Life-Cycle Fertility and Human Capital Accumulation *

George-Levi Gayle
GSIA, Carnegie Mellon University

Robert A. Miller
GSIA, Carnegie Mellon University

Previous Version: January 22, 2003
Present Version: April 30,2003

Abstract

This paper develops and implements a semiparametric estimator for investigating, with panel data, the importance of human capital accumulation, non-separable preferences of females and child care costs on females life-cycle fertility and labor supply behaviors. It presents a model in which the agents' expectations are correlated with their future choices and provides a set of conditions under which statistical inferences are possible from a short panel. Under the assumption that observed allocations are Pareto optimal, a dynamic model of female labor supply, labor force participation and fertility decision is estimated. In that model, experience on the job raises future wages, time spent nurturing children affects utility, while time spent off the job in the past directly affects current utility(or, indirectly through productivity in the non-market sector). This paper then uses the estimates from the model to conduct different policy simulations which shows that human capital accumulation is the most important determinant of life-cycle fertility behavior.

1. INTRODUCTION

In this paper we estimate a choice theoretic model of female labor supply and fertility behavior with panel data, and use our estimates to predict how changes in family policy would affect the behavior of these variables over the life cycle. The motivation for investigating dynamic interactions between fertility and female labor supply comes from broad trends in aggregate behavior and also empirical results from previous empirical work using cross sectional and panel data.

At the aggregate level the trends in birth rates and female labor supply and the wages of females in developed countries are striking. Measures of annual total fertility rates (TFRs) provide a useful way of summarizing trends in fertility at the aggregate level. There are two basic measures, by period and/or by cohort. A period TFR predicts the total life time number of births if a representative woman realized the age-specific fertility rates that prevailed in a given year. A cohort TFR measures the number of children born to a particular birth cohort. Both measures used for the U.S. and most developed countries show there has been a substantial

*The first author acknowledges financial support from the Andrew Mellon Research Fellowship. We thank participants at seminars at, the University of Essex, the University of Illinois, the University of Kansas, SOLE European meetings, SUNY-Stony Brook and the University of Pittsburgh. All errors are naturally ours.

decline in total completed fertility from the beginning to the end of the twentieth century.[1] For example at the beginning of this century a typical woman in the U.S. who reached the age of 45 bore, on average, four children over her lifetime, but by the end of the century, that number has fallen to only 1.9. The period TFR was 4.0 in Sweden in 1905 and had declined to 1.4 by the mid 1980's.[2]. Although the majority of women in developed countries eventually bear a child, an increasing fraction of them bear no children. The incidence of childlessness has fluctuated over the twentieth century but seems to have increased towards the end of the century. For example, Hotz et al (1996) report that over the last 20 years the incidence of childlessness has almost doubled in the U.S., going from 9% of women who reached age 40-45 in 1978 to 18% for comparably aged women in 1994. [3] The changes in total fertility rates over the last century have been accompanied by changes in the life-cycle timing of childbearing by women who were of fecundity age during the era. The so called baby boom of the U.S. was in essence fueled by women shifting their childbearing to earlier ages, and the subsequent bust was largely the result of the tendency for childbearing to be delayed.[4]

Parallel to this marked decline in childbearing has been a rise over time in female labor force participation in both developed and developing countries.[5] In the US the participation of all wives increased by 36% over the last 25 years, the rates of mothers with children under the age of three increased by 83%, and by 91% for women with children one year old or younger. The rise in female participation in the labor force has been accompanied by a decline in the difference between male and female earnings of full time workers.

These aggregate trends can be rationalized by simple economic models of household choices. The parents' demand for children depend on the prices of the inputs used in raising children, and the levels of household income and/or wealth. The opportunity cost for the mother's care and nurture for her offspring is the female wage rate, and as this increases female labor supply increases and the fertility rate falls. Controlling for the wage, a simple model predicts that families with more wealth have more children. These simple models also predict that the prices of other goods and services, including child quality, and the mother's market wage rate(s) can also explain families' demand for children. However, the sign and magnitudes of these effects are most often not unambiguously indicated by theory, which means that these issues are empirical questions.

Empirical investigations of micro data bear out these predictions. The mother's wage is negatively related to the demand for children in all types of models of fertility (Ward and Butz (1979, 1980), Hotz and Miller (1988)).[6] The relationship between household wealth and fertility, controlling for the opportunity cost of the mother's time, is harder to document. Schultz (1976) reports many earlier studies that found a positive relationship between family income and/or consumption and parental fertility. Several more recent studies also found a positive relationship between parental fertility and husband income or other household income (Ward and Butz (1980), Wolpin (1984), Hotz and Miller (1988)). On the other hand Willis (1973) finds

[1] See Hotz et al (1996) and reference there in for a survey on this issue.

[2] See Walker (1995) for details.

[3] Some of this increase reflects the decline in the fraction of women who are married over the same period. But the incidence of childlessness has risen among married women as well. See Hotz et al (1996) and reference therein for more details.

[4] See Hotz et al (1996) for details on this shift.

[5] See Heckman and Killingsworth (1996).

[6] See Paul Schultz (1976) for a summary of some the earlier studies.

a U-shape relationship between completed fertility and husband income, while Heckman and Walker(1990) found that there weakis a, if any relationship between husband's income (and also female wages) and the incidence of childlessness.

A further complication to interpreting the evidence about the effects of higher female wages on fertility was introduced by Becker (1965), who argued that parents not only choose the quantity of offspring but also their quality. Thus highly educated parents might choose to have a lower number of children, but invest more inputs in them. This modification to the basic model provides another reason why women with higher opportunity costs have lower fertility rates. Therefore the almost unanimous finding, that parental demand for children are negatively related to the educational level of the mothers, is hardly surprising. This result holds in static models of completed fertility (Willis (1973))[7], reduced form dynamic models (Walker (1996), Hotz and Miller (1988), Hotz, Heckman and Walker (1990)), and structural dynamic models of fertility and contraceptive practices (Wolpin (1984), Hotz and Miller (1993)). In a more direct test of the role of the mother's inputs on measures of childhood achievement, Michaels (1992) finds that after controlling for labor supply the offspring of more highly educated women perform better at school.

While the basic models of fertility can be used to explain the relationship between total fertility rates and measures of female labor supply and household wealth, they have much less to say about the timing of births and how this is related to female labor supply over the life cycle. Empirical dynamic models of fertility find that the time costs young children impose on their mothers help to rationalize the spacing of later births (Hotz and Miller (1988). Similarly there is strong evidence from dynamic models of labor supply and human capital accumulation that, in addition to providing wages, work experience is a form of investment in human capital that increases the future wage rate (Eckstein and Wolpin (1989), Miller and Sanders (1997), Altug and Miller (1998)). Thus the costs of staying home to raise children are significantly greater than the current wages foregone.

These empirical results suggest that the patterns of investing in the work force through current labor force participation is intertwined with decisions about the timing and the amount of offspring a household chooses to have. This study then is an attempt to combine both forms of human or family capital within a unified framework. Only by capturing both kinds of choices can one reasonably expect to answer policy questions that bear upon how households will change their contraceptive and labor force behaviors in response to changes in provisions for maternity leave, child care facilities and the tax treatment of dependents, to name just three examples of topical interest.

Recently there have been a number of paper looking at the importance of female's fertility behavior on their labor supply, Angrist and Evans(1998) looked at the effect of children on there mothers' labor supply however this was done in a static reduced form model which did not allow them to examine the issue of timing of birth, time cost,and human capital accumulation, simultaneously on labor supply and the joint effect of supply on fertility behavior. In this vein similar to this paper, Francesconi(2002) estimated a dynamic model joint fertility and work decision of married women. However they only have full time versus part work, did not looked directly on the human capital accumulation mechanisms, the different costs of having children nor the timing and spacing of birth over the life-cycle.

[7] See Paul Schulz (19..) survey for a comprehensive summary of these results.

The next two sections provide the theoretical underpinnings to our empirical investigations. Section 2 lays out a life cycle model of labor supply and fertility. Then in Section 3 we derive the conditions implied by dynamic optimization that form the basis for identification and estimation. Sections 4 through 7 are the heart of the estimation. First, we explain our estimation strategy in Section 4. Then we briefly summaraize the sample of households used in our empirical work, which is drawn from the Panel Study of Income Dynamics (PSID). In Section 6 we report our estimates of the wage equation from wage and labor supply data. The wealth effects of the household are estimated in Section 7 from data on consumption. In Section 8 we estimate from data on labor supply and births the parameters that determine preferences for children, as well as the direct intertemporal effects of labor supply or leisure on household utility. The last two sections of the paper explore the quantitative implications of our model. They use the estimates obtained from the body of the estimation to conduct some policy simulations and summarize our findings.

2. A FRAMEWORK

The model is set in discrete time, and measures the woman's age beyond adolescence with periods denoted by $t \in \{0, 1, \ldots, T\}$. It analyzes the accumulation of two kinds of human capital, offspring and labor market experience.

Female labor market experience for the n^{th} household in our sample is embodied in the wage rate, denoted w_{nt}, and depends on labor market experience and demographic variables. The latter, denoted by z_{nt}, include such variables as age, formal education, regional location, ethnicity and race. It is assumed that z_{nt} is independently distributed over the population with cumulative probability distribution function $F_0(z_{nt+1} \mid z_{nt})$. Let h_{nt} denote the proportion of time worked in period t as a fraction of the total time available in the period, let d_{nt} denote participation in period t, that is an indicator if $h_{nt} > 0$. We assume that the mapping from experience to the current wage rate is given by:

$$w_{nt} = g(d_{nt-\rho}, ..., d_{nt-1}, h_{nt-\rho}, ..., h_{nt-1}, z_{nt}) \tag{2.1}$$

for some positive integer ρ. Thus Equation 2.1 shows that, in addition to the demographic variables, the current wage depends on past participation and past hours up to ρ periods ago.

The birth of a child at period t is denoted by the indicator variable $b_{nt} \in \{0, 1\}$. It contributes directly to household utility. We assume that the spacing of births is related to preferences by the household over the age distribution of its children, as captured by interactions in the birth dates of successive children. More specifically, let γ_0 denote the additional lifetime expected utility a household receives for its first child, let $\gamma_0 + \gamma_k$ denote the utility from having a second child when the first born is k years old, let $\gamma_0 + \gamma_k + \gamma_j$ denote the utility from having a third child when the first two are aged k and j years old, and so on. Thus the deterministic benefits from offspring to the n^{th} household in period t can be summarized by the random variable, U_{0nt}, defined as:

$$U_{0nt} = b_{nt}(\gamma_0 + \sum_{k=1}^{M} \gamma_k b_{nt-k} + \gamma_M \sum_{k=M+1}^{T} b_{nt-k}) \tag{2.2}$$

Raising children requires market expenditure and parental time. We assume that the discounted cost of expenditures of raising a child is π, a parameter that varies with household demographics, and that a k year old requires nurturing time of ρ_k. Letting c_{nt} denote the fraction of time the n^{th} household spend nurturing children in the household, our assumption about nurturing implies:

$$c_{nt} = \sum_{k=0}^{t} \rho_k b_{n,t-k} \tag{2.3}$$

Equation (2.3) nests several specifications of the child care costs (or the related child care technology) considered in the literature. Hill and Stafford(1980) found, using data from time diaries, that maternal time devoted to child care declines as the children age. To capture this latter pattern, suppose that the demands a child makes on its mother's time declines geometrically with age after a give age M, so that

$$\rho_k = \begin{cases} \rho_k & \text{if } k \leq M \\ \rho_M \delta^{k-1} & \text{if } k > M \end{cases} \tag{2.4}$$

where $0 < \delta < 1$. Our specification for child care costs ignores two issues which have been examined in the literature. First, it implies that the child care process exhibits constant returns to scale in the number of existing children. The evidence on the importance of such scale economies is mixed; Lazear and Michael(1980) find evidence of large scale economies while Espenshade(1984) finds them to be small. Second, by assuming that the coefficients in equation (2.3) are fixed weights, no substitution is allowed between market and maternal time in the care of children. However, by interpreting these cost as the minimum time required nurturing time and allowing π to vary according to individual characteristics we can capture such effects in this model.

Leisure in period t, denoted l_{nt}, is defined as the balance of time not spent at work or nurturing children. It follows that the time allocated between nurturing children, market work and leisure must obey the constraint:

$$1 = h_{nt} + l_{nt} + c_{nt} \tag{2.5}$$

Apart from having utility for children, household utility also comes from its consumption of market goods, denoted x_{nt}, and leisure, denoted l_{nt}. We assume that preferences are additive in consumption and leisure, but not separable with respect to leisure at different dates. To model this dependence, define $z_{nt}^* = (b_{nt-M}, ..., b_{nt-1}, h_{nt-\rho}, ..., h_{nt-1}, z'_{nt})$, where the first ρ elements of z_{nt}^* capture the dependence of the current household state on lagged labor supply and birth choices, and the remaining elements are the set of observed demographics. Letting U_{1nt} represent the fixed utility costs of the n^{th} female from working in period t, we assume:

$$U_{1nt} = u_1(z_{nt}^*, l_{nt}) + u_2(z_{nt}^*, d_{nt})$$

This formulation incorporates both fixed and variable utility costs associated with working. It models the variable costs of working as a mapping of observed household characteristics alone, but allows participation in the work force to be determined by observed factors, entering through $u_2(z_{nt}^*, d_{nt})$. We assume that $u_1(z_{nt}^*, 1 - h_{nt} - c_{nt})$ is a concave increasing function in l_{nt}.

Let us first recast the decision process of the individual by defining the following indicator variables.

$$\mathbb{I}_{0nt} = \left\{ \begin{array}{cc} 1 & \text{if } d_{nt} = 0 \text{ and } b_{nt} = 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{2.6}$$

$$\mathbb{I}_{1nt} = \left\{ \begin{array}{cc} 1 & \text{if } d_{nt} = 1 \text{ and } b_{nt} = 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{2.7}$$

$$\mathbb{I}_{2nt} = \left\{ \begin{array}{cc} 1 & \text{if } d_{nt} = 0 \text{ and } b_{nt} = 1 \\ 0 & \text{otherwise} \end{array} \right\} \tag{2.8}$$

$$\mathbb{I}_{3nt} = \left\{ \begin{array}{cc} 1 & \text{if } d_{nt} = 1 \text{ and } b_{nt} = 1 \\ 0 & \text{otherwise} \end{array} \right\} \tag{2.9}$$

Let ε_{ntk} $(k = \{0, ..., 3\})$ demographic and psychological factors that determine the precise timing of birth and participation in the labor force that is unobserved by the econometrican, and that $(\varepsilon_{nt1}, ..., \varepsilon_{nt1})$ is identically and independently distributed across (n, t) with multivariate probability distribution $F_{14}\left(\varepsilon_{nt1}, ..., \varepsilon_{nt4}\right)$.

The third component in utility is derived from current consumption. We denote by $U_{3nt} \equiv u_3(x_{nt}, z_{nt}, \varepsilon_{5nt})$ the current utility from consumption by household n in period t, and assume $u_2(x_{nt}, z_{nt}, \varepsilon_{5nt})$ is concave increasing in x_{nt} for all values of the observed and unobserved demographic variables $(z_{nt}, \varepsilon_{5nt})$. Analogous to the assumptions made for the other unobserved variables, we assume ε_{5nt} is identically and independently distributed across (n, t) with bivariate probability distribution $F_5\left(\varepsilon_{5nt}\right)$.

The period t utility for household n utility is found by summing over the three components. Let $\beta \in (0, 1)$ denote the common subjective discount factor, and write $E_t(.)$ as the expectation conditional on information available to household n at period t. The expected lifetime utility of household n is then:

$$E_0 \left\{ \sum_{t=0}^{T} \beta^t \left[\sum_{k=0}^{3} \mathbb{I}_{knt}(U_{0ntk} + U_{1ntk} + U_{3nt} + \varepsilon_{ntk}) \right] \right\} \tag{2.10}$$

Table 1 displays the notation defining the main elements of our model.

3. Optimal Decision Making

In this paper we assume there are no distortions within the labor supply and consumer goods markets. This approach was recently utilized by Altug and Miller (1998) to estimate a life-cycle model of how work experience affects female wages and labor supply. Indeed, over the last decade an empirical literature has emerged that tests for deviations from Pareto optimal allocations using panel data on households. (See, for example, Altug and Miller (1990), Altonji, Hayashi and Kotilikoff (1995), Cochrane (1991), Mace (1991), Miller and Sieg (1997), and Townsend (1994)) Taken together, this body of work shows that the restrictions imposed by Pareto optimal allocations are quite hard to reject with panel data on households, unless one assumes very limited forms of population heterogeneity, and also that preferences are strongly additive, two assumptions that are widely regarded by microeconomists as being implausible. In addition, the limited empirical work that exists on incomplete markets cannot easily be generalized beyond the highly stylized frameworks that are investigated. So while few economists believe that the real world supports a rich set of Arrow Debreu securities spanning the commodity space, in the absence of clear guidance about precisely how gains from trade are left unfulfilled, the assumption of ignoring such impediments to trade is a useful working hypothesis. In this case the assumption, that observed allocations are Pareto optimal, allows us to derive optimality tractable conditions from a model in which households deal with the complex interactions that arise from spacing births, given the time commitment to their young, while simultaneously determining labor supply in a world where labor force attachment impacts on future wages.

The Pareto optimal allocations are derived as the solution to a social planner's problem for a large population of households n in a cohort defined on the $[0, 1]$ interval, in which the integral of the weighted, expected discounted utilities of each household are maximized subject to an aggregate feasibility or resource constraint. Therefore, the objective function for the social planner is formed from the individual utilities defined by Equation (2.10), and the social weights attached to each individual n, which we denote by η_n^{-1}. The planner is constrained by the time available to each household n in the period $t \in \{1, 2, \ldots, T\}$ that cohort is active, which is Equation (2.5), and must respect the time required to nurture children, as indicated by Equation (2.3). At the margin consumption goods are produced, the value of marginal product function for labor, which is Equation (2.1), and net transfers between members of the cohort and others (including other income generating family members and public transfers) are exogenously set to W_t in period t. We now state the social planner's constrained optimization problem more formally.

The aggregate feasibility condition equates aggregate consumption at each date t to the sum of output produced by all individuals $n \in [0, 1]$ and the aggregate endowment W_t. Defining $\mathcal{L}$ as the Lebesque measure which integrates over the cohort population, this constraint requires:

$$\int_0^1 (x_{nt} + \pi b_{nt} - w_{nt} h_{nt})\, d\mathcal{L}(n) \leq W_t \tag{3.1}$$

The Pareto optimal allocations are found by maximizing

$$E_0 \left\{ \int_0^1 \sum_{t=0}^{T} \eta_n^{-1} \beta^t [\sum_{k=0}^{3} \mathbb{I}_{knt}(U_{0ntk} + U_{1ntk} + U_{3nt} + \varepsilon_{ntk})] \right\} d\mathcal{L}(n) \tag{3.2}$$

subject to aggregate budget inequality 3.1 and also the individual household time constraints 2.5 with respect to sequences (of random variables that are successively measurable with respect to the information known at periods $t = 0, 1, \ldots$) for consumption and labor supply $\{x_{nt,} h_{nt}, b_{nt}\}_{t=0}^{T}$ chosen for all the cohort members $n \in [0, 1]$. For future reference we denote the optimal choices by $\{x_{nt,}^o h_{nt}^o, b_{nt}^o\}_{t=0}^{T}$ writing $d_{nt}^o = 1$ whenever $h_{nt}^o > 0$.

The necessary conditions characterizing the optimal consumption, labor supply and birth allocations are the basis for the estimation procedure. Turning to the derivation of optimal consumption first, we define $\beta^t \lambda_t$ as the Lagrange multiplier associated with the aggregate feasibility constraint. Differentiating the Lagrangian formed from equation(3.2) and equation(3.1), the optimal consumption allocations satisfy the necessary conditions:

$$\frac{\partial u_3(x_{nt,}^o, z_{nt}, \varepsilon_{5nt})}{\partial x_{nt}} = \eta_n \lambda_t, \tag{3.3}$$

for all $n \in [0, 1]$ and $t \in \{0, 1, ...\}$. Notice that η_n corresponds to the inverse of the social weight for each household n. In our empirical work we estimate Equation (3.3) using data on consumption and household demographics to obtain estimates of the social weight by inverting an estimate of the estimated marginal utilities of wealth.

At the heart of the decision-maker's trade off between working career and family, is the fact that births and work reduce the amount of time left for leisure time, so they cannot be solved independently. Since each person's labor supply and birth decisions contribute only infinitesimally to aggregate output, they are valued at a constant rate each period λ_t by the social planner. Thus we may represent the conditional valuation functions for household n associated with each discrete decision in period t as:

$$V_{jnt}+\varepsilon_{jnt}\equiv \max_{\{h_{nr}(\mathbb{I}_{knr})_{k=0}^{3}\}_{r=t}^{T}} E_t \left[\begin{array}{c} \sum_{r=t}^{T} \beta^{r-t}(\sum_{k=0}^{3} \mathbb{I}_{knt}(U_{0ntk} + U_{1ntk} + \eta_n \lambda_r w_{nr} h_{nr} \\ -\eta_n \lambda_r \pi_k(z_{nr}) b_{nr} + \varepsilon_{ntk})) \mid \mathbb{I}_{jnt} = 1 \end{array} \right] \quad (3.4)$$

$for\ j \in \{0, 1, 2, 3\}$.

Up to the household specific factor of proportionality η_n, the term $V_{knt} + \varepsilon_{knt}$, denotes the social value from n choosing option k at date t, conditional upon all the information available to the social planner (and the household) at the beginning of time t. Whether each individual choice is optimal or not depends on:

$$\mathbb{I}_{knt}^{o} = \begin{cases} 1, & \text{if } V_{knt} + \varepsilon_{knt} \geqslant V_{jnt} + \varepsilon_{jnt}\ \forall j \neq k \\ 0, & \text{otherwise} \end{cases} \quad (3.5)$$

Upon defining p_{knt} as the conditional choice rate in period t, we obtain the probability of making choice k by the n^{th} female in period t as:

$$p_{knt} = E[\mathbb{I}_{knt}^{o} = 1 \mid z_{nt}^{*}] \quad (3.6)$$

This definition shows that if a representation for $V_{knt} - V_{jnt}$ can be readily obtained in terms of the variables and parameters that characterize the household's problem, the parameters can be estimated using standard approaches to estimating discrete choice models with labor supply and other demographic data, including data on births. Our estimation approach uses the fact that p_{knt} can be estimated nonparametrically and that V_{knt}'s have the recursive representations:

$$V_{jnt}\equiv \max_{h_{nt}>0} E_t \left[\begin{array}{c} U_{0ntk} + U_{1ntk} + \eta_n \lambda_t w_{nt} h_{nt} - \eta_n \lambda_t \pi(z_{nt}) b_{nt} \\ +\beta E_t[\sum_{k=0}^{3}\ p_{knt+1}(V_{knt+1}+\varepsilon_{knt+1})] \mid \mathbb{I}_{jnt} = 1 \end{array} \right] \quad (3.7)$$

We also use the fact that an interior solution for those participating in the labor force requires $\partial V_{1nt} \backslash \partial h_{nt} = 0$ or $\partial V_{3nt} \backslash \partial h_{nt} = 0$. Thus if $\mathbb{I}_{knt}^{o} = 1$ for $k = \{1, 3\}$, then h_{nt}^{o} solves:

$$\frac{\partial U_{knt}}{\partial h_{nt}} + \eta_n \lambda_t w_{nt} = -\beta E_t \left\{ \sum_{k=0}^{3} [\ p_{knt+1} \frac{\partial (V_{knt+1}+\varepsilon_{knt+1})}{\partial h_{nt}} + (V_{knt+1}+\varepsilon_{knt+1})\ \frac{\partial\ p_{knt+1}}{\partial h_{nt}}] \right\} \quad (3.8)$$

The left side of Equation (3.8) gives the current benefits and costs of spending a marginal hour working, comprising a utility cost in terms of leisure foregone, and the value of the extra goods and services produced. The right side shows the expected future benefits. Marginally adjusting current hours worked directly affects future productivity as well as the benefits of future leisure. Moreover, supposing the probability of working next period increases next period from this adjustment, the net benefits of working next period should be applied to the increase. This is captured in the second expression on the right side of Equation (3.8).

4. An Estimation Strategy

This framework is amenable to a multi-stage estimation strategy. First, there is contemporaneous separability of consumption from labor supply and birth in the utility function. Second, wages are assumed to be noisy measures of individual-specific marginal products of labor, which are determined by our two forms of human capital accumulation, namely formal education, past labor market participation and number of hours worked, plus other individual characteristics. Provided the measurement error in wages is uncorrelated with current and past labor supply and birth choices (an assumption we can readily test for providing a set of overidentifying instruments exist), the consumption and wages equations can be estimated separately from the hours, participation and birth equations to provide estimates of the determinants of household consumption and the effects of human capital accumulation on the individual wages.

The representation of individuals' valuation functions defined by equations (3.4), and (3.8) imply that the fixed cost of participation, benefit of a birth and cost of a birth can be recovered from a model in which the income generated by the decision to work (jointly with the decision to have a birth) is evaluated using the product of shadow value of consumption λ_t and the time-invariant individual-specific effect η_n. However, the existence of fixed costs of participation, birth benefits, birth and the effect of endogenous labor market participation, birth decision, and the optimal choice of hours implies that techniques developed for dynamic discrete choice models must be used to estimate the hours, participation and birth conditions. In principle one could use one of the many maximum likelihood estimation (ML) procedures available (see e.g. Miller (1984), Wolpin (1984), Pakes (1986), Rust (1987), etc.). This, however, involves the derivation of the valuation function as a mapping of the state and parameter space to calculate the probability of the sample outcome. Our model is very complicated in that it allows nonseparability of the both birth and labor supply decision. This would make the computational cost of employing ML in this setting very prohibitive. For this reason, we adapted a conditional choice probability (CCP) estimator, which does not require us to solve the valuation functions.

The CCP estimator forms an alternative representation for the conditional valuation functions that enter individuals' optimizing conditions by multiplying current utilities, evaluated at respective state for a given parameter value and corrected for dynamic selection bias, with the probability that the state in question occurs, and then summing over all states. The probabilities are estimated non-parametrically and then substituted into a criterion function that is optimized over the structural parameters. Although CCP estimators are far more tractable than ML, the computational burden of estimating conditional choice probabilities at every node in the decision tree is great. We exploit the property of finite state dependence, enjoyed by our model, to use the representation results from Altug and Miller (1998).

4.1. Conditional Choice Prabibilities Estimation

The starting point for our CCP estimation is the discrete component of model, the euler equation can easily be estimated using standard GMM procedures. Without loss of generality, let $H_{nt} \equiv (z'_{nt}, \mu_n\eta_n\lambda_t\omega_t, h_{nt-\rho}, ..., h_{nt-1}, d_{nt-\rho}, ..., d_{nt-1}, b_{nt-\overline{\rho}}, ..., b_{nt-1})$ represent the agent's relevant history as of the beginning of period t.In each period t, there is a current utility or payoff, U_{tj}, associated with each cjoice j.Suppose we let $U_j^*(H_{nt}) = E(U_{tj} \mid H_{nt})$ denote the conditional expectation of U_{tj}, given H_{nt}.We can then reformulate the model specified section 2 such that

$$U_{tj} = U_j^*(H_{nt}) + \varepsilon_{tj} \tag{4.1}$$

where the stochastic utility component, ε_{tj}, is by construction, conditionally independent of H_{nt}.Let $\widetilde{U}(H_{nt}) = (U_0^*(H_{nt}), ..., U_3^*(H_{nt}))'$ and $\widetilde{\varepsilon}_t = (\varepsilon_{t0}, ..., \varepsilon_{t3})'$, repectively, denote 3×1 vectors of deterministic and stochastic utility components. We write the distribution fuction of $\widetilde{\varepsilon}_t$,given the the assumption of Type I Extreme Value distribution as:

$$G(\widetilde{\varepsilon}_t \mid H_{nt}) = \exp\{-\sum_{j=0}^{3} \exp(-\varepsilon_{tj})\} \tag{4.2}$$

which has a well-defined, joint density function $dG(\widetilde{\varepsilon}_t \mid H_{nt})$. Then the discrete choice which leads to the conditional valuation function in equation (3.5) can be stated as follows. First let us redefine the utility to fit in the above framework.

$$U_j^*(H_{nt}) = \begin{cases} u_1(z_{nt}, l_{nt}^{(0)}) + u_2(z_{nt}, 0) & j = 0 \\ u_1(z_{nt}, l_{nt}^{(1)}) + u_2(z_{nt}, 1) + \eta_n \lambda_t w_{nt} h_{nt}^*, & j = 1 \\ u_1(z_{nt}, l_{nt}^{(2)}) + u_2(z_{nt}, 0) - \eta_n \lambda_t z_{nt}' \pi & \\ +(\gamma_0 + \sum_{s=1}^{\rho} \gamma_s b_{nt-s} + \gamma_S \sum_{s=\rho+1}^{t} b_{nt-s}) & j = 2 \\ u_1(z_{nt}, l_{nt}^{(3)}) + u_2(z_{nt}, 1) + \eta_n \lambda_t w_{nt} h_{nt}^* & \\ +(\gamma_0 + \sum_{s=1}^{\rho} \gamma_s b_{nt-s} + \gamma_S \sum_{s=\rho+1}^{t} b_{nt-s}) - \eta_n \lambda_t z_{nt}' \pi, & j = 3 \end{cases} \tag{4.3}$$

where

$$l_{nt}^{(j)} = \begin{cases} 1 - \sum_{k=1}^{t} \rho_k b_{nt-1} & j = 0 \\ 1 - h_{nt}^* - \sum_{k=1}^{t} \rho_k b_{nt-1} & j = 1 \\ 1 - \sum_{k=0}^{t} \rho_k b_{nt-1} & j = 2 \\ 1 - h_{nt}^* - \sum_{k=0}^{t} \rho_k b_{nt-1} & j = 3 \end{cases} \tag{4.4}$$

is the leisure from choosing the diferent options.

Pareto optimality implies that the decentralized problem is equivalent to the individual problem, then the agent can be viewed as sequentially choosing $\{\widetilde{\mathbb{I}}_t\}_{t=0}^{T}$ to maximize the objective function:

$$E_0(\sum_{t=0}^{T} \sum_{j=0}^{3} \mathbb{I}_{tj} \beta^t [U_j^*(H_{nt}) + \varepsilon_{tj}]) \tag{4.5}$$

By letting $\widetilde{\mathbb{I}}_s^o = (\mathbb{I}_{s1}^o, \mathbb{I}_{s2}^o, \mathbb{I}_{s3}^o)'$ denote the agent's optimal in period s.We define the conditional valuation function associated with choosing j in period t defined in equation(3.4) as:

$$V_j(H_{nt}) = U_j^*(H_{nt}) + E_t(\sum_{r=t+1}^{T} \sum_{j=0}^{3} \mathbb{I}_{rj}^{(o)} \beta^{r-t} [U_j^*(H_{nt}) + \varepsilon_{tj}] \mid H_{nt}, \mathbb{I}_{tj}^{(o)} = 1) \tag{4.6}$$

Then conditional on history H_{nt},the probability the agent chooses action j is therefore:

$$p_j(H_{nt}) = \Pr[I_{\{V_j(H_{nt}) + \varepsilon_{tj} \geq V_k(H_{nt}) + \varepsilon_{tk}, \vee k \neq j\}}] \tag{4.7}$$

Let $\widetilde{p}(H_{nt}) \equiv (p_1(H_{nt}), p_2(H_{nt}), p_3(H_{nt}))'$ denote the 3 dimensional vector of conditional choice probabilities associated with the last 3 actions in period t.

4.2. An Alternative Representation of Conditional Value Functions

In general, the conditional valuation function, $V_j(H_{nt})$, does not have a closed form solution. The standard practice is to exploit Bellman's(1957) equation and use backward recursion methods to obtain one. This section provides an alternative representation of $V_j(H_{nt})$ which will prove convenient when estimating this model in a multistage procedure.

To derive this representation, note(4.2) and (4.7) imply that the conditional probability of making choice $j \in \{1, 2, 3\}$, can be written as:

$$p_j(H_{nt}) = E(\mathbb{I}^o_{tj} = 1 \mid H_{nt}) = \frac{e^{V_j(H_{nt})}}{\sum_{k=0}^{3} e^{V_k(H_{nt})}}\text{[8]} \tag{4.8}$$

For each $j \in \{1, 2, 3\}$, the expression corresponding to (4.8) is a positive, real-valued, mapping from the differences in conditional valuation functions associated with the optimal choice and the alternative actions. We now show that these differences can be expressed as functions of the conditional choice probabilities. Let $\widetilde{V}(H_{nt}) = (V_0(H_{nt}), V_1(H_{nt}), V_2(H_{nt}), V_3(H_{nt}))'$ be a 4-dimensional vector. For each period t and $j \in \{1, 2, 3\}$, define the real-valued function, $Q_j(\widetilde{V}, H_{nt})$, as:

$$Q_j(\widetilde{V}, H_{nt}) \equiv \frac{e^{V_j(H_{nt})}}{\sum_{k=0}^{3} e^{V_k(H_{nt})}} \tag{4.9}$$

and $\widetilde{Q}(\widetilde{V}, H_{nt})$, a 3-dimensional vector function as:

$$\widetilde{Q}(\widetilde{V}, H_{nt}) = (Q_1(\widetilde{V}, H_{nt}), Q_2(\widetilde{V}, H_{nt}), Q_3(\widetilde{V}, H_{nt}))' \tag{4.10}$$

If $\widetilde{V}$ comprises the differences in the conditional valuation functions, namely,

$$\widetilde{V}' = (V_1 - V_0, V_2 - V_0, V_3 - V_0)' \equiv \widetilde{V}'(H_{nt}) \tag{4.11}$$

then $\widetilde{p}(H_{nt}) = \widetilde{Q}(\widetilde{V}', H_{nt})$. The cornstone of this estimation strategy is the express $\widetilde{V}'$ as a function of $\widetilde{p}(H_{nt})$. This requires$\widetilde{Q}(\widetilde{V}, H_{nt})$ to be invertible in $\widetilde{V}'$. By proposition 1 of Hotz and Miller(1993, p 501). This enables us to express $V_j(H_{nt})$ in terms of the choice probabilities, transition probabilities and expected(per period) payoffs associated with future histories. To estimate the model we proceed in several steps. First we exploit the limited state dependency structure of our model to derive an alternative representation of our value function. Inorder better characterize the expected (per period payoffs associated with future histories we define the following four possible histories.To illustrate this, define the $(\rho + K + M + 1)-$ dimensional vectors $H^{(s)}_{0nt}$ $H^{(s)}_{1nt}$ $H^{(s)}_{2nt}$ and $H^{(s)}_{3nt}$ as

$$H^{(s)}_{0nt} \equiv (z'_{nt+s}, \mu_n \eta_n \lambda_{t+s\omega_{t+s}}, h_{n,t-\rho+s}, ..., h_{n,t-1}, b_{n,t-M+s}, ..., b_{n,t-1}0, ..., 0, 0, ..., 0)' \tag{4.12}$$

$$H^{(s)}_{1nt} \equiv (z'_{nt+s}, \mu_n \eta_n \lambda_{t+s} \omega_{t+s}, h_{n,t-\rho+s}, ..., h_{n,t-1}, b_{n,t-M+s}, ..., b_{n,t-1}, h^*_{nt}, ..., 0, 0, ..., 0)' \tag{4.13}$$

[8] See Hotz and Miller(1993, p.500) and McFadden(1981, p.204) for example.

$$H_{2nt}^{(s)} \equiv (z'_{nt+s}, \mu_n \eta_n \lambda_{t+s\omega_{t+s}}, h_{n,t-\rho+s}, ..., h_{n,t-1}, b_{n,t-M+s}, ..., b_{n,t-1}, 0, ..., 0, 1, ..., 0)' \quad (4.14)$$

$$H_{3nt}^{(s)} \equiv (z'_{nt+s}, \mu_n \eta_n \lambda_{t+s} \omega_{t+s}, h_{n,t-\rho+s}, ..., h_{n,t-1}, b_{n,t-M+s}, ..., b_{n,t-1}, h^*_{nt}, ..., 0, 1, ..., 0)' \quad (4.15)$$

for $s = 1, ...\overline{\rho}$, (i.e. $\overline{\rho} \equiv \max(\rho, M)$) where h^*_{nt} is the fraction of time a woman chooses to spend at work conditional on participating, $b_{n,t} = (b_{nt}, b_{nt-1}, ..., b_{nt-M})$. . The state vector $H_{0nt}^{(s)}$ is the state for a woman at date $t+s$ who has accumulated work and birth histories, $(h_{n,t-\rho+s}, ..., h_{n,t-1})'$ and $(b_{n,t-M+s}, ..., b_{n,t-1})'$ respectively, up to period t and then chooses not to work or have a child at date t and for $s-1$ periods following period t. The state vector $H_{0nt}^{(\overline{\rho})}$ corresponds to the labor market and birth histories in which the woman does not participate in the labor force or have a child between t and $t+\overline{\rho}$. Likewise $H_{1nt}^{(s)}$ is the state vector for a woman at time $t+s$ who accumulates work and birth histories, $(h_{n,t-\rho+s}, ..., h_{n,t-1})'$ and $(b_{n,t-M+s}, ..., b_{n,t-1})'$ respectively, up to period t, chooses to participate in the labor force in at date t but chooses not to have a child at date, and then chooses not to participate in the labor force or have a child for $s-1$ periods following period t. While $H_{2nt}^{(s)}$ is the state vector for a woman at time $t+s$ who accumulates work and birth histories, $(h_{n,t-\rho+s}, ..., h_{n,t-1})'$ and $(b_{n,t-M+s}, ..., b_{n,t-1})'$ respectively, up to period t, chooses to have a child at date t but chooses not to participate in the labor force at date, and then chooses not to participate in the labor force or have a child for $s-1$ periods following period t. Finally, $H_{3nt}^{(s)}$ is the state vector for a woman at time $t+s$ who accumulates work and birth histories, $(h_{n,t-\rho+s}, ..., h_{n,t-1})'$ and $(b_{n,t-M+s}, ..., b_{n,t-1})'$ respectively, up to period t, chooses to have a child at date t and at the same time chooses not to participate in the labor force at date, and then chooses not to participate in the labor force or have a child for $s-1$ periods following period t. Since we assume the we have limited state dependencies only histories up to $\overline{\rho}$ are going to be relevant for decisions in the current period. Also let

$$\varpi_j(p_j(H_{nt})) = E(\varepsilon_{tj} \mid H_{nt}, \mathbb{I}_{tj}^{(o)} = 1) \quad (4.16)$$

Then by forward recusion it follows that we can rewrite equation(4.6) as:

$$V_j(H_{nt}) = U_j^*(H_{nt}) + \beta E_t(\textstyle\sum_{k=0}^{3} p_k(H_{jnt}^{(1)})[V_k(H_{nt}^{(1)}) + \varpi_k(p_k(H_{jnt}^{(1)}))]) \quad (4.17)$$

We can derive an alternative expression for $V_j(H_{nt})$ by using the definitions of $Q_j^{-1}(p_j(H_{nt}))$ for $j \in \{1,2,3\}$, substituting for $Q_j^{-1}(p_j(H_{nt}))$ into equation(4.17) and rearranging we obtain:

$$\begin{aligned} V_k(H_{nt}) &= U_j^*(H_{nt}) + \beta E_t\{\sum_{s=1}^{\overline{\rho}.}[V_0(H_{jnt}^{(1)}) + p_0(H_{jnt}^{(1)})\varpi_0(p_0(H_{jnt}^{(1)}))] \quad (4.18) \\ &\quad + \sum_{k=1}^{3} p_k(H_{jnt}^{(1)})[Q_k^{-1}(p_k(H_{jnt}^{(1)})) + \varpi_k(p_k(H_{jnt}^{(1)}))]\} \end{aligned}$$

We then add and substract $\sum_{k=1}^{3} p_k(H_{jnt}^{(1)})\varpi_0(p_0(H_{jnt}^{(1)}))]$ and obtain

$$\begin{aligned} V_j(H_{nt}) &= U_j^*(H_{nt}) + \beta E_t\{V_0(H_{jnt}^{(1)}) + \varpi_0(p_0(H_{jnt}^{(1)}))] \\ &\quad + \sum_{k=1}^{3} p_k(H_{jnt}^{(1)})[Q_k^{-1}(p_k(H_{jnt}^{(1)})) + \varpi_k(p_k(H_{jnt}^{(1)})) - \varpi_0(p_0(H_{jnt}^{(1)}))]\} \end{aligned} \tag{4.19}$$

The above equation implies that

$$\begin{aligned} V_0(H_{jnt}^{(1)}) &= U_0^*(H_{jnt}^{(1)}) + \beta E_t\{V_0(H_{jnt}^{(2)}) + \varpi_0(p_0(H_{jnt}^{(2)}))] \\ &\quad + \sum_{k=1}^{3} p_k(H_{jnt}^{(2)})[Q_k^{-1}(p_k(H_{jnt}^{(2)})) + \varpi_k(p_k(H_{jnt}^{(2)})) - \varpi_0(p_0(H_{jnt}^{(2)}))]\} \end{aligned} \tag{4.20}$$

Substituting equation (4.20) into equation(4.19) gives

$$\begin{aligned} V_j(H_{nt}) &= U_j^*(H_{nt}) + \beta E_t\{U_0^*(H_{knt}^{(1)}) + \beta E_t[V_0(H_{jnt}^{(2)}) + \varpi_0(p_0(H_{jnt}^{(2)}))]] + \varpi_0(p_0(H_{jnt}^{(1)})) \\ &\quad + \sum_{k=1}^{3} p_k(H_{jnt}^{(2)})[Q_k^{-1}(p_k(H_{jnt}^{(2)})) + \varpi_k(p_k(H_{jnt}^{(2)})) - \varpi_0(p_0(H_{jnt}^{(2)}))]] \\ &\quad + \sum_{k=1}^{3} p_k(H_{jnt}^{(1)})[Q_k^{-1}(p_k(H_{jnt}^{(1)})) + \varpi_k(p_k(H_{jnt}^{(1)})) - \varpi_0(p_0(H_{jnt}^{(1)}))]\} \end{aligned} \tag{4.21}$$

Performing a $\overline{\rho}$-step induction yields:

$$\begin{aligned} V_j(H_{nt}) &= U_j^*(H_{nt}) + E\{\textstyle\sum_{s=1}^{\overline{\rho}} \beta^s[U_0^*(H_{nt}^{(s)}) + \varpi_0(p_0(H_{jnt}^{(s)})) + \displaystyle\sum_{k=1}^{3} p_k(H_{jnt}^{(s)})[Q_k^{-1}(p_k(H_{jnt}^{(s)})) \\ &\quad + \varpi_k(p_k(H_{jnt}^{(s)})) - \varpi_0(p_0(H_{jnt}^{(s)}))]] + \beta^{\overline{\rho}+1}[V_0(H_{jnt}^{(\overline{\rho}+1)}) + \varpi_0(p_0(H_{jnt}^{(\overline{\rho}+1)})) \\ &\quad + \sum_{k=1}^{3} p_k(H_{jnt}^{(\overline{\rho}+1)})[Q_k^{-1}(p_k(H_{jnt}^{(\overline{\rho}+1)})) + \varpi_k(p_k(H_{jnt}^{(\overline{\rho}+1)})) - \varpi_0(p_0(H_{jnt}^{(\overline{\rho}+1)}))]]\} \end{aligned} \tag{4.22}$$

Using equation(4.9) it follows from proposition 1 in Hotz and Miller(1993, p. 501) that $Q_k^{-1}(\widetilde{p}(H_{nt})$ is :

$$Q_j^{-1}(\widetilde{p}(H_{nt}) = \ln[p_j(H_{nt})/p_0(H_{nt})] \tag{4.23}$$

To complete the expression for $V_j(H_{nt})$, we need to characterise the form of the $\varpi_j(p_j(H_{nt}))$ functions associated with ε_{tj}. Given the the assumed distribution for $\varepsilon_{tj}'s$, the function takes the form:

$$\varpi_j(p_j(H_{nt})) = \gamma - \ln[p_j(H_{nt})] \tag{4.24}$$

for $j \in \{0,1,2,3\}$, where γ is Euler's constant (≈ 0.577). In this case, $Q_j^{-1}(\widetilde{p}(H_{nt}) = -[\varpi_j(p_j(H_{nt})) - \varpi_0(p_0(H_{nt}))]$, which implies that the conditional valuation functions do not depend on terms that show which choices will be optimal in the future. As a result, the dynamic selection terms, $Q_k^{-1}(p_k(H_{jnt}^{(s)})) + \varpi_k(p_k(H_{jnt}^{(s)})) - \varpi_0(p_0(H_{jnt}^{(s)}))$, drops out under this parameterization of the unobservables.

Finally inorder to estimate the model we will form moment conditions using equation4.23, i.e.

$$\ln[p_j(H_{nt})/p_0(H_{nt})] = V_j - V_0 \tag{4.25}$$

Substituting for V_j and V_0 from equation(4.22) and not all terms in that expression that involues histories $\overline{\rho}+1$ are that same for all options,we obtain

$$E_t\{\ln[\frac{p_j(H_{nt})}{p_0(H_{nt})}] - U_j^*(H_{nt}) + U_0^*(H_{nt}) - \sum_{s=1}^{\overline{\rho}} \beta^s [U_0^*(H_{jnt}^{(s)}) - U_0^*(H_{0nt}^{(s)}) + \ln[\frac{p_0(H_{0nt}^{(s)})}{p_0(H_{jnt}^{(s)})]}/] \mid \mathbb{I}_{tj}^{(o)} = 1\} = 0 \tag{4.26}$$

for $j = 1, 2, 3$ and $t = 1, ..., T$.

We then derive addtional moment conditions which responds to the continuous decision of number hours to work. In this way we are able to combine both the continuous and decrete decision in the same framework. Note that equation (4.22) also gives a alternative representation for the Euler equations for labor supply

$$\begin{aligned} 0 \quad = \quad & \frac{\partial U_j^*(H_{nt})}{\partial h_{nt}} + E_t\{\sum_{s=1}^{\overline{\rho}.} \beta^s \frac{\partial[U_0^*(H_{nt}^{(s)}) + \varpi_0(p_0(H_{jnt}^{(s)})]}{\partial h_{nt}}\} \\ & + E_t\{\sum_{s=1}^{\overline{\rho}.} \beta^s \sum_{k=1}^{3} p_k(H_{jnt}^{(s)}) \frac{\partial[Q_k^{-1}(p_k(H_{jnt}^{(s)})) + \varpi_k(p_k(H_{jnt}^{(s)})) - \varpi_0(p_0(H_{jnt}^{(s)}))]}{\partial h_{nt}}\} \\ & + E_t\{\sum_{s=1}^{\overline{\rho}.} \beta^s \sum_{k=1}^{3} [Q_k^{-1}(p_k(H_{jnt}^{(s)})) + \varpi_k(p_k(H_{jnt}^{(s)})) - \varpi_0(p_0(H_{jnt}^{(s)}))] \frac{\partial p_k(H_{jnt}^{(s)})}{\partial h_{nt}}\} \end{aligned} \tag{4.27}$$

for $j = \{1, 3\}$, conditional on these choices being optimal.

In order to evaluate the terms $\frac{\partial p_k(H_{jnt}^{(s)})}{\partial h_{nt}}$ which appears in the Euler equation, define

$$f_{1nt}^{(s),j} = f_1(H_{jnt}^{(s)} \mid \mathbb{I}_{jn,t+s} = 1) \tag{4.28}$$

as the probability density function for$H_{jnt}^{(s)}$, conditional on choosing option j at date $t+s$. Let

$$f_{nt}^{(s),j} = f(H_{jnt}^{(s)}) \tag{4.29}$$

as the related probability density that does not condition on choosing option j in period $t+s$. This both for $s = 1, ..., \overline{\rho}$. Denote their derivatives with respect to h_{nt} by $f_{1nt}^{\prime(s)}$ and $f_{nt}^{\prime(s)}$. Then

$$\frac{\partial p_k(H_{jnt}^{(s)})}{\partial h_{nt}} = \left[\frac{f_{1nt}^{\prime(s)}}{f_{1nt}^{(s)}} - \frac{f_{nt}^{\prime(s)}}{f_{nt}^{(s)}} \right] p_k(H_{jnt}^{(s)}) \tag{4.30}$$

The moment conditions defined in equations 4.26 and 4.27 now define moment restrictions which we can use to estimate the model, except for the incidental parameters in the form of the conditional choice probabilities , future transitional probabilities, their derivatives, the individual effects and aggregate components. In that respects we first estimate the these incidental parameters in separate stages (nonparametrically or parametrically) and substitute their estimate into the above moment conditions. As such we have a semiparametric estimation procedure, we derive the asymptotic properties in the appendix. In what follows we first describe the data and then go into more details for each stage of the estimation procedure..

5. Data

The data for this study are taken from the Family-Individual File, Childbirth and Adoption History File and the Marriage History File of the Michigan Panel Study of Income Dynamics (PSID). The variables used in the empirical study are h_{nt}, the annual fraction of hours work by individual n at date t; $\widetilde{w}_{nt}$, her reported real average hourly earnings at t; x_{nt}, real household food consumption expenditures; FAM_{nt}, the number of household members; $YKID_{nt}$, the number of children less than six years of age; $OKID_{nt}$, the number of children of ages between six and fourteen; AGE_{nt}, the age of the individual at date t; EDU_{nt}, the years of completed education of the individual at time t; $HIGH.SCH_{nt}$, completion of high school dummy; $BLACK$ and $HISPANIC$ race dummies for blacks and Hispanics, respectively; NE_{nt}, NC_{nt},SO_{nt}, which are region dummies for northeast, northcentral, and south, respectively, and MAR_{nt}, denoting whether a woman is married or not. The construction of our sample and the definition of the variables is described in greater detail in Appendix 3.

Table 1 contains summary statistics of our main variables. The sample has aged, household size has declined, and the decline is most pronounced amongst young children. The steep decline in household size over the two decades, and the aging evident in the sample, relative to aggregate trends in the US, largely reflects the sampling mechanism of the PSID. Thus we cannot infer any aggregate trend in fertility from this table. Household income has increased somewhat, but household consumption of food has declined. However, both food consumption and income per capita has increased over the sample period. More striking is the rise in female income, which greatly outstrips increases in household income. This is due to both higher wages and greater hours. Because schooling has not increased over the sample period, the number of years of formal education is not a factor in explaining aggregate trends in female wages and labor supply, or any changes that might have occurred in fertility.

6. Wages

We assume the wage rate, or value of marginal product function, Equation (2.1) can be parameterized as:

$$\begin{aligned} & g(d_{nt-\rho},...,d_{nt-\rho},...,d_{nt-1},h_{nt-\rho},...,h_{nt-1},z_{nt}) \\ = \quad & \omega_t\mu_n \exp[\sum_{s=1}^{\nu}(\delta_{1s}h_{n,t-s}+\delta_{2s}d_{n,t-s})+z_{nt}'B_3)] \end{aligned} \tag{6.1}$$

where μ_n and ω_t are an unobserved individual-specific effect and aggregate time-specific wage, respectively. We further assume that the reported wage rate, denoted $\widetilde{w}_{nt}$ (for the n^{th} household in period t) measures the woman's marginal product in the market sector with error, so that:

$$\widetilde{w}_{nt} = g(\widetilde{A}_{nt})\exp(\widetilde{\varepsilon}_{nt}) \tag{6.2}$$

where the multiplicative error term in equation (6.2) is conditionally independent over people, the covariates in the wage equation and the labor supply decision. Taking logarithms on both sides of Equation (6.2), and then differencing, yields:

$$
\begin{aligned}
\triangle\widetilde{\varepsilon}_{nt} &= \triangle\ln(\widetilde{w}_{nt}) - \sum_{s=1}^{\nu}(\delta_{1s}\triangle h_{n,t-s} + \delta_{2s}\triangle d_{n,t-s}) - \triangle z'_{nt}B_3 - \triangle\omega_t && (6.3)\\
&\equiv \triangle\ln(\widetilde{w}_{nt}) - Z_{nt}\Theta_1 && (6.4)
\end{aligned}
$$

where $\Theta_1 \equiv (\delta_{11}, ..., \delta_{1\nu}, \delta_{21}, ..., \delta_{2\nu}, B'_3)$ denotes the $(2\nu + k + T - 1)-$ dimensional vector of identifiable parameters, and Z_{nt} is the vector of covariates.

An instrumental variables estimator was used to estimate Equation (6.3). Defining Y_n and Z_n as:

$$
\begin{aligned}
Y_n &\equiv (\triangle\ln(\widetilde{w}_{n2}), ..., \triangle\ln(\widetilde{w}_{nT}))'\\
Z_n &\equiv (Z_{n2}, ..., Z_{nT})'
\end{aligned}
$$

we estimated Θ_1 with:

$$
\Theta_1^N = [N^{-1}\sum_{n=1}^{N} Z'_n\widetilde{W}_n^{-1}Z_n]^{-1}[N^{-1}\sum_{n=1}^{N} Z'_n\widetilde{W}_n^{-1}Y_n] \tag{6.5}
$$

where $\widetilde{W}_n$ is a consistent estimator of:

$$
W_n \equiv E[(Y_n - Z'_n\Theta_1)(Y_n - Z'_n\Theta_1)' \mid Z_n].
$$

Assuming the regressors are valid instuments for the wage equation, that is to say $E(\triangle\widetilde{\varepsilon}_{nt} \,| Z_n) = 0$ for each t, then Θ_1^N has the lowest asymptotic covariance within the class of GMM estimators.

Our estimates of the wage equation, displayed in Table III, are comparable to those reported in Miller and Sanders (1997) for the National Longitudinal Survey for Youth (NLSY), Altug and Miller (1998) also using the PSID, and others. All the coefficients are significant. Working an extra hour increases the wage rate up to four years hence, although in diminishing amounts. The effect is nonlinear, and this is captured by the participation variables. Age has a quadratic effect, eventually leading to declining productivity, and additional education mitigates the onset of the decline. We note that the linear terms on age are not identified.

The estimate quantitative magnitudes of past experience are also plausible. Recent working experience is more valuable than more distant experience: at 2000 hours per year, the wage elasticity of hours lagged once is about 0.18, but the wage elasticity of hours lagged twice is only 0.03. Also the further back the work experience is, the less the timing matters; an extra hour worked one year in the past has about twice the effect on current wages as an extra hour worked two years in the past, but the difference between the wage effects of an extra hour worked three and four years in the past, respectively, is less than 40%.

Another measure of the effect of past labor supply on wages: consider the total change in wages for a woman who has not worked up to date $t - \rho$ and then works the sample average of hours for those women who work, denoted $\overline{h}_t$. Then this measure is given by $\sum_{s=1}^{4}[\delta_{1s}\overline{h}_{t-s} + \delta_{2s}] = 0.12$. Much of this long-term effect is due to hours worked in the past year. Specifically, the growth in wages between $t-1$ and t for a woman who does not participate from $t-\rho$ to $t-2$, but works the sample average at $t-1$ is $\delta_{11}\overline{h}_{,t-s} + \delta_{21} = 0.08$. On the other hand, women who worked less than 1000 hours the previous year do not receive this increase in wages, this may be capturing the effect of discouragement normally found in the standard job search model. It should be noted that we do not explicitly model this type of search cost in our model, however, we can pick up the lower bound of this effect. This means that not everybody gets the benefit from past job experience, there is a threshold number of hours of about 1500 for this positive effect to kick in. This will impact fertility behavior even more than if there were positive benefit from all levels of past hours, since a mother could reduce her hours and still continue to enjoy the benefit of higher future wages. We will come back to this point in the empirical findings section when we will have estimates of the fraction of time a mother spends nuturing her new born.

The estimated change in aggregate wages over our sample period is displayed in Figure I, along with its 99% confidence interval. The most striking feature of that plot is that although the magnitude of the changes fluctuate over the sample period, the signs are always positive. This shows that over time the aggregate females wage has been increasing. This is not a surprising finding, given the fact the wage gap between males and females having been closing over time. However it does raise an interesting issue as to whether the attachment of females to the labor force, in term of their persistence in labor participation, is having an aggregate effect. For example, suppose by more females working more hours and participating on a more consistent level equivalent to men, then the employers in the aggregate are willing to pay females higher wages closer to males. This higher wages, some would argue, would then cause females to work more and have less children. Our approach can also disentangled such a result by controlling for aggregate shock, and then seeing the relative importance of the wage effect.

7. Preferences over Consumption and Wealth Effects

In our model, the effects of differences in wealth across households on their fertility and labor supply decisions is determined a single parameter, their weight in the social planner's problem. The inverse of their social weight is their marginal utility of wealth, and it can be estimated with household data on consumption This section reports our estimates of the parameters determining the utility from consumption and the marginal utility of wealth parameter, to be used in the labor supply and fertility equations that follow.

7.1. The first order condition

We assume that preferences over consumption take the parametric form:

$$u_3(x_{nt}, z_{nt}, \varepsilon_{5nt}) = \exp(z'_{nt}B_2 + \varepsilon_{5nt})x^{\alpha}_{nt}/\alpha \tag{7.1}$$

where the concavity parameter $\alpha < 1$. Substituting equation (7.1) into equation (3.3), taking logarithms and then first differencing yields:

$$(1-\alpha)^{-1} \triangle \varepsilon_{5nt} = \triangle \ln(x_{nt}) - (1-\alpha)^{-1} \triangle z'_{nt}B_2 + (1-\alpha)^{-1} \ln(\lambda_t) \tag{7.2}$$

Let Θ_2 denote the $(K+T-1)$ dimensional vector of parameters to be estimated, defined:

$$\Theta_2 = \begin{pmatrix} (1-\alpha)^{-1}B_2 \\ (1-\alpha)^{-1}\ln(\lambda_2) \\ \dots \\ (1-\alpha)^{-1}\ln(\lambda_T) \end{pmatrix}$$

We also define $Y_n = (\triangle\ln(x_{n2}),...,\triangle\ln(x_{nT}))'$ as a vector of endogenous variables, and Z_n the exogenous variables as

$$Z_n = \begin{bmatrix} \triangle z'_{n2} & D_2 & ... & 0 \\ . & . & . & . \\ \triangle z'_{nT} & 0 & ... & D_T \end{bmatrix}$$

where D_t denotes a time dummy for $t \in \{2,\dots,T\}$. The assumptions in Section 2 imply that the unobserved variable ε_{5nt} is independent of individual specific characteristics. Therefore $E((1-\alpha)^{-1}\triangle\varepsilon_{5nt}\,|z_{nt}) = 0$. Substituting for $(1-\alpha)^{-1}\triangle\varepsilon_{5nt}$ using equation(7.2) one can obtain a set of orthogonality conditions:

$$E\left[(Y_n - Z_n\Theta_2)\,Z_n\right] = 0$$

which can be exploited here to estimate Θ_2 using a similar method to the regression procedures that estimated the wage function.

The estimates of the consumption equation are based on the main sample of females for the years 1968 to 1992. Consumption for a given year in our study is measured by taking 0.25 of the value of the different components for year $t-1$ and 0.75 of it for year t. This is explained in more detail in the data appendix. The elements of z_{nt} used in this stage of the estimation are defined as FAM_{nt}, $YKID_{nt}$, $OKID_{nt}$, AGE^2_{nt}, NC_{nt} and SO_{nt}. The estimates in Table 4 show that consumption increases with family size and children consume less than adults, since the coefficients on children between the ages of zero and fourteen are negative and smaller in absolute magnitude than the coefficient on total household size. Furthermore, the behavior of consumption over the life-cycle is concave since the coefficient on age squared is negative. All the other coefficients are significant. Figure IV shows the estimated aggregate component of shadow value of consumption, along with its 99% confidence interval. This shows that these components are estimated very precisely. In fact, there is also significant variation over time as the test statistic for the null hypothesis that $(1-\alpha)^{-1}\Delta\ln(\lambda_t) = (1-\alpha)^{-1}\Delta\ln(\lambda_{t-1})$ for $t = 1969,...,1992$ is 395. Under the null hypothesis, it would be distributed as a χ^2 with 23 degrees of freedom, implying rejection of the null at 99% significance levels.

7.1.1. Individual-specific Effects

Estimation of the labor supply and fertility equations also requires estimates of individual-specific effects $\eta_n\mu_n$ and η_n. There are two approaches that we shall employ for estimating there quantities, the traditional fixed effects estimators and the regression approach of Macurdy (1982), extended by Altug and Miller (1998) to handle nonlinearities using nonparametric estimation techniques. The traditional estimators are simple to compute. They are, however, subject to small sample bias arising from short panel length, although the limited Monte Carlo evidence provided in Hotz and Miller (1988) suggests that small sample bias might not greatly affect the estimates of the other parameters. On the other hand, the nonparametric estimator achieves consistency of the cross section of the panel data set, but can only deal with any unobserved permanent characteristic that is a mapping of observed random variables.

The traditional fixed effect are estimated as follows. Let T_1 denote the number of time periods for which the wage equation is estimated, and T_2 be the number of time periods for which the marginal utility of consumption equation is estimated.

Let:

$$\begin{aligned}\phi_{1n} \quad &\equiv \quad \sum_{t=1}^{T_1}\left[\ln\left(w_{nt}\right) - \sum\nolimits_{s=1}^{\nu}\left(\delta_{1s}h_{n,t-s} + \delta_{2s}d_{n,t-s}\right) - z'_{nt}B_3\right]/T_1 \\ &\quad + \sum_{t=1}^{T_1}[\ln(x_{nt}) - (1-\alpha)^{-1}z'_{nt}B_2 + (1-\alpha)^{-1}\ln(\lambda_t)]/T_2 \end{aligned} \tag{7.3}$$

and

$$\phi_{2n} \equiv \sum_{t\in T_1}[\ln(x_{nt}) - (1-\alpha)^{-1}z'_{nt}B_2 + (1-\alpha)^{-1}\ln(\lambda_t)]/T_2 \tag{7.4}$$

Then estimates of $\eta_n\mu_n$ and η_n are then estimated by simple time averages of the estimated residuals of the consumption and wage equations.

Suppose that $\eta_n\mu_n$ and η_n can be expressed as functions of a Q-dimensional vector of regressors z_n, which is assumed to represent the permanent characteristics of individual n. Let the vector z_n be observed and satisfies the conditions, $E[z_n(\phi_{1n} - \eta_n\mu_n)] = 0$ and $E[z_n(\phi_{2n} - \eta_n)] = 0$. Here z_n could include such observed observable demographic characteristics as religion, marital status, the age distribution of children, home ownership, educational level and geographical location. Let δ_N denote the bandwidth of the proposed kernel estimator and J the normal kernel on R^Q. Then our estimators are

$$\eta_n^N\mu_n^N = \frac{\sum_{m=1,m\neq n}^{N}\phi_{1m}J[\delta_N^{-1}(z_m - z_n)]}{\sum_{m=1,m\neq n}^{N}J[\delta_N^{-1}(z_m - z_n)]} \tag{7.5}$$

and

$$\eta_n^N = \frac{\sum_{m=1,m\neq n}^{N}\phi_{2m}J[\delta_N^{-1}(z_m - z_n)]}{\sum_{m=1,m\neq n}^{N}J[\delta_N^{-1}(z_m - z_n)]} \tag{7.6}$$

The distribution of the estimated fixed-effects estimates for the wage equation, the consumption equation, and the combined individual effects and time effects of the consumption, are displayed in figures II, III, IV and V respectively. The most stricking features of these plots is that the nonparametric estimates are significantly smoother than the traditional fixed effects. The range of the distribution of the shadow price of consumption all lies above zero, while only a very small portion of the estimated (inverse) social weights lies below zero in the traditional case.

8. Participation, Hours and Birth

8.1. A Parametrization

The remaining components of the utility function u_1 and u_2 are parametrized as:

$$u_2(z^*_{nt}, d_{nt}) = d_{nt} z'_{nt} B_0$$

and

$$u_1(z^*_{nt}, l_{nt}) = z'_{nt} B_1 l_{nt} + \sum_{s=0}^{\rho} \delta_s l_{nt} l_{nt-s}$$

In these expressions, B_0 are parameters that characterize the fixed-costs of participating in the work force, B_1 shows the effect of exogenous time-varying characteristics on the marginal utility of leisure. Preferences are concave, decreasing in time spent in lesure if $z'_{nt} B_{11} l_{nt} + 2\delta_0 l_{nt} + \sum_{s=1}^{\rho} \delta_s l_{nt-s} > 0$ and $\delta_0 < 0$. The parameters δ_s for $s = 1, ..., \rho$ capture intertemporal non-separabilities in preferences with respect to leisure choices. A value of $\delta_s < 0$ for $s = 1, ..., \rho$ means that leisure s periods ago increases the marginal utility of leisure, and results in less work and child cre time today. Equivalently, a finding of $\delta_s < 0$ implies that current and past leisure time are substitutes where as $\delta_s > 0$ implies that current and past leisure time are complements.

The distributional assumption for the idiosyncratic stocks and the parameterization of the utility function implies that the Euler equation can be written as

$$E_t\{\eta_n \lambda_t w_{nt} - z'_{nt} B_1 - 2\delta_0 l^{(1)}_{nt} - \sum_{s=1}^{\rho} \delta_s l_{nt-s} - E_t[\sum_{s=1}^{\overline{\rho}} \beta^s [p_0(H^{(s)}_{1nt})^{-1} \frac{\partial p_0(H^{(s)}_{1nt})}{\partial h_{nt}}] \mid \mathbb{I}^o_{1nt} = 1\} = 0 \tag{8.1}$$

$$E_t\{\eta_n \lambda_t w_{nt} - z'_{nt} B_1 - 2\delta_0 l^{(3)}_{nt} - \sum_{s=1}^{\rho} \delta_s l_{nt-s} - \sum_{s=1}^{\overline{\rho}} \beta^s [p_0(H^{(s)}_{3nt})^{-1} \frac{\partial p_0 H^{(s)}_{3nt})}{\partial h_{nt}}] \mid \mathbb{I}^o_{3nt} = 1\} = 0 \tag{8.2}$$

In this parametization, equation 4.26 gives the following three moment conditions:

$$E_t \left\{ \begin{array}{c} \ln\left(\frac{p_{1nt}}{P_{0nt}}\right) - z'_{nt} B_0 + z'_{nt} B_1 (l^{(0)}_{nt} - l^{(1)}_{nt}) + \delta_0 (l^{(0)2}_{nt} - l^{(1)2}_{nt}) \\ + \sum_{s=1}^{\rho} \delta_s (l^{(0)}_{nt} - l^{(1)}_{nt}) l_{nt-s} - \eta_n \lambda_t w_{nt} h^*_{nt} - \sum_{s=1}^{\overline{\rho}} \beta^s \ln(p_0(H^{(s)}_{0nt})/p_0(H^{(s)}_{1nt})) \end{array} \right. |\mathbb{I}^o_{1nt} = 1\} = 0 \tag{8.3}$$

$$E_t \left\{ \begin{array}{c} \ln\left(\frac{p_{2nt}}{P_{0nt}}\right) + z'_{nt} B_1 (l^{(0)}_{nt} - l^{(2)}_{nt}) + \delta_0 (l^{(0)2}_{nt} - l^{(2)2}_{nt}) \\ + \sum_{s=1}^{\rho} \delta_s (l^{(0)}_{nt} - l^{(2)}_{nt}) l_{nt-s} + \eta_n \lambda_t z'_{nt} \pi \\ -(\gamma_0 + \sum_{s=1}^{\rho} \gamma_s b_{nt-s} + \gamma_S \sum_{s=\rho+1}^{t} b_{nt-s}) - \sum_{s=1}^{\overline{\rho}} \beta^s \ln(p_0(H^{(s)}_{0nt})/p_0(H^{(s)}_{2nt})) \end{array} \right. \mid \mathbb{I}^o_{2nt} = 1\} = 0 \tag{8.4}$$

$$E_t \left\{ \begin{array}{c} \ln\left(\frac{p_{3nt}}{P_{0nt}}\right) - z'_{nt} B_0 + z'_{nt} B_1 (l^{(0)}_{nt} - l^{(3)}_{nt}) + \delta_0 (l^{(0)2}_{nt} - l^{(3)2}_{nt}) \\ + \sum_{s=1}^{\rho} \delta_s (l^{(0)}_{nt} - l^{(3)}_{nt}) l_{nt-s} - \eta_n \lambda_t w_{nt} h^*_{nt} + \eta_n \lambda_t z'_{nt} \pi \\ -(\gamma_0 + \sum_{s=1}^{\rho} \gamma_s b_{nt-s} + \gamma_S \sum_{s=\rho+1}^{t} b_{nt-s}) - \sum_{s=1}^{\overline{\rho}} \beta^s \ln(p_0(H^{(s)}_{0nt})/p_0(H^{(s)}_{3nt})) \end{array} \right. \mid \mathbb{I}^o_{3nt} = 1\} = 0 \tag{8.5}$$

8.2. Nonparametric estimation of the conditional choice probabilities

Then the probabilities p_{knt} can be computed as nonlinear regressions of participation index $\mathbb{I}_{knt}$ on the current state H_{nt}^{N}, where the N superscript denotes an estimated quantity. Define $K[\delta_N^{-1}(H_{nt}^{N}-H_{mr}^{N})]$ as a given kernel , where δ_N is the bandwidth associated with each argument. Then the nonparametric estimate, p_{knt}^{N} , is then computed using kernel estimator

$$p_{knt}^{N} \equiv \frac{\sum_{m=1,m\neq n}^{N}\sum_{r=1,r\neq t}^{T}\mathbb{I}_{kmr}K[\delta_N^{-1}(H_{nt}^{N}-H_{mr}^{N})]}{\sum_{m=1,m\neq n}^{N}\sum_{r=1,r\neq t}^{T}K[\delta_N^{-1}(H_{nt}^{N}-H_{mr}^{N})]} \tag{8.6}$$

The conditional choice probabilities $p_j(H_{knt}^{(s)})$ are also estimated as nonlinear regressions of a choice index on the appropriate state variables. Define the variable

$$\mathbb{I}_{knt}^{(s)} = \mathbb{I}_{kn,t-s}\prod_{\ell=1}^{s-1}(1-\mathbb{I}_{kn,t-\ell}), \quad k\in\{0,1,2,3\} \tag{8.7}$$

Notice that $\mathbb{I}_{knt}^{(s)}=1$ if the person choose option k at $t-s$, but then did not choose option k for $s-1$ periods. Thus, $\mathbb{I}_{knt}^{(s)}$is an index variable that allows us to condition on the behavior of individuals with the labor market and birth histories defined by $z_{knt}^{(s)}$. The conditional choice probabilities $p_j^N(H_{knt}^{(s)})$ are then computed as

$$p_j^N(\Psi_{knt}^{(s)}) \equiv \frac{\sum_{m=1,m\neq n}^{N}\sum_{r=1,r\neq t}^{T}\mathbb{I}_{jmr}\mathbb{I}_{kmr}^{(s)}K[\delta_N^{-1}(H_{knt}^{(s)N}-H_{kmr}^{(s)N})]}{\sum_{m=1,m\neq n}^{N}\sum_{r=1}^{T}\mathbb{I}_{kmr}^{(s)}K[\delta_N^{-1}(H_{knt}^{(s)N}-H_{kmr}^{(s)N})]} \tag{8.8}$$

where $H_{knt}^{(s)N}\equiv(z_{knt}^{(s)\prime},\mu_n^N\eta_n^N)'$ for $k\in\{0,1,2,3\}$ is the state vector for individual n.

8.3. Hours, participation and birth Conditions

So combining the estimate we obtain in section 7.1.1 gives

$$\xi_{nt} = (1-\alpha)^{-1}\widehat{\ln(\lambda_t\eta_n)} \tag{8.9}$$

Note that we can obtain an estimate of $\eta_n\lambda_t$ and η_n, respectively, as:

$$\widehat{\eta_n\lambda_t} \equiv \exp((1-\alpha)\xi_{nt}) \tag{8.10}$$

The remaining unknown parameters of the model consist of the parameters in the hours, participation ond birth equations, the discount factor β and the risk aversion parameter from the utility function for consumption, α.Define the $(3K+\rho+\overline{\rho}+2)-$dimemensional vector $\Theta_3\equiv(B_0',B_1,'\delta_0,...,\delta_\rho,\gamma_0,...,\gamma_{\overline{\rho}},\gamma_S,\alpha,\pi)'$. We do not estimate the discount factor because of problem with identitifying it, this is a standard problem most dynamic structural models so instead we estimate the model using different value of β. This fact motivates an estimator of Θ_3, conditional on β. For this purpose, the idiosyncratic errors or moment restrictions associated with the Euler and discrete choice equations can be written as

$$\begin{aligned}\widehat{m}_{1nt}(\Theta_3) \quad = \quad & \mathbb{I}_{1nt} \times \left(\exp((1-\alpha)\xi_{nt})\widehat{w}_{nt} - z'_{nt}B_1 - 2\delta_0 l^{(1)}_{nt} - \sum_{s=1}^{\rho} \delta_s l_{nt-s} - \sum_{s=1}^{\overline{\rho}} \beta^s \widehat{p}_0(H^{(s)}_{1nt})^{-1} \frac{\widehat{\partial p_0(H^{(s)}_{1nt})}}{\partial h_{nt}} \right) \\ & + \mathbb{I}_{3nt} \times \left(\exp((1-\alpha)\xi_{nt})\widehat{w}_{nt} - z'_{nt}B_1 - 2\delta_0 l^{(3)}_{nt} - \sum_{s=1}^{\rho} \delta_s l_{nt-s} - \sum_{s=1}^{\overline{\rho}} \beta^s \widehat{p}_0(H^{(s)}_{3nt})^{-1} \frac{\widehat{\partial p_0 H^{(s)}_{3nt})}}{\partial h_{nt}} \right) \end{aligned} \tag{8.11}$$

$$\widehat{m}_{2nt}(\Theta_3) = \mathbb{I}_{1nt} \times \begin{pmatrix} \ln(\frac{\widehat{p_{1nt}}}{\widehat{P_{0nt}}}) - z'_{nt}B_0 + z'_{nt}B_1(l^{(0)}_{nt} - l^{(1)}_{nt}) + \delta_0(l^{(0)2}_{nt} - l^{(1)2}_{nt}) \\ + \sum_{s=1}^{\rho} \delta_s(l^{(0)}_{nt} - l^{(1)}_{nt}) l_{nt-s} - \exp((1-\alpha)\xi_{nt})\widehat{w}_{nt} h_{nt} - \sum_{s=1}^{\overline{\rho}} \beta^s \ln(\widehat{p}_0(H^{(s)}_{0nt})/\widehat{p}_0(H^{(s)}_{1nt})) \end{pmatrix} \tag{8.12}$$

$$\widehat{m}_{3nt}(\Theta_3) = \mathbb{I}_{2nt} \times \begin{pmatrix} \ln(\frac{\widehat{p_{2nt}}}{\widehat{P_{0nt}}}) + z'_{nt}B_1(l^{(0)}_{nt} - l^{(2)}_{nt}) + \delta_0(l^{(0)2}_{nt} - l^{(2)2}_{nt}) \\ + \sum_{s=1}^{\rho} \delta_s(l^{(0)}_{nt} - l^{(2)}_{nt}) l_{nt-s} + \exp((1-\alpha)\xi_{nt}) z'_{nt}\pi \\ -(\gamma_0 + \sum_{s=1}^{\rho} \gamma_s b_{nt-s} + \gamma_S \sum_{s=\rho+1}^{t} b_{nt-s}) - \sum_{s=1}^{\overline{\rho}} \beta^s \ln(\widehat{p}_0(H^{(s)}_{0nt})/\widehat{p}_0(H^{(s)}_{2nt})) \end{pmatrix} \tag{8.13}$$

and

$$\widehat{m}_{4nt}(\Theta_3) = \mathbb{I}_{3nt} \times \begin{pmatrix} \ln(\frac{\widehat{p_{3nt}}}{\widehat{P_{0nt}}}) - z'_{nt}B_0 + z'_{nt}B_1(l^{(0)}_{nt} - l^{(3)}_{nt}) + \delta_0(l^{(0)2}_{nt} - l^{(3)2}_{nt}) \\ + \sum_{s=1}^{\rho} \delta_s(l^{(0)}_{nt} - l^{(3)}_{nt}) l_{nt-s} - \exp((1-\alpha)\xi_{nt})\widehat{w}_{nt} h_{nt} + \exp((1-\alpha)\xi_{nt}) z'_{nt}\pi \\ -(\gamma_0 + \sum_{s=1}^{\rho} \gamma_s b_{nt-s} + \gamma_S \sum_{s=\rho+1}^{t} b_{nt-s}) - \sum_{s=1}^{\overline{\rho}} \beta^s \ln(\widehat{p}_0(H^{(s)}_{0nt})/\widehat{p}_0(H^{(s)}_{3nt})) \end{pmatrix} \tag{8.14}$$

Let $\widehat{m}_{nt}(\Theta_3) \equiv (\widehat{m}_{1nt}, \widehat{m}_{2nt}, \widehat{m}_{3nt}, \widehat{m}_{4nt})'$ and let T_3 denote the set of periods for which the hours and discrete participation conditions are valid. Define $\widehat{m}_n(\Theta_3) \equiv (\widehat{m}'_{n1}, ..., \widehat{m}'_{nT_3})'$ as the the vector of moment restrictions for a given individual over time. Similarly, define $\Phi_n \equiv E_t[\widehat{m}_n(\Theta_3)\widehat{m}_n(\Theta_3)']$. Notice that the matrix Φ_n is block diagonal with diagonal elements defined as $\Phi_{nt} \equiv E_t[\widehat{m}_{nt}(\Theta_3)\widehat{m}_{nt}(\Theta_3)']$, and off-diagonal elements that are zero because $E_t[\widehat{m}_{nt}(\Theta_3)\widehat{m}_{ns}(\Theta_3)'] = 0$ for $s \neq t$, $s < t$. The 4×4 conditional heteroscedasticity matrix $\widehat{\Phi}_{nt}$ associated with the indivdiual-specific errors $\widehat{m}_{nt}(\Theta_3)$ is evaluated using a nonparametric estimator based on the estimated residuals, $\widehat{m}_{nt}(\Theta_3)$, using an initial consistent estimator of Θ_3.This estimator is simalar to Robinson(1987) estimator except we use a kernel based nonparametric regressions instead of a Nearest neighbor regression approach. To ensure none zero variance we trimmed the data. The optimal GMM estimator for, Θ_3 satisfies

$$\widehat{\Theta}_3 = \arg\min_{\Theta_3}[1/N \sum_{n=1}^{N} \widehat{m}_n(\Theta_3)]\widehat{\Phi}_n[1/N \sum_{n=1}^{N} \widehat{m}_n(\Theta_3)]'. \tag{8.15}$$

Appendix A.3 derives that asymptotic properties of a general class of estimators which can be used to show consistency and asymptotic nomality of the estimator $\widehat{\Theta}_3$.

9. Empirical Findings

Tables V, VI, VII and VIII contain estimates of alternative estimators of our participation cost,nurturing cost, utility of leisure and birth equation. Column (1) reports estimates of the birth preference and cost parameters that are based on nonparametric estimates of individual effects η_n and $\mu_n\eta_n$, while estimates in column (2) are based on the standard effects estimators of η_n and $\mu_n\eta_n$. The most striking feature of our results is the similarities between both sets of estimates. Given these similarities, we will focus in the discussion that follows on the nonparametric estimates.

9.1. Fixed Cost of Participation

Table V contains estimates of the fixed cost of participation. First the constant term is negative, which means that particiaption in the labor force has a fixed utility cost instead of a benefit, which is what standard economic theory would predict. Age reduces this cost of participation in the labor, but this reduction is at a decreasing rate as the parameter estimate on the AGE^2 is negative. Education increases the cost associated with age. There is a negative sign on the estimates of AGE$\times$EDUC which implies that a more educated female has a higher cost of participation for a given age than a less educated female. To understand the overall effect of age and education on the fixed cost of participation, we investigate what is the shape of this function conditional on education. Females with less that 7.7 years of education have a concave function in age, while females with more that 7.7 years of education have the cost of participation increasing at an increasing rate in age. One possible explanation for this result maybe the fact that less educated females will earn less over their life-time and as they get older will not have the discretion of whether to work are not. So in the data we would expect to observe that older females with small levels of education to be participating in the labor force at a higher rate. Married women have a higher cost of participation while blacks and hispanics have a lower cost of participation for a given age and education level. Again these results are not surpising since the standard literature has documented similar results(see for example Altug and Miller (1998))..

9.2. Nurturing Cost

Table VI contains the results from the estimation of the fraction of time spent nurturing a child. The risk aversion parameter is very reasonable for a CES utility function. These estimates seems quite reasonable. For example, a new birth seems to require about 35% of the mother's time, and this falls to about 16% for a five years old child. These are similar results to those found by Hotz and Miller (1984) which found that these parameters follow off a geometric rate. This is very important in our model, since with the nonlinearity observed in the estimates of the wage equation, this implies that if a female reduces her time in the labor force to have a child, then they would not benefit from the increases in wages as a result of human capital accumulation in terms of their previous labor supply. So holding all other things constant, this would make having children less desirable for a female who is on a high wage trojectury. This combined with the estimates of the risk aversion parameter means that females would like to smooth more there consumption, hence working more in earlier years and delaying child-bearing to later years. This would mean that working females would have less children than nonworking female.

9.3. Utility Cost of Leisure

Table VII contains the estimates for the utility cost of leisure. Although the sign on the l_{nt} is negative this do not mean that females in our sample obtain a negative utility from leisure, which would contradict theory, we have to look at the over all first order effect of l_{nt}. For example, for an average women in our sample, the direct effect on the utility from leisure is positive .The second order effects of leisure(i.e. term on the the squared of leisure) is negative giving us the standard concave utility function in our results. Our estimates suggest that leisure is intertemporily nonseparable. Past leisure are compliments with for current leisure These is similar to what is found in Altug and Miller (1998), among other, about the separability of leisure. Another, supprising results we found is that the sign on marriage in our results is negative. At first glance, this would imply that married females love leisure less. One explanation for this effect could be simple the fact that married females are working more than before and is still having children. Since we do not allow at the moment for the utility of birth or the time cost of raring a child to depend on such demographics, as marital status, then the only way they found then be having children and still working is if they as a group love leisure less. Another explanation may be due to the welfare system.In the era of our sample, a subsistence income (AFDC) is available to unmarried mothers, but (basically) only conditional on them not working. Married females do not face a similar tradeoff. Since welfare participation among female heads is quite common in this era (roughly around one-third), this is definitely an important enough phenomenon to account for this results.In short, the "leisure" time of female heads is highly subsidized, and they may well have similar preferences as wives.[9] This is some thing that we will explore further.

9.4. Birth Effects

We concurred with the classical literature that children are good and not bad, since we find a positive utility up to the 6th birth. The parameter on the timing of births for example, would imply that the optimal space of a two-child family would be 3 to 4 years apart. So, having children too close or too far apart is less desirable. Turning to the cost of a child, we find that both sets of estimates give similar results. There is a positive cost discounted life-time cost to having a child. We find that having at least a high school education significantly increases that cost. After controlling for education, we find that Blacks and Hispanics have a significantly lower cost than White. The fact that education significantly increases the cost of having a birth coincides with our earlier hypothesis, and can help explain the unanimous empirical finding that number of children is negatively related to level of education.

10. Policy Simulations

There are many ways in which public policy over the last century has affected the costs and benefits of having children. From child labor laws to the public provision of schooling, from the subsidizing of health care to local taxes that support amenities such as swimming pools, as

[9] We would like to thank Elizabeth Powers for pointing out this very insightful possibility to us.

well as sporting and other events for children, raising children depends on social infrastructure that is often taken for granted in modern developed societies. Over the last several decades, greater attention has been paid to jointly determining fertility and female labor supply. Part of the concern about the falling rates of fertility are related to the long-term viability of the social security system in many developed countries, especially in Western Europe.

This section considers a variety of policies that subsidize fertility to investigate how responsive women are to changes in the incentives they factor in between market work and raising a family. Our study shows that different policies not only have different aggregate or average effects on fertility and female labor supply, but also have very significant compositional effects, or incidence across this heterogeneous population. We hasten to add that our contribution is postitive, not normative, seeking to provide quantitative analysis against which different policy options can be evaluated.

10.1. Overview of the simulations

We substituted the parameters obtained from our estimation procedures into the utility function, the equation characterizing the returns to experience, and the child care cost equation and solved the decision-maker's problem. We conducted simulations for a wide range of female types in the population, but they are not exhaustive. We stratified the population, breaking down the groups according to a three-way classification scheme, by race, marriage and education, and considered an individual whose unobserved fixed effects correspond to the estimated means of the distributions. Three racial types were considered, namely Black, White and Hispanic (respectively abbreviated B, H and M in Tables IX and X below). Marriage was a dichotomous variable partitioning women by marital status at age 25, where M denotes she was married at age 25 or before, and U if not. We considered three educational groups, those who completed some years at college (denoted by the inequality sign $>$), those who completed some years at high school but not college (denoted by HS), and those with less education than that (denoted by a $<$ sign). Thus our simulations apply to women in the 18 categories whose marginal utilities' of wealth, and whose endowed marginal product of labor (controlling for schooling and experience), correspond to the estimated sample means.

The models we simulated are slightly less complex than the estimation framework itself in three ways. The first simplification was to limit the choice set. Rather than assuming that workers made a discrete choice about whether to participate in the labor force or not with a continuous hours choice, we discretized the labor supply choice set facing workers, limiting them to 10 equally spaced choices in the $[0, 1]$ interval. Second, we linearized the value of marginal consumption around the marginal utility of consumption achieved in the current regime. Thus in the objective function (2.9), U_{3ntk} is replaced with $\widetilde{U}_{3ntk} \equiv \eta_n^{-1}$. Third, we investigated an economy where there are no aggregate shocks. As a practical matter, the quantitative significance of aggregate demographic shocks (such as the baby boom in the U.S., the AIDS crisis in Botswana and other countries, the effects on fertility of immigration both legal and illegal into U.S. and parts of Western Europe) is difficult to overstate, and we think that excluding them is the main reason why our results should be treated cautiously.

The model was solved for each group under five policy regimes. The benchmark regime, labelled Estimation, is the current one, which may be compared with the conditional sample means from the data set. In the first two alternative regimes we analyze the subsidy to having children does not vary with the recipient, although the value a mother places on the scheme depends on her wealth and wage rate. In the regime labelled Expenses, the state pays all the estimated monetary costs associated with raising children, removing the wedge in the marginal utility of wealth between households that have children and those that do not. Under the Daycare policy, maternal time is replaced with publicly funded child care centers. In the other two regimes the payment mothers receive depends on her wages and hours she worked before taking time off to have a child. The Wages policy would pay the mother the wages she would have received if she had decided against having her child. If the Retraining policy is adopted, mothers are given retraining upon reentering the workforce that fully restore the human capital from lost workforce experience.

In our model there are three costs associated with childcare: the lifetime discounted cost of market inputs used up raising a child, the direct time cost in terms of the required for nurturing, and the human capital accumulation cost stemming from the experience acquired from working that is not used when women quit the labor force to have children. We will provide the costs of each policy in a future version of this paper.

10.2. Solving the Model

We first simulated the prediction of the model for females in each of the categories described above over the 25 years of a partial life cycle starting at age 20, for use as a bench mark case. This requires us to solve 18 valuation functions for the optimization problem each type solved, obtain the optimal decision rules, and thus compute the probabilities of observing any given decision, as a mapping of the state variables, which in this case are the vector of lagged labor supplies and a vector for the ages of the offspring. An appendix describes the algorithm in detail. Briefly, we combined the use of both policy function iteration (using Newton steps) with value function iteration (using the contraction operator on the value function). Convergence to the solution of the infinite horizon problem occurred relatively quickly, typically within seven iterations.

The labor force participation rate and expected fertility rate over this period (essentially the TFR) for each type is reported in the second column of Tables IX and X under the heading of Estimation. A sense of how representative our groups are is found by comparing the simulated results for our estimated model with their corresponding sample means in the first column, headed Actual. Note that the numbers are not very close, although many of the inequalities within each column are preserved. This is attributable to two factors. The first is estimation error. The second is that the sample means do not condition on the values of the unobservables, which enter in a highly nonlinear way into the participation and fertility choices. To separate out these separate influences, we will nonparametrically estimate the same set of statistics for that person in the group with the estimated mean fixed effects, which simply weights the data used to obtain the averages in the first column by how close each observation is to the mean estimated fixed effect vector.

Table IX shows most of the types have fertility rates below the replacement rate of 2. For example, the TFR of all the college educated groups are all below the replacement rate. College educated white females bear the least number of children (1.1 for the group as a whole and 1.2 at the mean fixed effects), and black married females with less than high school education the most (2.1 for the overall group and 2.4 at the mean fixed effects).

In most, but not all groups, those married by 25 bear more children than those who had not married by then. Table X shows that, with the notable exception of college educated whites, unmarried women are more likely to participate in the labor force. At 0.93, the labor force participation rate for a married college educated white female with the mean fixed effects exceeds all other groups, closely followed by unmarried college educated black women (at 0.91). Across education achievement and marital status but within race categories, blacks exhibit the biggest range in labor force participation rates. The exact derivation is presented in more details in Appendix 1.

10.3. Childcare Support

There are many ways to subsidize fertility by having the state pay for the discounted lifetime cost of children. For example, it could be achieved though tax credits at upper income levels and child support payments for those who do not receive enough taxable income. In this framework this is equivalent to imposing the constraints $\pi_0 = 0$ and $\pi_1 = 0$ in the expression for child care costs:

$$\pi\left(z_{nt}\right) = \pi_0 + z'_{nt}\pi_1$$

The total fertility and labor force participation rates that are induced by this subsidy are shown in the third columns of Tables IX and X. Paying the market goods inputs for raising children has a substitution and wealth effects. In a static model, the substitution effect induces women to have more children and reduce their own consumption of leisure and other goods, while the wealth effect induces them to increase their consumption of leisure and children. The results of the dynamic simulations lend support to this intuition. In 16 of the 18 groups labor force participation declines, and in all but one instance fertility rises, 6 groups (compared to 4) now settling above the replacement rate. The 3 types whose fertility behavior is most sensitive to this policy shift are the married non-college educated black female and the unmarried lowest educated black female. By way of contrast the biggest reduction in labor force participation rate is amongst unmarried high school educated whites.

10.4. Daycare

Rather than pay for market inputs directly, another public policy for subsidizing fertility is to expand the availability of child care services for the mothers of infants and preschool age children, by financially supporting centers, or reimbursing mothers who place their children in them. In our framework a policy that eliminates the maternal time inputs altogether would set $\rho_i = 0$ for $i \in \{1, \ldots, 5\}$. This increases the amount of time mothers of young children have for leisure and work. In a static model of fertility and labor supply, fertility increase in response to a reduction in one of its factor inputs, maternal time. Furthermore, the time freed up from looking after children is distributed between extra leisure, and working for more goods and services over and above those used up by the additional children. Consequently, one predicts that both fertility and labor supply would increase, the latter less than the amount of time released from child care.

The fourth column shows the labor force participation and fertility outcomes from solving the optimization problem under the Daycare policy. As expected all the group exhibit higher fertility rates, 12 now at or above the replacement rate of 2.0, with married high school educated white females registering the biggest increase (from a TFR of 1.52 to 2.30). Comparing the effects on TFR across different groups, we see that switching from subsidizing market inputs to replacing maternal time inputs has a far greater impact on females with some college education than those who did not complete high school. Indeed in just one group, married blacks who did not complete high school, TFR would actually fall from 2.63 to 2.41 if subsidizing market inputs were replaced with subsidizing maternal inputs. This finding demonstrates that the type of subsidy to child care helps determine not just the aggregate level of births, but also their composition within different types of households.

The change in labor force participation rates are more ambiguous, in fact puzzling. Since returns from experience on the job is likely to strengthen attachment to the labor force beyond that predicted by the static model, we are further investigating this counter-intuitive result.

10.5. Paid Maternity Leave

Paying females wages when they take maternity leave is a third way of promoting higher fertility. A distinguishing feature of this policy is that women with high wages receive greater payment than those receiving lower wages. (Note that if the payment is a fixed allowance, then the analysis of Expenses policy applies.) In contrast to the two previous schemes, (each of which has only one degree of freedom, the proportion of costs or time covered), this scheme has two, what percentage of her market wage a mother is paid while on maternity leave, and the maximum eligibility period per child. Under the Wages policy, mothers are paid the wage they would have received if they had not given birth, and the maximum eligibility period is the amount of time they would have withdrawn from the workforce in the absence of the subsidy. These variables are for the most part negatively correlated, and therefore affect the total payment in offsetting directions.

In particular, suppose the woman gives birth at period t, let $h^o_{n,t+s}(b_{nt}=0)$ denote the woman's labor supply s periods after the birth had she not left the workforce to give birth, let $w^o_{n,t+s}(b_{nt}=0)$ denote her wage rate had she not given birth, and let τ^0_n denote the number of periods she would have taken off if there were no provisions for paid maternity leave. Then in this policy regime the wage payment she receives upon having a child is:

$$\sum\nolimits_{s=0}^{\tau_n} \lambda_{n,t+s} w^o_{n,t+s}(b_{nt}=0)\, h^o_{n,t+s}(b_{nt}=0)$$

In a static framework, paid maternity leave induces women to reduce their labor supply and have larger families. In our dynamic framework paying wages does not fully compensate a mother for taking maternity leave, because job market experience acquired before giving birth depreciates over the time spent out of the labor force. Consequently, females who decide to have a child because of the paid maternity leave may simply exit the labor force permanently if their market capital has depleted sufficiently quickly. This scenario certainly arises when, in the absence of the paid leave policy, women essentially choose between having a career and having a family.

Our preliminary simulation results are displayed in the fifth columns of Tables IX and X. They show that in 13 out of the 18 cases the labor supply participation falls, because of the substitution effect into child rearing activities, and the compounding effect of human capital depletion. Although total fertility rates increase in all categories, this policy is not as effective as directly paying for the time inputs; in every category fertility rates under subsidized Daycare exceed those in attained when there is paid maternity leave as mandated in Wages.

10.6. Retraining

In our framework mothers lose human capital from temporarily withdrawing from the labor force. The last counter factual regime we consider does not make any payments to mothers, but offers partial compensation by putting women returning to work from maternity leave on an equal footing with those who chose not to have children. The policy scheme simulated in Retraining restores them to the wage trajectory they would have been on if they not withdrawn from the workforce to have children. In our framework the labor force experience over the previous ν periods helps determine the current wage. Thus, if the female in Model 4 reenters τ_n^4 periods after she has her birth, the natural logarithm of her wages increases by:

$$\sum\nolimits_{s=0}^{\min\left\{\nu,\tau_n^4\right\}} \left[\delta_{1s} h_{n,t-s}^o \left(b_{nt}=0\right) + \delta_{2s} d_{n,t-s}^o \left(b_{nt}=0\right)\right]$$

The last columns of Tables IX and X display the results, which in some ways are the most dramatic. The total fertility rate of every group except the unmarried white females with less than high school education rises above the replacement rate, and for one group, married black females with high school education, reaches 3.

11. Conclusion

This paper develops a dynamic model of female labor supply and fertility behavior and estimates its structural parameters. Previous empirical research on female labor supply had shown that current labor supply choices affect future wages and utility through intertemporal nonseparabilites in the production function (such as through learning by doing or staying in practice), and in utility (for example, through the household production function and also possibly due to the intertemporal nature of utility from leisure). In addition, there are a small number of studies of fertility behavior that suggest the timing of later births is partly determined by economic factors. Our study nests both kinds of dynamic interactions within a unified structural model.

Our empirical results reaffirm findings from previous work, and provide a set of parameters that capture the costs and benefits of having children within a dynamic structural framework. More specifically, our estimates reaffirm the importance of nonseparabilites in labor supply choices. Wages are increased by experience up to four years in the past, recent experience counting the most. In addition, we reject the null hypothesis that leisure is intertemporally separable, our estimates suggesting that there is also learning by doing in home-making activities.

With regards to fertility previous work estimating linear index functions and fertility hazards had found that the timing of later births depended on the ages of older siblings. Although there is not much in the literature with which we can directly compare our findings, our estimated costs and benefits of children are plausible. They imply that households view children as good, not bad, thus suggesting that households limit their size because of the time and money costs associated with raising offspring, not because adults do not like having offspring. Our estimates show that there is an optimal gestation period with respect to births that is partly determined by the same economic factors that have played a role in reducing TFR over the past generation or two, a finding which is consistent with previous work on the spacing of births over the life cycle.

With respective the aggregate wage effect on labor force participation and fertility behavior of females, we can reasonability conclude that although it is important, it is not the most significant feature driving our results. This is based on the fact that the wage effect in the simulations seems to have very small impact on female behavior relative to others, say, human capital accumulations.

Our study was motivated by the fact that the decline in childbearing coincides with higher wages for females, who are raising fewer children, participating in the labor force in greater numbers and are working longer hours. Our model is uniquely suited to analyzing whether shifts in public policy towards child support could affect these trends, and in what ways. As such we conducted different policy simulations which we use to analyze the effects on such trends and found that the effect differ depending on the social economic group that we are looking at. This suggests that any such fertility policy must be undertaken with a great degree of care, to ensure that we are not only subsidizing a group of individuals who would have children in any event, and hence does not change the population growth rates.

References

[1] Ahn, Namkee(1995), *Measuring the Value of Children by Sex and Age Using a Dynamic Programming Model*, Review of Economic Studies 62, 361-79

[2] Altug, Sumru and Robert A. Miller (1998), *The Effect of Work Experience on Female Wages and Labour Supply*, Review of Economic Studies, 45-85.

[3] Altug, Sumru and Robert Miller(1990), *Household Choices in Equilibrium*, Econometrica, Vol. 58. no. 3., 543-570.

[4] Angrist, J. and W. N. Evans(1998), *Children and Their Parents' Labour Supply: Evudence from Exogenous Variation in Family Size*, American Economic Review 88, 450-77.

[5] Arroyo, Cristino R. and Junsen Zhang (1997), *Dynamic Microeconomic Models of Fertility Choice: A Survey.*, Journal of Population Economics, Vol.10, 23-65.

[6] Becker, G. (1965), *A Theory of Allocation of Time, Economic, Economic Journal, 75, 496-517.*

[7] Becker, Gary S., Kevin M. Murphy and Robert Tamura (1990), *Human Capital, Fertility, and Economic Growth*, Journal of Political Economy, Vol. 98, no.2, S12-S37.

[8] Del Boca, Daneila (2002), *The Effect of Child Care and Part-Time Opportunities on Participation and Fertility Decisions in Italy,* Journal of Population Economics, Vol.

[9] Butz, William P. and Michael P. Ward (1980), *Completed Fertility and its Timing,* Journal of Political Economy, Vol.88, no. 5, 917-940.

[10] Butz, William P. and Michael P. Ward (1979), *The Emergence of Countercyclical U.S. Fertility*, American Economic Review, Vol. 69, no. 3 , 318-328.

[11] Card D. (1990), *Labour Supply With a Minimum Threshold*, Carnegie-Rochester Series on Public Policy, 33, 137-168.

[12] Eckstein anf K. Wolpin (1989), *Dynamic labour Force Participation of married womem and Endogenous Work Experirence*, Review of Economic Studies, Vol. 56, 1989, 375-590.

[13] Ejrnaes, Mette and Astrid Kunze (2002), *Wage Dips and Drops around First Birth*, Working Paper, University of Copenhagen.

[14] Francesconi, M. (2002), *A Joint Dynamic Model of Fertiliy and Work of Married Women*, Journal of Labor Economics, 20, 336-380

[15] Heckman, James J. , V. Joseph Hotz and James R. Walker, *New Evidence on the Timing and Spacing of Births*, American Economic Review, Vol. 75, no. 2 (1985) 179-184.

[16] Heckman, James J. and James R. Walker (1990), *The relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data*, Econometrica, Vol. 58, no. 6, 1411-1441.

[17] Hotz, V. Joseph, Jacob Alex Klerman and Robert J. Willis (1997), *The Economics of Fertility in Developed Countries, in Handbook of Population and family Economics*, ed. by M. Rosenzweig and O Stark. North Holland .

[18] Hotz, V. Joseph and Robert A. Miller (1993), *Conditional Choice Probabilities and the Estimation of Dynamic Models of Discrete Choice*, Review of Economic Studies, 60, 497-429.

[19] Hotz, V. Joseph and Robert A. Miller (1988), *An Empirical Analysis of Life Cycle Fertility and Female Labour Supply*, Econmetrica, Vol. 56. no. 1, 91-118.

[20] Hotz, V. Joseph, Robert A. Miller , S. Sanders and J Smith (1994), *A Simulation Estimator for Dynamic Models of Discrete Choice*, Review of Economic Studies, 61 , 265-289.

[21] Krammer, Walter and Klaus Newusser (1984), *The Emergence of a Countercyclical U.S. Fertility: Note,* American Economic Review, Vol. 74, no. 1, 201-202.

[22] Lewis, Frank D. (1983), *Fertility and Saving in the United States: 1830-1900*, Journal of Political Economy, Vol. 91, no. 5, 825-840.

[23] Lundlolm and Ohlsson(1998), *Wages, Taxes and Publicly Day Care,* Journal of Population Econmics, Vol. no.

[24] Mace, B. (1991), *Full Insurance in the Presence of Aggregate Uncertainty*, Journal of Political Economy, vol. 89, 1059-1085.

[25] Macunovich, Diane J. (1998), *Fertility and the Easterlin Hypothesis: An assessment of the literature*, Journal of Population Economics, Vol. 11, 53-111.

[26] Merrigan, Philip and Yvan St.-Pierre (1998), *An Econometric and Neoclassical Analysis of the Timing and Spacing of Births in Canada from 1950 to 1990*, Journal of Polpulation Economics, Vol. 11 , 29-51.

[27] Miller , Robert A. (1997), *Estimating Models of Dynamic Optimization with Microeconomic Data*, in Pesaran M. and Schmidt, P. (eds) Handbook of Applied Econometrics 2. Microeconemetrics (London: Basil Blackwell), 247-299.

[28] Miller, Robert A. (1984), *Job Matching and occupational Choice*, Journal of Political Economy, 92, , 1086-1120.

[29] Miller. Robert A. and Sanders, S. (1997), *Human Capital development and Welfare Participation*, Carnegie-Rochester Conference Series on Public Policy, 46, 237-253.

[30] Newman, John L. (1983), *Economic Analyses of the Spacing of Births*, American Economic Review, Vol. 73, no. 2, 33-37.

[31] Olsen, Randall J. (1983), *Mortality Rates, Mortality Events, and the Number of Births*, American Economic review, Vol. 73, no. 2, 29-32.

[32] Powell ,James L., James H. Stock, Thomas M. Stoker(1989),*Semiparametric Estimation of Index Coefficients,* Econometrica, Vol. 57, No. 6., pp. 1403.

[33] Schultz, Paul T. (1994), *Human Capital, Family Planning, and Their Effects on Population Growth*, American Economic Review, Vol. 84, no. 2, 255-260.

[34] Van Der Klaauw, Wilbert (1996), *Female Labour Supply and Marital Status Decisions: A :Life Cycle Model.*, Review of Economic Studies, Vol. 63, no. 2, 199-235.

[35] Waldfogel, Higuchi and Abe (1999), *Family Leave Policies and Women's Retention After Childbirth: Evidence from the United States, Britain and Japan,* Journal of Population Economics, Vol.

[36] Walker, J. (1996), *Parental Benefits, Employment and Fertility Dynamics,* Research in Population Economics, Vol. 8

[37] Willis, Robert J. (1973), *A New Approach to the Economic Theory of Fertility Behavior*, Journal of Poltical Economy, Vol. 81, no. 2, S14-S64.

[38] Wolpin, Kenneth I. (1984), *An Estimable Dynamic Stochastic Model of fertility and Child Mortality*, Journal of Political Economy, Vol. 92, no. 5, 852-874

12. Appendix 1

In this appendix we defines a class of conditional choice probability (CCP) estimators, to which the estimator used in Section 8.3 belongs, and show consistency and asymptotic normality of these estimators.

12.0.1. A class of CCP estimators.

The estimators (Θ_3^N, Γ^N) defined by equation (**??**) and (**??**) are examples of CCP estimators, in which the individual-specific effects $\mu_n^N \eta_n^N$, time-specific effects $\omega_t^N \lambda_t^N$ and the conditional choice probabilities p_{knt}^N for $k = 0, .., 3$ and $p_{0nt}^{(s,k,N)}$ for $s = 1, ..., \overline{\rho}$ enter as incidental parameters. This estimator falls within a class of CCP estimators that can be described as follows.

Let $D_n(\Theta, \mu_n, p_n)$ be a $q \times 1$ vector function such that $\Theta_0 \equiv (\Theta_{30}, \Gamma_0)$ is the unique root of $E[D_n(\Theta, \mu_n, p_n)]$. For each, $n \in \{1, 2, ...\}$ and $\Theta \in \Xi$, let μ_n^N be a kernel or traditional estimator which converges uniformly to μ_n, let $p_n^N(\mu_n)$ be a kernel estimator which converges uniformly to $p_n(\mu_n)$. We define Θ^N as any solution to

$$\frac{1}{N}\sum_{n=1}^{N} D_n(\Theta^N, \mu_n^N, p_n^N(\mu_n^N)) = 0 \tag{12.1}$$

The proof of proposition 1 below shows that Θ^N is asymptotically normal, but is not centered on zero. While an asymptotically unbiased estimator could be calculated following the procedure in Hotz and Miller (1993) by forming a linear combination of the estimators which are based on different bandwidths for the incidental parameters, the limited empirical evidence suggests that the asymptotic bias is unimportant.[10]

Proposition 1: Θ^N converges to Θ_0 and $\sqrt{N}(\Theta^N - \Theta_0)$ is asymptotic normal with mean $-E(v_n)/2$ and covariance matrix $(D_0')^{-1}S_0 D_0$, where v_n D_0 and S_0 are defined by equations (12.10), (12.17) and (12.18).

Proof.

For ease of notation, we assume that μ_n^N and $p_n^N(\mu_n^N)$ take the form of nonparametric kernel estimators weighted or unweighted probability density functions of the form

$$\mu_n^N = \sum\nolimits_{m=1, m\neq n}^{N} \phi_m \delta^{-q} J[\delta_N^{-1}(x_m - x_n)] \tag{12.2}$$

and

$$p_n^N(\mu_n^N) = \sum\nolimits_{m=1, m\neq n}^{N} d_m \delta^{-q} J[\delta_N^{-1}(k(z_m, \mu_m^N) - k(z_n, \mu_n^N))] \tag{12.3}$$

where $k(z_m, \mu_m^N)$ is mapping that defines the distance between the observations. The proof that Θ^N converges in probability to Θ_0 is standard, relying on the uniform convergence of the incidental parameters to their true values, so that the approximating sample moments obtained by substituting the incidental parameter estimates for their respective true values only affect the resulting structural parameter estimates by an $o_p(1)$ term.[11]

[10] For evidence on the magnitude of this asymptotic bias, see the Monte Carlo simulations in Powell, Stock and Stoker (1989) and the fertility application in Hotz and Miller (1993).

[11] See Hotz and Miller(1993) for a consistency proof of a very similar semiparametric estimator.

To establish the mean, covariance, and bias, we first consider anotherestimator denoted by $\widetilde{\Theta}^N$, and show that this has the same asymptotic distributional properties as Θ^N. For ease of notation, let $D_n \equiv D_n(\Theta_0, \mu_n, p_n)$, $p_n \equiv p_n(\mu_n)$

$$D_{0n} \equiv \left[\frac{\partial D_n(\Theta_0, \mu_n, p_n)}{\partial \Theta}\right]$$

$$D_{1n} \equiv \left[\frac{\partial D_n(\Theta_0, \mu_n, p_n)}{\partial \mu_n} + \frac{\partial D_n(\Theta_0, \mu_n, p_n)}{\partial p_n} . \frac{p_n(\mu_n)}{\partial \mu_n}\right]$$

and

$$D_{2n} \equiv \left[\frac{\partial D_n(\Theta_0, \mu_n, p_n)}{\partial p_n}\right]$$

The estimator $\widetilde{\Theta}^N$ satisfies the equation

$$-N^{-1} \sum\nolimits_{n=1}^{N} [D_n + D_{0n}(\widetilde{\Theta}^N - \Theta_0)] \tag{12.4}$$

$$= \quad N^{-1} \sum\nolimits_{n=1}^{N} [D_{1n}(\mu_n^N - \mu_n) + D_{2n}(p_n^N(\mu_n^N) - p_n(\mu_n))] \tag{12.5}$$

Define the quantities

$$v_{1mn}^N \equiv D_{1n}[\phi_m \delta^{-q} J[\delta_N^{-1}(x_m - x_n)] - \mu_n] + D_{1m}[\phi_n \delta^{-q} J[\delta_N^{-1}(x_m - x_n)] - \mu_m] \tag{12.6}$$

$$\begin{aligned} v_{2mn}^N \quad \equiv \quad & D_{2n}[d_m \delta^{-q} J[\delta_N^{-1}(k(z_m, \mu_m^N) - k(z_n, \mu_n^N))] - p_n] & (12.7) \\ & + D_{2m}[d_n \delta^{-q} J[\delta_N^{-1}(k(z_m, \mu_m^N) - k(z_n, \mu_n^N))] - p_m] & (12.8) \end{aligned}$$

$$v_{mn}^N \equiv v_{1mn}^N + v_{2mn}^N \tag{12.9}$$

$$v_n = f(x_n)[D_{1n}(\mu_n + \phi_n) + D_{2n}(p_n + d_n)] - D_{1n}\mu_n - D_{2n}p_n \tag{12.10}$$

where $f(x_n)$ is the density of x_n.

Expanding the first expression on the right-side of 12.4 using the definition of the nonparametric estimator for μ_n yields

$$\begin{aligned} & N^{-1} \sum\nolimits_{n=1}^{N} [D_{1n}(\mu_n^N - \mu_n) & \\ = \quad & N^{-1} \sum\nolimits_{n=1}^{N} D_{1n}[\sum\nolimits_{m=1, m \neq n}^{N} \phi_m \delta^{-q} J[\delta_N^{-1}(x_m - x_n)] - \mu_n] & (12.11) \\ = \quad & N^{-1} \sum\nolimits_{n=1}^{N} \sum\nolimits_{m=1, m \neq n}^{N} D_{1n}[\phi_m \delta^{-q} J[\delta_N^{-1}(x_m - x_n)] - \mu_n] & \\ = \quad & N^{-1}(N-1)^{-1} \sum\nolimits_{n=1}^{N-1} \sum\nolimits_{m=n+1}^{N} v_{1mn}^N & (12.12) \end{aligned}$$

Similarly, the second expression on the right side of 12.4 may be written as

$$N^{-1} \sum\nolimits_{n=1}^{N} D_{2n}[p_n^N(\mu_n^N) - p_n(\mu_n)] = N^{-1}(N-1)^{-1} \sum\nolimits_{n=1}^{N-1} \sum\nolimits_{m=n+1}^{N} v_{2mn}^N \tag{12.13}$$

Following Hotz and Miller (1993), it is straight forward to show that $E[\left\|v_{imn}^{N}\right\|^2] = o(N)$ for $i = 1, 2$. Then appealing to lemma 3.1 of Powell, Stock and Stoker (1989), p.1410

$$N^{-1}(N-1)^{-1}\sum_{n=1}^{N-1}\sum_{m=n+1}^{N} v_{mn}^{N} = \frac{E[v_{mn}^{N}]}{2} + (N-1)^{-1}\sum_{n=1}^{N-1}\{E[v_{mn}^{N} \mid n] - E[v_{mn}^{N}]\} + o_p(1) \tag{12.14}$$

The right-side of 12.14 depends on N. To derive the asymptotic distribution of $\widetilde{\Theta}^N$. Lemma 1 derives the appropriate limit for the right side of 12.14 as

$$N^{-\frac{1}{2}}\frac{E[v_{mn}^{N}]}{2} + N^{-\frac{1}{2}}\sum_{n=1}^{N}\{E[v_{mn}^{N} \mid n] - E[v_{mn}^{N}]\} = N^{-\frac{1}{2}}\frac{E[v_n]}{2} + N^{-\frac{1}{2}}\sum_{n=1}^{N}\{v_n - E(v_n)\} + o_p(1) \tag{12.15}$$

The conditions that define $\widetilde{\Theta}^N$ can now be written as

$$-N^{-\frac{1}{2}}\sum_{n=1}^{N}[D_n + D_{0n}(\widetilde{\Theta}^N - \Theta_0)] = N^{-\frac{1}{2}}\frac{E[v_n]}{2} + N^{-\frac{1}{2}}\sum_{n=1}^{N}\{v_n - E(v_n)\} + o_p(1) \tag{12.16}$$

The Central Limit Theorem implies that the right-side of 12.16 converges in distribution to a normal random variable with mean $-\frac{E[v_n]}{2}$. Hence, $\sqrt{N}(\widetilde{\Theta}^N - \Theta_0)$ converges to a normal random variable with mean $-\frac{E[v_n]}{2}$ and covariance $(D_0')^{-1}S_0D_0^{-1}$ where

$$D_0 \equiv E[D_{0n}] \tag{12.17}$$

and

$$S_0 \equiv E[(D_n + v_n - E(v_n))(D_n + v_n - E(v_n))'] \tag{12.18}$$

We complete the proof of this proposition with lemma 2 provided below, which implies that Θ^N and $\widetilde{\Theta}^N$ have the same asymptotic distribution, that is, $\sqrt{N}(\widetilde{\Theta}^N - \Theta^N)$ is $o_{p(1)}$ Q.E.D.

Lemma 1: $N^{-\frac{1}{2}}\frac{E[v_{mn}^{N}]}{2} + N^{-\frac{1}{2}}\sum_{n=1}^{N}\{E[v_{mn}^{N} \mid n] - E[v_{mn}^{N}]\}$
$= N^{-\frac{1}{2}}\frac{E[v_n]}{2} + N^{-\frac{1}{2}}\sum_{n=1}^{N}\{v_n - E(v_n)\} + o_p(1)$

Proof.
Consider v_{1mn}^{N} which has the form

$$\begin{aligned} v_{1mn}^{N} \equiv\ & D_{1n}\phi_m\delta^{-q}J[\delta_N^{-1}(x_m - x_n)] \\ & - D_{1n}\mu_n + D_{1m}\phi_n\delta^{-q}J[\delta_N^{-1}(x_m - x_n)] - D_{1m}\mu_m \end{aligned} \tag{12.19}$$

Taking the first on the right-side of 12.19

$$
\begin{aligned}
E[D_{1n}\phi_m \delta^{-q} J[\delta^{-1}(x_m - x_n)] \quad | \quad & x_n] \\
= \quad & D_{1n} \int \mu(x)\delta^{-q} J[\delta^{-1}(x - x_n)] f(x) dx \\
= \quad & D_{1n} \int \mu(x_n + \delta u) J(u) f(x_n + \delta u) du \\
= \quad & \int D_{1n}\{\mu(x_n) f(x_n) + \mu(x_n + \delta u) f(x_n + \delta u) \\
& -\mu(x_n) f(x_n)\} J(u) du \\
= \quad & D_{1n}\mu(x_n) f(x_n) + D_{1n} t_n(\delta)
\end{aligned}
$$

where $t_n(\delta) \equiv \int[\mu(x_n + \delta u) f(x_n + \delta u) - \mu(x_n) f(x_n)] J(u) du$. Furthermore,

$$
\begin{aligned}
E[t(\delta)^2] \quad &= \quad E\left\{\phi_n^2 \left[\int [D_1(x_n + \delta u) f(x_n + \delta u) - D_1(x_n) f(x_n)] J(u) du\right]^2\right\} \\
&= \quad E\left\{\phi_n^2 \left[\int_{x_n}^{x_n + \delta u} \frac{\partial (D_1 f)(x)}{\partial x} J(u) du\right]^2\right\} \\
&\leq \quad E\left[\phi_n^2 \int \delta^2 u^2 \left\|\frac{\partial (D_1 f)(x)}{\partial x}\right\| J(u) du\right] \\
&= \quad E\left[\phi_n^2 \delta^2 \left\|\frac{\partial (D_1 f)(x)}{\partial x}\right\| \sigma_u^2\right] \\
&= \quad o_{p(1)}.
\end{aligned}
$$

Thus, $t_n(\delta)$ has a neglible effect because its variance asymptotes to zero and it has a mean of zero. As a consequence,

$$
\begin{aligned}
N^{-\frac{1}{2}} \sum_{n=1}^{N} \{E[D_{1n}\phi_m \delta^{-q} J[\delta^{-1}(x_m - x_n)] \quad | \quad & x_n] \\
& -E[D_{1n}\phi_m \delta^{-q} J[\delta^{-1}(x_m - x_n)]]\} \\
= \quad & N^{-\frac{1}{2}} \sum_{n=1}^{N} \{D_{1n}\mu_n f(x_n) \\
& -E[D_{1n}\mu_n f(x_n)]\} + o_{p(1)}.
\end{aligned}
$$

Similarly, considering the third term in 12.4

$$
\begin{aligned}
N^{-\frac{1}{2}} \sum_{n=1}^{N} \{E[D_{1m}\phi_n \delta^{-q} J[\delta^{-1}(x_m - x_n)] \quad | \quad & x_n] \\
& -E[D_{1m}\phi_n \delta^{-q} J[\delta^{-1}(x_m - x_n)]]\} \\
= \quad & N^{-\frac{1}{2}} \sum_{n=1}^{N} \{D_{1n} f(x_n)\phi_n \\
& -E[D_{1n} f(x_n)\phi_n]\} + o_{p(1)}.
\end{aligned}
$$

It now follows that

$$\begin{aligned} N^{-\frac{1}{2}} \sum\nolimits_{n=1}^{N} \{E[v_{1mn}^{N} \quad &| \quad n] - E[v_{1mn}^{N}]\} \\ &= \quad N^{-\frac{1}{2}} \sum\nolimits_{n=1}^{N} \{D_{1n} f(x_n)(\mu_n + \phi_n) - D_{1n}\mu_n \\ & \quad\quad -E[D_{1n} f(x_n)(\mu_n + \phi_n) + D_{1n}\mu_n] + o_{p(1).} \end{aligned} \tag{12.20}$$

By a similar argument

$$\begin{aligned} N^{-\frac{1}{2}} \sum\nolimits_{n=1}^{N} \{E[v_{2mn}^{N} \quad &| \quad n] - E[v_{2mn}^{N}]\} \\ &= \quad N^{-\frac{1}{2}} \sum\nolimits_{n=1}^{N} \{D_{2n} f(x_n)(p_n + d_n) - D_{2n}p_n \\ & \quad\quad -E[D_{2n} f(x_n)(p_n + d_n) + D_{2n}p_n] + o_{p(1).} \end{aligned} \tag{12.21}$$

Q.E.D.

Lemma 2: $\sqrt{N}(\widetilde{\Theta}^N - \Theta^N)$ is $o_{p(1)}$.
Proof.
Expanding the right-side of 12.1 about the true structural parameters, Θ_0 and the true incidental parameters, we obtain

$$\begin{aligned} & -\frac{1}{N} \sum\nolimits_{n=1}^{N} [D_n + \widetilde{D}_{0n}(\Theta^N - \Theta_0)] \\ = \quad & \frac{1}{N} \sum\nolimits_{n=1}^{N} [\widetilde{D}_{1n}(\mu_n^N - \mu_n) + \widetilde{D}_{2n}(p_n^N(\mu_n^N) - p_n(\mu_n))] \end{aligned} \tag{12.22}$$

where $\sim$ indicates that the appropriate partial derivatives are evaluated at points on the line segment joining (Θ_0, μ_n, p_n) and $(\Theta^N, \mu_n^N, p_n^N)$. Substracting 12.22 from 12.4 gives

$$\begin{aligned} & -\frac{1}{N} \sum\nolimits_{n=1}^{N} [\widetilde{D}_{0n}(\Theta_0 - \Theta^N) - D_{0n}(\Theta_0 - \widetilde{\Theta}^N)] \\ = \quad & \frac{1}{N} \sum\nolimits_{n=1}^{N} [(\widetilde{D}_{1n} - D_{1n})(\mu_n^N - \mu_n) \\ & +(\widetilde{D}_{2n} - D_{2n})(p_n^N(\mu_n^N) - p_n(\mu_n))] \end{aligned} \tag{12.23}$$

consider the following asymptotic expansion

$$\begin{aligned} & \frac{1}{N} \sum\nolimits_{n=1}^{N} \{\widetilde{D}_{0n}(\Theta_0 - \Theta^N) - D_{0n}(\Theta_0 - \widetilde{\Theta}^N)\} \\ = \quad & \frac{1}{N} \sum\nolimits_{n=1}^{N} \{D_{0n}(\Theta_0 - \Theta^N) - D_{0n}(\Theta_0 - \widetilde{\Theta}^N) + (\widetilde{D}_{0n} - D_{0n})(\Theta_0 - \Theta^N)\} \\ = \quad & \frac{1}{N} \sum\nolimits_{n=1}^{N} D_{0n}(\widetilde{\Theta}^N - \Theta^N) + o_{p(1)}(\Theta_0 - \Theta^N) \\ = \quad & \{E[D_{0n}] + o_p(1)\}(\widetilde{\Theta}^N - \Theta^N) + o_{p(1)}(\Theta_0 - \Theta^N) \\ = \quad & \{E[D_{0n}] + o_p(1)\}(\widetilde{\Theta}^N - \Theta^N) + o_{p(1)} \end{aligned} \tag{12.24}$$

Considering the second expression in 12.23

$$\frac{1}{N}\sum\nolimits_{n=1}^{N}[(\widetilde{D}_{1n}-D_{1n})(\mu_n^N-\mu_n)=o_p(1)\frac{1}{N}\sum\nolimits_{n=1}^{N}(\mu_n^N-\mu_n) \tag{12.25}$$

where the right of 12.25 follows from the fact that $\widetilde{D}_{1n}$ converges in probability to D_{1n} uniformly in n. Similar U-statistic arguments to that used to justify the asymptotic normality of $\sqrt{N}(\widetilde{\Theta}^N-\Theta_0)$, show that $N^{-\frac{1}{2}}\sum_{n=1}^{N}(\mu_n^N-\mu_n)$ converges in distribution to a normal random variable which is $o_p(1)$. Therefore, 12.25 is $o_p(N^{\frac{1}{2}})$. Finally the third expression in 12.25 can be written as

$$\begin{aligned}
&\frac{1}{N}\sum\nolimits_{n=1}^{N}[(\widetilde{D}_{2n}-D_{2n})(p_n^N(\mu_n^N)-p_n(\mu_n))]\\
=\;& o_p(1)\frac{1}{N}\sum\nolimits_{n=1}^{N}(p_n^N(\mu_n^N)-p_n(\mu_n))\\
=\;& o_p(1)\frac{1}{N}\sum\nolimits_{n=1}^{N}(p_n^N(\mu_n^N)-p_n(\mu_n^N))\\
&+o_p(1)\frac{1}{N}\sum\nolimits_{n=1}^{N}(p_n(\mu_n^N)-p_n(\mu_n))
\end{aligned} \tag{12.26}$$

This mean that

$$\frac{1}{\sqrt{N}}\sum\nolimits_{n=1}^{N}(p_n^N(\mu_n^N)-p_n(\mu_n^N))$$

and

$$\frac{1}{\sqrt{N}}\sum\nolimits_{n=1}^{N}(p_n(\mu_n^N)-p_n(\mu_n))$$

are asymptotically normal. Then using the results obtained for 12.24, 12.25 and 12.26 in 12.22. We thus establish that

$$0=\{E[D_{0n}]+o_p(1)\}\sqrt{N}(\widetilde{\Theta}^N-\Theta^N)+op(1) \tag{12.27}$$

Noting that $E[D_{0n}]$ is nonsingular, 12.27 implies that $\sqrt{N}(\widetilde{\Theta}^N-\Theta^N)$ is $op(1)$ as claimed. Q.E.D.

13. Appendix 2

In this appendix we outline in more detail the methods we used to solve the valuation funcion used in the policy experiments.To solve the for the valuation function we shall used a hybrid method which is a combination of the the finite horizon and infinite horizon contraction .

13.0.2. Optimization

We first rewrite the per period utility to depend on $k\in\{0,1,2,3\}$. Let

$$U_{nt0}(z_{nt}^*,x_{nt})=z_{nt}'B_1+\sum_{s=0}^{\rho}\delta_s l_{nt-s}+\varepsilon_{0nt}$$

$$U_{nt1}(z_{nt}^*,x_{nt})=z_{nt}'B_0+z_{nt}'B_{11}(1-h_{nt})+\sum_{s=0}^{\nu}\delta_s(1-h_{nt})l_{nt-s}+\eta_n\lambda_t w_{nt}h_{nt}^*\ +\varepsilon_{1nt}$$

$$
\begin{aligned}
U_{nt2}(z_{nt}, x_{nt}, h_{nt}) &= z_{nt}' B_{11}(1 - c_{nt_{-1}} - \rho_0) + \sum_{s=0}^{\nu} \delta_s (1 - c_{nt_{-1}} - \rho_0) l_{nt-s} \\
&+(\gamma_0 + \sum_{k=1}^{M} \gamma_k b_{nt-k} + \phi_1 N_{nt} + \phi_2 N_{nt}^2) \\
&-\eta_n \lambda_t \pi(z_{nt}) + \varepsilon_{2nt}
\end{aligned}
$$

$$
\begin{aligned}
U_{nt3}(z_{nt}, x_{nt}, h_{nt}) &= z_{nt}' B_0 + z_{nt}' B_{11}(1 - c_{nt_{-1}} - \rho_0 - h_{nt}) \\
&+ \sum_{s=0}^{\nu} {}_s\delta_s (1 - c_{nt_{-1}} - \rho_0 - h_{nt}) h_{nt-s} \\
&\eta_n \lambda_t w_{nt} h_{nt}^* + (\gamma_0 + \sum_{k=1}^{M} \gamma_k b_{nt-k} + \phi_1 N_{nt} + \phi_2 N_{nt}^2) \\
&-\eta_n \lambda_t \pi(z_{nt}) + \varepsilon_{3nt}
\end{aligned}
$$

Then let

$$Z^* \equiv [z_{nt}', h_{nt}, ..., h_{nt+1-\rho}, b_{nt}, ..., b_{nt+1-M}, \mu_n \eta_n]$$

and

$$\Pr[Z_{nt+1}^* = Z^* \mid Z_{nt}^*, x_{nt}, h_{nt}, \mathbb{I}_{ntk} = 1] \equiv F_{ntk}(Z^{**} \mid Z_{nt}^*, x_{nt}, h_{nt}).$$

Let x_{nt}^o, h_{nt}^o and $\mathbb{I}_{ntk}^0$ be the optimal action conditional on the current state Z_{nt}^*. Then we can recast the problem recursively. To that end, the value function, $V_{nt}(Z^*)$ is defined for each $(t, Z^*) \in \{0, ..., T\} \times \mathbb{Z}^*$ by substituting the optimal decision rule back into the expected lifetime utility function:

$$V_{nt}(Z^*) \equiv E_t[\sum_{s=t}^{T} \sum_{k=0}^{3} \mathbb{I}_{nsk}^o \beta^{t-s} U_{nsk}(z_{ns}, x_{ns}^o, h_{ns}^o)]$$

For notational convinence define the reduced form utilities

$$U_{ntk}(Z^*) \equiv U_{ntk}(x_{knt}^o, h_{knt}^o, Z^*)$$

and the transition probabilities

$$F_{ntk}(Z^{**} \mid Z_{nt}^*) \equiv F_{ntk}(Z^{**} \mid Z_{nt}^*, x_{knt}^o, hk_{nt}^o).$$

The optimal discrete choice $\mathbb{I}_{ntk}^0$ maximizes:

$$U_{ntk}(Z^*) + \int V_{nt+1}(Z^{**}) dF_{ntk}(Z^{**} \mid Z_{nt}^*)$$

over $\{0, 1, 2, 3\}$, patently a finite discrete choice problem. Since

$$Z^{**} \equiv [z_{nt+1}', h_{nt}, ..., h_{nt+1-\rho}, b_{nt}, ..., b_{nt+1-M}, \mu_n \eta_n]$$

and since $\Pr[Z^*_{nt+1} = Z^* \mid Z^*_{nt}, x_{nt}, h_{nt}, \mathbb{I}_{ntk} = 1] \equiv F_{ntk}(Z^{**} \mid Z^*_{nt}, x_{nt}, h_{nt})$ then the transition density is deterministic in our model. So let Z^{**}_k the future state variable, then $\mathbb{I}^0_{ntk}$ is the optimal choice that maximizes:

$$U_{ntk}(Z^*) + V_{nt+1}(Z^{**}_k)$$

13.0.3. Finite Horizon Problem

The standard solution method is the Bellman's (1957) perspective of the backward induction. Suppose the problem has a finite horizon T, and consider the choices facing an individual entering the last period with state variables Z^*_{nT} her valuation function is simple $V_T(Z^*_{nT}) = \max\{U_{nTk}(Z^*_{nT})\}$. Taking expectation of $V_T(Z^*_{nT})$ one period before when her state variables on Z^* yields

$$\begin{aligned} g_{T-1}(Z^*) &= E[V_T(Z^*_{nT}) \mid Z^*] \\ &= E[\max\{U_{nTk}(Z^*_{nT})\} \mid Z^*] \end{aligned}$$

Let's for notational convenient denote

$$U_{ntk}(Z^*_{nt}) \equiv \overline{U_{ntk}(Z^*_{nt})} + \varepsilon_{knt}$$

Then

$$g_{T-1}(Z^*) = \sum_{k=0}^{3} p_k(Z^*_{nt})[\overline{U_{ntk}(Z^*_{nt})} + E(\varepsilon_{knt} \mid Z^*_{nt}, \mathbb{I}_{knt} = 1)]$$

By the extreme value assumption

$$\begin{aligned} p_k(Z^*_{nt}) &= \Pr[\overline{U_{ntk}(Z^*_{nt})} + \varepsilon_{knt} > \overline{U_{ntj}(Z^*_{nt})} + \varepsilon_{jnt}, \forall k \neq j] \\ &= \frac{\exp(\overline{U_{ntk}(Z^*_{nt})})}{\sum_{s=0}^{3} \exp(\overline{U_{nts}(Z^*_{nt})})} \end{aligned}$$

and

$$E(\varepsilon_{knt} \mid Z^*_{nt}, \mathbb{I}_{knt} = 1) = \zeta - \ln(p_k(Z^*_{nt}))$$

These assumptions imply that

$$\begin{aligned} g_{T-1}(Z^*) &= \zeta + \frac{\sum_{k=0}^{3} \exp(\overline{U_{ntk}(Z^*)}) \ln[\sum_{j=0}^{3} \exp(\overline{U_{ntj}(Z^*)})]}{\sum_{s=0}^{3} \exp(\overline{U_{nts}(Z^*)})} \\ &= \zeta + \ln[\sum_{j=0}^{3} \exp(\overline{U_{ntj}(Z^*)})] \end{aligned}$$

Having calculated $g_{T-1}(Z^*)$, at the beginning of the period $T-1$ the person chooses the maximum over the different options and obtains a valuation function of:

$$V_{T-1}(Z^*) = \max\{\overline{U_{ntk}(Z^*_{nt})} + \varepsilon_{knt} + \beta g_{T-1}(Z^{**})\}$$

In a iterative fashion lets

$$g_{T-2}(Z^*_{nT-2}) = E[V_{T-1}(Z^*_{nT-1})].$$

By successively solving for the functions $g_{T-1}(Z^*)$ through $g_1(Z^*)$, the value function $V_t(Z^*)$ is derived numerically for all $t \in \{0, ..., T\}$.

13.0.4. Infinite Horizon

Here we will use the contraction mapping theory to extends the idea of iteration on the value function to the infinite horizon case. Let $H_0(Z^{**})$ be any real value bounded continuous function defined on the coordinate Z^* and define the real valued mapping as:

$$C[H(Z^{**})] = E[\max_k\{\overline{U_k(Z^*)} + \varepsilon_k + \beta H(Z^{**})\}]$$

where the integration is over ε_k , an extreme value Type I random variable. The mapping $C[.]$ is a contraction mapping, and satisfies the fixed point property that a unique $H(Z^{**})$ solves:

$$H(Z^{**}) = C[H(Z^{**})]$$

In this case $H(Z^{**})$ is interpreted as the expected value function for the infinite horizon problem:

$$H(Z^{**}) = E[V(Z^*)].$$

By a corrollary of the contraction mapping fixed point theorem, we obtain an upper bound on the sequence of iteration approximating function. In particular for an initial $H_0(Z^{**})$, then

$$\|C^s[H_0(Z^{**})] - H(Z^{**})\| \leq (1-\beta)^{-1} \left\|C^s[H_0(Z^{**})] - C^{s-1}[H_0(Z^{**})]\right\|$$

In our specific case with the extreme value Type I assumption on the errors, we have:

$$\begin{aligned} C[H(Z^{**})] &= \zeta + \frac{\sum_{k=0}^3 \exp(\overline{U_{ntk}(Z^*)} + \beta H(Z^{**})) \ln[\sum_{j=0}^3 \exp(\overline{U_{ntj}(Z^*)} + \beta H(Z^{**}))]}{\sum_{s=0}^3 \exp(\overline{U_{nts}(Z^*)} + \beta H(Z^{**}))} \\ & \zeta + \ln[\sum_{j=0}^3 \exp(\overline{U_{ntj}(Z^*)} + \beta H(Z^{**}))] \end{aligned}$$

We can then use a Newton fixed method to numerically solve for the fixed point. The Newton iteration is of the form:

$$H^{s+1} = H^s - \frac{C[H^s]}{C'[H^s]}, \qquad s > 0$$

this gives the following

$$H^{s+1} = H^s - \beta^{-1} \ln[\sum_{j=0}^3 \exp(\overline{U_j} + \beta H^s)], \qquad s > 0$$

Although convergence is global, an intelligent choice for the intial function $H_0()$ reduces the number of iterations required to achieve convergence. One such choice , is to combine the finite horizon problem with the infinite horizon problem and set

$$H_0() = (1-\beta) g_{T-1}(Z) \tag{13.1}$$

which is the discounted life time utility for a household one period before the terminal horizon. This choice works very well in our application and achieve convergence in 4 to 7 iterations.

14. Appendix 3

In part B of this appendix, we describe in more detail the construction of our sample and the construction of the variables used in our study. We used data from the Family-Individual File , Childbirth and Adoption History File, and the Marriage History File of the Michigan Panel Study of Income Dynamics (PSID). The Family- Individual File contains a separate record for each member of all households included in the survey in a given year. The Childbirth and Adoption History File contains information collected in 1985-1992 waves of PSID regarding histories of childbirth and adoption. The file contains details about childbirth and adoption events of eligible people living in a PSID family at the time of the interview in any wave from 1985 through 1992. Each set of records for a specified individual contains all known cumulative data about the timing and circumstances of his or her childbirth and adoption experience up to and including 1992, or those waves during that period when the individual was in a responding family unit. If an individual has never had any children, one record indicates that report. Note that "eligible" here means individuals of childbearing age in responding families. Similarly, the 1985-1992 Marriage History file contains retrospective histories of marriages for individuals of marriage-eligible age living in a PSID family between 1985 and 1992. Each set of records for a specified individual contains all known cumulative data about the timing and circumstances of his or her marriages up to and including 1992, or those waves during that period when the individual was in a responding family unit.

Our sample selection started from the Childbirth and Adoption history file, which contains 24,762 individuals. We initially selected women by setting "sex of individual" variable equal to two. Out of an initial sample of 24,762 individuals included in the Childbirth and Adoption file, this initial selection produced a sample of 12,784 female. We then drop any individual who was in the survey for four years or less, this selection criteria eliminated a further 1,946 individuals from our sample. We then drop all individuals who were older than 45 in 1967, this eliminated an additional 1,531 individuals. We then drop all individuals that were less than 14-years-old in 1991, this eliminated an additional 385 individuals.

The corresponding number of observations for the interviewing year 1968 through 1992 are given by 5,429,5,608, 5,793,5,970, 6,197, 6,346, 6,510, 6,696, 6,876, 7,094, 7,236, 7,320, 7,393, 7,455, 7,551, 7,634, 7,680, 7,761, 7,712, 7,666, 7,618, 7,574, 7,532, 7,378 and 7,233, respectively.

Since individuals who had become non-respondents as of 1992, either because they and their families were last to the study or they were mover-out non-respondents in years prior to the 1992 interviewing year, are not in the twenty-five Family-Individuals Respondents File, the number of observations increases with the interviewing years.

There were coding errors which occurred for the different measures of consumption in the PSID from which we construct our consumption measure. In particular, our measure of food consumption expenditures for a given year is obtained by summing the values of annual food expenditures for meals at home, annual food expenditures for eating out, and the value of food stamps received for the year. We measured consumption expenditures for year t by taking 0.25 of the value of this variable for the year $t-1$ and 0.75 of its value for the year t. The second step was taken to account for the fact that the survey questions used to elicit information about household food consumption is asked sometime in the first half of the year, while the response is dated in the previous year.

The variables used in the construction of the measure for total expenditures are also subject to the problem of truncation from above in the way they are coded in the 1983 PSID data tapes. The truncation value for the value of food stamps received for that year is \$999.00, while the relevant value for this variable in the subsequent years and for the value of food consumed at home and eating out is \$9,999.00. Taken by itself, the truncation of different consumption variables resulted in a loss of 467 person-years. We also use variables describing various demographic characteristics of the women in our sample. The dates of birth of the women were obtained from the Child Birth and Adoption file. The age variable resulted in a loss of 162 individuals.

The race of the individual or the region where they are currently residing were obtained from the Family portion of the data record. We defined the region variable to be the geographical region in which the household resided at the time of the annual interview. This variable is not coded consistently across the years. For 1968 and 1969, the values 1 to 4 denote the regions Northeast, Northcentral, South and West. For 1970 and 1971, the values 5 and 6 denote the regions Alaska and Hawaii, and foreign country, respectively. After 1971 a value of 9 indicates missing data but no person years were lost due to missing data for these variables.

We used the family variable "Race of The Household Head" to measure the race variable in our study. For the interviewing years 1968-1970, the values 1 to 3 denote White, black, and Puerto Rican or Mexican, respectively. 7 denotes other (including Oriental and Philippino), and 9 denotes missing data. For 1971 and 1972, the third category is redefined as Spanish-American or Cuban and between 1973-1984, just Spanish American. After 1984, the variable was coded in such a way that 1-6 correspond to the categories White, Black, American Indian, Aleutian or Eskimo, and Asian or Pacific Islander, respectively. A value of 7 denotes the other category, a value of 9 denotes missing. We used all available information for all the years to assign the race of the individual for years in the sample when that information was available.

We used a combination of individual and family level variables to construct our measure of educational attainment. This was because the variable for the individual does not contain data for the head of the household or wife, this we obtained from the family level files.

The marital status of a women in our subsample was determined by using the marriage history file. The number of individuals in the household and the total number of children within that household were also determined from the family level variables of the same name. In 1968, a code for missing data (equal to 99) was allowed for the first variable, but in the other years, missing data were assigned. The second variable was truncated above the value of 9 for the interviewing years 1968 and 1971. After 1975, this variable denotes the actual number of Children within the family unit.

We constructed some additional variables. The variable showing the value of home-ownership was constructed by multiplying the value of a household's home by an indicator variable determining home ownership. A similar procedure was followed to generate value of rent paid and rental value of free housing for a household. Mortgage payment and Principal of Mortgage outstanding were obtained from the family variables of the same names. Finally, household income was measured from the PSID variable total family money income, which included taxable income of head and wife total transfer of head and wife, taxable income of others in the family units and their total transfer payments.

We used two different deflators to convert such nominal quantities as average hourly earnings, household income, and so on to real. First, we defined the (spot) price of food consumption to be the numeraire good at t in the theoretical section. We accordingly measured real food consumption expenditures and real wages as the ratio of the nominal consumption expenditures and wages and the annual Chain-type price deflator for food consumption expenditures published in table t.12 of the National Income and Products. On the other hand, we deflated variables such as the nominal value of home ownership or nominal family income by the Chain-type price deflator for total personal consumption expenditures.

Table I: Notations

w_{nt}	individual marginal product of labor
x_{nt}	consumption of market goods
z_{nt}	demographic variables
h_{nt}	proportion of time worked in period t
	as a fraction of the total time available in the period
l_{nt}	leisure in period t :balance of time not spent at work or nurturing children
d_{nt}	labor force participation dummy
b_{nt}	indicator of the birth of a child at period t
γ_0	additional lifetime expected utility a household receives for its first child
$\gamma_0 + \gamma_k$	utility from having a second child when the first born is k years old
$\gamma_0 + \gamma_k + \gamma_j$	utility from having a third child when the first two are aged k and j years old
U_{0nt}	benefits from offspring to the n^{th} household in period t
U_{1nt}	utility costs of the n^{th} female from working in period t
U_{3nt}	current utility from consumption by household n in period t
π	discounted cost of expenditures of raising a child
ρ_k	nurturing time required by a k year old child
c_{nt}	fraction of time the n^{th} household spend nurturing children in the household
η_n^{-1}	social weight attached to each individual n
W	aggregate endowment or output from the exogenous production process
d_{nt}^o	optimal labor forceparticipation decision at date t
h_{nt}^*	optimal labor supply
h_{nt}^o	optimal labor supply conditional on participation
b_{nt}^0	optimal birth decision
λ_t	shadow value of consumption
μ_n	time-invariant individual-specific effect of marginal product

TABLE II

Characteristic of main sample

Years	1972	1973	1974	1975	1976	1977	1978	1979
No.of obs.	6,133	6,326	6,490	6,674	6,855	7,071	7,210	7,282
Demographics								
Age	18.10	19.00	19.92	20.86	21.81	22.79	23.78	24.78
Family size	5.13	4.89	4.74	4.60	4.49	4.42	4.31	4.15
No.of child. <6	1.68	1.58	1.49	1.40	1.31	1.22	1.15	1.09
No.of child. 6 to 14	.466	.451	.429	.411	.400	.387	.381	.374
Prop.married	.379	.392	.396	.395	.398	.398	.399	.412
Years of compl. edu.	11.19	11.17	11.15	11.12	11.11	11.11	11.11	11.13
Prop. black	.430	.429	.426	.425	.423	.420	.419	.418
Prop. hisp.	.045	.046	.046	.046	.046	.046	.046	.046
Hours,Earnings, Participation and Consumption								
Household income	32,597	33,942	35,213	34,143	32,772	34,157	34,881	35,608
Total labor income	1,224	1,327	1,516	1,702	1,908	2,112	2,351	2,718
Annual hrs worked	470	475	505	506	519	535	558	598
Hourly earnings	2.55	2.72	2.94	3.26	8.07	3.80	4.17	4.32
Prop. who worked	.556	.543	.550	.549	.546	.538	.547	.577
Food Consumption	7,187	7,618	7,777	7,244	7,125	6,612	6,409	6,469

TABLE II

Characteristic of main sample (to be continued)

Years	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
No.of obs.	7,353	7,408	7,491	7,562	7,600	7,662	7,612	7,566	7,511	7,459	7,341	7,200	7,063
Demographics													
Age	25.78	26.78	27.78	28.78	29.78	30.78	31.78	32.78	33.78	34.78	35.78	35.78	35.78
Family size	4.05	3.97	3.87	3.79	3.73	3.64	3.56	3.48	3.42	3.37	3.30	3.25	3.18
No.of child. <6	1.03	.972	.913	.850	.794	.734	.673	.619	.566	.508	.455	.405	.351
No.of child. 6 to 14	.374	.376	.379	.387	.390	.395	.413	.430	.442	.458	.477	.487	.504
Prop.married	.421	.425	.430	.440	.448	.457	.462	.467	.474	.484	.492	.497	.504
Years of compl. edu.	11.15	11.17	11.19	11.22	11.24	11.26	11.28	11.31	11.34	11.38	11.38	11.38	11.40
Prop. black	.418	.416	.415	.414	.412	.411	.409	.405	.402	.399	.397	.395	.392
Prop. hisp.	.046	.046	.046	.045	.046	.046	.046	.046	.046	.046	.046	.047	.047
Hours,Earnings, Participation and Consumption													
Household income	35,276	34,092	32,482	31,393	32,447	33,776	34,012	34,667	35,102	36,426	36,175	35,481	34,898
Total labor income	3,109	3,446	3,843	4,174	4,631	5,196	5,775	6,323	6,960	7,752	8,298	8,945	9,550
Annual hrs worked	615	637	648	659	680	750	773	801	836	874	899	931	947
Hourly earnings	4.82	5.18	5.64	5.97	6.71	6.64	7.11	7.35	7.86	8.25	8.81	9.20	9.72
Prop. who worked	.594	.607	.607	.611	.625	.656	.665	.687	.711	.728	.755	.772	.773
Food Consumption	6,268	6,099	5,840	5,722	5,866	5,946	5,821	5,756	5,828	5,702	5,616	5,552	5,632

Table III. Wage Equation

$\ln(w_{nt}) = \ln(\omega_t) + \ln(\mu_n) + \sum_{s=1}^{\nu}(\delta_{1s}h_{n,t-s} + \delta_{2s}d_{n,t-s}) + z'_{nt}B_3$

Variable	Parameter	Estimate
Lags of hours worked		
$\triangle h_{n,t-1}$	δ_{11}	14.1011 (0.2337)
$\triangle h_{n,t-2}$	δ_{12}	10.9974 (0.2471)
$\triangle h_{n,t-3}$	δ_{13}	8.8360 (0.2437)
$\triangle h_{n,t-4}$	δ_{14}	5.4729 (0.2227)
Lags of participation		
$\triangle d_{n,t-1}$	δ_{21}	-6.8664 $(4.01e-02)$
$\triangle d_{n,t-2}$	δ_{22}	-4.4241 $(4.46e-02)$
$\triangle d_{n,t-3}$	δ_{23}	-2.8986 $(4.44e-02)$
$\triangle d_{n,t-4}$	δ_{24}	-1.6065 $(3.92e-02)$
Socioeconomic Variables		
$\triangle AGE_{nt}^2$	B_{31}	-0.0114 $(3.0e-04)$
$\triangle(AGE_{nt} \times EDU_{nt})$	B_{32}	0.0161 $(3.1e-03)$

† Standard Errors in parenthesis

Table IV: Consumption Equation

$\ln(x_{nt}) = 1/(1-\alpha)[z'_{nt}B_2 - \ln(\eta_n\lambda_t) + \epsilon^c_{nt}]$

Variable	Parameter	Estimate
Socieconomic variables		
$\triangle FAM_{nt}$	$(1-\alpha)^{-1}B_{21}$	$3.19e-02$ $(3.0e-04)$
$\triangle YKID_{nt}$	$(1-\alpha)^{-1}B_{22}$	$-3.33e-02$ $(1.6e-03)$
$\triangle OKID_{nt}$	$(1-\alpha)^{-1}B_{23}$	$-1.12e-02$ $(1.2e-03)$
$\triangle AGE^2_{nt}$	$(1-\alpha)^{-1}B_{24}$	$-1.0e-04$ (0.0000)
Region Dummies		
$\triangle NC_{nt}$	$(1-\alpha)^{-1}B_{25}$	$-3.7e-03$ $(3.3e-03)$
$\triangle SO_{nt}$	$(1-\alpha)^{-1}B_{26}$	$-1.19e-02$ $(3.2e-03)$

† Standard Errors in parenthesis

Table V: Sample Averages of Nonparametric Estimates of Conditional Choice Probabilities

Variable	Sample mean †	Sample std. dev †	Sample mean ‡	Sample std. dev ‡
p_{0nt}	0.4369	0.4044	0.4377	0.4052
p_{1nt}	0.5191	0.4036	0.5182	0.4036
p_{2nt}	0.0153	0.0721	0.0153	0.0721
p_{3nt}	0.0288	0.1128	0.0288	0.1128
$p_0(H_{0nt}^{(1)})$	0.7344	0.1658	0.7338	0.1675
$p_0(H_{0nt}^{(2)})$	0.2460	0.1254	0.2477	0.1278
$p_0(H_{0nt}^{(3)})$	0.1745	0.0909	0.1744	0.0927
$p_0(H_{0nt}^{(4)})$	0.1154	0.0662	0.1162	0.0674
$p_0(H_{0nt}^{(5)})$	0.1400	0.0634	0.1405	0.0637
$p_0(H_{1nt}^{(1)})$	0.1581	0.1274	0.1592	0.1296
$p_0(H_{1nt}^{(2)})$	0.4007	0.1956	0.4012	0.1973
$p_0(H_{1nt}^{(3)})$	0.5978	0.2455	0.5975	0.2464
$p_0(H_{1nt}^{(4)})$	0.6656	0.2370	0.6656	0.2362
$p_0(H_{1nt}^{(5)})$	0.6863	0.1699	0.6840	0.1685
$p_0(H_{2nt}^{(1)})$	0.7199	0.1993	0.7231	0.1915
$p_0(H_{2nt}^{(2)})$	0.5626	0.2175	0.5632	0.2116
$p_0(H_{2nt}^{(3)})$	0.4804	0.2261	0.4913	0.2236
$p_0(H_{2nt}^{(4)})$	0.4970	0.1961	0.4935	0.1952
$p_0(H_{2nt}^{(5)})$	0.4146	0.1488	0.4238	0.1464
$p_0(H_{3nt}^{(1)})$	0.2777	0.2275	0.2824	0.2310
$p_0(H_{3nt}^{(2)})$	0.3103	0.2431	0.3127	0.2425
$p_0(H_{3nt}^{(3)})$	0.3300	0.2600	0.3261	0.2553
$p_0(H_{3nt}^{(4)})$	0.2730	0.2102	0.2756	0.2075
$p_0(H_{3nt}^{(5)})$	0.2181	0.0635	0.2179	0.0684

†Time-averaged individual effects

‡Nonparametrically estimated individual effects

Table V(cont'd): Sample Averages of Nonparametric Estimates of Conditional Choice Probabilities

Variable	Sample mean †	Sample std. dev †	Sample mean ‡	Sample std. dev ‡
$\partial p_0(H_{1nt}^{(1)})/(\partial h_{nt})$	-0.0714	0.0474	-0.0716	0.0474
$\partial p_0(H_{1nt}^{(2)})/(\partial h_{nt})$	-0.1405	0.0637	-0.01399	0.0642
$\partial p_0(H_{1nt}^{(3)})/(\partial h_{nt})$	-0.1370	0.0963	-0.1365	0.0947
$\partial p_0(H_{1nt}^{(4)})/(\partial h_{nt})$	-0.1318	0.1070	-0.1312	0.1067
$\partial p_0(H_{3nt}^{(1)})/(\partial h_{nt})$	-0.0939	0.0613	-0.0924	0.0608
$\partial p_0(H_{3nt}^{(2)})/(\partial h_{nt})$	-0.1089	0.0693	-0.1073	0.0673
$\partial p_0(H_{3nt}^{(3)})/(\partial h_{nt})$	-0.1126	0.0766	-0.1109	0.0728
$\partial p_0(H_{3nt}^{(4)})/(\partial h_{nt})$	-0.1471	0.0812	-0.1487	0.0794

†Time-averaged individual effects

‡Nonparametrically estimated individual effects

Table VI: Fixed Utility of Labour Force Participation

$u_{10}(z_{nt}) = d_{nt}B_0 z'_{nt}$

Variable	Parameter	Nonparametric	Traditional
$CONSTANT$	B_{00}	-12.23 (6.02)	-10.95 (6.39)
AGE_{nt}	B_{01}	0.635 (0.223)	0.555 (0.300)
AGE^2_{nt}	B_{02}	-0.007 (0.003)	-0.007 (0.004)
$AGE_{nt} \times EDUC_{nt}$	B_{03}	0.002 (0.0002)	0.0004 (0.001)
$MART.STATUS_{nt}$	B_{04}	0.311 (0.1001)	0.260 (0.122)
$BLACK_{nt}$	B_{05}	-0.582 (0.182)	-0.712 (0.343)
$HISPANIC_{nt}$	B_{06}	1.145 (0.421)	1.190 (0.354)

† Standard Errors in parenthesis.

Table VII: Nurturing Time Cost

$c_{nt} = \sum_{k=0}^{t} \rho_k b_{n,t-k}$

Variable	Parameter	Nonparametric	Traditional
NuturingTime	ρ_0	0.133 (0.012)	0.127 (0.023)
	$\rho 1$	0.021 (0.0003)	0.013 (0.002)
	ρ_2	0.013 (0.0002)	0.004 (0.002)
	ρ_3	0.019 (0.0001)	0.0002 (0.0002)
	ρ_4	0.016 (0.001)	0.005 (0.003)
	ρ_5	0.013 (0.0012)	0.005 (0.003)
	ρ_M	0.002 (0.0002)	0.0003 (0.0001)
RISK AV.	α	0.996 (0.328)	0.888 (0.481)

† Standard Errors in parenthesis.

Table VIII: Utility of Leisure

$$U_{11}\left(z_{nt}^{*}, l_{nt}\right) = z_{nt}' B_{11} l_{nt} + \sum_{s=0}^{\rho} \theta_s l_{n,t-s} l_{nt}$$

Variable	Parameter	Nonparametric	Traditional
l_{nt}	B_{110}	15.81 (3.45)	14.53 (2.97)
$AGE_{nt} \times l_{nt}$	B_{111}	0.066 (0.002)	0.110 (0.034)
$AGE_{nt}^2 \times l_{nt}$	B_{112}	-0.001 (0.0003)	-0.002 (0.0001)
$AGE_{nt} \times EDUC_{nt} \times l_{nt}$	B_{113}	-0.0003 (0.00001)	-0.0002 (0.0001)
$MART.STATUS_{nt} \times l_{nt}$	B_{114}	-0.082 (0.002)	-0.009 (0.003)
$BLACK_{nt} \times l_{nt}$	B_{115}	0.093 (0.024)	0.130 (0.102)
$HISPANIC_{nt} \times l_{nt}$	B_{116}	0.196 (0.023)	0.168 (0.121)
l_{nt}^2	θ_0	-4.077 (0.789)	-3.838 (1.328)
$l_{nt} l_{nt-1}$	θ_1	-4.304 (2.123)	-3.952 (2.125)
$l_{nt} l_{nt-2}$	θ_2	-2.142 (1.043)	-1.781 (0.756)
$l_{nt} l_{nt-3}$	θ_3	-1.155 (0.236)	-1.195 (0.876)
$l_{nt} l_{nt-4}$	θ_4	-0.413 (0.345)	-0.845 (0.642)

† Standard Errors in parenthesis.

Table IX: Birth Effects

$$U_{0nt} = b_{nt}\left(\gamma_0 + \sum_{k=1}^{M}\gamma_k b_{nt-k}\right) + b_{nt}\gamma_M \sum_{k=M+1}^{T} b_{nt-k}$$

$$\pi_{nt}\left(z_{nt}\right) = \pi_0 + z_{nt}'\pi_1$$

VARIABLE	PARAMETER	NONPARAMETRIC	TRADITIONAL
b_{nt}	γ_0	514.83 (48.78)	523.936 (300.23)
$b_{nt}b_{nt-1}$	γ_1	-3.088 (1.234)	-3.328 (2.134)
$b_{nt}b_{nt-2}$	γ_2	0.816 (0.346)	0.323 (0.187)
$b_{nt}b_{nt-3}$	γ_3	1.585 (0.362)	1.195 (0.872)
$b_{nt}b_{nt-4}$	γ_4	-1.162 (0.239)	-0.674 (0.078)
$b_{nt}b_{nt-5}$	γ_5	-0.570 (0.206)	-0.758 (0.492)
$b_{nt}\sum_{k=M+1}^{T} b_{nt-k}$	γ_M	-1.112 (0.419)	-1.065 (0.974)
$CONSTANT$	π_0	424.44 (78.98)	235.944 (98.67)
$HIGH\ SCH_{nt}$	π_1	0.481 (0.004)	0.846 (0.078)
$BLACK_{nt}$	π_2	1.078 (0.036)	0.770 (0.129)
$HISPANIC_{nt}$	π_3	3.635 (0.956)	3.23 (1.245)

† Standard Errors in parenthesis.

TABLE IX

Completed Fertility Simulation Outcome

Marital	Education	Actual	Estimation	Expenses	Daycare	Wages	Retraining
Black							
M	<	2.12	2.45	2.63	2.41	2.57	2.69
	HS	1.93	2.03	2.60	2.8	2.19	3.00
	>	1.35	1.68	1.71	2.3	1.66	2.50
U	<	2.15	2.35	2.56	2.58	2.41	2.57
	HS	1.82	1.97	2.04	2.1	1.98	2.05
	>	1.23	1.17	1.26	1.85	1.37	2.24
Hispandic							
M	<	2.08	2.19	2.23	2.31	2.25	2.02
	HS	1.83	1.79	1.89	2.03	1.87	2.35
	>	1.55	1.46	1.50	1.87	1.49	2.03
U	<	2.00	2.15	2.23	2.26	2.23	2.31
	HS	1.78	1.87	1.96	2.12	1.89	2.38
	>	1.46	1.56	1.67	2.00	1.72	2.30
White							
M	<	1.78	2.04	2.12	2.16	2.09	2.07
	HS	1.34	1.52	1.63	2.30	1.67	2.45
	>	1.12	1.23	1.32	1.97	1.24	2.03
U	<	1.47	1.56	1.54	1.78	1.58	1.87
	HS	1.25	1.31	1.56	1.90	1.67	2.08
	>	1.11	1.24	1.39	1.78	1.48	2.03

TABLE X

Simulation outcomes for Annual Labour Force Participation Rate

Marital	Education	Actual	Estimation	Expenses	Daycare	Wages	Retraining
Black							
M	$<$	0.570	0.452	0.436	0.421	0.476	0.423
	HS	0.673	0.772	0.722	0.732	0.724	1.723
	$>$	0.781	0.729	0.745	.742	0.732	0.732
U	$<$	0.678	0.616	0.606	0.627	0.601	0.591
	HS	0.723	0.763	0.751	0.749	0.761	0.763
	$>$	0.897	0.912	0.913	0.916	0.915	0.921
Hispandic							
M	$<$	0.612	0.634	0.632	0.625	0.623	0.618
	HS	0.722	0.745	0.739	0.738	0.737	0.735
	$>$	0.823	0.856	0.842	0.845	0.835	0.812
U	$<$	0.732	0.742	0.692	0.693	0.695	0.683
	HS	0.752	0.765	0.745	0.746	0.748	0.746
	$>$	0.824	0.878	0.867	0.857	0.856	0.872
White							
M	$<$	0.678	0.693	0.687	0.598	0.662	0.597
	HS	0.897	0.876	0.874	0.873	0.878	0.871
	$>$	0.912	0.927	0.921	0.928	0.926	0.923
U	$<$	0.753	0.734	0.727	0.714	0.701	0.692
	HS	0.857	0.876	0.767	0.798	0.845	0.855
	$>$	0.866	0.857	0.867	0.849	0.867	0.856

Figure I: Estimated Change in Aggregate Wage

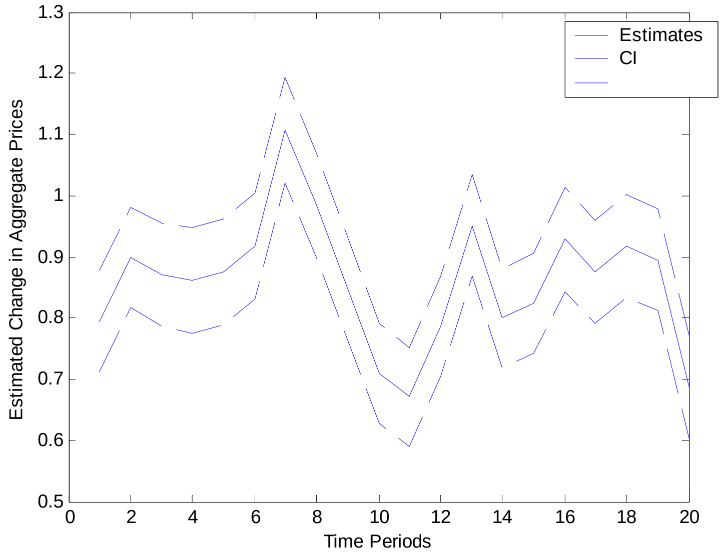

Figure II-a: Traditional Estimates of Fixed Effects of Marginal Products

Figure II-b: Nonparametric Estimates of Fixed Effects of Marginal Products

Figure III: Aggregate Prices

Figure IV -a: Traditional Fixed Effects Estimates of Social Weights

Figure IV-b: Nonparametric Fixed Effects Estimates of Social Weights

Figure V-a: Traditional Shadow Prices of Consumption

Figure V-b: Nonparametric Shadow Prices of Consumption