Early entry to fatherhood

estimated from men's and women's survey reports in combination

Michael S. Rendall*, Ryan Admiraal**, Mark S. Handcock**, and Kara Joyner***

October 9, 2006

Paper for presentation at the National Survey of Family Growth Research Conference,

October 19-20, 2006

* RAND and Pennsylvania State University, email: mrendall@rand.org, ** Center for Statistics and the Social Sciences, University of Washington, *** Department of Policy Analysis and Management, Cornell University

Acknowledgements: This work was funded by grants by the National Institute of Child Health and Human Development under investigator grants R01-HD043472-01 and P01-HD045610, and under center grants to the RAND Labor and Population program (R24-HD050906), the Penn State University Population Research Institute (R24-HD41025), and to the University of Washington Center for Studies in Demography and Ecology (R24-HD41025). Address correspondence to the first author at mrendall@rand.org.

Abstract

While underreporting of fatherhood is a widely acknowledged problem, satisfactory methods for its correction have yet to be developed. In the present study, we investigate methods of correction that are specific to marital status at the time of the birth and at the time of retrospective reporting, focusing on fatherhood under age 30. Matched women's and men's survey reports of births, in each case reported by marital status and age of the father, form the basis for our corrections. Male age-specific fertility rates are estimated from these survey data by using women's reports for the births numerator and men's reports for the exposed-years denominator. These are shown to match well to male age-specific fertility rates estimated from population data sources. When marital births in the men's and women's survey data are differentiated by whether the birth is within a current or previous marriage, only for births in previous marriages is there a male reporting deficit. Further, this deficit is completely explained by under-representation of men's exposed years in previous marriages. We find no evidence of underreporting of births for those exposed years. These results are used to develop a constrained maximum likelihood estimator in which male fertility is constrained by age and marital status, with a focus on correcting for underreported non-marital fertility.

INTRODUCTION

While the early childbearing literature has concentrated on the age of the mother and subsequent disadvantaged outcomes for her and her children (e.g., Wu and Martinson 1993), the father's age and his outcomes have also received attention (Cooney, Pedersen, Indelicato, and Palkovitz 1993; Nock 1998; Marsiglio, Amato, Day, and Lamb 2000; Pears et al 2005). A problem to overcome, however, is that men's survey reports underestimate fatherhood overall (Garfinkel et al 1998; Elo, King, and Furstenberg 1999; Rendall et al 1999), and at younger ages in particular (Lindberg et al 1998; Martinez et al 2006; Rendall et al 2006).

Rendall et al (2006) compared NSFG men's reports to both NSFG women's reports of father's age, and to father's ages reported in the birth registration system. Their findings on the age pattern of NSFG men's underreporting of fatherhood were consistent between these two alternative sources: the younger the father, the greater the underreporting. These results are broadly consistent with patterns of underreporting in the NSFG 2002 found by Martinez et al (2006) when comparing to the birth registration system. Those authors found large underestimates for 15-19 year old men reporting births between 1997 and 2001. Martinez et al, however, found no significant differences among men aged 20 and above, while Rendall et al found the tendency towards underreporting extended through men's mid-20s. This discrepancy between the studies may be due to the reported fatherhood events being closer to the survey date in the Martinez et al study (1997-2001) than in Rendall et al's (1991-2000). In addition to more general recall problems with greater length of time between the survey and the retrospectively-reported event, births closer to survey date are more likely to be within unions that are still intact and with the children present in the household. Underreporting is unlikely under those conditions.

Higher survey underreporting at younger ages presents potentially serious problems for the analysis of the factors involved in early fatherhood. In an attempt to correct for this in a regression estimation of the male fertility hazard, Rendall et al (2006) used a constrained logistic regression model in which the constraints were age-specific male fertility rates derived from population data. A problem in making these age-specific corrections is that the correction factors are assumed to apply equally to all men of that age. In violation of this assumption, Rendall et al (2006) found underreporting to be greater among black than non-black men, and found different age patterns of underreporting between black and nonblack men: for black men, underreporting was similar between births occurring in their late teens and early 20s and births occurring in their mid 20s; for nonblack men, underreporting decreased with age. As a result, applying a uniform age-specific correction resulted in over-correction of nonblack men's fertility especially in the mid-20s, and under-correction of black men's fertility, again especially in the mid-20s.

Previous work suggests higher rates of non-marital fertility may be an important factor explaining these differential rates of male underreporting. Further, underrepresentation of unmarried and previously-married men in surveys may account for a substantial proportion of the deficit in births when using men's reports (Rendall et al 1999). This motivates the extensions to previous work that are explored in the present study.

Specifically, we use women's versus men's survey reports of births by age of father and marital status to first analyze the sources of the overall age-specific male fertility deficits at younger ages, and second to develop correction methods that incorporate marital status in addition to age.

Our method for correcting for men's underreporting is based entirely on the matching of women's to men's survey reports of births and fertility exposure. Survey respondents' recollection of events, even important events, has been found to deteriorate as time elapses. However, the dates of rehearsed events, such as birth dates of children and marriage anniversaries, are less likely to be forgotten (Wu, Martin, and Long 2001). Comparing retrospective and contemporaneous reports of marital history, Peters (1988) found substantial agreement about the dates of events. Recent studies, however, suggest that respondents have difficulty recalling information about cohabiting relationships. Based on in-depth interviews, Manning and Smock (2005) found that because cohabitation is a gradual transition, many respondents had trouble identifying the start date of cohabiting relationships. Using Fragile Families data, Teitler and colleagues (2006) compared contemporaneous and retrospective reports of baseline cohabitation status (i.e., status at the time of birth) and found that a substantial number of respondents provide inconsistent reports, and typically "upgrade" their reports one year later to indicate they were cohabiting at the time of their birth.

DATA AND METHOD

Data

The survey data for both men and women are from the 2002 National Survey of Family Growth (NSFG), collected by the National Center for Health Statistics. The NSFG was designed to provide information on factors influencing fertility, union formation, and reproductive health based on a sample of U.S. men and women ages 15-45 (Lepkowski et al 2005). The 2002 NSFG is the first NSFG cycle to collect samples of both men (n = 4,928) and women (n = 7,643). These are independent samples, as only one member of any given household is interviewed. In addition to including nationally representative samples of men and women, the NSFG incorporated over-samples of blacks and Hispanics, as well as individuals aged 15 to 24. Importantly, the NSFG collected dates of birth for all biological children of men aged 15-44, and it asked women to provide information on the age of fathers at the time of birth. The NSFG used CAPI technology to collect fertility histories as part of the sexual partner histories. Research has shown that this strategy can improve the quality of male fertility data (Lindberg et al 1998).

Fertility data were collected differently for men and women in the NSFG. The survey instruments for women collected information on births separately from information about cohabitation and marriage. Specifically, women were asked about the dates that each pregnancy ended in a live birth, in addition to the beginning and end dates of all cohabitating relationships and marriages. Using these dates, the NSFG constructed a variable indicating the marital and cohabitation status of respondents at each birth. In contrast, the survey instruments for men collected male fertility data in the context of relationships in order to obtain higher reports of non-marital fertility (Lindberg et al 1998). For men who were married to the parent of their child, the NSFG asked whether they were married at the time of birth if it was not clear from the dates whether the marriage or birth came first. For men who were not married at the time of birth, and men who had cohabitated with the parent, the NSFG asked whether they were living with the parent at the time of birth. It is important to note that the technique of matching dates for women does not ensure that the cohabiting or marital partner is the father of the child. While we anticipate being able to overcome these problems in subsequent analyses that match cohabitation reports between men and women in the NSFG, we limit the analysis in the current study to one based on marital status alone.

Estimation

The goal of the methodological analyses of the present study is to be able to estimate a male fertility regression equation that adjusts for men's underreporting of fatherhood events (births). This regression equation uses the same basic data on men's fertility as used in Rendall et al (2006), consisting of men aged between 18 and 27 years old in 1991 to 2000. Because the NSFG is retrospective, the variables that can be identified as preceding fatherhood are more limited than for panel data. Explanatory variables in the regression may nevertheless potentially extend to men's reports of their parity, race and ethnicity, employment and income at the time of the survey, and family background variables including mother's education and family structure experienced over childhood.

While it is ultimately the goal to include as many of these variables as possible, the small subset of them used in the present study facilitates the prerequisite methodological development. Thus the only substantive variables additional to age and marital status are the man's race/ethnicity (black, Hispanic, and white/other) and parity (any previous children or not). Year dummies are also included as regressors, to control for both period trends and for differential underreporting with length of time between the survey and the reported event. As in Rendall et al (2006), men's reports are adjusted for underreporting by using constrained maximum likelihood estimation. The adjustment method of the present study, however, uses NSFG data from women to adjust for men's underreporting, while the former study used birth registration data. The main advantage of using the NSFG data is that more variables may then be matched between the men's reports and the women's reports. In particular, correcting men's reports using NSFG women's reports as a standard allows for marital status at the birth and at the survey to be used.

The substantive focus is on the associations between early fatherhood and variables that are not reported (for the father) by the women. Bias in the regression coefficients for these variables may be reduced when the fatherhood probabilities are corrected using variables that are correlated with them, although the magnitudes of such corrections based on indirect associations are likely to be very small (Handcock et al 2005). However, by adjusting the baseline male fertility hazard by age and marital status upwards for underreporting, the quantitative impact of regression coefficients for these variables, as measured by change in the predicted fatherhood probabilities, may increase substantially. That is, a given proportionate increase in the male fertility rate, as derived from the regression coefficient, will result in a greater absolute increase in the male fertility probability.

Estimation is restricted to fatherhood at ages 18 to 27 years old in the period 1991 to 2000. These age and period ranges are chosen for several reasons. First, it is expected that data problems will be greatest at younger ages, and therefore that more could be gained from using additional data and estimation procedures to correct these problems. Second, because of the sample restrictions in the NSFG (respondent ages capped at 45), fertility exposure over a greater range of ages is available closer to the 2002 survey year. This allows for women of a greater range of ages to report men's fertility between the ages of 18 and 27. Third, it is expected that both the relatively recent period of the 1990s and the younger fatherhood ages would be of particular research and policy interest.

A discrete-time hazard model with constraints, similar in its statistical structure to that used by Handcock, Rendall, and Cheadle (2005), is employed in the present study. But while they used population data for the constraints, the present study uses survey data. Women's reports are used to provide a constraining numerator of births to men by the father's age and her marital status (married or unmarried, and further whether the marriage is still intact) at the time of the birth, assuming that the father's marital status is the same as that of the mother's. Based on the results of prior studies cited above using the NSFG 2002 and other survey datasets, women's reports of these variables are expected to provide a more complete births numerator for fathers than are men's own reports.

The main statistical consequence of using survey instead of population data for the constraints is that the sampling variability of the constrained estimates will be larger when using survey data in the constraints than when using population data in the constraints. In the appendix, we describe the method for incorporating this additional sampling error. While the main purpose of the constraints here is bias reduction, because the female sample size is larger than the male sample size, there may also be efficiency gains to be realized through the use of the female data in the constrained estimation.

The response variable Y has two levels: 0 denotes no birth, and 1 denotes a birth, during the year $[t-1, t)$. We ignore the complication of multiple fatherhood events in a year, noting that these account for a very small percentage of all cases of at least one birth reported in a year (results not shown). The conditional expectation of this response variable may then be interpreted as a probability of fathering a child at a given single-year age. It is modeled with a logistic regression on covariates $X = \{X_1, X_{2,} ..., X_q\}$ for age, parity (0, 1+), marital status (married, unmarried), and race/ethnicity (black, Hispanic, white/other):

$$\text{logit}[P(Y=1 \mid X_1 = x_1, X_2 = x_2 ..., X_q = x_q, \theta_0)] = \sum_{k=1}^{q} \theta_{0k} x_k .$$

In addition to the main-effect regressor dummies, interactions of parity with marital status, and linear interactions of race/ethnicity dummies and parity with age, are included among the regressors.

The women's information on births, combined with the men's on exposure, together give single-year age-, race-, and marital-status-specific male fertility probabilities (further details below). These are introduced to the estimation as constraints. They are expressed as functions of the weighted conditional expectations of Y for a given age j, where the weights are given by the proportion of the sample aged j that has a given set of covariate values $X = x$. Let the i^{th} regressor be the man's age, and let population value ϕ_{ij} be the rate of giving birth *(Y=1)* in the population with age $X_i = j$. The corresponding constraint functions $\mathrm{C}_{\mathrm{ij}}(\theta)$ are each of the form:

$$\phi_{ij} \ = \ C_{ij}(\theta) \ = \ E_\theta[g_{ij}(Y,X)] \ = \ \sum_{x:x_i=j} P(Y=1 \mid X=x,\ \theta)\ \pi(X=x \mid \mathrm{X}_{\mathrm{i}}=j)$$

where

$$g_{ij}(y,x) = \begin{cases} 1 & y=1 \text{ and } x_i = j \\ 0 & \text{otherwise} \end{cases}$$

The constraint function is expressed as the sum over of two product terms (on the right hand side of the constraint function). The first term is the probability of a birth conditional on the value of the regressor vector x. The second term is the proportion π of the population with a specific set of values on the regressor variables given that age $X_i = j$. The effect of the constraints is to force the predicted probabilities of fatherhood for the NSFG sample men to equal those of the fatherhood probabilities estimated by combining the female and male data.

Eestimation of the Constraint

At each year t=1991-2000, a number *N(x,r,t)* of men of age x were married and the remaining *N(x,u,t)* were unmarried at that age and year. The number *N(x,r,t)* may be estimated alternately from the Male Respondent File and from the Female Respondent File, in each case using the 'FINALWGT' variable to weight to the US population.

Denote the weighted respective quantities $N^m(x,r,t)$ and $N^f(x,r,t)$. To keep sampling error as low as possible, the numbers are aggregated over the 10 years (1991-2000) to $N^m(x,r)$ and $N^f(x,r)$. Those married years may be divided into (1) those for a marriage that is the *current* marriage as at the 2002 survey date, $N(x,c)$; and (2) those for a *previous* marriage, $N(x,p)$ ---- that is, a marriage that ended before the 2002 survey date. These numbers can also be estimated alternately from the Male Respondent File and from the Female Respondent File. Superscripting these again with *m* for male reports and *f* for female reports, the quantities are $N^m(x,c)$ and $N^f(x,c)$ for births within current marriages and $N^m(x,p)$ and $N^f(x,p)$ for births in previous marriages.

Births are also reported by year, father's age, *and* the (joint) marital status of the father and mother at the time of the birth. Therefore it is also possible to estimate the same quantities from both men's and women's reports, and to additionally estimate a quantity for "unmarried" fathers from both men's and women's reports. Denote these quantities $B^m(x,r)$ and $B^f(x,r)$; $B^m(x,c)$ and $B^f(x,c)$; $B^m(x,p)$ and $B^f(x,p)$; and additionally $B^m(x,u)$ and $B^f(x,u)$. Note that no reports of men's unmarried fertility exposure are available from NSFG women, unless that exposure occurs within a cohabiting union. We discuss the ways that such information can be used at the end of the paper. Women's and men's reports are used to form the numerators and denominators of the age-specific marital and non-marital fertility constraints as follows:

$f^m(x,r) = B^f(x,r) \,/\, N^m(x,r)$ for marital fertility; and

$f^m(x,u) = B^f(x,u) \,/\, N^m(x,u)$ for non-marital fertility.

Constraints may similarly be estimated that distinguish fertility in current and previous marriages, as follows:

$f^m(x,c) = B^f(x,c) \,/\, N^m(x,c)$ for current marital fertility; and

$f^m(x,p) = B^f(x,p) \,/\, N^m(x,p)$ for non-marital fertility.

For all but non-marital fertility, women's reports may be match to men's reports of their fertility exposure. This allows for evaluation of the extent to which men's reports of births men's births can be attributed to under-representation (if N^f exceeds N^m), versus to under-reporting (if B^f exceeds B^m after accounting for any differences in N^f and N^m).

RESULTS

Our ultimate objective was to develop methods for correcting for men's fertility underreporting. To this end, we first used women's reports of birth by father's age to estimate a male fertility rate by single-year age from ages 18 to 27. We compared this both to a rate using a male-reported births numerator, and to Rendall et al's (2006) estimates based on population data (see Figure 1). Reassuringly, the fit of the sample estimates using women's reports was very close to that of the population male fertility rates. This match between survey and population sources provides the major empirical justification for proceeding further with use of the survey source to correct for the male reporting deficits, using women's reports of men's births.

[FIGURE 1 ABOUT HERE]

The remainder of our analyses use NSFG data only. We first compare men's and women's reports of births (Table 1) and fertility exposure (Table 2). In both tables, age refers to the man (or father). The main results are expressed in the "ratio" columns. Because the ages in each case refer to the men, the ratio should be 100 to the extent that men and women at the time of the survey are equally likely to be in the household (non-institutional) population.

[TABLES 1 AND 2 ABOUT HERE]

The ratios for births (Table 1) show that men's reports of nonmarital births at ages 15 to 29 are only two thirds (65 percent) of the nonmarital births reported by women for those same fathers' ages. Men's reports of marital births, meanwhile, are higher than those for married women (in a ratio of 109 to 100). When marital births are divided between those in current and previous marriages, however, men are seen to report births to previous marriages in a ratio of 93 births for every 100 births reported by women in previous marriages, while 116 births in current marriages are reported by men for every 100 births reported by women. While these results await statistical tests of the effects of sampling error, they are suggestive of over-representation of married men, or over-representation of married men with higher achieved fertility.

In Table 2, we find that married men are not under-represented overall, but that previously-married men are under-represented. All married years are reported in a ratio of 99 men's married years for every 100 women's married years. Some under-representation of men’s previously-married years is seen, where the ratio is 82 to 100. Because the ratio of previously married men to previously married women (82 per 100) is lower than the ratio of men’s births in previous marriages to women’s births in previous marriages (93 per 100), under-representation of men’s previously married years is sufficient to explain all of the previous-marriage men’s reporting deficit. Even adjusting for possible over-representation of married men with higher fertility, the ratio of ratios in previous-marriages and in current-marriages are similar between births and exposure: 93:116 and 82:99 respectively. Thus conditional on being sampled and responding to the NSFG, there is no evidence here that men under-report births in their previous marriages. For this reason, we do not adjust for any previous marriage underreporting in our regression analysis.

We use a simple non-marital/marital fertility distinction to develop methods to correct for the most obvious and serious problem (non-marital fertility underreporting). The non-marital fertility constraint developed by using women’s reports of non-marital births by father’s age for the numerator, and men’s reports of unmarried years for the denominator, is shown in Figure 2. To reduce the effects of sampling error in this “constraint” on male fertility regression estimates, we used a simple linear regression with age and age squared.

[FIGURE 2 ABOUT HERE]

We now use this constraint, and a similar marital fertility constraint, in a regression estimation of men's annual birth probabilities at ages 18 to 27 years old. As was noted in the method section, the main gains are expected to be in the correction of fertility levels. We therefore focus on the predicted probabilities from the regression estimates, and for non-marital fertility in particular. In figures 3 and 4, predicted birth probabilities are presented respectively for unmarried men with no previous children (reported) and for unmarried men with at least one child. In each case separate predictions are made for (non-Hispanic) white, black, and Hispanic men. Interestingly, only for first births (men at parity 0) are the there any differences by ethnic group. Linear interactions by ethnicity and ethnicity and parity were used in the regression.

[FIGURES 3 AND 4 ABOUT HERE]

There is some evidence in Figure 3 that a polynomial age interaction (age and age squared) might have produced a better fit to the data by ethnicity, as there appears to be a curvilinear relationship of fertility by age in the observed data for black men and, to a lesser extent, Hispanic men. The main effect of constraining to the women's reports of men's fertility, however, is a reasonably uniform lifting of the fertility probabilities by age over the 18 to 27 year old range.

SUMMARY AND DISCUSSION

In the present study, we took advantage of the NSFG women's reports of fertility by marital status and age of the father to first evaluate men's reports by marital status and age, and second to use these results to correct for deficiencies in men's reporting in a regression analysis. When births are differentiated in the survey source by the parents' marital status at the time of the birth, the difference between the male fertility rate using men's versus women's reports was found to be much larger for non-marital births, amounting to a third fewer births when reported by men than when reported by women. When marital births were differentiated by whether the birth is within a marriage that is still intact, the only men's marital reporting deficit is found to be due to births in previous marriages. These deficits of births in previous marriages are low, however, at around 10 percent only. Moreover, they are completely explained by under-representation of previously married men. Differential reporting of previous years in marriage by men and women cannot be discounted, however, and the results for both exposure and births await statistical testing for the effects of sampling error. At this stage, we conclude that there is as yet no evidence of underreporting of births within men's previous marriages among surveyed men.

The above results were then used to develop a constrained maximum likelihood estimator of male fertility that focused on addressing non-marital fertility underreporting. The marital-status-specific constraints (for the male non-marital and marital age-specific fertility rates) resulted in corrections that did not introduce any clear patterns of bias in age patterns by race/ethnic group. In this way, this constraining of fertility rates using survey data appears to be an improvement on our earlier efforts (Rendall et al 2006) to apply population data constraints that were by age alone.

We indicated that a next step is to include cohabiting unions. The expectation is that, as for marital unions, reporting within current unions will be unbiased, while male reporting deficits are likely where the children were born into unions that have subsequently dissolved. If so, including cohabitational unions in the correction will allowed for corrections that focus even more narrowly on the births believed to be most at risk of going unreported in the men's data. Years in a non-marital cohabiting union could be treated in any of three different ways. Currently, they are treated simply as exposure to non-marital fatherhood. This doesn't allow for a probable greater likelihood of men's reporting of children born within a co-residential union, especially if that union is still intact at the time of the survey. A second way is to treat cohabiting unions as a separate state. The sampling error in estimates of cohabiting unions, however, will be high. This will make it difficult to distinguish between real reporting differences and differences due to the relatively small samples of men and women reporting cohabiting unions and births. This will be especially true if cohabiting unions are differentiated by whether they are current or previous. But failing to differentiate the unions in this way will miss an important factor in the differential reporting of births. The third way to treat cohabiting unions is by creating a state of "in union," combining marital and non-marital cohabiting unions, and differentiating them from "out of union." This would allow for further diffentiating by current and previous unions, and would allow for the maximum amount of exposed years to be matched between men and women (as a cohabiting woman implies a cohabiting man).

APPENDIX: ESTIMATION OF STANDARD ERRORS INCLUDING SAMPLING ERROR IN THE CONSTRAINTS

We would like to consider age-specific constraints for births for a specified father's age. To do so, we must estimate the size of the population of men for that specific age as well as the total number of men that age who became fathers that year. The size of the population of men for a specific age can be estimated from the final weights for men of that age from the NSFG. For our purposes, we will assume that this estimate is in fact the population size. The more difficult problem is determining the total number of men of that age who became fathers that year. To do this, we use information for women from the NSFG to estimate the total number of births for a specified father's age. Clearly, this constraint is not fixed or known but is simply an estimate, so we will need to account for the variability in this constraint when calculating the variability in the parameter estimates for the main model. A brief summary of the steps necessary in solving this problem is given below:

1. Estimate $N(a)$, the number of men of age a, using the men's NSFG data.
2. Estimate $B(a)$, the number of women who have a child with a father of age a, using the women's NSFG data. Denote the estimate by $\hat{B}(a)$. (From this point on, estimates will be represented by hats.)
3. Use each $\frac{\hat{B}(a)}{N(a)}, a = 18, \ldots, 27$ as a constraint.
4. Estimate $\Delta_h \equiv V\left(\frac{\hat{B}(a)}{N(a)}\right)$, the variability in each constraint.
5. Compute V_s, the covariance of the parameter estimates for the main *unconstrained* logistic model (i.e, the UCMLEs) estimated at the UCMLEs. This model

uses the male data.

6. Estimate H, the Jacobian of the constraint functions. This is computed by taking the partial derivatives of the constraint functions with respect to the parameters and then evaluating them at the CMLEs.

7. Estimate Δ_g, the covariance in the proportion of the population with a set of regressor values specified in the constraint functions.

8. Using $\hat{\Delta}_h, V_s, \hat{H}$, and $\hat{\Delta}_g$, we estimate the covariance of the parameter estimates for the main *constrained* logistic model (i.e., the CMLEs).

With this outline in mind, we now provide a more detailed explanation of each step. Note that, for the most part, i will be an index reserved for men, and j will be an index reserved for women. In addition, Z and θ will be the regressors and parameters, respectively, reserved for the constraint function models, while X and β will be the regressors and parameters reserved for the main model. Then we use the following procedure to determine the appropriate constraint and measure of variability:

1. Using the men's NSFG data, we compute $N(a), a = 18,...,27,$ the population size of men of age a. If we let A_i denote the age and w_i^{men} the final weight for man i, and if we let L be the number of men sampled, then we can estimate $N(a)$ by means of

$$N(a) = \sum_{i=1}^{L} w_i^{men} \times 1_{[A_i = a]}. \tag{1}$$

This is simply the sum of the final weights for all men in the sample who are of age a. For our purposes, we will assume that each computed $N(a)$ is in fact the true population size.

2. Using the women's NSFG data, we estimate $B(a), a = 18, \ldots, 27,$ the estimated number of men of age a in the population who became fathers that year. Let Y_j indicate whether woman j has given birth in the past year, let A_j denote the age of the father as reported by the mother, let w_j represent the woman's final weight, let Z_j denote the vector of women-specific covariates for woman j used in predicting incidence of birth, and let M be the total number of women sampled. Then, for a fixed father's age $A_j = a$, we can predict the probability of birth for woman j through

$$\hat{\pi}_j(a) = P\left(Y_j(a) = 1\right) = \frac{\exp\left(Z_j\hat{\theta}\right)}{1 + \exp\left(Z_j\hat{\theta}\right)}, \tag{2}$$

where $Y_j(a) = Y_j \times 1_{[A_j = a]}$. This is 1 if the woman gave birth and the father's age was a, and 0 otherwise. This means that we will need to consider a separate logistic regression of the form of (2) for each father's age under consideration. This is not quite right, however, as each woman represents w_j individuals, so we instead use weighted logistic regressions. Using the predicted probabilities $\hat{\pi}_j(a)$ for a father of age a and final weights w_j, we can estimate $B(a)$, the total number of births for men of age a, via

$$\begin{aligned}\hat{B}(a) &= \sum_{j=1}^{M} w_j \hat{\pi}_j(a) \\ &= w^T \hat{\pi}(a). \end{aligned} \tag{3}$$

3. Once we have $\hat{B}(a)$ and $N(a)$ for each age $a = 18, \ldots, 27,$ we use $\frac{\hat{B}(a)}{N(a)}$ as our constraints. Then, for a constrained version of our main model, we can obtain estimates $\hat{\beta}_C$ for our parameters β using constrained maximum likelihood. All that remains to be

done is estimating the variance of $\hat{\beta}_C$.

4. We estimate Δ_h, the variability induced by the constraint. We calculate this by

$$\Delta_h \equiv V\left(\frac{\hat{B}(a)}{N(a)}\right) = \frac{1}{[N(a)]^2}\sum_{j=1}^{M} w_j^2 Cov\left(\hat{\pi}_j(a)\right)$$

$$= \frac{1}{[N(a)]^2} w^T Cov\left(\hat{\pi}(a)\right) w, \qquad (4)$$

where $Cov\left(\hat{\pi}(a)\right)$ is an $M \times M$ covariance matrix with ij^{th} element the covariance between $\hat{\pi}_i(a)$ and $\hat{\pi}_j(a)$. (Here we will drop the a–notation, so $\hat{\pi}(a)$ will simply be $\hat{\pi}$.)

In order to calculate $Cov\left(\hat{\pi}\right)$, recall that the predicted probabilities are given by

$$\hat{\pi} = \frac{\exp\left(Z\hat{\theta}\right)}{1+\exp\left(Z\hat{\theta}\right)}.$$

We can use the standard estimate of $Cov\left(\hat{\theta}\right)$ from the logistic regression in (2), so calculating $Cov\left(\hat{\pi}\right)$ simply requires that we use the delta method. (For details, see Agresti (2004).)

$$\widehat{Cov}\left(\hat{\pi}\right) = \left[\mathrm{diag}\left(\hat{\pi}\right) - \hat{\pi}\hat{\pi}^T\right] Z Cov\left(\hat{\theta}\right) Z \left[\mathrm{diag}\left(\hat{\pi}\right) - \underline{s}\hat{\pi}\hat{\pi}^T\right] (5)$$

$$= \left[\mathrm{diag}\left(\hat{\pi}\right) - \hat{\pi}\hat{\pi}^T\right] Z \left\{Z^T\left[\mathrm{diag}\left(\hat{\pi}\right) - \hat{\pi}\hat{\pi}^T\right] Z\right\}^{-1} Z \left[\mathrm{diag}\left(\hat{\pi} ght\right) - \hat{\pi}\hat{\pi}^T\right]$$

Finally, an estimate of Δ_h is

$$\hat{\Delta}_h = \frac{1}{[N(a)]^2} w^T \times \widehat{Cov}(\hat{\pi}) \times w. \qquad (6)$$

5. We compute V_s, the covariance of $\hat{\beta}$, using the male data. This is the covariance of the parameter estimates for an unconstrained version of our main model. Computation of V_s is straightforward and will not be explained here.

6. We estimate H, the Jacobian of the constraint function. Using the notation of Handcock et al (2005), let

$$C(\beta) = E_\beta[g(Y, X)], \qquad (7)$$

where $C(\beta)$ is a constraint and $g(Y, X)$ is a constraint function. Note that Y now represents the response variable of the main model. Then H is given by

$$H = \left[\frac{\partial C_i(\beta)}{\partial \beta_j} \right], \qquad (8)$$

and is estimated by

$$\hat{H} = H |_{\beta = \hat{\beta}_C}. \qquad (9)$$

7. We estimate Δ_g. Let

$$g(x;\beta) = E_{Y|X=x;\beta}[g(Y, X = x)] \qquad (10)$$

Then

$$\Delta_g \equiv Cov(g(X;\beta)), \qquad (11)$$

and is estimated by

$$\hat{\Delta}_g = \Delta_g |_{\beta = \hat{\beta}_C}. \qquad (12)$$

8. Using $\hat{\Delta}_h, V_s, \hat{H}$, and $\hat{\Delta}_g$, we can compute $Cov\left(\hat{\beta}_C\right)$. Following Handcock et al

(2005), in the case where $C(\beta)$ is fixed,

$$V \equiv Cov\left(\hat{\beta}_C\right) = \left[V_s^{-1} + H^T \Delta_g^{-1} H\right]^{-1}$$

$$= V_s - V_s H^T \left[H V_s H^T + \Delta_g\right]^{-1} H V_s. \qquad (13)$$

Note that if the population values of $g(X;\beta)$ are known, then (13) simplifies, as the Δ_g term drops out.

Our situation is different in that $C(\beta)$ is no longer known exactly. Instead of having a fixed constraint and a set of data to fit to this constraint, we now have two sets of data—one to determine constraints, and the other to fit to the constraints. We have previously defined M to represent the number of woman-years from the data used to estimate the constraint from the women's NSFG data. Now let N represent the number of man-years from the male NLP data used in the constraint function. Imbens and Lancaster (1994) show that

$$V\left[\sqrt{N}\left(\hat{\beta}_C - \beta\right)\right] \rightarrow \left[V_s^{-1} + H^T \left(\Delta_g + \tfrac{N\Delta_h}{M}\right)^{-1} H\right]^{-1},$$

$$\rightarrow V_s - V_s H^T \left[H V_s H^T + \Delta_g + \tfrac{N\Delta_h}{M}\right]^{-1} H V_s,$$

as $\frac{M}{N} \rightarrow K > 0$ and $N \rightarrow \infty$. Hence, we approximate

$$V\left(\hat{\beta}_C\right) \approx \frac{1}{N}\left\{\hat{V}_s - \hat{V}_s \hat{H}^T \left[\hat{H}\hat{V}_s\hat{H}^T + \hat{\Delta}_g + \frac{N\hat{\Delta}_h}{M}\right]^{-1} \hat{H}\hat{V}_s\right\}, \qquad (14)$$

Notice that this has the same form as the case when $C(\beta)$ is constant, except an additional term $\left(\frac{N\hat{\Delta}_h}{M}\right)$ has been added to account for the variability in our constraint.

REFERENCES

Agresti, A. (2004) Categorical Data Analysis New York:Wiley.

Cooney, T.M., F.A. Pedersen, S. Indelicato, and R. Palkovitz. (1993). Timing of fatherhood: Is “on time” optimal? Journal of Marriage and the Family 55:205-215.

Elo, I. T., R. B. King, and F. F. Jr. Furstenberg (1999) Adolescent Females: Their Sexual Partners and the Fathers of Their Children Journal of Marriage and the Family 61(1): 74-84.

Garfinkel, I., S. S. McLanahan, and T. L. Hanson. (1998). A Patchwork Portrait of Nonresident Fathers. In Fathers Under Fire: The Revolution in Child Support Enforcement, edited by I. Garfinkel, S. McLanahan, D. Meyer, and J. Seltzer. New York: Russell Sage.

Handcock, M.S., M.S. Rendall, and J.E. Cheadle (2005) Improved regression estimation of a multivariate relationship with population data on the bivariate relationship Sociological Methodology 35(1)291-334.

Imbens, G. W. and T. Lancaster (1994) Combining micro and macro data in microeconometric models Review of Economic Studies 61(4):655-680.

Lepkowski, JM, WD Mosher, K Davis, R Groves, S Heeringa, J VanHoewyk, T Adams, J Willem (2005) Sample Design, Sampling Weights, Imputation, and Variance Estimation in the 2002 National Survey of Family Growth Vital and Health Statistics Series 2. Hyattsville, MD: National Center for Health Statistics.

Lindberg, L. D., F. L. Sonenstein, G. Martinez, and J. Marcotte. (1998) Completeness of young fathers' reports of fertility Journal of Economic and Social Measurement 24(1):15-23.

Manning, W.D. and P.J. Smock. Measuring and Modeling Cohabitation: New Perspectives From Qualitative Data. Journal of Marriage and Family, 67:989-1002.

Marsiglio, M., P. Amato, R.D. Day, and M.E. Lamb (2000) Scholarship on fatherhood in the 1990s and beyond Journal of Marriage and the Family 62:1173-1191.

Martinez, G.M., A. Chandra, J.C. Abma, J. Jones, and W.D. Mosher (2006) Fertility, contraception, and fatherhood: Data from Cycle 6 of the 2002 National Survey of Family Growth Vital and Health Statistics Series 23, Number 26. Hyattsville, MD: National Center for Health Statistics.

Nock, S.L. (1998) The consequences of premarital fatherhood American Sociological Review 63(2):250-263.

Pears, K.C., S.L. Pierce, H.K. Kim, D.M. Capaldi, and L.D. Owen (2005) The timing of entry into fatherhood in young, at-risk men. Journal of Marriage and the Family 67:429-447.

Peters, H. E. (1988) Retrospective versus panel data in analyzing life-cycle events Journal of Human Resources 23(4):488-513.

Rendall, M.S., L. Clarke, H. E. Peters, N. Ranjit, and G. Verropoulou (1999) Incomplete Reporting of Male Fertility in the United States and Britain: A Research Note. Demography 36(1):135-144.

Rendall, M.S., K. Joyner, H.E. Peters, F. Yang, M.S. Handcock, and R. Admiraal (2006) A Bayesian Approach to Combining Population and Survey Data for Male Fertility Estimation. Paper presented at the 2006 Annual Conference of the Population Association of America, Los Angeles.

Teitler, J., N. Reichman, and H. Koball (2006). Contemporaneous Versus Retrospective Reports of Cohabitation in the Fragile Families Survey Journal of Marriage and Family 68:469-477.

Wu, L. L., S. Martin, and D. Long (2001) Comparing Data Quality of Fertility and First Intercourse Reports. Journal of Human Resources 36(3):520-555.

Wu, L.L., and B.C. Martinson (1993) Family structure and the risk of a premarital birth American Sociological Review 58:210-232.

Table 1 Births by Marital Status of Parents and Age of Father: NSFG Men between 1991 and 2000

Weighted	Unmarried			Married			Previous Marriages			Current Marriages		
reported by:	Men	Women		Men	Women		Men	Women		Men	Women	
father's age	*$B^m(x,u)$*	*$B^f(x,u)$*	*ratio*	*$B^m(x,r)$*	*$B^f(x,r)$*	*ratio*	*$B^m(x,c)$*	*$B^f(x,c)$*	*ratio*	*$B^m(x,p)$*	*$B^f(x,p)$*	*ratio*
15	49,339	14,516	*340*	12,075	0	-	12,075	0	-	0	0	-
16	127,722	181,980	*70*	0	0	-	0	0	-	0	0	-
17	230,989	349,588	*66*	57,621	24,309	*237*	46,957	24,309	*193*	10,663	0	-
18	348,723	658,122	*53*	130,988	142,951	*92*	120,790	81,140	*149*	10,198	61,811	*16*
19	432,913	936,848	*46*	190,480	223,230	*85*	115,713	146,410	*79*	74,767	76,821	*97*
20	586,228	976,824	*60*	341,984	312,858	*109*	102,363	131,935	*78*	239,621	180,923	*132*
21	626,161	964,638	*65*	409,761	595,307	*69*	159,865	268,209	*60*	249,896	327,099	*76*
22	709,623	747,565	*95*	616,955	700,324	*88*	289,773	287,745	*101*	327,182	412,579	*79*
23	829,159	947,617	*87*	921,231	848,510	*109*	207,593	297,870	*70*	713,638	550,641	*130*
24	790,050	844,981	*93*	1,264,411	1,076,859	*117*	255,909	265,648	*96*	1,008,501	811,211	*124*
25	425,169	875,834	*49*	1,493,927	1,313,813	*114*	322,691	328,124	*98*	1,171,236	985,689	*119*
26	470,448	976,509	*48*	1,631,945	1,151,321	*142*	230,604	331,650	*70*	1,401,341	819,672	*171*
27	480,779	749,304	*64*	1,529,342	1,635,607	*94*	321,693	362,863	*89*	1,207,649	1,272,744	*95*
28	389,288	691,464	*56*	1,964,834	1,340,965	*147*	310,793	298,927	*104*	1,654,041	1,042,038	*159*
29	321,718	533,028	*60*	1,665,230	1,818,695	*92*	396,810	295,687	*134*	1,268,420	1,523,008	*83*
ages 15-29	6,818,308	10,448,818	*65*	12,230,784	11,184,749	*109*	2,893,629	3,120,514	*93*	9,337,155	8,064,235	*116*

Unweighted	Unmarried		Married		Previous Marriages		Current Marriages	
reported by:	Men	Women	Men	Women	Men	Women	Men	Women
father's age	*$B^m(x,u)$*	*$B^f(x,u)$*	*$B^m(x,r)$*	*$B^f(x,r)$*	*$B^m(x,c)$*	*$B^f(x,c)$*	*$B^m(x,p)$*	*$B^f(x,p)$*
15	4	3	1	0	1	0	0	0
16	13	29	0	0	0	0	0	0
17	26	54	3	2	2	2	1	0
18	31	122	8	22	7	14	1	8
19	39	154	12	33	8	23	4	10
20	54	142	27	42	14	21	13	21
21	47	153	33	80	18	37	15	43
22	60	141	44	94	21	44	23	50
23	57	150	48	114	14	42	34	72
24	54	149	68	132	19	39	49	93
25	42	138	93	161	33	47	60	114
26	35	143	75	154	22	48	53	106
27	33	113	88	193	28	51	60	142
28	39	105	101	153	26	32	75	121
29	26	90	88	202	25	41	63	161

Notes: Marital status at time of birth. Source: authors tabulations from NSFG 2002

Table 2 Fertility Exposure by Marital Status and Age of Men: NSFG Men and Women between 1991 and 2000

Weighted	Unmarried	Married			Previous Marriages			Current Marriages		
	Men	Men	Women		Men	Women		Men	Women	
age	$N^m(x,u)$	$N^m(x,r)$	$N^f(x,r)$	*ratio*	$N^m(x,c)$	$N^f(x,c)$	*ratio*	$N^m(x,p)$	$N^f(x,p)$	*ratio*
18	19,111,759	201,383	292,123	*69*	23,752	110,999	*21*	177,630	181,124	*98*
19	18,762,934	548,172	659,863	*83*	171,476	162,157	*106*	376,696	497,706	*76*
20	18,099,962	1,157,048	1,242,296	*93*	288,840	329,382	*88*	868,208	912,914	*95*
21	17,391,968	1,758,015	2,023,727	*87*	404,281	566,039	*71*	1,353,734	1,457,688	*93*
22	16,397,387	2,732,800	3,007,772	*91*	585,860	793,456	*74*	2,146,940	2,214,316	*97*
23	15,456,983	4,168,650	4,131,507	*101*	838,142	905,386	*93*	3,330,508	3,226,121	*103*
24	14,835,675	5,335,483	5,238,491	*102*	1,012,832	1,153,021	*88*	4,322,651	4,085,470	*106*
25	13,885,043	6,696,299	6,285,968	*107*	1,085,249	1,453,369	*75*	5,611,050	4,832,600	*116*
26	13,002,499	7,595,747	7,438,317	*102*	1,221,972	1,566,691	*78*	6,373,775	5,871,626	*109*
27	12,139,284	8,455,790	8,517,484	*99*	1,277,873	1,743,145	*73*	7,177,916	6,774,339	*106*
28	11,713,505	9,096,811	9,536,309	*95*	1,494,782	1,799,046	*83*	7,602,030	7,737,264	*98*
29	9,096,811	10,091,543	10,286,301	*98*	1,894,540	1,935,319	*98*	8,197,004	8,350,982	*98*
ages 18-29	179,893,812	57,837,741	58,660,158	*99*	10,299,599	12,518,009	*82*	47,538,142	46,142,149	*103*

Unweighted		Married		Previous Marriages		Current Marriages	
	Men	Men	Women	Men	Women	Men	Women
age	$N^m(x,u)$	$N^m(x,r)$	$N^f(x,r)$	$N^m(x,c)$	$N^f(x,c)$	$N^m(x,p)$	$N^f(x,p)$
18	1627	15	40	4	17	11	23
19	1554	42	89	15	28	27	61
20	1468	80	170	29	58	51	112
21	1386	119	268	42	86	77	182
22	1285	174	393	57	114	117	279
23	1198	238	530	72	140	166	390
24	1152	306	658	89	169	217	489
25	1089	374	768	99	189	275	579
26	1038	422	872	116	202	306	670
27	1033	469	967	128	221	341	746
28	960	511	1058	145	241	366	817
29	511	542	1115	154	248	388	867

Notes: Marital status at middle of age at risi Source: authors tabulations from NSFG 2002

Figure 1: Male fertility rate, 1991-2000, by data source

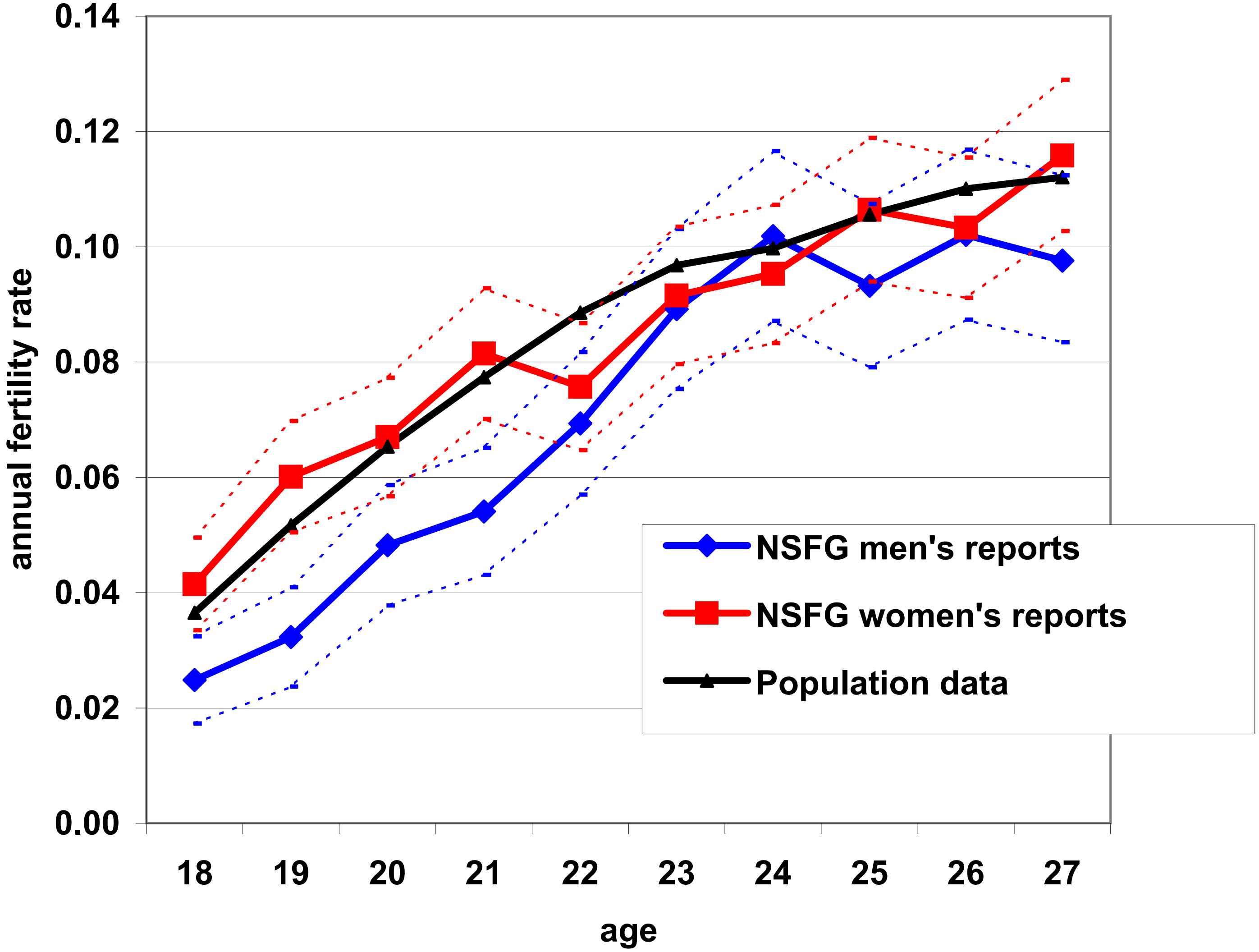

Figure 2: Fitting the male non-marital fertility constraint

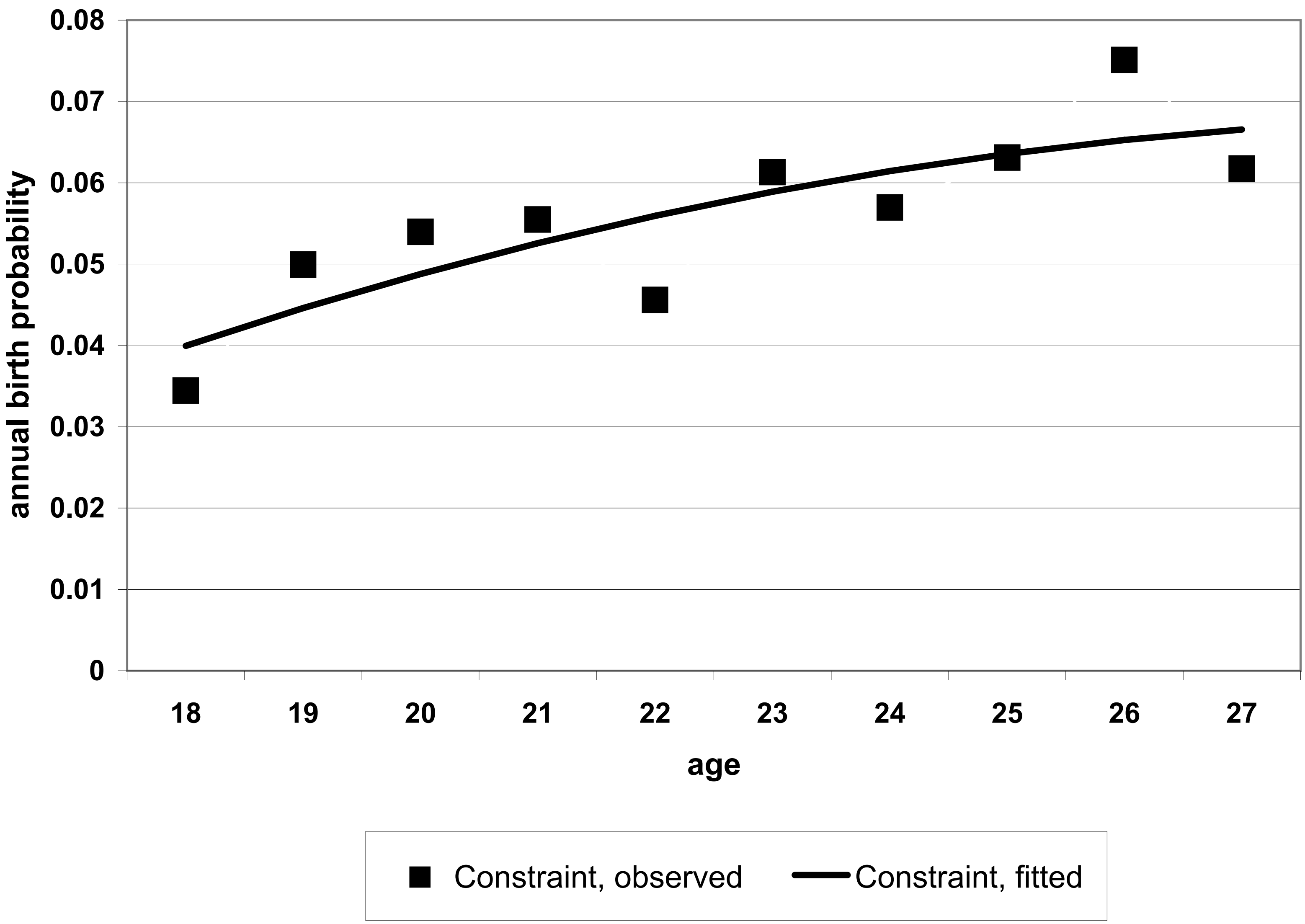

Figure 3: Male non-marital fertility, blacks, whites, Hispanics; parity 0

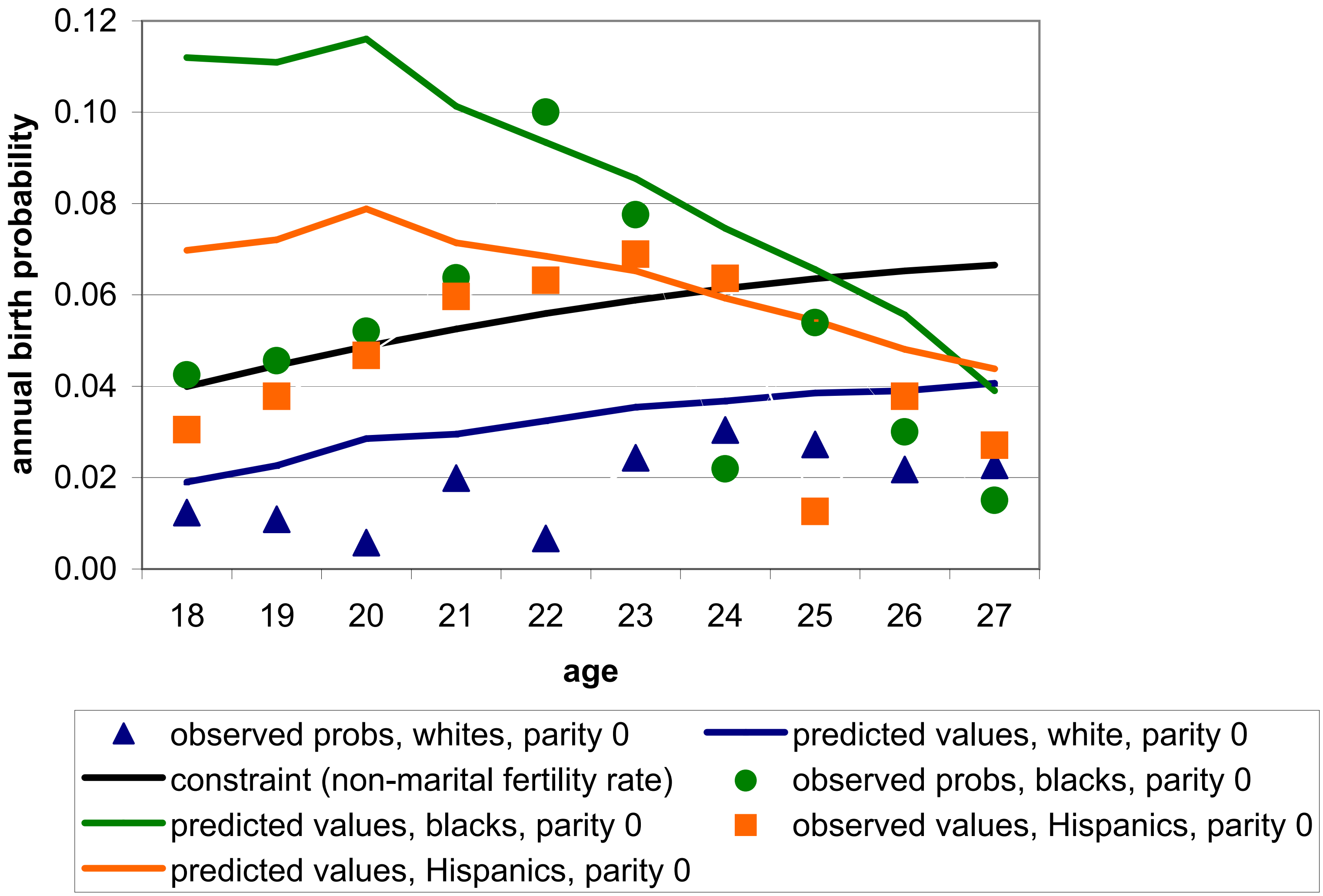

Figure 4: Male non-marital fertility, blacks and whites; parity 1+