1 Introduction

Estimating the effect children have on women's labor supply has a long and varied history in economics.[1] One of the main insights from this research is that, given the dynamic nature of the problem, estimating the effect of fertility on the dynamics of female labor supply is an extremely difficult problem. Both the number and timing of children are variables that are controlled, at least in part, by women. Heterogeneity in the demand for children and the taste for work will affect women's investment in human capital long before the birth of any children, which will in turn affect labor supply prior to the birth of children. In addition, children will have both a direct and indirect effect on a women's supply of labor to the market after they are born. The direct effect is a result of children raising the non-market opportunity of women thereby reducing their labor supply. The indirect effect derives from the fact that when women leave the labor market to have children they experience a fall in their investment in human capital and therefore in their propensity to work in the future. Estimating models that are general enough to incorporate all of these elements not only requires panel data, but also requires the use of dynamic response models along with information identifying exogenous variation in the demand for children and the supply of labor.

The goal of this paper is to estimate the effect of children on the dynamic labor supply of women using panel data from the 1979 National Longitudinal Survey of Youth (NLSY79). We use Markov chain Monte Carlo (MCMC) methods to estimate the dynamic labor supply behavior of women, while simultaneously estimating the determinants of fertility. From these estimates we then decompose the effect of fertility on labor supply into the direct and indirect components and examine how these effects vary by personal characteristics. We also use these estimates to examine the effect of expected future fertility on the current labor supply of women.

Our paper extends the recent literature on the effect of children on female labor supply in a number of directions. First, following the suggestion in Browning (1992), we divide our analysis into a pre-birth and a post-birth period and examine the effect of children on women's labor supply in both periods. As far as we know, we are the first to explicitly examine the effect of expected future fertility on women's labor supply. Second, for multiple children, we estimate separate effects for each child. Previous research on the causal effect of children on labor supply has exploited exogenous variation in the second or third child to measure the causal effect of children (e.g. Carrasco, 2001; Angrist and Evans, 1998; Rosenzweig and Wolpin 1980). If, as seems likely to us, the effect of children is declining in the number of children, these previous estimates of the effect of children will understate the effects of the first child. Our estimates should provide a more precise measure of the effect of the first child on a woman's labor supply as well as showing how the effect varies with the number of children. Third, following the suggestions in Browning (1992) and Nakamura

[1] Browning (1992) and Nakamura and Nakamura (1992) both provide reviews of the history of this liturature.

and Nakamura (1992) we decompose the total effect of children into the direct and indirect effect and examine how these effects change over time. Fourth, in contrast to the previous papers that model labor market participation using two states (Carrasco, 2001; Angrist and Evans, 1998; Hyslop, 1999; Chib and Jeliazkov, 2003) or three states (Buddlemeyer and Voicu, 2003) we model labor market participation using four states: full time, full time part year, part time and nonparticipation. If one of the primary effects of children on women's labor supply is through the number of hours worked, this four state model allows us to better capture the dynamics of labor market participation. This may be particularly important around the birth of a child (Blank, 1989 and 1998). Finally, we explicitly model and estimate the determinants of fertility.

We also extend the literature on estimating dynamic discrete choice models in two directions. First, through the use of a more general four state multi-period multinomial probit model with auto-correlated error terms to model labor force decisions, along with a probit model with state dependence and an auto-correlated error term to model fertility decisions. Second, by employing a mixed effects structure that combines fixed and random effects. The mixed-effect structure accomplishes several goals: a) it allows time-invariant individual heterogeneity to simultaneously affect participation and fertility behavior; b) it allows both time invariant tastes for children and the timing of births to vary across individuals; and c) it captures heterogeneity in the dynamics of the level of labor market involvement around birth. The cross-equation correlation of the random coefficients is important for the endogenous modelling of participation and fertility decisions and for the multinomial probit representation of labor market decisions.

We estimate this model using data on women from the NLSY79. Because the NLSY79 data consist of individuals who were 14 to 21 in 1979, we have a fairly complete picture of both these women's labor market and fertility dynamics. In addition, the NLSY79 data contain a rich set of family background variables, such as parent's labor market status and education, and sibling's fertility. These variables, along with the panel structure of the data, help us identify heterogeneity in the taste for and the cost of children across the women in our data.

We find that children affect the labor market dynamics of women in post-birth periods and that expected future fertility affects the labor market dynamics of women in the pre-birth period, with much of the difference in labor market participation prior to birth due to unobserved heterogeneity. We also find that the direct and indirect effects of children are of similar size immediately after birth, but that the indirect effect declines quickly over time, suggesting that the loss of human capital resulting from a woman exiting the labor market to have a child is fairly small. The direct effect declines with the age of children. Both the direct and indirect effects are smaller for the second child than for the first. We find that the size of the direct and indirect effects vary with personal characteristics. Relative to white women, the direct effect of children are smaller for black and Hispanic women. For the first child both the direct and indirect effects are smaller for more educated women, but the opposite is true for the second child. Finally, we find that our four state model of labor force participation better captures women's labor force dynamics. There are significant qualitative difference between working full time part year and working full time or part time.

The remainder of the paper is structured as follows. Section 2 contains a theoretical background and a description of the data. The empirical specification, the estimation method, and the estimation results are presented in section 3. In section 4 we describe our pre-birth and post-birth simulations that we use to study the effects of children, we present the formal definition of the direct and indirect effects, and we discuss selected results. Concluding remarks follow.

2 Theoretical background and data

The existing literature on women's labor supply suggests two basic facts. First, children have a negative effect on women's labor supply and second, this effect fades away as children grow older. Many different factors account for these findings. Women's physical capacity for performing market work is sharply diminished during the period surrounding birth; rearing children is time-intensive and initially involves a taxing personal and family adjustment process. As children grow, caring for them requires less time and more market choices become available. Families also adjust to the new state. This effect can be formalized and studied using various models. The neoclassical labor supply theory assumes that individuals make employment decisions by comparing the utility of working with the utility of not working. The value of working relative to not working increases as the child ages (Mincer 1962, Heckman 1980, Leibowitz, Klerman, and Waite 1992). In a job-search framework (Mortensen, 1986) the value of time in alternative (non-work) states can be assumed to vary with the number of children and their ages. The birth of the child will raise the value of time in alternative use and, through it, the reservation wage. As a result, the probability of employment will decline.

The second fact is that sequential employment decisions of women are correlated. As a result, labor market interruptions are associated with a decline in the probability of employment in subsequent periods. Heckman and Willis (1977) have defined two sources of dependence: a) unobserved heterogeneity generated by different preferences, and b) state dependence. There are multiple sources of state dependence. Human capital theory predicts that skills accumulated through experience raise the probability of working in the future. Fixed costs of entering the labor force (search costs, for example) make future participation more likely for individuals already working. Job matching models where employers and employees learn about the quality of the match induce state dependence even if investment in firm-specific human capital does not take place.

The presence of state dependence is very important in studying the effect of fertility on labor supply. Maternity-related work interruptions lead to a decline in human capital investment, and possibly to depreciation of the human capital stock. Also search costs and information on the quality of the match may be lost. Longer interruptions are more detrimental in the human capital framework.

These two phenomena provide a framework for studying the effect of children on women's labor supply. They imply first that a women's investment in human capital prior to entering the labor market, as well as her labor supply in every period, will be affected by expectations about future fertility. They also imply that a women's post-birth employment likelihood should be driven by the increased demand placed on her time by newborn children and by the length of the maternity-related work interruption. The first component should be fading with a child's age. The second component should be stronger the longer the interruption. In this paper we use the broad labels direct and indirect effects for these two mechanisms. The measures of the direct and the indirect effect depend on the events for which they are measured. In the following sections we describe formally the statistical model we use to estimate the effects of children on women's labor both prior to and after a birth and provide definitions of the direct and indirect effects for particular events.

The data we use in this analysis come from the 1979 National Longitudinal Survey of Youth (NLSY79). The NLSY79 contains a representative sample of individuals between 14 and 21 years old in 1979. These individuals are surveyed every year between 1979 and 1994, and every other year thereafter. We use data from the nonmilitary sample of the 1979-1994 surveys. We restrict the sample to women who are not married and are childless in 1979, get married after 1979, and remain married until 1994, only have children while married, and only have biological children in the household over the period of our data.[2] We exclude women who live on a farm larger than 100 acres at any point in the period. In order to abstract from the trade-off between schooling and working, we only consider a woman at risk to work or to have a child once she has been out of school for at least 18 months continuously.[3] Finally, we require at least five years of data for each woman.[4] This selection results in an unbalanced panel of 880 women observed between 1979 and 1994. Through this selection we ensure that initial conditions are identical across individuals in the sample–all women are childless and have no labor market history in the first period.

The following series of tables and figures provide an overview of some of the main variables used in the analysis. Table 1 presents summary statistics for variables that do not change over time, what we refer to as background variables. The numbers in this table are for all 880 women who are in our sample. Roughly 70 percent of our sample is white, with the remainder evenly split between Hispanic and black. Respondents' fathers tend to work full time while about 1/3 of their mothers work full time and 1/3 do not work at all.[5]

[2] This latter criteria eliminates women who adopt children or who marry men who have children who live with them.

[3] Once a women leaves school we consider her still at risk even if she returns to school.

[4] Relaxing this requirements adds 13 more women to the sample and has no effect on our results.

[5] Information about parents are asked in the 1979 survey. Information about labor market status refers to parents' labor market status in 1978.

On average, women in the sample have slightly over three siblings, the siblings have 13 years of schooling on average and siblings' average number of children is 1.3.[6]

Table 2 presents summary statistics, by year, for the time varying variables used in the analysis.[7] Column 2, which presents the number of women considered at risk in a given year, shows the unbalanced nature of the data. In 1979 only 169 women are considered at risk, by 1988 all 880 women are considered at risk. This table shows that no women had any children prior to 1981. In 1981 two percent of the at-risk women in the sample had children age 0 to 1. By 1993, seven percent of the sample had children 10 years old or older, while 37 percent of the sample had children between six and nine years old.

Figure 1 presents the labor market status for all women in the sample in a given year. We measure labor market status using hours worked in a given year, weeks worked in a given year, and income. To be considered working a woman must have both positive hours worked and positive income. Women who worked more than 1750 hours in a year are classified as full time. Women who work between zero and 1750 hours, but who work on average more than 35 hours a week, are considered full time part year. Women who work between zero and 1750 hours, but who work on average less than 35 hours a week, are considered part time.[8] Women who work zero hours or who have zero income are considered not working. Figure 1 shows that the percentage of women working full time and working full time part year declines over time while the percentage of women not working rises. The percentage of women working part time remains fairly constant.

Figure 2 plots the birth rate for women in the sample in a given year. The birth rate is simply the proportion of at risk women who give birth to a child in the year. The birth rate rises fairly dramatically in the early years,[9] seems to reach a peak around 1990, and appears to decline slightly toward the end of the period.

Figures 3-5 present the labor force status for women in the years surrounding the birth of a child. Figure 3 presents these numbers for women who have one child by 1993, Figure 4 for women who have two children by 1993, and Figure 5 for women who have three children by 1993. In these figures 0 on the X-axis represents the year of the child's birth, the negative numbers represent the years prior to the birth and the positive numbers represent the years after the birth.

[6] Information on siblings comes from questions asked in the 1992 survey. We use infomation on the age of the children to calculate when the sibling first had the child. All data on years of schooling are truncated at 20 years in the NLSY79.

[7] All of our income numbers have been deflated using the CPI-U and are in 1979 dollars. We imputed income for a number of cases where it was missing. Our exact imputation procedure is described in the data appendix.

[8] We measure hours worked using information on hours worked per week avaliable in the NLSY79. We compute the average hours worked per week by summing the hours worked per week for the year and dividing by the number of weeks worked during the year. Due to problems with missing and invalid values we are forced to impute hours worked for a few women. We describe the imputation process in the data appendix.

[9] This is true partly by construction since women have zero children when they enter the panel.

The top panel in Figure 4 shows this information around the birth of the first child while the bottom panel presents this information around the birth of the second child. In Figure 5 the top panel again is based on the birth of the first child and the bottom panel uses data from around the birth of the third child.

Figure 3 shows that, prior to the birth of their first child, almost 75 percent of women work full time, with the rest either working part time or full time part year. Very few women do not work at all. In the year of the birth the percentage of women working full time drops considerably, while there is a jump up in the percentage working full time part year as well as an increase in the percentage working part time or not working. After the birth of the child the percentage of women working full time part year returns to the pre-birth level while there is a continual increase in the percentage of women in all other labor market states. In every period full time contains the largest number of women.

Comparing the labor market behavior prior to the birth of their first child for women who have two children (Figure 4) or three children (Figure 5) we see a similar pattern to that seen in Figure 3; most women work full time prior to the birth of the first child and the rate remains fairly steady up to the year of the birth. However, comparing the percentage of women working full time across Figures 3-5 shows that women who have two or three children are 10 percentage points less likely to work full time prior to the birth of their first child than women who have only one child. While far from conclusive, this is evidence that expected future fertility affects pre-birth labor market behavior.[10] One goal of our empirical analysis will be to see whether we can account for this difference using observable characteristics, such as education, race, or spouse's earnings, and unobserved heterogeneity.

Examining the labor market status across these three figures after the birth of the last child also reveals significant differences. In the bottom panels of Figures 4 and 5 we see that after the birth of the last child women are equally likely to work full time, part time, or to not work at all, with approximately 30 percent of women in each state. In Figure 3 we see that around 50 percent of women who have one child work full time in the periods following the birth with 20 percent of these women in the part-time or non-work state. This is evidence that the cost of having children does increase with the number of children. Another goal of the subsequent analysis will be to divide these costs into the direct and indirect components and to estimate the marginal cost of additional children.

These figures demonstrate that our four-state model of the labor market does a better job capturing the behavior of women, particularly around birth of a child. They also show that there are differences in labor market behavior that are related to subsequent fertility, suggesting that fertility is endogenous with respect to labor supply. In the next section we develop a statistical model that incorporates these findings.

[10] It could be that women who are less sucessful in the labor market have more children.

3 The empirical model

The effect of children on women's labor supply has been the object of numerous applied and theoretical panel data econometric studies. Applied papers frequently assume implicitly or explicitly that, once they control for individual heterogeneity, children variables are exogenous. Hyslop (1999), for example, makes this assumption noting that there is no evidence against the exogeneity in settings which account for the dependence of sequential labor market decisions. On the other hand, theoretical panel data papers (Chamberlain, 1984; Arellano and Honore, 2001; Carrasco, 2001; Arellano and Carrasco 2003) have underscored the fact that exogeneity of children variables is a strong assumption even when controlling for measured characteristics since it requires that all variation in labor supply plans is explained or that labor supply plans have no effects on fertility decisions at any point in the life cycle.

In estimating models with continuous dependent variables, endogeneity has been addressed by employing instrumental-variables techniques based on the lack of correlation between future errors in first differences and lagged values of variables (Anderson and Hsiao, 1981; Holz-Eakin, Newey and Rosen, 1988; Arellano and Bond, 1991) and between first differences in predetermined variables and errors in levels (Arellano and Bover, 1995). In discrete choice models endogeneity has been modelled through correlation between regressors and an individual specific error component. Several methods of estimation are available. The conditional approach is based on defining a set of probabilities which do not depend on the individual-specific error component (Honore and Kiriazidou, 2000). Alternative approaches are based on specifying the distribution of the random effects, either completely as in Chamberlain (1984) or semi-parametrically as in Carrasco (2001) and Arellano and Carrasco (2003). While Chamberlain's approach requires strictly exogenous regressors, Carrasco (2001) and Arellano and Carrasco (2003) are applicable when regressors are predetermined. In a notable departure from the literature, Carrasco (2001) jointly estimates participation and fertility decisions by employing a switching probit model where the equations underlying the participation decision contain a semi-parametrically defined individual component.

In this paper we propose a model for labor market and fertility decisions that is more general than alternatives found in the literature. Labor market decisions and fertility decisions are generated by a sequential optimization process. At the beginning of each period an individual decides on the level of labor market involvement for the current period while simultaneously making a fertility decision. The level of labor market involvement is chosen from a set of four alternatives–full-time work (FT), full-time part-year work (FP), part-time work (PT), non work(NW)–by comparing their associated value functions. In every time period fertility choices are driven by the comparison of value functions corresponding to having a child and not having a child, respectively. Let the value functions associated with each labor market state be denoted by U_{it}^{FT}, U_{it}^{FP},U_{it}^{PT}, and U_{it}^{NW}, respectively, and the difference of the relevant fertility value functions be U_{it}^{F}.

In reduced form, sequential labor market decisions are represented by a multinomial probit model with auto-correlated error terms while fertility decisions are represented by a probit model with state-dependence and auto-correlated error terms.[11] Since the choice of a level of labor market involvement depends only on differences of value functions, we transform the model by considering only values relative to the nonwork state.

$$U_{it}^1 = U_{it}^{FT} - U_{it}^{NW} = K_{it}\alpha^1 + X_{it}^{LM}\beta^1 + Z_{it}^1\gamma + \sum_m K_{it}\delta_{ml(i,m)}^1 + u_{it}^1$$
$$U_{it}^2 = U_{it}^{FP} - U_{it}^{NW} = K_{it}\alpha^2 + X_{it}^{LM}\beta^2 + Z_{it}^2\gamma + \sum_m K_{it}\delta_{ml(i,m)}^2 + u_{it}^2$$
$$U_{it}^3 = U_{it}^{PT} - U_{it}^{NW} = K_{it}\alpha^3 + X_{it}^{LM}\beta^3 + Z_{it}^3\gamma + \sum_m K_{it}\delta_{ml(i,m)}^3 + u_{it}^3$$
$$U_{it}^F = K_{it}\alpha^F + X_{it}^F\beta^F + \sum_m K_{it}\delta_{ml(i,m)}^F + u_{it}^F$$

where the subscript i indicates individuals, $i = 1, ..., N$ and subscript t indicates time periods, $t = 1, ..., T_i$. The vectors $K_{it}, X_{it}^{LM}, X_{it}^F$ contain variables describing the number of children and their age distribution, personal characteristics relevant to labor market decisions, and personal characteristics relevant to fertility decisions, respectively, while $Z_{it}^1, Z_{it}^2, Z_{it}^3$ are expected wages in each of the alternative labor market states. In this model the $\alpha's$ and $\beta's$ are vectors of global (fixed effect) parameters.

The transformed latent variables define the labor market and fertility decisions the following way:

$$\begin{aligned} y_{it}^{FT} &= 1 \text{ if } U_{it}^1 > U_{it}^2 \text{ and } U_{it}^1 > U_{it}^3 \text{ and } U_{it}^1 > 0; 0 \text{ otherwise} \\ y_{it}^{FP} &= 1 \text{ if } U_{it}^2 > U_{it}^1 \text{ and } U_{it}^2 > U_{it}^3 \text{ and } U_{it}^2 > 0; 0 \text{ otherwise} \\ y_{it}^{PT} &= 1 \text{ if } U_{it}^3 > U_{it}^1 \text{ and } U_{it}^3 > U_{it}^2 \text{ and } U_{it}^3 > 0; 0 \text{ otherwise} \\ y_{it}^{NW} &= 1 \text{ if } U_{it}^1 \leq 0 \text{ and } U_{it}^2 \leq 0 \text{ and } U_{it}^3 \leq 0; 0 \text{ otherwise} \end{aligned}$$

The fertility decision remains

$$y_{it}^F = 1 \text{ if } U_{it}^F > 0; 0 \text{ otherwise}$$

The vector of personal characteristics X_{it}^{LM} includes marital status, spouses wage, other income, region and urban/rural character of residence which are assumed to be exogenous. Determinants of fertility decisions, X_{it}^F, are other income, region and urban/rural character of residence, as well as the number of siblings with children which captures family background characteristics

[11] By comparison, Carrasco (2001) specifies state-dependent participation decisions but static fertility decisions, while Vella and Verbeek(1998, 1999) incorporate state dependence in the auxiliary (selection) equation but allow for dynamic effects in the main equation.

potentially important to fertility decisions. The number of siblings with children is excluded from the participation equation while spouse's wage and the respondent's wage are excluded from the fertility equation. These exclusion restrictions, while not crucial, help identify the parameters of the model.

Following the suggestions of Nakamura and Nakamura (1992) we employ a somewhat richer specification to capture the effects of children than is typical. The vector K_{it} includes the number of children in three age categories (0-1, 2-4, 5 and older) and a constant term. Age is measured as age at last birthday. The construction of the children variables defines the dynamics of the fertility decisions and the relationship between fertility decisions and labor market decisions. Children variables, which are transformations of lagged values of fertility, describe the entire history of fertility decisions–how many children have been born and how far in the past–thus rendering current fertility and participation decisions dependent on the entire fertility history. This construction allows a more accurate representation of the relationship between fertility and participation decisions and generates a particular structure for the system of equations. Fertility decisions may result in the birth of twins. Had we defined fertility as the occurrence of a birth in the current period and included it in the participation equation we would have lost the information on the number of children born in a period, which is important in estimating the cost of children. Our use of the children variables to capture past fertility decisions is different from the structure typically used in the literature estimating dynamic panel data models where the contemporaneous dependent variables enter each equation.[12]

Personal characteristics and children variables vary over i and t but are constant across labor market states. Multinomial probit models such as this are frequently difficult to identify due to flat spots in the likelihood function. Therefore, we follow the suggestion of Geweke, et al. (1997) and include the variable Z_{it} in the model, which is the hourly wage of women in each labor market state.[13] The variable Z_{it} varies over i, t, and labor market state, the coefficient on Z_{it} is constrained to be the same across states. When constructing Z_{it} we use the hourly wage for the labor market state the women is observed in and impute the hourly wage in the alternative states.[14]

We incorporate individual heterogeneity in the model by assuming that the constant term and the coefficients on the children variables in all equations differ across individuals. Specifically, we assume that there are $m = 1, ..., M$ independent sources of heterogeneity affecting individuals' decisions. Each source of heterogeneity has l_m levels. We use the following five sources of heterogeneity: individuals' time invariant personal characteristics (education, race) and

[12] For example, Carrasco (2001) includes the occurence of a birth in the participation equation. Vella and Verbeek (1998), who use panel data to estimate the union wage premium, include current union membership in the wage equation while simultaneously estimating the determinants of union status.

[13] See Geweke, et al. (1997) and Geweke and Keane (2001) for a further discussion of this issue and their proposed solution.

[14] We impute hourly wages in the alternative states using the coefficients from a standard wage regression estimated using observations on all women in the NLSY between 1979 and 1994.