) are Poisson distributed with parameter *F*(*b*)-*F*(*a*). Complete birth histories like #1 are special cases, in which all births have known times and all gaps therefore contain zero births.

Let us now establish a general system of notation for (possibly) incomplete histories and their likelihoods. Consider a CPS survey taken at exact time *y* that records each woman's age (*A*), parity (*B* for "births"), and exact maternal age (or equivalently, exact dates) at up to *n* births (ages $a_1 \ldots a_n$). Because low-parity women may not need all of the *n* available age variables, define dummy indicators δ_i that equal one if there is an age at birth recorded in a_i, and equal zero otherwise. The number of "dated" births is thus $D=\Sigma[\delta_i]$, with $D \leq n$. If parity exceeds *n*, then by convention a_n records age at last (rather than *n*th) birth, and ages for births $n \ldots B\text{-}1$ are omitted. All women therefore have *B*-*D* births during age interval $[a_{n-1}, a_n]$, but for many women all births are dated and *B*-*D*=0.

If fertility hazards *f*(*a*,*t*) vary with both age and time, then the log likelihood of an individual history is

Equation 1

$$\begin{aligned} L = & -\int_0^A f(x, y-A+x)\,dx \\ & + \sum_{i=1}^{n} \delta_i \ln f(a_i, y-A+a_i) \\ & + (B-D)\ln\left[\int_{a_{n-1}}^{a_n} f(x, y-A+x)\,dx\right] - \ln(B-D)! \end{aligned}$$

and the log likelihood for a sample of independent histories would be sum of individual log likelihoods with this form. The expression is more complex than those above because summations are over cohort paths through (*a*,*t*) pairs rather than over age only, but is has the same essential form – minus the cumulative hazard, plus the log fertility rates at all dated births, plus Poisson probability terms for undated births.

In practice, the CPS measures ages and times in discrete integer units. To simplify calculations while retaining the analytical advantages of continuous-time notation, we discretize ages and dates as described in Appendix A. For foreign-born women we use the reported date of immigration (typically recorded as a range of several calendar years) to censor any part of the reported birth histories that could have occurred outside the US.

Parametric Specification for *f*(*a*,*t*)

We use the constrained quadratic spline (QS) model described in Schmertmann (2003) to describe period ASFR schedules. The QS model has four graphical parameters: the age at which fertility first rises above zero (α), the age at which fertility reaches its peak level (*P*), the age at which fertility falls to half its peak level (*H*), and the peak level of fertility (*R*). We assume that these four parameters vary over time, so that the period ASFR schedule at time *t* is

Equation 2

$$\boxed{f(a,t) \;=\; f^{*}[\, a \;\mid\; R(t),\ \alpha(t),\ P(t),\ H(t)]}$$

where $f^{*}(\,)$ is a QS model schedule. For parameters R, α, P, and H we modeled smooth time series as

Equation 3

$$
\begin{aligned}
R(t) &= \mathbf{a}_R + \mathbf{b}_R\, t + \mathbf{c}_R\, t^2 + \mathbf{d}_R \left(t - 1970\right)_+^2 + \mathbf{e}_R \left(t - 1985\right)_+^2 \\
\alpha(t) &= \mathbf{a}_\alpha \\
P(t) &= \alpha(t) + 1 + \frac{20}{1 + \exp\left(\mathbf{a}_P + \mathbf{b}_P\, t + \mathbf{c}_P\, t^2\right)} \\
H(t) &= P(t) + 1 + \frac{20}{1 + \exp\left(\mathbf{a}_H + \mathbf{b}_H\, t + \mathbf{c}_H\, t^2\right)}
\end{aligned}
$$

The 12 time series parameters ($a_R \ldots c_H$) therefore generate a complete $f(a,t)$ surface via (Equation 3) and (Equation 2). The time series for level parameter R is a quadratic spline with knots at 1970 and 1985. We selected the two dates arbitrarily, but results appear to be insensitive to the exact number or placement of knots. Time series functions for P and H ensure that they fall in demographically appropriate age ranges (between 1 and 21 years above α and P, respectively); this slightly complicated specification avoids numerical instabilities, by ensuring that the search algorithm cannot wander into demographically impossible parameter spaces.

Our specification allows considerable variability in level parameter R over the period under study (1960-2000), somewhat less variability in P and H, and no variability in the initial age parameter α. Fitting results for contemporary Swedish fertility schedules in Schmertmann (2003) suggest that period fertility levels vary more rapidly than shape parameters, and that the specification in (Equation 3) can generate a variety of realistic $f(a,t)$ surfaces.

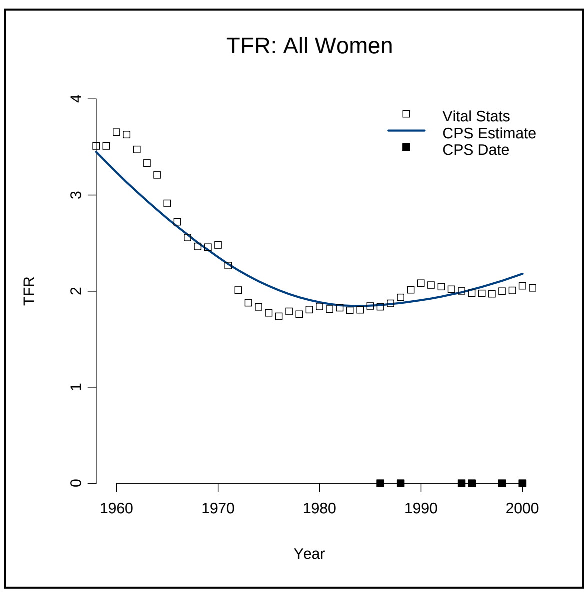
The model's core variables are the 12 time series parameters ($a_R \ldots c_H$). Fertility hazards affecting the likelihood of CPS sample histories are generated via Equations (Equation 3) and (Equation 2), and we use a nonlinear optimization program (specifically, PROC NLP in SAS) to find the time series parameters that maximize the log likelihood of a pooled sample of CPS fertility histories taken from June 1986 through June 2000. Our procedure therefore answers the following question:

> *What historical patterns in the level and shape of period fertility schedules best match the data on timing and number of births to women in different cohorts?*

CPS Estimates vs. Vital Statistics, for Subpopulations with Known Fertility

Our ultimate objective is to estimate fertility histories for subpopulations not covered by vital statistics. As an intermediate step toward that goal, we first compare results from the CPS model to vital statistics for groups whose fertility rates *are* included in official publications. We can be more confident about CPS estimates for "exotic" subpopulations if CPS estimates for "standard" subpopulations match available data from other sources.

Figure 1

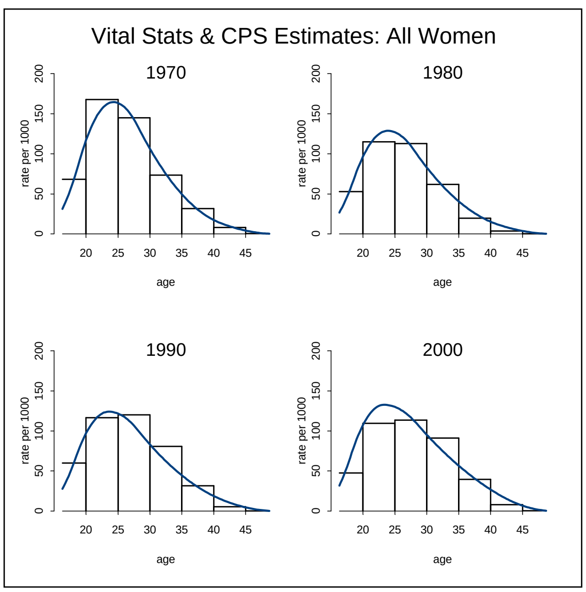


We begin by comparing vital statistics to CPS estimates for all women residing in the US. Figure 1 displays the time series of TFR from vital statistics (VS) as open squares (data from NCHS 2003, Table 4), and estimates from the pooled CPS histories as a solid line. CPS sample dates are marked on the horizontal axis with solid squares.

CPS estimates match the main trends in VS, particularly the fall in TFR over the 1960s, followed by a slow rise. The two time series are also quantitatively similar, with TFR near 3.5 at the start of the series and near replacement level at the end.

The overall impression in Figure 1 is that the CPS estimate is a smoothed version of the VS time trend. As such, it misses some interesting details in the VS series. Assuming VS data are correct, the CPS model time series underestimates fertility levels in the early 1960s, overestimates in the early 1970s, and misses the positive blip around 1990. Much of the discrepancy arises because VS data indicate a gradually rising TFR beginning in the mid-1970s, but the CPS-fitted TFR does not begin rising until the early 1980s. As stated earlier, the estimated CPS time trend is insensitive to the selection of knots for the quadratic spline time trend in the fertility level *R*. The difference in the timing of the fertility trough therefore appears to be a real feature of the CPS data, rather than an artifact of modeling decisions.

Figure 2



Because we are estimating complete $f(a,t)$ functions, rather than period TFRs, it is also important that the pooled CPS estimates yield plausible time series for age-specific fertility rates. Figure 2 and Figure 3 below provide information on age-specific model fits. Figure 2 displays both ${}_5f_x$ values from VS data (NCHS 2003) and the QS model schedules estimated from CPS data, for selected years. These data are for all US women.

Model schedules estimated from partial CPS histories match the empirical data well, but of course not perfectly. CPS fits are generally good, but they only partially capture the gradual shift toward births at higher maternal ages: in both 1990 and 2000, for example, the QS model appears to underestimate fertility in the 30-34 age range.

Figure 3

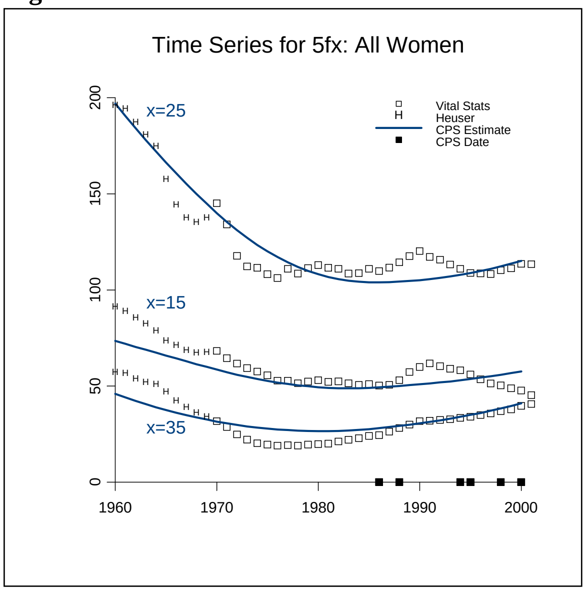
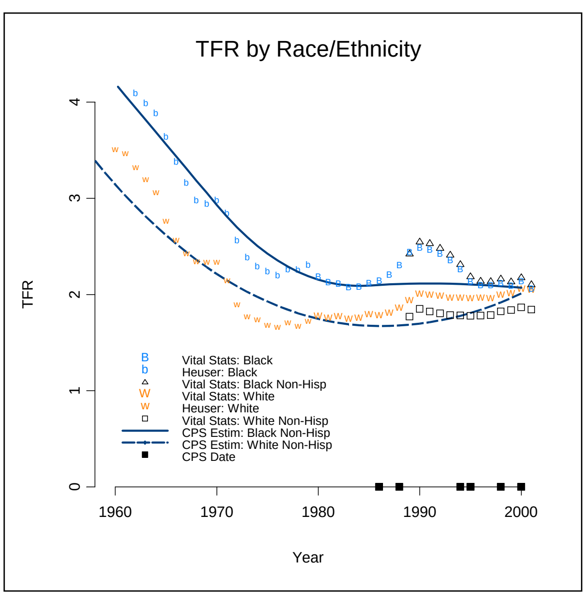


Figure 3 provides a more complete time series view, with slightly less age detail than Figure 2. Figure 3 shows time series for ${}_5f_{15}$, ${}_5f_{25}$, and ${}_5f_{35}$, over single years from 1960-2000. The blue lines represent QS model fits from CPS data, and the open squares represent VS data for 1970-2000 from NCHS (2003). Because NCHS (2003) data go back only to 1970, we have supplemented the vital statistics with period ${}_5f_x$ values calculated from Heuser (1976); these supplemental points are indicated by *H*s in the plot. As with earlier comparisons, Figure 3 shows that CPS model fits are generally very good, but hardly perfect. Rate differences between CPS estimates and other data sources tend to be small. The CPS model fits reflect the major time trends in age-specific rates, but they may also tend to smooth out some interesting year-to-year nuances.

We emphasize that our estimation method does *not* attempt to fit the histograms in Figure 2 (or the time series in Figure 1 and Figure 3) directly. Instead, the estimates in the figures correspond to the parametric time trends in period ASFR schedules that maximize the likelihood of dated and undated CPS births. The close correspondence with historical VS data suggests that CPS data are of high quality, and that our parametric specification is appropriate for describing historical trends.

We ultimately wish to estimate *f*(*a*,*t*) values for subpopulations, rather than for all women. To this end, we display in Figure 4 the CPS estimates of the TFR for Black Non-Hispanic and White Non-Hispanic women (top solid and bottom dashed lines, respectively). Vital statistics TFRs for these exact groups are included in the figure as triangles and squares for 1989-2000. Because disaggregation by Hispanic ethnicity is impossible in VS data before 1989, the plot includes supplemental information for race- (but not ethnicity-) specific TFRs for 1970-2000, supplemented by Heuser's race-specific time series for 1960-1969 (Heuser 1976). See the figure's key for details. Once again, the plot demonstrates that a model fit to pooled CPS data replicates known fertility patterns fairly accurately. CPS fits generally exhibit the same levels and patterns found in more complete data from other sources.

Figure 4



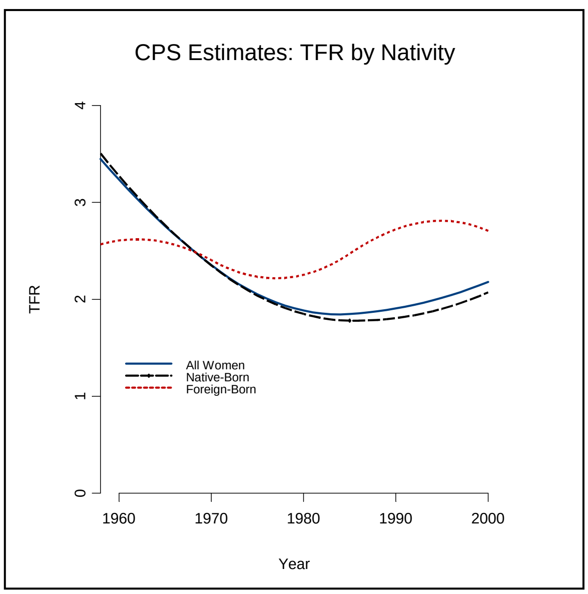
The overall conclusion that we draw from these comparisons is straightforward: fitting the relatively small (12-parameter) time series model to pooled CPS data from 1986-2000 produces a very good picture US fertility trends from 1960-2000. Estimated TFRs and age-specific rates are quite accurate, and the procedure also appears to capture trends within and differences between subpopulations. The model's ability to reproduce known time trends for familiar populations and subpopulations gives us confidence that it will also produce good estimates for the "exotic" subpopulations that we address next.

CPS Estimates for Immigrant Subpopulations

Fertility of Native-Born and Foreign-Born Women

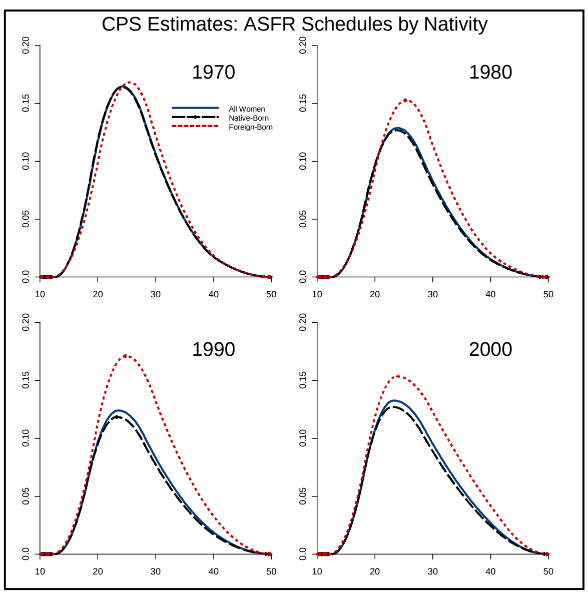
We now turn to CPS time series estimates for fertility among immigrants in the US. Data on the fertility rates of foreign-born women are fairly scarce even for recent periods, primarily because the Census Bureau does not produce the requisite denominators (intercensal population estimates by age, sex, ethnicity, and place of birth). Figure 5 reproduces the time series of TFR estimates for all women (solid blue line), and decomposes it by maternal place of birth (native-born women as a black dashed line, foreign-born as a red dotted line).

Figure 5



Several interesting features in the figure merit comment. The estimated total fertility of native-born women has tracked that of all women fairly closely, falling from Baby Boom highs near 3.5 in the late 1950s, and then rising slowly to near-replacement levels in the late 1990s. Estimated TFR among foreign-born US residents was less volatile, starting near 2.6 in 1960, reaching a low near 2.2 in the mid-1970s, peaking near 2.8 in the mid-1990s, and falling very slightly to approximately 2.7 in 2000. These estimates suggest that the presence of foreign-born women in the US probably *lowered* overall fertility slightly during the late Baby Boom years, but has *raised* overall fertility since the late 1960s. These effects are visible in the divergence between the time series for *Native-Born* and *All* women. The positive effect of foreigners' fertility on US TFR increased after 1970 not only because of an increasing gap between foreign-born and native TFR, but also because of the increasing proportion of foreigners in the US population (Schmidley 2001, Figure 1-1).

Figure 6



CPS estimates also allow investigation of immigration's contribution to age-specific fertility rates. Figure 6 provides examples, by depicting the estimated fertility schedules for native-born, foreign-born, and all US women in four selected years from 1970 to 2000. (Remember from Figure 5 that by 1970 foreign-born TFR was already slightly higher than native TFR). As in Figure 5, one can see the increasing influence of foreign-born fertility on US averages. Figure 6 shows clearly that the estimated foreign-born fertility schedule was slightly "older" than the native-born schedule in 1970: the mean age of childbearing from the 1970 estimated schedules is 26.9 for natives, and 27.4 for foreigners. Slightly later childbearing by foreigners makes intuitive sense, because of possible delays in other life cycle events (esp. marriage) associated with a move to the US as a young adult. It is difficult to compare the shapes of the native and foreign-born schedules once their levels after 1970, but the difference in mean age persists and even grows slightly larger by 2000 (when estimated mean age equals 27.4 for natives and 28.3 for foreigners).

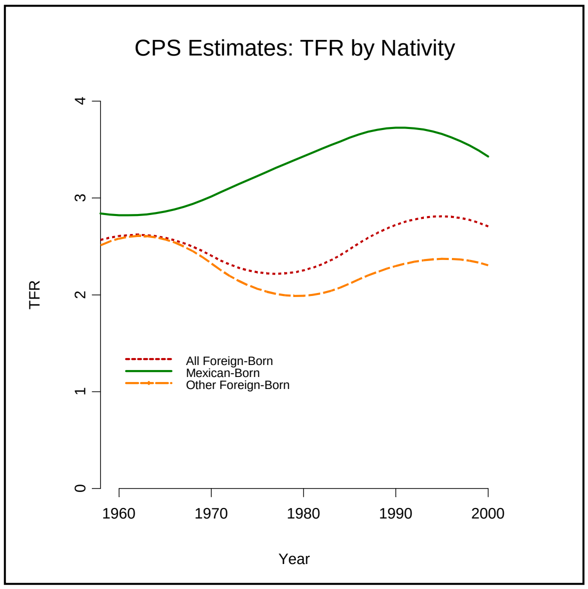
It is well known that changes in birth timing affect period TFR levels, even when the lifetime average number of births per woman remains constant (Ryder 1964). Interestingly, CPS estimates suggest that the increasing fertility contribution of foreign-born mothers has caused the US mean age of childbearing to increase slightly more rapidly than it would have otherwise. Timing changes would therefore tend to depress period TFR for all three groups in Figure 5. We emphasize, however, that shape changes in Figure 6 are small, as are any effects on period TFR.

Fertility of Mexican-Born and Other Foreign-Born Women

The composition, as well as the size, of the US foreign-born population changed significantly over the 1960-2000 period covered by the pooled CPS estimates. In 1960 approximately 85% of US foreign-born residents were European or Canadian, 9% were Latin American, and 5% were Asian. By 2000 the figures were 18% European or Canadian, 51% Latin American, and 26% Asian. (Schmidley 2001, Figure 2-2). More than one quarter of the foreign-born population in 2000 came from a single country: Mexico. (Schmidley 2001: 12). Among the more narrowly defined population of foreign-born women 15-49, the place-of-origin distribution probably changed even more dramatically over 1960-2000.

The large changes in the composition of the foreign-born population complicate the interpretation of time series such as those in Figure 5. For example, it is probable that foreign-born TFR was below native TFR in the 1960s because immigrant women of childbearing age were mainly from European countries that did not have sustained, US-style Baby Booms after World War II. The estimated increase in foreign-born TFR from the mid-1970s to the early 1990s may be due to the increasing proportion of immigrant women from relatively high-fertility Latin American countries (esp. Mexico).

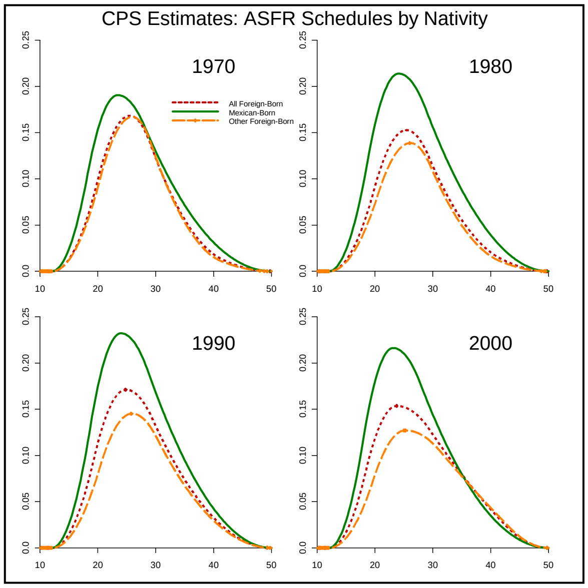
Figure 7



In this subsection we illustrate the potential of the CPS model to address such compositional questions, by further disaggregating the foreign-born population into "Mexican-born" and "Other". Figure 7 shows the time series of estimated TFR for all foreign-born women (identical to that in Figure 5), as well as the estimated TFR series for Mexican-born and Other foreign-born residents. The figure illustrates that the fertility level of Mexican women in the US has been high and, for much of the last 40 years, climbing. Estimated Mexican-born TFR was near 2.8 in 1960, climbed to near 3.7 in 1990, and has fell slightly to near 3.4 in 2000. The fertility of other immigrant women remained fairly stable over 1960-2000, starting near 2.6, falling to a minimum near 2.0 in the late 1970s, and rising again to approximately 2.3 in 2000. As expected, Mexican women exerted a steadily increasing influence on the fertility levels of the foreign-born;

this can be seen in the way the "all foreign-born" curve has increasingly been pulled toward that for Mexican-born women.

Figure 8



We can also decompose the age patterns of foreign-born fertility into "Mexican" and "Other" components, as illustrated in Figure 8. This figure depicts estimated ASFR schedules for four selected years, once again using the groups in Figure 7. We note that the Mexican-born fertility schedule has been consistently "younger" than that for other foreign-born women. Although it is not especially easy to perceive in the graph, the age schedule for other foreign-born women has shifted rightward over time (similar to the shift for native-born and for all women). Interestingly, the combination of (1) a rightward-shifting schedule for "other", (2) a consistently younger schedule for Mexican-borns, and (3) an increasing proportion Mexican, has resulted in a fairly stable age pattern in the foreign-born ASFR curve.

Conclusion

We have attempted to demonstrate the feasibility of estimating historical age- and period-specific fertility rates for relatively small subpopulations from pooled CPS samples. We faced two main obstacles: small sample sizes and incomplete birth histories.

Pooling June CPS samples from different years creates samples that may contain several thousand women in categories of interest, such as "women born in Mexico". These subsamples are still rather small for estimating age- and period-specific fertility $f(a,t)$, even if one uses five-year increments of age and time. Our approach uses parametric models to fill the data breach. Specifically, we assume that period fertility schedules belong to the family of quadratic spline models described in Schmertmann (2003), and that period fertility schedules change smoothly over time. For each subpopulation that we study (all women, foreign-born women, Mexican-born women, etc.), our model summarizes fertility rates over a 41*x*41 grid of single-year (a,t) values by using 12 parameters to describe time paths for level and shape parameters of period ASFR schedules.

We deal with the second obstacle, incomplete birth histories, by modeling birth times as outcomes of a nonhomogeneous Poisson process, in which all women in a cohort face an identical schedule of age-specific fertility hazards. The Poisson model allows estimation of likelihoods for both complete and incomplete birth histories in a unified mathematical framework. It also allows use of parity data, which is a significant advantage over other approaches, particularly for CPS samples in which dates are available for only one or two of a woman's births.

The combination of the Poisson timing model with a relatively low-dimensional parametric description of $f(a,t)$ appears to produce good estimates from pooled CPS data. Estimated time trends match 1960-2000 vital statistics data well for all women, and produce good quantitative and qualitative descriptions of race-specific fertility trends. The model's success at fitting known fertility patterns suggests that using the same procedures with other groups (groups *not* covered by vital statistics) will provide useful new fertility data.

Our preliminary experiments with immigrant women are promising. We are able to produce comprehensive US fertility histories for several groups of interest, such as foreign-born and Mexican-born women. We are also able to assess the effects of those group's fertility patterns on overall US fertility over time. The work reported in this paper is mainly exploratory, but it is encouraging. In the future we hope to refine our methods further, and to produce a more complete catalog of immigrant fertility histories. This includes fertility estimates for native-born women with immigrant parents, Asian-born women, and many other groups.

We have concentrated on estimating the fertility of immigrants after arrival in the US, but the modeling approach that we have tested in this paper is quite general. It could be used for estimating fertility in any subpopulation of moderate size that can be identified in the CPS.

REFERENCES

RL Heuser, 1976. *Fertility tables for birth cohorts by color*. U.S. Department of Health, Education, and Welfare. DHEW Publication No. (HRA)76-1152. Data downloaded from http://opr.princeton.edu/archive/cpft .

SP Morgan, N Botev, R Chen, and J Huang, 1999. "White and Nonwhite Trends in First Birth Timing: Comparisons using Vital Registration and Current Population Surveys". *Population Research and Policy Review* 18(4):339-356.

National Center for Health Statistics (NCHS), 2003. *Births: Final Data for 2002*. National Vital Statistics Reports, Vol 52, No 10, 17 Dec 2003. Available at http://www.cdc.gov/nchs/data/nvsr/nvsr52/nvsr52_10.pdf

RR Rindfuss, SP Morgan, and K Offutt, 1996. "Education and the Changing Age Pattern of American Fertility: 1963-1989". *Demography* 33(3):277-290.

SM Ross, 1983. *Stochastic Processes*. Wiley & Sons, New York.

NB Ryder, 1964. "The Process of Demographic Translation". *Demography* 1(1):74-82.

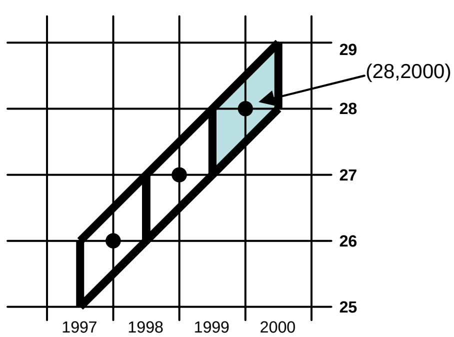
CP Schmertmann, 2003. "A System of Model Fertility Schedules with Graphically Intuitive Parameters". *Demographic Research* [online] Vol. 9, No. 5, pp. 81-110. Available at http://www.demographic-research.org/volumes/vol9/5/9-5.pdf.

AD Schmidley, 2001. *Profile of the Foreign-Born Population of the United States: 2000*. US Census Bureau, Current Population Reports Series P23-206. Available at http://www.census.gov/prod/2002pubs/p23-206.pdf .

CG Swicegood, SP Morgan, and RR Rindfuss, 1984. "Measurement and Replication: Evaluating the Consistency of Eight U.S. Fertility Surveys". *Demography* 21(1):19-33.

Appendix A

In this Appendix we briefly describe a version of the continuous-time, nonhomogeneous Poisson model that allows use of discrete, integer-valued age and time data. The main idea is to assume that fertility hazard functions are piecewise-constant, at levels corresponding to certain exact (a,t) combinations in the continuous model. Specifically, we assume that fertility hazards are constant within single-year period-cohort cells, indexed as in the figure. For example [28,2000] refers to the period [1 Jul 1999, 30 Jun 2000], and to individuals who are in age range [28,29] at the end of that period. We assume that the fertility hazard is constant everywhere within this [a,t] cell, at a level given by $f(a,t)$ from the continuous-time QS model. For the part of the likelihood calculations that involve undated births, we treat births in a cell [a,t] as if they all occur at exact age a and time t. This notation is sensible for data gathered in June surveys; the only awkwardness is that births occurring in the second half of a calendar year (e.g., Jul-Dec 1999) belong to the cell for the *next* calendar year (e.g. 2000).



In discrete notation (denoted by uppercase letters), the log likelihood of a history taken from a women in single-year age category *AGE* in June of calendar year *Y* with maternal birth ages $A_1 \ldots A_n$ is therefore

$$
\begin{aligned}
L = & -\sum_{x=0}^{AGE} f(x, Y-AGE+x) \\
& + \sum_{i=1}^{n} \delta_i \ln f(A_i, Y-AGE+A_i) \\
& + (B-D)\ln\Delta - \ln(B-D)!
\end{aligned}
$$

where Δ represents the difference in cumulative hazards between A_{n-1} and A_n :

$$
\Delta = \sum_{x=A_{n-1}}^{A_n} f(x, Y-AGE+x) - \frac{1}{2}\left[f(A_{n-1}, Y-AGE+A_{n-1}) + f(A_n, Y-AGE+A_n) \right]
$$

Appendix B

Appendix Table B1 reports time series parameter estimates and standard errors for the subpopulations discussed in the text. High standard errors for individual coefficients are not especially worrisome, because the time series "independent variables" (t, t^2, and the quadratic spline terms for R in Equation 3) are highly collinear. Bivariate correlations between the independent variables range from +.71 to +.98. Put another way, basis functions for the time series are not well designed for precise estimation of the separate coefficients. Our objective is to estimate the overall time trends, not the component parts.

Table B1: Time Series Parameter Estimates

Coefficient Estimates

		ALL	WNH	BNH	MEX	OFB	All FB
LEVEL [R]	a	23.9	24.3	23.5	18.6	16.9	17.5
	b	-7.0	-8.3	-4.3	-1.6	4.8	2.8
	c	0.7	1.3	-0.7	1.7	-4.1	-2.8
	d	1.0	0.1	3.1	-1.7	6.3	5.1
	e	-1.2	-0.4	-3.1	-1.9	-4.0	-4.7
START [alpha]	a	12.5	12.8	12.4	12.0	12.1	12.2
PEAK [P]	a	-0.25	-0.24	0.20	0.28	-0.43	-0.37
	b	0.09	0.04	0.40	-0.39	-0.19	-0.15
	c	-0.00	-0.01	-0.09	0.08	0.04	0.05
HALF [H]	a	0.71	0.81	-0.60	0.34	0.68	0.70
	b	0.05	0.12	0.39	-0.00	0.40	0.23
	c	-0.05	-0.06	-0.03	-0.00	-0.14	-0.10

Approximate Standard Errors

		ALL	WNH	BNH	MEX	OFB	All FB
LEVEL [R]	a	*4.7*	*5.9*	*5.8*	*8.0*	*5.7*	*4.9*
	b	*9.2*	*11.6*	*11.4*	*17.8*	*11.8*	*10.4*
	c	*4.4*	*5.5*	*5.5*	*9.1*	*5.9*	*5.2*
	d	*5.1*	*6.5*	*6.6*	*11.3*	*7.1*	*6.3*
	e	*2.1*	*2.9*	*3.5*	*6.2*	*3.3*	*3.0*
START [alpha]	a	*0.1*	*0.1*	*0.1*	*0.5*	*0.4*	*0.3*
PEAK [P]	a	*0.19*	*0.23*	*0.26*	*0.57*	*0.38*	*0.30*
	b	*0.19*	*0.26*	*0.25*	*0.48*	*0.38*	*0.28*
	c	*0.04*	*0.07*	*0.05*	*0.10*	*0.09*	*0.06*
HALF [H]	a	*0.46*	*0.52*	*0.97*	*1.05*	*0.78*	*0.63*
	b	*0.43*	*0.51*	*0.72*	*0.91*	*0.75*	*0.59*
	c	*0.09*	*0.11*	*0.13*	*0.18*	*0.17*	*0.13*

Notes:
1. Time variable t is scaled such that t=(year-1958)/10
2. Approximate standard errors from SAS Proc NLP are based on second derivatives evaluated at the final coefficient estimates.
3. ALL=all women, WNH=White Non-Hispanic, BNH=Black Non-Hispanic, MEX=Mexican-born, OFB=Other foreign-born, All FB= all foreign-born