DRAFT: DO NOT QUOTE

The Effect of a First Child on Female Labor Supply: Evidence from Women in Fertility Treatments*

Julian P. Cristia

August 1, 2006

Abstract

Estimating the causal effect of fertility on female labor supply is complicated due to the endogeneity of the fertility decision. Ideally, this problem could be solved by running a social experiment where women are randomly assigned children (the treatment group) or not (the control group). In this paper, I use field data from the National Survey of Family Growth (NSFG) to mimic this hypothetical experiment by focusing on a sample of women that sought help to become pregnant. After a certain period since they started receiving help, some of these women are successful and some of them are not. In this instance, fertility appears to be exogenous to labor supply in that pre-treatment labor supply is uncorrelated with subsequent fertility. Using this empirical strategy, I estimate that having a first child younger than one year old reduces female employment by 26.3 percentage points. These estimates are close to OLS estimates obtained using Census data and to OLS and fixed-effects estimates from NSFG data. The results also indicate that the estimated short-term impact of fertility on female labor supply decreased 40 to 50 percent during the 1980 to 1990 period.

*Health and Human Resources Division, Congressional Budget Office, julian.cristia@cbo.gov, 202-226-5674. I would like to thank Julio Caceres Delpiano, Mark Duggan, William Evans, Eugenio Giolito, Amy Harris, Sandra Hofferth, BeomSoo Kim, Noah Meyerson, Ron Moore, Reed Olsen, John Sabelhaus, Seth Sanders, Jon Schwabish, Michael Simpson, Julie Topoleski, Bruce Vavrichek and workshop participants at the University of Maryland and the Congressional Budget Office. The views expressed in this paper are those of the author and should not be interpreted as those of the Congressional Budget Office.

1. Introduction

Estimating the effect of fertility on female labor supply has been a long standing problem in economics. Knowing how families optimize their labor supply decisions to the arrival of a child is important for several reasons. First, it is interesting to know how much of the increase in female labor supply during the postwar period can be explained by delayed childbearing and reduced fertility (Goldin, 1990). Second, some researchers believe that the interruption of work attributable to childbearing is responsible for a significant fraction of the female-male wage gap (Goldin and Polachek, 1987; Gronau, 1988; Fuchs, 1989; Korenman and Neumark, 1992) and the size of the impact of childbearing on female labor supply is an important variable in this calculation. Third, if declines in labor supply after childbearing correspond to increases in child care time, then knowing the effect of childbearing on female labor supply will provide information about time inputs invested in the child (Stafford, 1987; Blau and Grossberg, 1992). Finally, and above all, economists have been interested in this question from a basic desire to know the quantitative importance of different determinants of female labor supply.

Given the importance of this topic, it is not surprising that hundreds of published studies have examined the relationship between fertility and female labor supply. However, as Browning (1992) notes in his literature review on this topic, “Although we have a number of robust correlations, there are very few credible inferences that can be drawn from them.” The key problem researchers face is that there are theoretical reasons to believe that the fertility decision may be endogenous and therefore the strong negative correlations found between different measures of fertility and female labor supply cannot be interpreted as evidence of causal effects.

In trying to overcome the type of criticism highlighted by Browning (1992), two strategies have been proposed that exploit exogenous changes in family size in order to estimate the effect of fertility on female labor supply. The first (Rosenzweig and Wolpin, 1980; Bronars and Grogger, 1994; Jacobsen, Pearce and Rosenbloom, 1999), used the fact that twins in the first birth represent an exogenous change in family size in order to estimate the effect of having a second child. The second (Angrist and Evans, 1998), exploited parental preferences for mixed-sex siblings in order to estimate the effect of a third or higher order child.

Still, the question of how a first child affects female labor supply has not been addressed with a strategy that convincingly tackles the problem of the endogeneity of fertility. It could be argued that the effect of having a first child is the most important one, given that it applies to a vast majority of women, whereas the effect of having a second or higher order child only applies to a smaller subset of women.[1]

Ideally, this question could be answered by running a hypothetical social experiment where childless women were randomly assigned children. After the assignment of children, employment rates for the treated and control groups could be compared in order to estimate the causal effect of having a first child on female labor supply.

This paper focuses on a situation that mimics this hypothetical experiment. In particular, I construct a sample of childless women that sought help to achieve pregnancy. At the time of seeking help, all of them wanted to have a child but after a certain period some of these women were successful and gave birth while others did not. Then, I compare the employment

[1] In the 1990 Census, among women aged 45 to 55, 89 percent of them had at least one child, whereas 78.3 percent had at least two and 50.4 percent had at least three children.

rates of women that were "treated" (i.e., gave birth to a child) with those who were not.[2]

The contribution of this paper is that while analysis of twins and the preference for mixed siblings strategies, under certain conditions, can be used to identify the effect of a second or higher order child, the estimation strategy pursued here is able to identify the effect of a first child on female labor supply.

The proposed strategy eliminates the potential problem of fertility being an endogenous variable because all women wanted to have a child at the time they sought help. However, early success in fertility treatments is not expected to be completely random. Still, I provide several pieces of evidence that suggest that this strategy consistently estimates the parameter of interest. First, following Heckman and Hotz (1989), I find that pre-treatment labor supply is uncorrelated with subsequent fertility. Second, estimates are very robust to the set of covariates added to the main regression. Third, observable characteristics of the sample of women that sought help to achieve pregnancy while childless are quite similar to women who have their first child after age 18.

Using the exogenous assignment of children to women via infertility treatments as an identifying strategy, I estimate that having a first child younger than one year old reduces female employment by 26.3 percentage points. These estimates are close to OLS and fixed-effects estimates obtained from panel data from the National Survey of Family Growth (NSFG) . They are also close to OLS estimates obtained using similarly defined samples from the 1980 and 1990 Census. This finding is important because almost all previous studies that take into account the endogeneity of the fertility decision provide much smaller estimated impacts than those studies that assume exogenous fertility. Finally, I provide evidence of an

[2]In this paper "treatment" refers to having a child.

important reduction in the estimated short-term impact of childbearing on female labor supply of around 40 to 50 percent during the 1980 to 1990 period.

2. Previous Research

Interest in the question of the effect of fertility on female labor supply is illustrated in the long list of studies that have focused on this issue. These studies can be classified into four groups depending on how they have tackled the problem of the endogeneity of the fertility decision. The first group is illustrated by the studies of Gronau (1973), Heckman (1974) and Heckman and Willis (1977). They assumed that fertility was exogenous and established a strong negative correlation between female labor supply and fertility.

A second group of studies acknowledged the endogeneity of the fertility decision and tried to deal with this problem by estimating simultaneous equations models (Cain and Dooley, 1976; Schultz, 1978; Fleisher and Rhodes, 1979). These studies find a much smaller estimate when fertility is treated as an endogenous variable compared to when it is considered exogenous. The disadvantage of this approach is that it is difficult to find plausible exclusion restrictions that could identify the underlying structural parameters.

A third group of studies incorporated actual fertility as a regressor but added the lagged dependent variable (labor supply) in an effort to control for unobserved heterogeneity across women. Nakamura and Nakamura (1992) recommended this approach and it has been used by a number of authors (Even, 1987; Lehrer, 1992). Although adding the lagged dependent variable can be useful to control for unobserved heterogeneity, it still does not address the problem of the endogeneity of the fertility decision.

Finally, a fourth group of studies tackled the endogeneity of the fertility variable by exploiting exogenous sources of variation in family size. Rosenzweig and Wolpin (1980) first used this strategy comparing labor supply of mothers having twins on their first birth with those that had a single child. Subsequent studies by Bronars and Grogger (1994) and Jacobsen, Pearce and Rosenbloom (1998) employed the same strategy but managed to obtain more precise estimates by developing an algorithm to detect twin births using Census data.

In the same spirit as the twins' studies mentioned above, Angrist and Evans (1998) exploited the fact that parents typically prefer mixed-sex siblings in order to estimate the effect of a third or higher order child on female labor supply. For a sample of couples with at least two children they instrumented further childbearing (i.e., having more than two children) with a dummy for whether the sex of the second child matches the sex of the first. As sex mix is virtually random, this strategy allows for the identification of the effect of a third or higher order child.

My work is most similar to this last mentioned group of studies as it uses the fact that the biology of reproduction is intrinsically stochastic in order to identify exogenous changes in fertility. Still, there are two main differences between these studies and this paper. First, I estimate the effect of a first child on female labor supply whereas these other studies estimates the effect of a second or higher order child. Second, while the mentioned group of studies have used instruments for fertility and then computed Two Steps Least Squares estimates, I tackle the endogeneity of fertility by focusing on a sample of women for which fertility is plausibly exogenous and then estimate the impact by just using OLS.

This paper is related to other strands of economic literature. First, its results shed light on a number of studies that have tried to explain the postwar rise in female labor supply (Mincer, 1962; Goldin, 1990). Second, it is related to a line of research that tries to establish the effect of childbearing related withdrawals from the labor market on females' wages and earnings (Goldin and Polachek, 1987; Gronau, 1988; Fuchs, 1989; Korenman and Neumark, 1992; Miller, 2005).[3] Lastly, it is linked with studies focusing on how maternal work affects children's outcomes (Stafford, 1987; Desai et al., 1989; Blau and Grossberg, 1992).

3. Background: the Reproductive Process and Infertility

Reproduction is a very delicate process that requires the correct functioning of the male and female reproduction systems as well as ideally timed sexual intercourse. Conception takes place when a motile sperm from the man burrows into an egg (ovum) from the woman and fertilizes it. Fertilization occurs in one of the Fallopian tubes and the fertilized egg starts dividing itself as it travels through the Fallopian tube towards the uterus. There, it will settle in the lining of the uterus and hopefully it will grow until the baby is born.

Healthy couples having intercourse regularly have only a 20 percent chance of conceiving during a month. This implies that around 26 percent of healthy couples will not have conceived after six months of unprotected sex, and this number falls to about 7 percent after 12 months. Given these facts, couples are recommended to start receiving testing and treatment only after 6 to 12 months of trying to conceive without success. Moreover, the medical community defines a couple as infertile if they have not conceived after 12 months of unprotected sex.[4] The National Center for Health Statistics estimated that in 1995 there were 2.1 million infertile married couples in the United States and 6.1 million women aged 15-44 had

[3]As in this paper, Miller (2005) exploited biological fertility shocks. However, while I aim to estimate the impact of childbearing on female labor supply, Miller was primarily concerned about how changes in the age at first birth impact long-run earnings and future wages.

[4]The WHO defines a couple as infertile if they have not been able to conceive after 24 months of unprotected sex.

an impaired ability to have children.[5]

Medical researchers have identified a number of factors (besides the conditions mentioned above) that affects the prognosis of a couple. Female's age, education, smoking status, consumption of recreational drugs and obesity as well as sexual frequency are important predictors of the probability of conception (Baird and Wilcox, 1985; Dunson et all, 2004).

Given the stochastic nature of the reproduction process, physicians usually start treatment with simple and cheap procedures (like advice and testing) and only start employing more invasive and expensive procedures as the simple procedures prove unsuccessful. As an example of this optimal sequential strategy, physicians typically only recommend in vitro fertilization methods after all other options have been exhausted or if they strongly believe that less invasive procedures will be unsuccessful.

4. Data

This paper uses data from the National Survey of Family Growth (NSFG), a survey conducted by the National Center for Health Statistics in 6 cycles (1973, 1976, 1982, 1988, 1995 and 2002). Cycles 1 to 5 were conducted at the homes of a national representative sample of women ages 15-44 years old. Cycle 6 also sampled men ages 15-44 years old. The main purpose of these surveys was to provide reliable national data on marriage, divorce, contraception, infertility and the health of women and infants in the United States.

Data from the NSFG Cycle 5 was chosen for this paper because it provides information about births, pregnancies, infertility services, demographic characteristics and in particular the complete work history for each individual.[6] In particular, the month in which each woman

[5] Source: http://www.cdc.gov/nchs/fastats/fertile.htm.

[6] Other cycles included all needed information except from monthly employment status for each woman. I

sought help for the first time to achieve pregnancy is provided, information that is critical for the strategy pursued in this paper. Other important variables included are age, race, ethnicity, educational attainment, school enrollment and smoking history. The survey also reports data on each full-time and part-time employment spell.

The NSFG Cycle 5 employs a multi-stage sampling design with an over-sample of Hispanic and black women. It took place between January and October, 1995 and the overall response rate was 79 percent. A total of 10,847 women were interviewed.

Data on fertility and employment are collected retrospectively. Although there are limitations of this type of design, Teachman, Tedrow and Crowder (1998) found the NSFG Cycle 5 data to be of high quality. They concluded that the employment information matches CPS data reasonably well, although the data on employment spells has not been validated using external records.

5. Empirical Strategy, Parameter of Interest and Sample Construction

5.1 Empirical Strategy

An ideal social experiment aimed at estimating the causal effect of childbearing on female labor supply would recruit women who wanted to have a child and then assign a child to a group of women (treated individuals) while not assigning a child to a second group (controls).[7] Given the stochastic nature of conception, this type of experiment can be approximated. To start with, we need a group of women who want to conceive a baby. Second, some of these women should receive babies in a way that is uncorrelated with baseline employment. Third,

cannot run this analysis without this information as I compare employment 21 months after each women sought help to become pregnant.

[7] To be precise, this experiment will estimate the effect of having a child on female labor supply for women that wanted to have a child, not for all women.

we need to observe female labor supply for both groups of women after the assignment of babies.

I aim to mimic the ideal social experiment and fulfill the three aforementioned conditions by focusing on the following situation. I construct a sample of women that sought help to have a first child (called the HELP sample). As women in this sample seek help to achieve pregnancy at different points in time, I normalize time by the month that they sought help for the very first time (denoted as month 0). Next, I classify these women depending on whether they had given birth to a child by month 21 and in this way I get two groups of women: treated and controls. Finally, I compare employment rates of these two groups of women in month 21 to estimate the causal effect of having a first child younger than a year on female labor supply.

I choose to compare employment in month 21 instead of other months for several reasons. First, at this point in time 97 percent of babies born are younger than a year old making more precise the definition of the treatment effect. Second, using a longer horizon could allow some women to have additional children, which would complicate the analysis.[8] Third, as time in treatment increases, women who are unsuccessful conceiving may start adopting. Finally, looking at this shorter time span, it is more plausible that women receive similar types of infertility treatments (e.g., in vitro fertilization treatments are typically not considered an option in the first 12 months after seeking help to achieve pregnancy).

Following this strategy, I tackle the endogeneity problem because all of the women in the HELP sample wanted to have children. Still, in order to consistently estimate the effect of childbearing on female labor supply, the assignment process of children must be uncorrelated

[8] Focusing on month 21 there are only six women that had two children. Five had given birth to twins initially; only one had given birth twice.

with baseline female labor supply. Clearly, this assumption is untestable. However, following Heckman and Hotz (1989), I provide evidence of the plausibility of this assumption in section 7, where I test whether pre-treatment labor supply is correlated with subsequent fertility.

A potential problem with this empirical strategy arises if women in the control group adopt a child or start cohabitating or marry an individual with children. In the treatment evaluation literature this is denoted as "substitution bias" and represents a situation where individuals in the control group receive close substitutes for the treatment in question (see Heckman and Smith (1995), pages 22-24). In the context of this paper, treatment is having a natural birth and a close substitute is adopting a child (or acquiring a step child). Even though substitution bias can be a problem in certain social experiments, it is not in this case.[9] Only 2.7 percent of women in the control group adopt or acquire a step child in the 21 months after they seek help to become pregnant (and only 0.5 percent in the treatment group).

5.2 Parameter of Interest

In this study, the parameter of interest is the average impact of having a first child younger than 12 months on female labor supply for women that want to have a child. It is important to note that it does not provide an estimate of the effect of having a first child for women whose child is unwanted. All the same, the parameter of interest that I am estimating applies to a fairly large population. Henshaw (1998) using data from the NSFG Cycle 5 found that 69 percent of births were planned for women aged 15-44 years old in 1994.

Throughout this study, I focus only on the short-term effects of having a first child

[9] In the case of the experimental evaluation of the training program JTPA, Heckman and Smith (1995) noted that 32 percent of control group members self-reported receiving training from other sources over the 18 months following random assignment.

(i.e., the estimated effect of having a child younger than one year old). It is clear that there are other treatment effects that are worthy of attention. However, for reasons already discussed, the strategy employed in this study is best suited to estimate this treatment effect.

Finally, an estimate of the impact of the first child younger than a year old is important for a number of reasons. First, as mentioned above, this effect will apply to a much wider population than estimates that focus on the effect of a second or higher order child. Second, there is consensus that the short-term effects of childbearing are substantially larger than the longer-term effects (Browning, 1992). Then, knowing the short-term effects is useful as it gives an upper bound for these longer-term effects. Third, Shapiro and Mott (1991) provide strong evidence that labor force status following the first birth is an important predictor of lifetime work experience. This implies that changes in the estimated short-term impact of having a first child on female labor supply could be predicting a substantial change in overall lifetime work experience for women. Finally, using this empirical strategy I can compare the estimated impacts obtained when tackling the endogeneity problem (i.e., using the HELP sample) with estimates from strategies that do not tackle this problem (e.g., OLS on Census data).

5.3 Sample Construction

The main sample used in this paper (the HELP sample) includes childless women who sought help to become pregnant while aged 19 to 38 years old.[10] Women that sought help less than 21 months before the interview are dropped from the HELP sample, because it is not

[10] I drop women younger than 19 years old at the time when they first sought help to become pregnant because work information is only reported since the women turned 18 and I want to know employment status one year before seeking help to become pregnant.

possible to observe their child and labor status at this key time.

Women that seek help to become pregnant are identified as those who answer affirmatively the question "Have you or your husband ever been to a doctor or other medical provider to talk about ways to help you become pregnant?" The wording of the question allows identifying a wide group of women that wanted to have children but were unsuccessful after trying for certain time. This fact explains why, as it will be seen later, women in the HELP sample are fairly representative of women that have their first child while being aged 19 to 38 years old.

Table 1 presents the algorithm employed in order to construct the HELP sample. This table shows that only 499 observations are included in the empirical analysis, a fact that may seem as an important limitation for this study. However, as shown in section 6, I precisely estimate the relevant coefficient of the effect of having a first child on female labor supply.

The basic empirical strategy of this paper consists in comparing women in the HELP sample that already had a baby by month 21 with those who did not. To identify these two groups of women, an indicator called *AnyChildren21* is defined that equals 1 if the woman had a baby by month 21. In this setting, women from the HELP sample for whom *AnyChildren21* equals 1 correspond to women that were "treated" and those for whom *AnyChildren21* equals 0 correspond to the "control" or comparison group.[11]

Descriptive statistics for women in the treatment and control groups are presented in Table 2. In the NSFG Cycle 5, respondents were asked about all their employment spells information, which I use in order to construct three employment variables. The variables

[11] For easy of exposition, along the paper I will denote as treated women to those that had a baby by month 21 whereas women that did not have a baby by month 21 are referred as "controls" (even though in the evaluation literature the word "controls" is used for subjects not receiving treatment in an experiment).

Employed21 and *Employed0* are dummy variables that equal 1 if the individual was employed in months 21 and 0 respectively. Similarly, *Employed_12* represents labor status in month -12 (i.e., 12 months before the woman sought help for the first time).

It is important to note that while employment rates in month 0 and -12 are similar between treated and control women, employment rates differ by 25.3 percentage points in month 21. Moreover, observable characteristics in month 0 between treated and control women are quite similar. As shown in Table 2, differences in means of key covariates between the treated and control groups are only statistically significant at the five percent significance level for the Hispanic and smoking dummies.[12]

A potential caveat for the strategy pursued in this paper is that, as typically is the case in social and medical experiments, the sample involved in the experiment may not be representative of the population of interest. In order to gauge the potential severity of this problem, Table 3 compares descriptive statistics of women in the HELP sample with those of women in the NSFG that had at least one child. For women in the HELP sample, time-varying variables are measured at the time they first sought help to achieve pregnancy, while for women that had at least one child these variables are measured at the time of first birth. In the second column of Table 3, statistics are presented for the set of women in the NSFG that had their first child while being 19 to 38 years old (as these age requirements were used to construct the HELP sample).

Comparing the second and third column of Table 3, we see that women in the HELP sample tend to be older, more educated, have higher employment, marriage and smoking rates, while a lower proportion of them are Hispanic or black, as compared to women from

[12]In section 7, I explore more deeply which variables predict fertility by month 21.

the NSFG that were aged 19 to 38 when they had their first child. Still, basic statistics for the HELP sample are not very different from those women in the NSFG that had a first child while being 19 to 38 years old. The last column of this table presents basic statistics for the HELP sample when observations are reweighted in order to match the distribution by age and year groups for 19-38 years old mothers in the NSFG. This adjustment makes the fraction of Hispanic and black very similar across the two samples while mean education also becomes closer.

Figure 1 compares the age distribution of women in the NSFG that gave birth when aged 19 to 38 years old to the age distribution of women in the HELP sample. This figure shows that the difference in mean age across these two groups is driven primarily by the group of women aged 19 to 21. This difference can be explained by the fact that some women in the NSFG group are having unplanned children and also by the fact that really young women would tend to delay their decision to seek help to achieve pregnancy.

6. Results

This section presents the main results of the empirical analysis. In essence, I will compare employment rates in month 21 for treated and control women in the HELP sample. The econometric model is represented by this simple OLS equation:[13]

$$Employed21_i = \alpha + \beta\, AnyChildren21_i + \gamma\, X_i + u_i$$

where the vector of covariates include black and Hispanic dummies, an indicator for insurance coverage of infertility treatments, year in which they sought help for the first time and the following variables measured in month 0: age, smoking status and years of education.

[13] Marginal effects results for probit and logit models are very similar to those obtained using OLS.

To gauge the potential importance of the problem of not having information on certain variables that may be simultaneously correlated with conception and labor supply, I run a number of regressions including separate sets of covariates. If the results were sensitive to the set of covariates added to the regression, this would raise some doubts about whether the identification strategy is consistently estimating the parameter of interest. Table 4 presents these regressions results.

In the model that includes all covariates (column 4), I estimate that having a first child younger than a year old decreases female employment by 26.3 percentage points. The results indicate that the estimated impact is remarkably robust to the set of covariates included in the regression. In particular, the estimated effect in a model with no covariates (column 1) is -0.253. That is, including the whole set of covariates, the estimated coefficient changes by just 1 percentage point or 4 percent of the estimated impact.

Column 5 presents linear probability estimates when observations are reweighted to match the age-year distribution for the sample of mothers in the NSFG that gave birth to their first child while aged 19 to 38 years old. The estimated impact is very close to the one obtained from original NSFG weights (column 4); this gives evidence that the obtained estimates could be generalized to the target population. Finally, in column 6 the model is augmented in order to check for varying treatment effects by age of mother and year in which they sought help to achieve pregnancy. While the treatment effect does not significantly change by age, the results suggest that the short-term effects of childbearing have decreased over time (this issue will be examined more deeply in subsection 8.2).

Women that have a child not only decide whether to have a job or not (the extensive margin) but also how many hours to work (the intensive margin). Unfortunately, the NSFG does not provide retrospective information on hours worked for women in the sample. Still, it provides information about whether the individual was working full-time or part-time and also the availability of maternity leave. Using this information, work status is determined among four categories (full-time, part-time, maternity leave and no job). Table 5 presents multinomial logit regression results of the impact of having a first child on work status. Having a child younger than a year old reduces the probability of working full-time by 43.1 percentage points while it raises the probability of being in the other three states. Interestingly, the increase in the probability of working part-time is quite small (4.8 percentage points).

7. Robustness of the Empirical Strategy

This section explores the robustness of the empirical strategy pursued. First, I try to identify which covariates can predict treatment and how much of the variation in the fertility variable is explained by these variables. Second, I test whether there are pre-treatment differences in the outcome variable (employment) between the treated and control groups. Finally, I check how robust the results are to changes in the specification of the econometric model.

To start with, I explore which variables in the data set predict early fertility success in the HELP sample. Table 6 shows that, as documented in the medical literature, female's age is one of the most important predictors of fertility. In this linear probability model, an increase in one year in the age of the woman decreases her expected probability of having a child by 1.6 percentage points. Smoking, also documented in the medical literature as having an effect on fertility, is a significant negative predictor of fertility success. Finally, Hispanics and more educated women are also more likely to be successful.

Even though there are several variables that can predict treatment, it should be noted that the adjusted R-squared is only 4.3 percent and there is much of the variation in the fertility variable that remains unexplained in this model.

Next, I tackle the issue of whether the significant differences in employment between treated and control women in month 21 can be interpreted as the effect of treatment or rather just heterogeneity in labor market attachment between groups. This is an important test for the empirical strategy pursued in the paper. Before presenting the regression results it is useful to look at Figure 2, which plots employment rates of the treatment and control groups for months -12 to 21 (again, month 0 corresponds to the month in which each woman first sought help to achieve pregnancy). Employment rates of both groups are quite similar for months -12 to 0 but they start diverging around month 3 and are far apart by month 21. The continuous decline in employment rates for the treated group corresponds to the fact that, as time goes by, more women are giving birth until in month 21 all of them had already given birth.

Table 7 presents results of regressions of employment status in month 0 (*Employed0*) on *AnyChildren21*. Several specifications are run in which I control for different sets of covariates in order to gauge the robustness of the results. The main conclusion from this table is that there are no statistically significant differences between the treated and control groups in employment rates in month 0.[14]

Finally, a number of additional regressions are run in order to check whether the results are robust to changes in the specification. First, I re-run the regressions whose results were presented in Table 4 but add an indicator for pregnancy in month 21. Second, the main independent variable *AnyChildren21* is replaced with another variable that equals to the number

[14]Similar results are obtained when regressing employment 12 months before seeking help to achieve pregnancy on *AnyChildren21*.

of children in month 21. Third, I replace *AnyChildren21* with two indicators for having one child or two children in month 21, respectively. Fourth, instead of running linear probability models of *Employed21* on *AnyChildren21*, I run probit and logit models using the same set of variables as in Table 4. In all these cases the estimated impacts are very similar to those reported in section 6.

8. Comparison to Estimates from NSFG and Census data

In his survey of the effect of children in the household, Browning (1992) concludes that studies that take fertility as exogenous typically found significant larger impact of fertility on female labor supply than those that treat it as endogenous and estimate simultaneous equations models. Angrist and Evans (1998) provides further evidence about this argument as they report that their 2SLS of the impact of having more than two children on female labor supply are statistically significantly smaller than their OLS estimates. This section compares estimates obtained using the HELP sample with those from similarly defined samples but without restricting them to women that sought help to become pregnant.

A problem faced in trying to replicate the HELP sample is that this dataset includes observations of fertility and labor supply for women that sought help to become pregnant at different points in time. This implies that in order to replicate the results from the HELP sample, I should construct comparable data sets with observations for individuals at different points in time: i.e., panel data or repeated cross-sections. Having this in mind, I compare estimates from the HELP sample to estimates from a panel data from the NSFG (in subsection 8.1) and to estimates from Census data for 1980 and 1990 (in subsection 8.2).

8.1 Comparison to estimates from NSFG panel data

I construct a panel data set from the NSFG Cycle 5 (called the NSFG panel data) following similar requirements to those used to construct the HELP sample. The unit of observation in this panel data is a woman-month. An observation is included in the NSFG panel data if the woman was aged 21 to 40 years old at that month, was childless or had children younger than a year old and was cohabitating or married.

As the HELP sample corresponds to a cross-section, in order to use the same source of variation when estimating both models, I construct a panel data set (called the HELP panel data) including for each individual in the HELP sample, observations for months -12 to 33 (remember that month 0 corresponds to when the individual first sought help to achieve pregnancy). As the goal is to estimate the impact of having a child younger than a year old, monthly observations for a woman are dropped when her child is older than this age. Finally, for women that by month 21 did not have a baby, monthly observations of later months are dropped if they give birth to a child.[15]

Table 8 presents summary statistics for the HELP panel data and the NSFG panel data. Mean values for key variables are similar and are only statistically significantly different for fractions employed and with children, calendar year and babies' age in months. Still, employment rates are not significantly different across samples once I condition for fertility status. With respect to differences in the fraction of women that have a child, this fact should be expected given that all individuals in the HELP panel data did not have children for months -12 up to (at least) month 7.

[15]Defining the sample in this way assures to have a balanced distribution of women with respect to their babies' age in months.

Linear probability estimates of the impact of having at least a child (younger than a year old) on the probability of having a job are presented in columns 1 and 3 of Table 8. In the first column, results are presented for the model estimated using the HELP panel data. The main independent variable is *AnyChildren* (equals 1 if the woman in that month had a child and 0 if not). The estimated impact (0.260) is very close to the estimates obtained in section 6. In the third column, results are presented for the same model estimated on the NSFG panel data. The key result of comparing columns 1 and 3 is that the estimated impact using the NSFG panel data (0.259) is notably similar to the one obtained using the HELP panel data.[16]

In order to gauge the robustness of these results, I estimate fixed-effects models on both panel data sets. Results are presented in columns 2 and 4 of Table 9. For the HELP panel data the estimated impact slightly decreases in absolute value to 0.234. In the case of the NSFG panel data, the estimated impact decreases in absolute value to 0.216. This result provides some evidence that women that have children tend to have lower employment rates in months previous to become pregnant. Still, both estimates are very similar and the t-value of the test of equality of coefficients is just -0.46.

Finally, I compare the estimated impact of having a child on work status (working full-time, part-time, maternity leave and no job) between these two panel data sets. Multinomial logit regression results are presented in Table 10. As this table shows, estimates of the marginal effect of having a child on the probability of being in each of the four work status categories are strikingly similar across the two data sets.

The fact that estimates from the HELP panel data are very similar to those from the NSFG panel suggests that the endogeneity problem of fertility is not very severe in regards

[16] The t-value of the test of equality of coefficients is 0.00.

to its effects on biasing estimates of treatment effects. Another explanation for this fact is that endogeneity does creates bias on estimates but both samples yield similar results because the bias is compensated by differences in treatment effects across samples (e.g. there may a positive bias in estimates on NSFG panel data but the true treatment effect in the NSFG panel data is larger than in the HELP panel data). However, given that statistics on observable characteristics across the two samples are very similar, the difference in treatment effects across samples should be originated entirely on differences in unobservables making the no endogeneity explanation more plausible.

8.2 Comparison to estimates using Census 1980 and 1990 data

In the HELP sample, fertility and other covariates are observed between 1972 and 1995. On average, these variables are observed in 1986 and the 10th and 90th percentiles correspond to years 1978 and 1993, respectively. In order to construct samples comparable to Census data, women in the HELP sample are assigned to two new samples, the EARLY and LATE HELP samples, depending on whether they sought help to become pregnant before or after 1985, respectively.[17]

Next, I construct two samples using the 5-percent Census Public Use Micro Samples for 1980 and 1990 (Ruggles et al, 2004). These samples (from now on PUMS 1980 and PUMS 1990) include married women aged 21 to 40 years old, childless or with children younger than a year old. Only married women are kept in the sample in order to get women that are "at risk" of having a child. To make these samples comparable to the HELP samples, I keep only married women in the HELP samples for the analysis performed in this subsection.[18]

[17]The threshold year is chosen as 1985 in order to construct two samples with roughly the same number of observations.

[18]Results obtained dropping the requirement of women in the HELP and PUMS samples to be married are

Table 11 presents descriptive statistics for the EARLY HELP, LATE HELP, PUMS 1980 and PUMS 1990 samples. In the case of the HELP samples, the variable *Employed* is an indicator that equals one if the woman had a job in month 21. For the PUMS samples, it equals 1 if the woman had a job during the previous week to the survey. The variables *AnyChildren*, *Age*, *Education*, *Hispanic* and *Black* are similarly defined in the four samples and are all measured in month 21 (for the HELP samples) or at the time of the survey (for the PUMS samples). *AnyChildren* equals 1 if the woman had at least a child. *Education* corresponds to the number of years of education. Finally, *Black* and *Hispanic* are dummy variables that equal to 1 if the woman belongs to each of these groups.

Results from Table 11 suggest that the 1980 and 1990 PUMS can be considered as sensible comparison data sets for the EARLY and LATE HELP samples, respectively. While women in the 1980 PUMS sample are surveyed in April 1980, those in the EARLY HELP sample are observed on average in June 1981. Similarly, while women in the 1990 PUMS sample are surveyed in April 1990, those in the LATE HELP sample are observed on average in January 1991. Moreover, basic statistics on education, fraction black and Hispanic are remarkably close. On the other hand, the fraction of women that have a child is significantly higher in the HELP samples. This should be expected given that presumably all women in the HELP samples wanted to have children. Finally, employment rates in the HELP samples, conditional on fertility status, are around 10 percent higher than the PUMS samples (this could stem from the fact that employment is not defined exactly in the same way in the NSFG compared to the Census).

Linear probability estimates of the impact of having a child (younger than a year old)

very similar to those presented in this subsection.

on employment are presented in Table 12. Comparing columns 1 and 2 we can see that the estimated impact is remarkably similar between the EARLY HELP sample and the 1980 PUMS sample (0.372 versus 0.365). Similarly, the estimated impact is also quite close when comparing the LATE HELP sample and then 1990 PUMS sample (0.182 versus 0.228). In both cases, t-tests of differences in the estimated impact cannot be rejected.

From this set of results two important conclusions can be drawn. First, the estimated impacts obtained for the sample for which I can identify an exogenous change in the fertility variable (the HELP sample) are very close to the estimates obtained using OLS on comparable samples from Census data for which I do not control for the endogeneity of the fertility variable (and they are also very close to estimates obtained using panel data from the NSFG as concluded in the previous subsection). Second, there is evidence of a significant reduction of about 40 to 50 percent in the short-term impact of childbearing on female labor supply in the 1980 to 1990 period.

9. Conclusions

This paper explores the issue of the causal effect of childbearing on female labor supply. This task is complicated by two factors. First, some researchers believe that women that have children at a certain age may have different baseline labor supply from women with similar observed characteristics that do not have children at that age (Browning, 1992). This expected unobserved heterogeneity across groups suggests the existence of bias in simple cross-section comparisons. As noted by Nakamura and Nakamura (1992), we can try to deal with this problem by adding to regressions of current labor supply on number of children, the lagged values of labor supply.

However, there is a second problem that complicates the estimation of the effect of childbearing on female labor supply and that cannot be solved by just using longitudinal data. This problem stems from the fact that the fertility decision may be endogenous to the woman and influenced by potential labor supply. Several studies starting with Rosenzweig and Wolpin (1980) have used the fact that having twins in the first birth changes (at least temporarily) family size. Angrist and Evans (1998) exploited the fact that parents typically prefer mixed-sex siblings in order to find exogenous variation to the fertility decision. Even though these papers have made a major contribution in answering the question posed, they are only able to estimate the effect of having a second or higher-order child.

In this paper, I estimate the short-term effects of having a first child on female labor supply. In order to deal with the problems of unobserved heterogeneity and endogeneity I restrict my attention to a group of women that sought help to achieve pregnancy. In this sample, all the women wanted to have children so that the problem of endogeneity is minimized. Moreover, as a major fraction of the fertility variable is random, I can suspect that results will not be contaminated by unobserved heterogeneity across groups. In fact, the attractiveness of the strategy pursued is that, focusing on this sample of women, I mimic an ideal social experiment in which for a group of women that wanted to have a child, some women are assigned children while others are not. I provide evidence in favor of the empirical strategy pursued as I find that pre-treatment labor supply is uncorrelated with subsequent fertility.

Following this empirical strategy, I estimate that having a first child younger than a year old reduces female labor supply by 26.3 percentage points. Interestingly, I obtain strong evidence that the estimates obtained using this strategy (which tackles the problem of the endogeneity of fertility) are very close to estimates derived from approaches that just assume the exogeneity of fertility.

Given that studies that assume the exogeneity of fertility typically find larger impacts of fertility on female labor supply than those that treat it as endogenous, a natural extension of this paper would be to attempt to understand why my empirical strategy reaches a different conclusion. One potential explanation is that there is not much selection when focusing women having a wanted first child after turning 19 years old. While Hotz, McElroy and Sanders (2005) found important differences in observable characteristics when comparing teen mothers to childless teenagers, for the NSFG and Census samples constructed in this paper, observable characteristics of women are quite similar when comparing mothers to childless women.

Another interesting question that is left unanswered in this paper is why fertility and baseline employment seem to be uncorrelated. There are many potential hypotheses that it is possible to lay out in order to predict problems with the identification strategy used here. For example, using this strategy I restricted the sample to women that are homogeneous in that all wanted to have a child at a certain point in time, but clearly they could differ in how much they wanted to have a child and this could be correlated with baseline labor force attachment.

A potential explanation for the evidence of subsequent fertility being uncorrelated with pre-treatment labor supply could be related to the fact that women in the HELP sample typically wait a number of months until they seek help to achieve pregnancy. This "waiting" scheme could be reducing the heterogeneity of individuals in the sample with respect to their baseline probability of being treated (where treatment refers to having a child). Individuals with very high probability of being treated receive treatment early and then they are not included in the sample if we restrict it to individuals that have not been treated after a certain period of time. As individuals in the sample have more similar probabilities of being treated, we tend to the ideal situation of random assignment which is characterized as one in which all individuals have *equal* probability of being treated. If evidence is found in favor of this hypothesis that "waiting" is a successful empirical strategy in the sense that it increases the similarity between the treated and control groups, then this same strategy could be applied to other evaluation problems where there is dynamic assignment of individuals to treatment.

References

Angrist, Joshua and Williams Evans (1998). “Children and their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.” *American Economic Review*, Vol. 88, No. 3 (June), pp. 450-477

Baird, Donna and Alex Wilcox (1985). “Cigarette Smoking Associated with Delayed Conception.” *Journal of the American Medical Association*, Vol. 253, No. 20 (September), pp. 2979-83

Blau, Francine and Adam Grossberg (1992). “Maternal Labor Supply and Children’s Cognitive Development.” *Review of Economic and Statistics*, Vol. 74, No. 3 (August), pp. 474-81

Bronars, Stephen and Jeff Grogger (1994). “The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment.” *American Economic Review*, Vol. 84, No. 5 (December), pp. 1141-1156

Browning, Martin (1992). “Children and Household Economic Behavior.” *Journal of Economic Literature*, Vol. 30, No. 3 (September), pp. 1434-1475

Cain, Glen and Martin Dooley (1976). “Estimation of a Model of Labor Supply, Fertility and Wages of Married Women.” *Journal of Political Economy*, Vol. 84, No. 4 (August), pp. S179-99

Desai, Sonalde, Lindsay Chase-Lansdale and Robert Michael. (1989) “Mother or Market? The Effects of Maternal Employment on the Intellectual Ability of 4-Year Old Children.” *Demography*, Vol. 26, pp. 545-61

Dunson, David, Donna Baird and Bernardo Colombo (2004). “Increased Infertility with Age in Men and Women.” *Obstetrics and Gynecology*, Vol. 103, No. 1 (January), pp. 51-56

Even, William (1987). “Career Interruptions Following Childbirth.” *Journal of Labor Economics*, Vol. 5, No. 2 (April), pp. 255-77

Fleisher, Belton and George Rhodes (1979). “Fertility, Women’s Wage Rate and Labor Supply.” *American Economic Review*, Vol. 69, No. 1 (March), pp. 14-24

Fuchs, Victor (1989). “Women’s Quest for Economic Equality.” *Journal of Economic Perspectives*, Vol. 3, No. 1 (Winter), pp. 25-41

Goldin, Claudia. (1990) “Understanding the Gender Gap.” *New York: Oxford University Press*

Goldin, Claudia and Solomon Polachek (1987). “Residual Differences by Sex: Perspectives on the Gender Gap Earnings.” *American Economic Review*, Vol. 77, No. 2 (May), pp. 143-51

Gronau, Reuben. (1973) “The Intrafamily Allocation of time: the Value of the Housewives’ Time.” *American Economic Review*, Vol. 63, pp. 634-51

Gronau, Reuben (1988). “Sex-Related Wage Differentials and Women’s Interrupted Careers - the Chicken or the Egg.” *Journal of Labor Economics*, Vol. 6, No. 3 (July), pp. 277-301

Heckman, James. (1974) “Shadow Prices, Market Wages and Labor Supply.” *Econometrica*, Vol. 42, pp. 679-94

Heckman, James and V. Joseph Hotz (1989). “Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: the Case of Manpower Training.” *Journal of the American Statistical Association*, Vol. 84, No. 408 (December), pp. 862-74

Heckman, James and Jeffrey Smith (1995). “Assessing the Case for Social Experiments.” *Journal of Economic Perspectives*, Vol. 9, No. 2 (Spring), pp. 85-110

Heckman, James and Robert Willis (1977). “A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women.” *Journal of Political Economy*, Vol. 85, No. 1 (February), pp. 27-58

Henshaw, Stanley (1998). “Unintended Pregnancy in the United States.” *Family Planning Perspectives*, Vol. 30, No. 1 (January), pp. 24-29

Hotz, V. Joseph, Susan McElroy and Seth Sanders (2005). “Teenage Childbearing and its Life Cycle Consequences: Exploiting a Natural Experiment.” *Journal of Human Resources*, Vol. 40, No. 3 (Summer), pp. 683-715

Jacobsen, Joyce, James Wishart Pearce III and Joshua Rosenbloom (1999). “The Effects of Childbearing on Married Women Labor Supply and Earnings: Using Twin Births as a Natural Experiment.” *Journal of Human Resources*, Vol. 34, No. 3 (Summer), pp. 449-474

Korenman, Sanders and David Neumark (1992). “Marriage, Motherhood and Wages.” *Journal of Human Resources*, Vol. 27, No. 2 (Spring), pp. 233-55

Lehrer, Evelyn (1992). “The Impact of Children on Married Women’s Labor Supply: Black-White Differentials Revisited.” *Journal of Human Resources*, Vol. 27, No. 3 (Summer), pp. 422-44

Miller, Amalia. (2005) “The Effect of Motherhood Timing on Career Path.” *Manuscript, University of Virginia*

Mincer, Jacob. (1962) “Labor Force Participation of Married Women: a Study of Labor Supply.” *In H. Gregg Lewis ed., Aspects of Labor Economics. Princeton, NJ: Princeton University Press*, pp. 63-105

Nakamura, Alice and Masao Nakamura (1992). "The Econometrics of Female Labor Supply and Children." *Econometrics Review*, Vol. 11, No. 1, pp. 1-71

Rosenzweig, Mark and Kenneth Wolpin (1980). "Life-Cycle Labor Supply and Fertility: Causal Inferences from Households Models." *Journal of Political Economy*, Vol. 88, No. 2 (April), pp. 328-48

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. (2004) "Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]." *Minneapolis, MN: Minnesota Population Center* [producer and distributor]

Schultz, T. Paul. (1978) "The Influence of Fertility on Labor Supply of Married Women: Simultaneous Equation Estimates." *Research in Labor Economics, Editor Ronald Ehrenberg*, pp. 273-351

Shapiro, David and Frank Mott (1994). "Long-Term Employment and Earnings of Women in Relation to Employment Behavior Surrounding the First Birth." *Journal of Human Resources*, Vol. 29, No. 2 (Spring), pp. 248-275

Stafford, Frank (1987). "Women's Work, Sibling Competition and Children's School Performance." *American Economic Review*, Vol. 77, No. 5 (December), pp. 972-80

Teachman, Jay, Lucky Tedrow and Kyle Crowder. (1998) "Evaluation of the 1995 NSFG." *Manuscript, Western Washington University*

Figure 1: Distribution of women by age at: a) first birth (mothers with first birth when aged 19 to 38), b) which first sought help to become pregnant (women in the HELP sample)

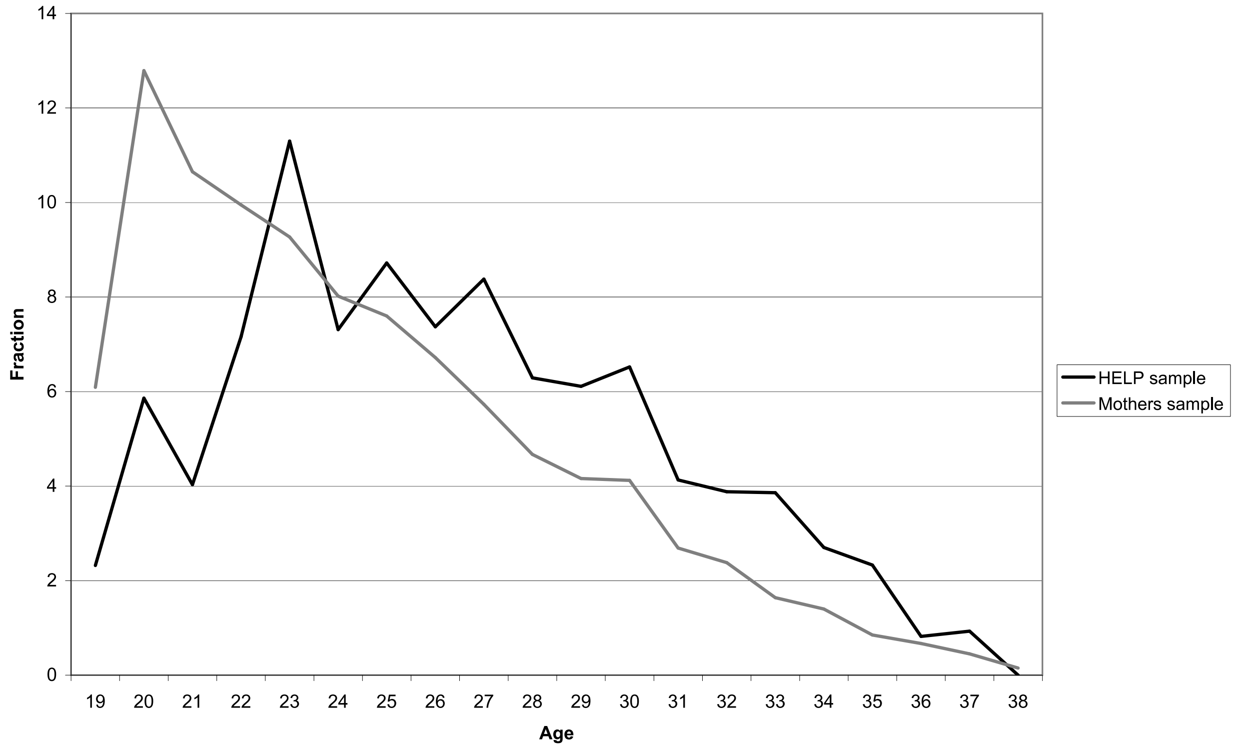

Figure 2: Employment rates by month for women that: a) had a baby by month 21 b) did not have a baby by month 21 - HELP sample

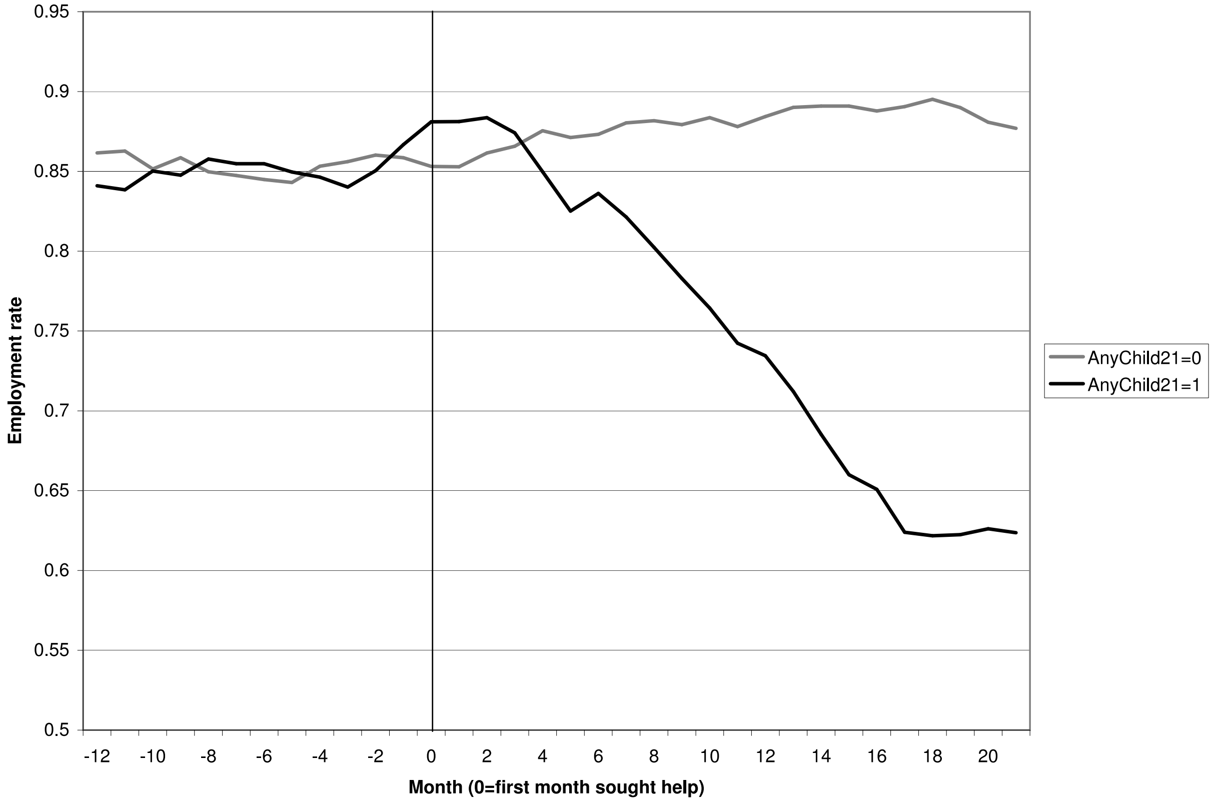

Table 1: Algorithm employed to construct the HELP sample

Step	Number of remaining observations
1- Start with the whole NSFG sample	10,847
2- Drop women that did not seek help to get pregnant	895
3- Drop women that sought help for the first time less than 21 months before the time of the interview	788
4- Drop women that were younger than 19 or older than 38 when sought help for the first time	745
5- Drop women that had already a child when sought help for the first time	553
6- Drop women that had adopted or step children when they first sought help to become pregnant	536
7- Drop women that were pregnant at some point of the month in which they sought help for the first time [a]	500
8- Drop a woman with missing information in the insurance coverage variable	499

[a] This group could include women that got pregnant right after seeking help for the first time (what occurred in the same month), or that were pregnant at the time when they sought help but did not know it. In fact 23 of the 36 reported as being pregnant the same month that first sought help to get pregnant got pregnant exactly in that month or in the previous one.

Table 2: Descriptive statistics – HELP sample

Variable	All women	Treated *AnyChildren21*=1[a]	Control *AnyChildren21*=0[a]
	Means and (standard deviations)		
Employed21 (=1 if employed in month 21)	0.798 (0.402)	0.624 ** (0.484)	0.877 (0.329)
Employed0 (=1 if employed in month 0)	0.862 (0.345)	0.881 (0.324)	0.853 (0.354)
Employed_12 (=1 if employed in month -12)	0.855 (0.352)	0.841 (0.366)	0.862 (0.345)
OwnChildren21 (number of own children in month 21)[b]	0.323 (0.491)	1.036 ** (0.185)	0.000 (0.000)
AnyOtherChildren21 (=1 if had adopted or step children in month 21)	0.020 (0.141)	0.005 (0.073)	0.027 (0.162)
Age0 (age in month 0)	26.3 (4.3)	25.9 (4.7)	26.5 (4.1)
Year0 (year in month 0 normalized as 1970=0)	14.7 (5.7)	15.0 (6.1)	14.5 (5.5)
Education0 (years of education in month 0)	13.6 (2.5)	13.8 (2.6)	13.5 (2.4)
Hispanic (=1 if Hispanic)	0.069 (0.254)	0.113 * (0.317)	0.050 (0.217)
Black (=1 if black)	0.087 (0.281)	0.078 (0.267)	0.091 (0.287)
Married0 (=1 if married in month 0)	0.884 (0.320)	0.884 (0.320)	0.884 (0.321)
Smoke0 (=1 if smoked in month 0)	0.370 (0.483)	0.286 * (0.452)	0.408 (0.492)
InsuranceCovered (=1 if insurance covered infertility treatments)	0.789 (0.408)	0.792 (0.406)	0.787 (0.409)
Number of observations	499	164	335

*,**: Significantly different from the mean of the control group at the 5%, 1% significance level.
[a] *AnyChildren21*=1 if the woman had at least an own child in month 21.
[b] There are six women that had two children. Five had given birth to twins and one had given birth twice.

Table 3: Comparison of HELP sample with mothers in the NSFG

Variables [a]	Means and (standard deviations)			
	NSFG – All Mothers [b]	NSFG –Mothers with first birth when aged 19 to 38 [c]	HELP sample	HELP sample re-weighted [d]
Age	22.9 ** (4.9)	24.5 ** (4.2)	26.3 (4.3)	24.5 (4.2)
Year	14.0 * (7.0)	15.0 (6.4)	14.7 (5.7)	14.7 (6.1)
Employed_12 [e]	N/A	0.787 ** (0.409)	0.855 (0.352)	0.835 (0.371)
Education	12.3 ** (2.6)	12.8 ** (2.5)	13.6 (2.5)	13.1 (2.4)
Hispanic	0.125 ** (0.331)	0.112 ** (0.316)	0.069 (0.254)	0.098 (0.298)
Black	0.150 ** (0.357)	0.110 (0.312)	0.087 (0.281)	0.112 (0.315)
Married	0.702 ** (0.457)	0.782 ** (0.413)	0.884 (0.320)	0.857 (0.350)
Smoke	0.336 (0.472)	0.329 (0.470)	0.370 (0.483)	0.420 (0.494)
Number of observations	6,911	5,150	499	499

*,**: Significantly different from the mean of the HELP sample at the 5%, 1% significance level.

[a] Variables for the two samples of mothers (second and third column) are measured at the month in which they gave birth to their first child (except from *Employed_12*). Variables for women in the HELP sample (last column) are measured in the month in which they first sought help to get pregnant (except from *Employed_12*).

[b] This sample is constructed selecting in the NSFG sample all women that had at least one child.

[c] Includes all women in the NSFG sample that gave birth their first child while being aged 19 to 38 years old.

[d] Statistics are computed re-weighting observations in the HELP sample in order to match the distribution by age and year groups in the sample of mothers in NSFG with first birth when aged 19 to 38.

[e] *Employed_12* equals to 1 if the woman was employed 12 months before her first birth (third column) or 12 months before she first sought help to get pregnant (fourth column). In the case of the NSFG – All mothers sample (second column) this variable cannot be computed as work status is asked in the survey only for months after the woman reaches 18 years old.

Table 4: Linear probability estimates. Impact of having a first child on employment. Dependent variable is *Employed21* - HELP sample

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
AnyChildren21	-0.253 (0.045)	-0.254 (0.044)	-0.261 (0.043)	-0.263 (0.043)	-0.283 (0.047)	-0.812 (0.265)
AnyChildren21 Age0*	-	-	-	-	-	0.011 (0.011)
AnyChildren21 Year0*	-	-	-	-	-	0.017 (0.008)
Age0	-	0.007 (0.005)	0.000 (0.005)	0.000 (0.005)	-0.005 (0.006)	-0.004 (0.006)
Year0	-	0.010 (0.004)	0.010 (0.004)	0.011 (0.004)	0.014 (0.004)	0.004 (0.004)
Smoke0	-	-	-0.045 (0.042)	-0.046 (0.041)	-0.025 (0.047)	-0.044 (0.040)
Education0	-	-	0.021 (0.007)	0.020 (0.007)	0.030 (0.008)	0.019 (0.007)
Hispanic	-	-	-0.131 (0.071)	-0.138 (0.069)	-0.087 (0.075)	-0.150 (0.070)
Black	-	-	0.014 (0.050)	-0.016 (0.051)	-0.149 (0.076)	-0.020 (0.051)
Married0	-	-	-	-0.089 (0.038)	-0.148 (0.046)	-0.104 (0.037)
InsuranceCovered	-	-	-	0.109 (0.049)	0.177 (0.055)	0.099 (0.048)
Constant	0.877 (0.019)	0.563 (0.126)	0.459 (0.156)	0.449 (0.149)	0.395 (0.162)	0.675 (0.169)
Adj. R-squared	0.0854	0.1190	0.1467	0.1666	0.2151	0.1904
Number of observations	499	499	499	499	499	499

For regressions (1) through (4) and (6), observations are weighted using weights from the NSFG. For regression (5) observations are re-weighted in order to match the distribution by age and year groups for mothers in the NSFG that had their first birth while aged 19 to 38. The mean of *Employed21* using NSFG weights is 0.798. For the re-weighted sample, the mean of *Employed21* is 0.771. Standard errors in parenthesis.

Table 5: Multinomial logit estimates. Impact of having a first child on work status HELP sample

	Marginal effects of changing *AnyChildren21* from 0 to 1 (and standard errors)	
	HELP sample	HELP sample re-weighted [a]
No Job	0.291 (0.047)	0.314 (0.052)
Maternity leave	0.092 (0.027)	0.083 (0.025)
Part-time [a]	0.048 (0.027)	0.054 (0.036)
Full-time [a]	-0.431 (0.050)	-0.450 (0.055)
Number of observations	499	499
Log pseudo-likelihood value	-374.30	-381.25
Pseudo R-squared	0.1738	0.2000

The dependent variable has four categories: no job, maternity leave, part-time and full-time. Covariates: *Age0*, *Year0*, *Smoke0*, *Education0*, *Hispanic*, *Black*, *Married0*, *InsuranceCovered*. Standard errors in parenthesis.
[a] Observations are re-weighted in order to match the distribution by age and year groups in the sample of mothers in the NSFG with first birth when aged 19 to 38.

Table 6: Linear probability estimates. Predicting fertility using selected covariates. Dependent variable is *AnyChildren21* - HELP sample

Independent variable	
Age0	-0.016 (0.006)
Year0	0.007 (0.005)
Smoke0	-0.102 (0.046)
Education0	0.013 (0.010)
Hispanic	0.198 (0.082)
Black	-0.052 (0.066)
Married0	-0.012 (0.051)
InsuranceCovered	0.022 (0.054)
Constant	0.472 (0.184)
Adj. R-squared	0.0427
P-value of F-test of joint significance	0.0020
Number of observations	499

The mean of *AnyChildren21* is 0.312. Standard errors in parenthesis.

Table 7: Linear probability estimates. Explaining employment in month 0 using fertility status in month 21. Dependent variable is *Employed0* - HELP sample

Independent variable	(1)	(2)	(3)	(4)	(5)
AnyChildren21	0.028 (0.035)	0.027 (0.035)	0.025 (0.034)	0.022 (0.034)	0.006 (0.041)
Age0	-	0.005 (0.004)	0.002 (0.005)	0.002 (0.004)	-0.003 (0.005)
Year0	-	0.007 (0.004)	0.007 (0.004)	0.008 (0.004)	0.011 (0.004)
Smoke0	-	-	-0.019 (0.041)	-0.020 (0.040)	0.011 (0.045)
Education0	-	-	0.010 (0.006)	0.009 (0.006)	0.021 (0.009)
Hispanic	-	-	-0.073 (0.067)	-0.075 (0.070)	-0.013 (0.067)
Black	-	-	0.003 (0.047)	-0.024 (0.048)	-0.107 (0.068)
Married0	-	-	-	-0.071 (0.037)	-0.093 (0.045)
InsuranceCovered	-	-	-	0.119 (0.052)	0.123 (0.056)
Constant	0.853 (0.022)	0.629 (0.112)	0.580 (0.142)	0.553 (0.146)	0.466 (0.164)
Adj. R-squared	0.0014	0.0242	0.0333	0.0592	0.0837
Number of observations	499	499	499	499	499

For regressions (1) through (4) observations are weighted using weights from the NSFG. For regression (5) observations are re-weighted in order to match the distribution by age and year groups for mothers in the NSFG that had their first birth while aged 19 to 38. The mean of *Employed0* using NSFG weights is 0.862. For the re-weighted sample, the mean of *Employed0* is 0.847. Standard errors in parenthesis.

Table 8: Descriptive statistics - HELP panel data and NSFG panel data

	Means (and standard deviations)	
Data	NSFG – Cycle 5 (1995)	NSFG – Cycle 5 (1995)
Sample	HELP panel data	NSFG panel data
Unit of observation	Woman-month	Woman-month
Employed	0.841 * (0.365)	0.808 (0.394)
AnyChildren	0.087 ** (0.281)	0.171 (0.376)
Age	27.0 (4.4)	27.0 (4.4)
Education	14.0 (2.5)	14.0 (2.6)
Married	0.873 (0.333)	0.891 (0.311)
Smoke	0.361 (0.480)	0.327 (0.469)
Year (1970=0)	14.8 ** (5.5)	15.6 (5.8)
Hispanic	0.059 (0.236)	0.066 (0.248)
Black	0.076 (0.264)	0.056 (0.230)
Baby age in months (for women with babies)	5.5 ** (3.5)	6.1 (3.7)
Number of observations	19,743	237,751
Number of women	467 [a]	4,786

*,**: Significantly different from the mean of the NSFG panel data at the 5%, 1% significance level.

[a] There are 32 women that are included in the HELP sample but that answered the NSFG less than 33 months after seeking help to get pregnant. These women are not included in this panel data set as it includes monthly observations for each woman in the 33 months after seeking help to get pregnant.

Table 9: Impact of having a first child on employment. Dependent variable is *employed*. HELP panel data and NSFG panel data

Data	NSFG – Cycle 5 (1995)		NSFG – Cycle 5 (1995)	
Sample	HELP panel data		NSFG panel data	
Unit of observation	Woman-month		Woman-month	
Regression model	OLS	Fixed effects	OLS	Fixed effects
AnyChildren	-0.260 (0.036)	-0.234 (0.034)	-0.259 (0.010)	-0.216 (0.010)
Pregnant	-0.092 (0.020)	-0.065 (0.017)	-0.074 (0.008)	-0.050 (0.007)
Age	0.003 (0.004)	0.004 (0.007)	0.000 (0.002)	-0.003 (0.003)
Education	0.011 (0.005)	0.032 (0.023)	0.005 (0.001)	0.029 (0.011)
Married	-0.037 (0.033)	-0.055 (0.025)	0.033 (0.011)	-0.020 (0.010)
Smoke	-0.026 (0.037)	0.091 (0.060)	0.016 (0.003)	0.007 (0.022)
Year (1970=0)	0.008 (0.003)	-	-0.073 (0.019)	-
Hispanic	-0.103 (0.066)	-	0.024 (0.018)	-
Black	-0.035 (0.048)	-	-0.020 (0.010)	-
Constant	0.568 (0.143)	0.321 (0.287)	0.576 (0.045)	0.616 (0.158)
Adj. R-squared	0.0813	0.6666	0.0880	0.5761
Number of observations	19,743	19,743	237,751	237,751

Fixed effects model for the HELP panel data includes dummies for individuals and months relative to the first time they sought help to become pregnant. Fixed effects model for the NSFG panel data includes dummies for individuals and calendar years. Observations clustered by individual. Standard errors in parenthesis.

Table 10: Multinomial logit estimates. Impact of having a first child on work status. HELP panel data and NSFG panel data

Data	NSFG – Cycle 5 (1995)	NSFG – Cycle 5 (1995)
Sample	HELP panel data	NSFG panel data
Unit of observation	Woman-month	Woman-month
	Marginal effects of changing *AnyChildren* from 0 to 1 (and standard errors)	
No Job	0.253 (0.038)	0.246 (0.009)
Maternity leave	0.115 (0.015)	0.116 (0.004)
Part-time [a]	0.010 (0.021)	0.005 (0.006)
Full-time [a]	-0.378 (0.036)	-0.368 (0.009)
Number of observations	19,743	237,751
Log pseudo-likelihood value	-13887.78	-195,128.98
Pseudo R-squared	0.0973	0.0904

[a] The dependent variable has four categories: no job, maternity leave, part-time and full-time. Covariates: *Age*, *Year*, *Smoke*, *Education*, *Hispanic*, *Black*, *Married*. Observations clustered by individual. Standard errors in parenthesis.

Table 11: Descriptive statistics - HELP and PUMS samples

	Means (and standard deviations)			
Sample	EARLY HELP	1980 PUMS	LATE HELP	1990 PUMS
Sample description	Married women in HELP sample that sought help before 1985	Married women aged 21 to 40 childless or with children younger than 1 year old	Married women in HELP sample that sought help on or after 1985	Married women aged 21 to 40 childless or with children younger than 1 year old
Variables measured	21 months after seeking help for the first time	in 1980	21 months after seeking help for the first time	in 1990
	(1)	(2)	(3)	(4)
Observation year	1981.5 ** (3.5)	1980.3 (0.0)	1991.0 ** (2.6)	1990.3 (0.0)
Employed	0.731 (0.443)	0.726 (0.446)	0.854 * (0.353)	0.796 (0.403)
AnyChildren	0.289 ** (0.453)	0.158 (0.364)	0.358 ** (0.479)	0.128 (0.334)
Age	26.1 ** (3.1)	27.2 (4.9)	30.2 ** (4.1)	29.3 (5.3)
Education	13.4 (2.4)	13.4 (2.6)	14.1 (2.6)	13.9 (2.5)
Hispanic	0.050 (0.218)	0.053 (0.223)	0.081 (0.272)	0.077 (0.266)
Black	0.065 (0.246)	0.061 (0.239)	0.069 (0.253)	0.061 (0.240)
Number of observations	216	287,292	224	301,371

*,**: Significantly different from the mean of the PUMS comparable samples at the 5%, 1% significance level. This means that the EARLY HELP sample is compared against 1980 PUMS and LATE HELP against 1990 PUMS.

Table 12: Linear probability estimates. Impact of having a first child on employment. HELP and PUMS samples

Sample	EARLY HELP	1980 PUMS	LATE HELP	1990 PUMS
Sample description	Married women in HELP sample that sought help before 1985	Married women aged 21 to 40 childless or with children younger than 1 year old	Married women in HELP sample that sought help on or after 1985	Married women aged 21 to 40 childless or with children younger than 1 year old
Variables measured	21 months after seeking help for the first time	in 1980	21 months after seeking help for the first time	in 1990
Mean of dependent variable – *Employed*	0.731	0.726	0.854	0.796
Independent variables	(1)	(2)	(3)	(4)
AnyChildren	-0.372 (0.072)	-0.365 (0.002)	-0.182 (0.055)	-0.228 (0.003)
Age	0.007 (0.012)	-0.004 (0.000)	-0.011 (0.007)	-0.001 (0.000)
Education	0.024 (0.013)	0.030 (0.000)	0.021 (0.008)	0.031 (0.000)
Hispanic	0.042 (0.096)	-0.047 (0.004)	-0.259 (0.099)	-0.087 (0.004)
Black	-0.024 (0.109)	-0.017 (0.003)	0.032 (0.062)	-0.033 (0.004)
Year0	0.014 (0.011)	-	-0.002 (0.008)	-
Constant	0.196 (0.270)	0.489 (0.006)	1.011 (0.272)	0.417 (0.007)
Adj. R-squared	0.2145	0.1222	0.1253	0.0814
Number of observations	216	287,292	224	301,371

Standard errors in parenthesis.