Discussion Paper 47

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

HUMAN CAPITAL, AGGREGATE SHOCKS, AND PANEL DATA ESTIMATION

Sumru Altug*

Institute of Empirical Macroeconomics
and University of Minnesota

Robert A. Miller*

Carnegie-Mellon University
and Economics Research Center/NORC

ABSTRACT

This paper analyses how the wage and employment decisions of females are affected by past workforce participation and hours supplied. Our estimation methods exploit the fact that, when markets are complete, the Lagrange multiplier for an agent's lifetime budget constraint always enters multiplicatively with the prices of (contingent claims to) consumption and leisure. Depending on the properties of the equilibrium price process, it is thus possible to predict the behavior of a wealthy agent by observing that of a poorer person living in a more prosperous world. This provides the key to estimating, nonparametrically, the expectations that enter the calculus of equilibrium decisionmaking, and ultimately the structural parameters which characterize preferences.

*This paper was first presented at the May 1990 "Conference on Empirical Applications of Structural Models" organized by the Econometric Society and SSRI at Madison, Wisconsin; the June 1990 "Conference on Specification Search and Robust Estimation in Micro Labor Markets" held at Center, Tilberg University, and 6th World Congress of the Econometric Society, Barcelona, Spain, August 1990. We have also benefited from seminar presentations at the World Bank, the Triangle Econometrics Workshop and Georgetown and Carnegie-Mellon Universities. All computations were done on the Cray-2 supercomputer on a grant from the Minnesota Supercomputer Institute.

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

1. INTRODUCTION

The apparent interdependence between wages and unemployment or nonparticipation, on the one hand, and human capital acquired through labor market experience, on the other, is difficult to ignore when studying female labor supply and participation. Due to greater demand for their time from their offspring, and perhaps because of sex discrimination, females supply less, but more variable, labor than males and earn lower wages. These effects are compounded by human capital acquired over the life cycle. With respect to older workers, females have less market experience than males, and this factor by itself yields lower wage and participation rates. Anticipating lower wages, more intermittent labor supply and earlier retirement, younger women choose jobs which have a smaller investment component. They also choose occupations that are more complementary with childrearing activities. Such choices increase their productivity at home and simultaneously retard depreciation. Contributions by Becker (1971, 1975), Ben-Porath (1967), Rosen (1972), Weiss and Gronau (1981), amongst others, have laid out the economic theory underlying this stylistic characterization.

Empirical studies have also addressed these issues. For example, Oaxaca (1973) regresses wages on characteristics observed by econometricians to determine how much can be explained by sex; Mincer and Ofek (1982) find interruptions in a woman's career do reduce her wages, more recent interruptions having a greater impact. Mincer and Polachek (1974) and Polachek (1981) provide some empirical results on the degree to which women make occupational choices with less learning on the job, the benefits of which may be more readily available at home, and with lower depreciation rates. Lazear (1979) finds one major impact of government programs to eliminate sexual discrimination is to increase investment in human capital by females. Related work in this area also includes Corcoran (1979), Corcoran and Duncan

(1979), Polachek (1975, 1979) and Sandell and Shapiro (1978). While informative, statistical inference from these studies is jeopardized by sample selection and endogeneity problems. Moreover, these studies do not account directly for the impact of changes in overall market conditions or aggregate shocks on the decision to acquire home versus market capital.

The issue of incorporating the effects of aggregate shocks in panel data estimation of dynamic equilibrium models was recently addressed by Altug and Miller (1990). In the environment considered there, time dummies could be used to capture the effects of aggregate shocks that are transmitted through equilibrium prices in observed or realized states of the world. However, in models with preferences that are not additively separable over time, or with human capital accumulation, job search and matching, as well as fertility, agents or households evaluate their utility in future states of the world that never occur as they make plans for the future. In this case, the method of inserting dummy variables to estimate prices of contingent claims for observed states of the world is insufficient. Since there exist goods (or assets) that yield consumption (or dividends) in states of the world that might have been realized but were not, the econometrician would like to know the probability distribution that characterizes the state variables, and estimate the prices as functions of the possible states that lie in the support of this probability distribution. Previous applications have resorted to a variety of solutions when faced with this problem. In his paper on job matching Miller (1984) assumes utility is linear in consumption; Wolpin (1984) and Hotz and Miller (1989) assume aggregate shocks are insignificant; the empirical study of Hotz, Kydland and Sedlacek (1988) on nonadditivities in leisure find the time dummies they insert are significant but lack a theoretical interpretation for them.

This paper estimates a dynamic, stochastic model with a constant interest rate but aggregate shocks, where time nonseparabilities arise from the specification of preferences and the human capital accumulation process. It seeks to investigate how past labor market participation and hours of work decisions by women affect their current wages and employment, and to estimate the depreciation that occurs when females temporarily or permanently withdraw from the labor market. The estimation method proposed here seeks to incorporate the effects of aggregate shocks in panel data estimation of models with time nonseparabilities by making three important assumptions. The first one concerns the market structure; we assume it is competitive and complete. The second specifies, up to a parameterization, the stochastic process for prices to follow and thus indirectly characterizes aggregate shocks; we assume proportional changes in contingent prices are independent of their level. The third assumption is that a sufficient statistic for individual wealth, or some transformation of it, is available on the cross section. As a consequence of these assumptions, it is possible to predict the behavior of a wealthy agent living in economic slumps by observing that of a poorer person living in a more prosperous world. Hence, the probability distributions describing the behavior of an individual in some future event can be inferred by nonparametrically estimating the current behavior of individuals she may later mimic weighted by the probability of this event occurring. Rather than estimating the whole probability distribution, however, the techniques of simulation estimation can be adapted to this nonparametric context.

The paper is organized as follows. Section 2 describes a framework for analyzing human capital and preferences that are not additively separable over time; it also explains why the Euler equation methods of Hansen and Singleton (1982) cannot be applied directly, or with minor modifications (such as inserting additive time dummies to the forecast error). Then in Section 3 we derive new representations for the Euler and participation equations, from which orthogonality conditions can be formed, to identify an estimator for this class of problems which can be implemented with panel data. The representations are similar to those previously developed by Hotz and Miller (1990) for sequential, discrete choice problems, but here adapted to account for aggregate shocks. Since the computational costs associated with integrating over all future aggregate shocks would be prohibitive, a simulation estimator is developed. Our motivation is related to that given for recent analyses of the (parametric) simulation estimators in McFadden (1989) and Pakes and Pollard (1989). Because the criterion function is differentiable in our case however, standard proof strategies can be applied, providing the incidental parameter problem (associated with estimating the conditional choice probabilities) is dealt with. The estimator itself is consistent in N, the number belonging to the cross section, converges at rate $N^{1/2}$, and is asymptotically normal. Section 4 contains a discussion of the data and the empirical specification we estimate. The last section estimates the model and reports our results.

2. A FRAMEWORK

Information, Preferences and Technology

This section describes the model we develop to address the empirical relationship between human capital, wages and female labor supply. Consider the following two sector competitive model of consumption and (female) labor supply. Let $c_{nt} \in [0,\infty)$ denote consumption by agent n in period t, and $l_{nt} \in [0,\infty)$ her hours of work. Previous work experience affects labor's productivity in market work via the $\rho \times 1$ vector of lagged hours $\underset{\sim}{l}_{nt} \equiv (l_{n,t-1},\dots,l_{n,t-\rho})'$. The economy runs from date 1 to date $\tau < \infty$, and the n^{th} individual is active between dates $\underline{n}$ and $\bar{n}$. Uncertainty is treated as the probability space $(\Omega, \mathcal{F}, P)$, the element signifying a particular realization of all (random) variables in this economy from 1 to τ. The increasing sequence of σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ characterizes how information accumulates over time; we assume all information is public; thus the random variables with t subscripts defined below are $\mathcal{F}_t$ measurable.

Each period $t \in \{\underline{n}, \ldots, \bar{n}\}$ the n^{th} woman chooses c_{nt} and contracts with firms to undertake tasks. The total amount of time they take to complete, l_{nt}, is only revealed to her when she begins working on them. More specifically, let $h_{nt} \in [0, \infty)$ be chosen by n at the beginning of period t, and let ε_{nt} denote a random variable, identically and independently distributed across (n,t) with probability distribution function $F(\varepsilon_{nt})$ defined on support $[0, \infty)$ which has at least two moments. We assume $l_{nt} = 0$ if $h_{nt} = 0$, but if $h_{nt} > 0$ then:

$$(2.1) \qquad l_{nt} = h_{nt} + \varepsilon_{nt}$$

For convenience, a participation indicator $d_{nt} \in \{0, 1\}$ is also defined, with $d_{nt} = 1$ if and only if $h_{nt} > 0$.

Females are identical up to a vector of characteristics $(x'_{nt}, y'_n, \underset{\sim}{l}_{nt}, \varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})$, where x_{nt} is an observed, time-varying $q \times 1$ vector independently distributed across the population and generated by the known probability transition $F_0(x_{n,t+1} | x_{nt})$; $y_n \equiv (y_{n1}, \ldots, y_{n\nu})'$ is an observed $\nu \times 1$ vector fixed over time; and $(\varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})$ is distributed independently across (n,t) drawn from $F_1(\varepsilon_{0nt}, \varepsilon_{1nt}, \varepsilon_{2nt})$. They obey the expected utility hypothesis and have rational expectations, preferences taking the time additive form:

$$(2.2) \qquad E_0\left[\sum_{t=\underline{n}}^{\bar{n}} \beta^t (u_{nt})\right]$$

where:

$$(2.3)\qquad u_{nt} \equiv u_{0nt} + u_{1nt} + u_{2nt}$$

$$u_{0nt} \equiv u_0(x_{nt},\ y_n,\ (1-d_{nt})\varepsilon_{0nt} + d_{nt}\varepsilon_{1nt})$$

$$u_{1nt} \equiv u_1(\underset{\sim}{l}_{nt},\ l_{nt},\ x_{nt},\ y_n)$$

$$u_{2nt} \equiv u_2(c_{nt},\ x_{nt},\ \varepsilon_{2nt})$$

Thus u_{0nt} represents the (reduction in) utility from participating in the labor force, u_{1nt} the (decremental) loss from working greater numbers of hours (which reduce leisure time), and u_{2nt} is the current utility from consumption. Notice that u_{0nt} and u_{2nt} are known at the beginning of period t when female n makes her consumption choices and work plans, but u_{1nt} is only revealed at the end of period t if she participates.

We denote by w_{nt} the marginal product of the n^{th} agent's labor working for a firm belonging to the first sector (goods) at time t, and let $w_t \equiv (w_t^{(1)},\ldots,w_t^{(\nu)})'$ represent the $\nu \times 1$ price vector for units of y characteristics in period t. We assume w_{nt} admits the multiplicative decomposition

$$(2.4)\qquad w_{nt} = \sum_{r=1}^{\nu} w_t^{(r)} y_{nr}\gamma_r(\underset{\sim}{l}_{nt},x_{nt})$$

Thus $\gamma_r(\underset{\sim}{l}_{nt},x_{nt})$ measures the efficiency units of y person n produces per unit time.

This specification of preferences and choices over leisure differs from those found in the existing literature in several important respects. In contrast to previous static models of female labor supply (discussed in Killingsworth [1983] for example), and their dynamic counterparts (estimated by Heckman and MaCurdy [1980, 1982]), investment in human capital and intertemporally nonseparable preferences play an important role here. Another distinctive feature of our formulation is that, conditional on working, labor supply plans depend on observables alone; the difference between those plans and actual hours worked, attributable to unanticipated market demands upon a woman's time, prevents stochastic singularities arising (between working women who share identical characteristics but supply different amounts of hours). The limited role unobservables play in selection, coupled with the prominence attached to dynamic considerations, is foreshadowed in Mroz (1987); he reports that "among potential specifications found unimportant are ... controls for self-selection when experience is treated as endogenous" (p. 795), whereas assuming the wife's wage is exogenous induces "upward bias in the estimated wage effect" (p. 795). Comparing our study with related work in male labor supply, Hotz, Kydland and Sedlacek (1988) and Shaw (1989) respectively investigate nonseparabilities in male labor supply and learning on the job, but whereas their work assumes an interior solution pertains, we investigate the participation decision as well. Again the reason for this difference is driven by data: the female workforce participation rate is substantially lower than the rate for males.

Optimization and Equilibrium

In a competitive equilibrium with complete markets, prices exist for all commodities. Accordingly, let $\Lambda_t \equiv (\Lambda_{0t},\ldots,\Lambda_{\nu t})'$, measures defined on F_t for each $t \in \{1,\ldots\}$, denote prices of contingent claims to consumption and efficiency units of labor in period t. Thus $\Lambda_{0t}(A_t)$ denotes the date 0 price of a consumption unit vector to be delivered on date t, contingent on $A_t \in \mathcal{F}_t$

being realized. We assume Λ_t is absolutely continuous with respect to P and denote by $\lambda_t \equiv (\lambda_{0t},\ldots,\lambda_{\nu t})'$ the Radon-Nikodym derivative of Λ_t with respect to P. Hence Λ_t admits the representation:

$$(2.5) \qquad \Lambda_t(A_t) = \int_{A_t} \lambda_t(\omega)P(d\omega)$$

Female n maximizes her expected utility at date 0 by choosing the $\mathcal{F}_t$ measurable vector (c_{nt}, h_{nt}) for $t \in \{\underline{n},\ldots,\bar{n}\}$ subject to a lifetime budget constraint. Define the exogenously determined quantity c_n as bequests net of inheritances. Assuming bundles of goods are valued via an inner product representation, (2.5) implies the lifetime budget constraint for the n^{th} household may be written as:

$$(2.6) \qquad E_0\left\{\sum_{t=\underline{n}}^{\bar{n}} \lambda_{0t}[c_{nt} - l_{nt}w_{nt}]\right\} \leq c_n$$

where $w_t^{(r)}\lambda_{0t} = \lambda_{rt}$ for $r \in \{1,\ldots,\nu\}$. For future reference, let η_n denote the Lagrange multiplier associated with the budget constraint (2.6), c^*_{nt} the optimal consumption of n in period t; also let d^*_{nt} characterize her optimal participation decision, and h^*_{nt} be the expected labor supply associated with the local interior optimum. Thus (d^*_{nt}, h^*_{nt}) determines her optimal time allocation plan, which results in $l^*_{nt} \equiv d^*_{nt}(h^*_{nt} + \varepsilon_{nt})$ hours of work.

We assume the real interest rate is constant, which implies $E_0(\beta\lambda_{0t}/\lambda_{0,t-1})$, the price of sure consumption in period t in terms of period t-1 goods, does not change. As an empirical matter, the lack of interest rate variation over time, coupled with the difficulty in measuring it (using nominal interest rates and data on inflation), makes the assumption relatively innocuous for short panels. Similar assumptions are made with respect to the other intertemporal prices; more specifically, we assume the stochastic process for

$$\psi_t \equiv (\lambda_{0t}/\lambda_{0,t-1}, \ldots, \lambda_{\nu t}/\lambda_{\nu,t-1})$$

is independently and identically distributed over time with distribution function $F_2(\psi_t)$. Although prices are specified exogenously, we note that, as in the representative consumer model of Novales (1990), the individual optimization problem described here coexists with a competitive equilibrium for a simple economic environment. Suppose there are some assets which yield an exogenous stream of dividends (in consumption units), and assume there also exists a stochastic, constant scale returns production technology for consumption using labor inputs alone; it is characterized by the law of motion for w_t. Letting D_t denote total dividends in period t and $H_t(x_{nt}, \underset{\sim}{l}_{nt}, y_n, \eta_n)$ the joint probability distribution of $(x_{nt}, \underset{\sim}{l}_{nt}, y_n, \eta_n)$ over the population at that time, clearance in the goods market means:

$$(2.7) \qquad D_t = \int \left[c^*_{nt} - w_{nt} d^*_{nt} h^*_{nt}\right] dH_t(x_{nt}, \underset{\sim}{l}_{nt}, y_n, \eta_n)$$

Suppose ownership claims to assets are distributed in any way to make n initially evaluate her wealth as c_n at prices $\{\Lambda_t\}_{t\in T}$. Then $\{\Lambda_t\}_{t\in T}$ is a competitive equilibrium if these prices support the consumption and leisure allocations generated by the production and dividend process. By defining the dividend process D_t to satisfy (2.7), it is thus established, by construction, that this is indeed the case.

The first order conditions for an interior solution to the agent's optimization problem for the equilibrium postulated above are:

(2.8) $$0 = \partial E_t\left[\sum_{s=1}^{\rho}\beta^s u_1(\underset{\sim}{l}_{n,t+s} l_{n,t+s}, x_{n,t+s}, y_n) + \eta_n \sum_{r=1}^{\nu} \lambda_{r,t+s}\, y_{nr}\, \gamma_r(\underset{\sim}{l}_{n,t+s}, x_{n,t+s}) l_{n,t+s}\right] / \partial h_{nt}$$

(2.9) $$\partial u_2(c_{nt}, x_{nt}, \varepsilon_{2nt})/\partial c_{nt} = \lambda_{0t}\eta_n$$

Then (c^*_{nt}, h^*_{nt}), the solutions to (3.1) and (3.2), denote the optimal interior policy. Whether agent n participates in period t or not depends on the difference in the social surplus she generates, because the competitive allocation is pareto optimal under the assumption that markets are complete. The social surplus associated with the labor supply decision accounts both for the effects on her utility, and also the value of her marginal product from working weighted by her marginal utility of wealth. Accordingly define the conditional valuation function for setting $d_{nt} = k \in \{0,1\}$ by:

(2.10)

$$v_{nt}^{(k,0)} = \max_{\{h_{ns}\}_{s=t+1}^{\bar{n}}} E_t\left[u_{1nt} + \eta_n w_{nt} l_{nt} + \sum_{s=t+1}^{\bar{n}} \beta^{s-t}(u_{0ns} + u_{1ns} + \eta_n w_{ns} l_{ns}) \,|\, h_{nt} = kh^*_{nt}\right]$$

The social surplus associated with planning to work h^*_{nt} is therefore $u_0(x_{nt}, y_n, \underset{\sim}{l}_{nt}, \varepsilon_{1nt}) + v_{nt}^{(1,0)}$, and the optimal participation rule d^*_{nt} is defined:

$$(2.11)\qquad d^*_{nt} = \begin{cases} 1 \text{ if } u_0(x_{nt}, y_n, \varepsilon_{1nt}) + v^{(1,0)}_{nt} \geq u_0(x_{nt}, y_n, \varepsilon_{0nt}) + v^{(0,0)}_{nt} \\ 0 \text{ otherwise} \end{cases}$$

Thus $(c^*_{nt}, d^*_{nt} h^*_{nt})$ summarizes the optimal contingent plan for woman n facing the problem (2.1) through (2.6).

Forecast Errors in Cross Sections

Our empirical methods exploit the Euler and participation equations, to estimate our model with panel data. We conclude this section by demonstrating why time series methods for estimating Euler equations should not be applied directly to (short) panels. Originally developed by Hansen and Singleton (1982) for analyzing time series data, Euler equation methods exploit the orthogonality between an agent's forecast error and elements within her information set. As Chamberlain's (1984, p. 1311) remarks suggest, this procedure cannot be applied directly here, because large sample averages of cross sections taken at a single point in time (or a small number of them) cannot typically margin out aggregate fluctuations.

To illustrate this point, we now temporarily specialize the model by replacing (2.1), (2.3) and (2.4) with:

$$(2.12)\qquad h_{nt} = l_{nt}$$

$$u_{0nt} + u_{1nt} + u_{2nt} = l_{nt} l_{n,t-1} - l^2_{nt} + \ln(c_{nt})$$

$$w_{nt} = w$$

respectively, and assuming ν_t, the only remaining aggregate shock in this specialization, is a standard normal variate, with distribution function $\Phi(\nu)$;

these additional assumptions leave only β to estimate. Without further loss of generality we normalize $\lambda_{0t} = 1$, so the period t labor supply function for individual n can then be expressed as $h^*_{nt} \equiv h(l_{n,t-1}, \eta_n)$. Let:

$$e_{nt} \equiv \beta h^*_{n,t+1} - \beta E_t(h^*_{h,t+1})$$

denote the forecast error associated with the Euler equation:

$$h_{n,t-1} - 2h^*_{nt} + \beta E_t\left[h^*_{n,t+1}\right] = \eta_n w$$

Averaging e_{nt} over the population of females $n \in \{1,\ldots,N\}$, all of whom are assumed to participate in this example, we obtain in the limit of N:

$$e_t^{(1)}(\nu_{t+1}) \equiv \beta\int\left\{h\left[h(l_{n,t-1},\eta_n),\ \eta_n\nu_{t+1}\right]\right.$$

$$\left. - \int h[h(l_{n,t-1},\eta_n),\ \eta_n\nu]d\Phi(\nu)\right\} dH_t(l_{n,t-1},\eta_n)$$

where $H_t(l_{n,t-1},\eta_n)$ is the population distribution function for $(l_{n,t-1},\ \eta_n)$. Unless labor supply is independent of wealth (making the second argument of $h(l,\eta_n)$ redundant), $e_t^{(1)}(\nu_{t+1})$, is a nontrivial function of ν_{t+1}.

Recognizing the importance of macroeconomic shocks in their own studies which have applied Euler equation methods to panel data, several authors, including Hotz, Kydland and Sedlacek (1988), Shaw (1989) and Zeldes (1989), inserted time dummies into the forecast errors; they then assumed the remaining components are independently distributed across the population and over time, and proceeded by adopting as instruments elements belonging to the information sets of agents thought to be correlated with their observed choices. Altug and Miller (1990) show that if there is no human capital accumulation and preferences are additively separable over time, the procedure can be justified in the presence of complete markets. (The idiosyncratic shock to the forecast error is then identically zero). Otherwise it is hard to rationalize, as Hotz, Kydland and Sedlacek [1988] acknowledge (p. 347). To see why, consider any valid instrument vector, such as $(1,\eta_n)' \in \mathcal{F}_0 \subseteq \mathcal{F}_t$. Averaging over $n \in \{1,\ldots,N\}$ the product of $(1,\eta_n)$ and e_{nt} converges to the vector $(e_t^{(1)}(\nu_{t+1}),\ e_t^{(2)}(\nu_{t+1}))$ where:

$$e_t^{(2)}(\nu_{t+1}) = \beta\int\eta_n\{h[h(l_{n,t-1},\eta_n),\eta_n\nu_{t+1}] - \int h[h(l,\eta_n),\eta_n\nu]d\Phi(\nu)\}dH_t(l_{n,t-1},\eta_n)$$

Contradicting the assumption made by the authors cited above $e_t^{(1)}(\nu_{t+1}) \neq e_t^{(2)}(\nu_{t+1})$, and both are nonzero functions. Therefore two sets of time dummies are required to correct the sample moments for their cross sectional bias (rather than just one). Consequently, there are three parameters to estimate $(e_t^{(1)},\ e_t^{(2)},\ \beta)$ from only two equations, so the system remains unidentified. Moreover adding extra instruments cannot identify β (the parameter of interest), because the number of time dummies to estimate also increases concomitantly.

3. ESTIMATION AND INFERENCE

Representations for the Optimality Conditions

This section modifies the existing approach to inference from Euler equations so that they can be applied to panel data. The modifications involve a new representation of the Euler equation, and the development of appropriate estimation techniques. As a biproduct, we show how estimating dynamic models of discrete choices by treating the conditional choice probabilities as incidental parameters, can be extended to incorporate aggregate shocks transmitted through prices.

Our analysis is prefaced by some remarks to motivate the estimator and clarify the role of certain assumptions. Since the stochastic process for ψ_t (changes in contingent prices in realized states of the world), is independent and identically distributed, and labor supply lagged more than ρ periods does not affect current utility or the wage rate, the finite dimensional vector $(\underset{\sim}{l}_{nt}, x_{nt}, y_n, \eta_n\lambda_t)$ is a sufficient statistic for choosing (c_{nt}, h_{nt}) optimally. Inspection of (2.1) through (2.6) reveals a further reduction in the state space is possible: the vector z_{nt} defined as:

$$z_{nt} \equiv (\underset{\sim}{l}_{nt}, x_{nt}, \eta_n\lambda_{0t}, \eta_n y_{1n}\lambda_{1t}, \ldots, \eta_n y_{\nu n}\lambda_{\nu t})$$

is (minimal) sufficient. Our assumptions imply $\eta_n y_n$ is dense in the positive orthant of $\mathbb{R}^{\nu+1}$. Hence, following a similar semiparametric procedure to Hotz and Miller (1989), cross sectional variation in z_{nt} can be used instead of time series data to nonparametrically estimate, as preliminary incidental parameters, the policy function evaluated at certain arguments. Intuitively, this is because an individual's response to any aggregate shocks (not necessarily observed in the data) systematically resembles some behavior of others actually observed. For example, supposing:

$$(x_{nt}, \underset{\sim}{l}_{nt}) = (x_{ms}, \underset{\sim}{l}_{ms}),$$

$$\eta_n(\lambda_{0t}, y_{1n}\lambda_{1t}, \ldots, y_{\nu n}\lambda_{\nu t}) = \eta_m(\lambda_{0s}, y_{1m}\lambda_{1s}, \ldots, y_{\nu m}\lambda_{\nu s}),$$

then $h^*_{nt} = h^*_{ms}$, so one can imagine inferring the behavior of a female m in some future period s who responds to an aggregate shock λ_s, by forming an appropriate comparison group of choices actually observed in (the earlier) period t.

This approach cannot be extended easily to accommodate more flexible price processes, as the following example illustrates. In contrast to the assumption in Section 2, now suppose the stochastic process for ψ_t depends on λ_t (as ARCH models suggest). Temporarily replacing $F(\psi_t)$ with $\tilde{F}(\psi_t|\lambda_t)$, the state space becomes (z_{nt},λ_t). Since the contingent prices applying to any realized state can be parameterized by a set of time dummies, our estimation methodology is valid subject to the important caveat that $\rho = 1$. If, however, more than one lagged labor supply enters the model (meaning $\rho \geq 2$), it is impossible to obtain consistent estimates of h^*_{ms} for $s \geq t + 2$ (to substitute into the Euler equation), unless coincidentally $\lambda_{s-1} = \lambda_t$.

The estimator is developed in stages. We first derive new representations for the Euler and participation equations upon which our estimator is based. They both hinge on the observation that the difference between the conditional valuation functions can be expressed as a mapping of the conditional choice probability. Let $p_{nt} \equiv p(z_{nt})$ denote the (nonlinear) regression of d_{nt} on z_{nt}, or the conditional probability of n participating on date t given attributes z_{nt}. Appealing to Proposition 1 of Hotz and Miller (1989), there exists a continuous, increasing mapping defined from $p \in (0,1)$ to $(-\infty,\infty)$, denoted q(p), such that:

$$(3.1) \qquad q(p_{nt}) = v_{nt}^{(1,0)} - v_{nt}^{(0,0)}$$

for all $z_{nt} \in Z$.

Equation (3.1) is a building block for characterizing expectations about the future in the participation and Euler equations. To see this, some extra notation is helpful. Define for any $\underset{\sim}{l}_{nt}$ the ρ dimensional row vector $\underset{\sim}{l}_{nt}^{(0,s)}$

as $(1_{n,t-\rho+s},\ldots,1_{n,t-1},0,\ldots,0)$; this corresponds to person n accumulating history $\underset{\sim}{1}_{nt}$ by time t but then not working for s periods. (Thus $\underset{\sim}{1}_{nt}^{(0,\rho)}$ is the ρ dimensional 0 vector.) Similarly let $\underset{\sim}{1}_{nt}^{(1,s)}$ denote the ρ dimensional row vector $(1_{n,t-\rho+s},\ldots,1_{n,t-1}, h^*_{nt} + \varepsilon_{nt}, 0,\ldots,0)$; in this case a nonemployment spell of s-1 periods follows period t when she works $h^*_{nt} + \varepsilon_{nt}$ given her t history $\underset{\sim}{1}_{nt}$.

We also adopt the following abbreviations. For $k \in \{0,1\}$ and $s \geq k$, let:

$$(3.2) \qquad u_{nt}^{(k,s)} = u_1(\underset{\sim}{1}_{nt}^{(k,s)}, 0, x_{n,t+s}, y_n)$$

$$p_{nt}^{(k,s)} = p(\underset{\sim}{1}_{nt}^{(k,s)}, x_{n,t+s}, \eta_n\lambda_{0,t+s},\ldots,\eta_n y_{\nu n}\lambda_{\nu,t+s})$$

$$q_{nt}^{(k,s)} = q(p_{nt}^{(k,s)})$$

Thus $u_{nt}^{(k,s)}$ represents the current utility to n of not participating in period t + s after a nonemployment spell of s-k periods, while $p_{nt}^{(k,s)}$ is the conditional probability she will resume working then. Finally let $u_{nt}^{(1,0)}$ denote the expected social surplus (in female n utils), accruing at time t, from n planning to work h^*_{nt}, gross of participation costs:

$$(3.3) \qquad u_{nt}^{(1,0)} = \int\left[u_1(\underset{\sim}{1}_{nt}, h^*_{nt} + \varepsilon_{nt}, x_{nt}, y_n) + (h^*_{nt} + \varepsilon_{nt})w_{nt}\eta_n\right]dF(\varepsilon_{nt})$$

This notation the representations which we will presently exploit in estimation are given concisely by the following.

Proposition 1

$$(3.4) \qquad 0 = \frac{\partial}{\partial h^*_{nt}} E_t\left\{u_{nt}^{(1,0)} + \sum_{s=1}^{\rho}\beta^s\left[u_{nt}^{(1,s)} + p_{nt}^{(1,s)}q_{nt}^{(1,s)}\right]\right\}$$

$$(3.5) \quad v_{nt}^{(1,0)} - v_{nt}^{(0,0)} = E_t\left\{u_{nt}^{(1,0)} - u_{nt}^{(0,0)} + \sum_{s=1}^{\rho}\beta^s\left[u_{nt}^{(1,s)} - u_{nt}^{(0,s)} + p_{nt}^{(1,s)}q_{nt}^{(1,s)} - p_{nt}^{(0,s)}q_{nt}^{(0,s)}\right]\right\}$$

Proof of Proposition 1

To prove this proposition we extend the definition of $v_{nt}^{(k,0)}$ given in (2.10). For each $s \in \{0,\ldots,\rho+1\}$, define $v_{nt}^{(k,s)}$ as:

$$v_{nt}^{(k,s)} = \max_{\{h_{nr}\}_{r=t+s+1}^{\bar{n}}} E_{t+s}\left[u_{nt}^{(k,s)} + \sum_{r=t+s+1}^{\bar{n}}\beta^{r-t-s}\left(u_{0nr}+u_{1nr}+\eta_n w_{nr} l_{nr}\right) \mid \underset{\sim}{l}_{nt}^{(k,s)}\right]$$

Intuitively, $v_{nt}^{(k,s)}$ is the expected utility, taken from a social planner's perspective but measured in the utils of woman n at time t + s if, having been out of the labor force for s - k periods, she does not participate in period t + s but thereafter optimally chooses her labor supply. Combining Bellman's (1957) principle of optimality with (3.1):

$$(3.6) \qquad v_{nt}^{(k,s)} = u_{nt}^{(k,s)} + \beta E_t\left[v_{nt}^{(k,s+1)} + p_{nt}^{(k,s+1)}q_{nt}^{(k,s+1)}\right]$$

Starting at $v_{nt}^{(k,s)}$, and repeatedly applying (3.6) for $s \in \{0,\rho+1\}$, then yields:

$$(3.7) \qquad v_{nt}^{(k,0)} = u_{nt}^{(k,0)} + E_t\left\{\beta^{\rho+1} v_{nt}^{(k,\rho+1)} + \sum_{s=1}^{\rho} \beta^s \left[u_{nt}^{(k,s)} + p_{nt}^{(k,s)} q_{nt}^{(k,s)}\right]\right\}$$

The Euler equation is found by setting k = 1 and differentiating $v_{nt}^{(1,0)}$ with respect to h_{nt}^*. Since $\underset{\sim}{1}_{nt}^{(1,\rho+1)} = 0$, it follows that $v_{nt}^{(1,\rho+1)}$ does not depend on h_{nt}^*. Therefore:

$$0 = \frac{\partial v_{nt}^{(1,0)}}{\partial h_{nt}^*}$$

$$= \frac{\partial}{\partial h_{nt}^*} E_t\left\{u_{nt}^{(1,0)} + \sum_{s=1}^{\rho} \beta^s \left[u_{nt}^{(1,s)} + p_{nt}^{(1,s)} q_{nt}^{(1,s)}\right]\right\}$$

as claimed in (3.4). The participation equation is derived by substituting for k = 1 and k = 0 in (3.7), and differencing the resulting two equations. Noting $v_{nt}^{(0,\rho+1)} = v_{nt}^{(1,\rho+1)}$, equation (3.5) obtains. ||

Notice (3.4) requires an estimate of the derivative of $\partial p_{nt}^{(1,s)}/\partial h_{nt}$; to obtain it, we exploit results (reviewed in Prakasa Rao [1983] for example) on nonparametric estimation of derivatives of probability density functions. In particular, let:

$$z_{nt}^{(k,s)} \equiv (\underset{\sim}{1}_{nt}^{(k,s)}, x_{n,t+s}, \eta_n(1, y_n) * \lambda_{t+s})$$

and define:

$$f_{1nt}^{(s)} \equiv f_1(z_{nt}^{(1,s)} | d_{n,t+1} = \ldots = d_{n,t+s-1} = 0 \text{ and } d_{n,t+s} = 1)$$

as the probability density function for $z_{nt}^{(s)}$ conditional on not participating in periods t + 1 through t + s - 1 but working in period t + s. Similarly let:

$$f_{0nt}^{(s)} \equiv f_0(z_{nt}^{(0,s)} | d_{n,t+1} = \ldots = d_{n,t+s-1} = 0)$$

be a related probability density function, which does not condition n on participating in period t + s. Their derivatives with respect to h^*_{nt} are denoted by $f_{1nt}^{(s)\prime}$ and $f_{0nt}^{(s)\prime}$ respectively. It now follows that:

$$(3.8) \qquad \frac{\partial(p_{nt}^{(1,s)} \, q_{nt}^{(1,s)})}{\partial h^*_{nt}} = \left[p_{nt}^{(1,s)} q'(p_{nt}^{(1,s)}) + q_{nt}^{(1,s)}\right] \frac{\partial p_{nt}^{(1,s)}}{\partial h_{nt}}$$

$$= \left[p_{nt}^{(1,s)} q'(p_{nt}^{(1,s)}) + q_{nt}^{(1,s)}\right] \left[\frac{f_{1nt}^{(s)\prime}}{f_{1nt}^{(s)}} - \frac{f_{0nt}^{(s)\prime}}{f_{0nt}^{(s)}}\right] p_{nt}^{(1,s)}$$

Thus (3.4) becomes:

$$(3.9) \qquad -w_{nt}\eta_n = E_t\left\{\sum_{s=1}^{\rho} \beta^s \left[\frac{\partial u_{nt}^{(1,s)}}{\partial h^*_{nt}}\right.\right.$$

$$\left.\left. + \left(p_{nt}^{(1,s)} q'(p_{nt}^{(1,s)}) + q_{nt}^{(1,s)}\right)\left(\frac{f_{1nt}^{(s)\prime}}{f_{1nt}^{(s)}} - \frac{f_{0nt}^{(s)\prime}}{f_{0nt}^{(s)}}\right) p_{nt}^{(1,s)}\right] + \frac{\partial u_1(\underset{\sim}{1}_{nt}, h^*_{nt} + \varepsilon_{nt}, x_{nt}, y_n)}{\partial h^*_{nt}}\right\}$$

Orthogonality Conditions

If certain incidental parameters (including the marginal utility of wealth and some conditional choice probabilities specified below) are known, then GMM techniques can be directly applied to produce a $N^{1/2}$ consistent and asymptotically normal estimator $\theta_1^{(N)}$. There are three steps to estimation, although the second two are taken together.

The first simulates, for each person $n \in N$, a hypothetical future path of ρ periods for the aggregate shocks and exogenous time varying characteristics, respectively denoted by $\{\psi_{nt}^{(s)}(\theta)\}_{s=1}^{\rho}$ and $\{x_{nt}^{(s)}\}_{s=1}^{\rho}$. Consequently the contingent price vector in period t + s associated with the simulated state for n is

$$\lambda_{nt}^{(s)}(\theta) \equiv \lambda_t \Pi_{r=1}^{s} \psi_{nt}^{(s)}(\theta),$$

where $\lambda_t \in \mathcal{F}_t$ is a time dummy to be estimated. Observe $F_0(x_{n,t+1}|x_{nt})$, the transition probability for x_{nt} is either known, or can be estimated prior to simulation. However $\psi_{nt}^{(s)}(\theta)$ is governed by the parametric distribution $F_2(\psi;\theta)$, which is identified by expectations people reveal via orthogonality conditions formed from the Euler and participation equations (discussed below). To simulate:

$$\psi_{nt}^{(s)}(\theta) \equiv (\psi_{0nt}^{(s)}(\theta),\ldots,\psi_{\nu nt}^{(s)}(\theta))$$

for each (n,t,s), we take $(\nu+1)$ random draws from the uniform [0,1] distribution, generically denoted ξ_r, express $F_2(\psi;\theta)$ as the product of a marginal distribution and ν conditional ones, namely:

$$F_2(\psi;\theta_0) \equiv F_{20}(\psi_0;\theta)\Pi_{r=1}^{\nu} F_{2r}(\psi_r|\psi_{r-1},\ldots,\psi_0;\theta_0),$$

and then recursively assign $\psi_{rnt}^{(s)}$ the value

$$F_{2r}^{-1}(\xi_r|\psi_{r-1,nt}^{(s)},\ldots,\psi_{0nt}),$$

where F_{2r}^{-1} is the inverse of the conditional distribution function $F_{2r}(\psi_r|\psi_{r-1},\ldots,\psi_0;\theta)$.

The second step estimates, nonparametrically, the marginal utilities of wealth, the conditional choice probabilities and policy functions appearing in Proposition 2, evaluated at appropriate points (which are jointly determined by the data and the simulations). The last step substitutes these incidental parameter estimates into Proposition 2, forms sample moments to orthogonality conditions, from which $\theta_1^{(N)}$ emerges as a $N^{1/2}$convergent, asymptotically normal, estimator.

Before discussing the second step, we show how our method would be applied in the absence of an incidental parameter problem. Both the labor supply of participants and the participation decision itself have information content. Here the interior solution for labor supply exploits (3.9). Let $p_n \equiv \{p_{nt}^{(0,s)}, p_{nt}^{(1,s)}, f_{1nt}^{(s)}, f_{1nt}^{(s)\prime}, f_{2nt}^{(s)}, f_{2nt}^{(s)\prime}\}_{s=0}^{\rho}$ and define $g_{1n}(\theta,p_n)$ as

$$(3.10) \qquad g_{1n}(\theta,p_n) \equiv w_{nt}\eta_n + \frac{\partial u_1(\underset{\sim}{l}_{nt},h_{nt}^{*}+\varepsilon_{nt},x_{nt},y_n)}{\partial h_{nt}}$$

$$+\sum_{s=1}^{\rho}\beta^s\left\{\frac{\partial u_{nt}^{(1,s)}}{\partial h_{nt}} + \left(p_{nt}^{(1,s)}q_{nt}^{(1,s)} + q_{nt}^{(1,s)}\right)\left(\frac{f_{1nt}^{(s)\prime}}{f_{1nt}^{(s)}} - \frac{f_{0nt}^{(s)\prime}}{f_{0nt}^{(s)}}\right)p_{nt}^{(1,s)}\right\}$$

The parameters affecting participation are estimated using (3.8) in Proposition 1. Define $g_{2n}(\theta,p_n)$ as

(3.11)

$$g_{2n}(\theta,p_n) \equiv \sum_{s=0}^{\rho}\beta^{s}\left[u_{nt}^{(1,s)} - u_{nt}^{(0,s)} + p_{nt}^{(1,s)}q_{nt}^{(1,s)} - p_{nt}^{(0,s)}q_{nt}^{(0,s)}\right] - q_{nt}^{(0,0)}$$

The third estimation step forms orthogonality conditions from Proposition 2, and then determines $\theta_1^{(N)}$ by setting their corresponding sample moments to zero. Following Hansen and Singleton (1982), the orthogonality conditions are constructed by multiplying certain forecast errors, with instruments belonging to the associated information set. Let y_{nt} denote an observed R × 1 vector (with $2R \geq Q$) in the period t information set, write $g_n(\theta,p_n)$ for $(g_{1n}(\theta,p_n),\ g_{2n}(\theta,p_n))'$, and define the estimator $\theta_1^{(N)}$ to solve:

$$(3.12) \qquad N^{-1}A_N\sum_{n=1}^{N} y_{nt} \otimes g_n(\theta,p_n) = 0$$

where A_N is a convergent Q × 2R matrix. Appealing to Hansen (1982), $\theta_1^{(N)}$ converges to θ_0 almost surely and $N^{1/2}(\theta_1^{(N)} - \theta_0)$ converges in distribution to a normal random variable with mean 0 and covariance matrix $\Sigma_1 \equiv D^{-1}A_NWA_ND^{-1\prime}$ where:

$$(3.13) \qquad D \equiv E\left[y_{nt} \otimes \frac{\partial g_n(\theta,p_n)}{\partial\theta}\right]$$

$$W \equiv E\left[\left(y_{nt} \otimes \frac{\partial g_n(\theta,p_n)}{\partial\theta}\right)\left(y_{nt} \otimes \frac{\partial g_n(\theta,p_n)}{\partial\theta}\right)'\right]$$

Incidental Parameters

In fact, the labor supply policy function, the conditional choice probabilities, and the marginal utility of wealth, are unknown; therefore $\theta_1^{(N)}$

is not viable as an estimator. The remainder of this section addresses the issues of estimating the incidental parameters, modifying the definition of the structural parameter estimator appropriately, and finally, investigating its large sample properties. We now assume η_n is a function of the observed fixed covariates y_n:

$$(3.14) \qquad \eta_n = f(y_n)$$

One way of estimating the marginal utility of wealth is to treat η_n as a household fixed effect and form a GMM estimator from (2.9), the FOC for consumption. The main limitation of this approach is well known; most panels only contain data for a large number of households N over a relatively small number of periods T, whereas the consistency of fixed effect estimators is defined with respect to T (the panel length), not N (its cross sectional size). This shortcoming motivates the following alternative estimator for η_n which is a nonparametric extension of MaCurdy (1981, pp. 1066-69) that achieves consistency in N. Suppose a random variable ϕ_n is observed, such that:

$$(3.15) \qquad \phi_n = \eta_n + \varepsilon_n$$

$$E(\varepsilon_{nt}) = E(y_n \varepsilon_n) = 0.$$

Now consider the nonparametric regression of y_n on ϕ_n. Let $\delta_N^{(\eta)} \in (0,\infty)$ denote the bandwidth of the proposed kernel estimator, and $J_\eta(y)$ a real valued bounded symmetric differentiable function defined on $\mathbb{R}^k$ which integrates to 1. Appealing to Assumption (3.2), our estimator for η_n is defined as:

$$(3.16) \qquad \eta_n^{(N)} = \frac{\sum_{m=1, m\neq n}^{N} \phi_n J_\eta\left(\dfrac{y_m - y_n}{\delta_N^{(\eta)}}\right)}{\sum_{m=1, m\neq n}^{N} J_\eta\left(\dfrac{y_m - y_n}{\delta_N^{(\eta)}}\right)}$$

Actually our application does not assume any such ϕ_n exists. Instead it exploits the first order conditions for consumption to generate a $N^{1/2}$ consistent estimator of ϕ_n, but this approximation error has no asymptotic consequences for the estimator of η_n.

The marginal utility of wealth is not the only place where nonparametric techniques are applied. The remaining incidental parameters pose several additional complications, because they depend on the simulated variables $(x_{nt}^{(s)}, \lambda_{nt}^{(s)})$. By assumption F_1, the law of motion for x_{nt}, is known, and in empirical applications where this is untrue (such as ours), the asymptotic standard errors for the other structural parameters can be readily corrected for estimation error in F_1. However the dependence of $p_{nt}^{(s)}$ on $\lambda_{nt}^{(s)}(\theta_0)$ poses a more challenging estimation problem because (as we mentioned above) θ_0, the parameters determining F_2, the probability distribution which governs aggregate shocks, must be inferred from (unobserved) expectations people hold about the future, along with the parameters characterizing preferences. To accomplish this, we nest a kernel estimation procedure within the GMM framework, and simultaneously estimate both. For a specified value of θ the kernel estimator for $p_{nt}^{(k,s)}(\theta)$ is

$$(3.17) \qquad p_{nt}^{(k,s,N)}(\theta) =$$

$$\frac{\Sigma_{m=1,m\neq n}^{N} d_{mt} J_p[(x_{nt}^{(s)}-x_{mt})/\delta_N, (\underset{\sim}{1}_{nt}^{(k,s)}-\underset{\sim}{1}_{mt})/\delta_N, \eta_n^{(N)}\lambda_{nt}^{(s)}(\theta)-\eta_m^{(N)}]/\delta_N]}{\Sigma_{m=1,m\neq n}^{N} J_p[x_{nt}^{(s)}-x_{mt})/\delta_N, (\underset{\sim}{1}_{nt}^{(k,s)}-\underset{\sim}{1}_{mt})/\delta_N, \eta_n^{(N)}\lambda_{nt}^{(s)}(\theta)-\eta_m^{(N)})/\delta_N]}$$

Nonparametric estimates of the densities $f_{1nt}^{(s)}$ and $f_{2nt}^{(s)}$, as well as their derivatives, $f_{1nt}^{\prime(s)}$ and $f_{2nt}^{\prime(s)}$, are obtained similarly. (Appendix B contains the details.) In this manner $\rho_n(\theta)$ is formed.

Substituting the nonparametric estimates for their true values into (3.10), the modified equation system given by using $g_n(\theta,\rho_n(\theta))$ rather than $g_n(\theta,\rho_n)$ defines an operational estimator, denoted $\theta_2^{(N)}$.

<u>Proposition 2</u>

$\theta_2^{(N)}$ converges in probability to θ_0, and $N^{1/2}(\theta_2^{(N)} - \theta_0)$ is asymptotically normal.

The slow rate at which the incidental parameters converge to their true values implies $\theta_2^{(N)}$ is not necessarily $N^{1/2}$ consistent. Nevertheless the proof to Proposition 3 shows how to construct asymptotically unbiased estimators as in Hotz and Miller (1989), by taking a linear combination of estimators like $\theta_2^{(N)}$, which only differ in the bandwidth used for the incidental parameters. This section concludes with the main proposition, proved by construction using Proposition 2, which demonstrates an asymptotically normal, $N^{1/2}$ consistent estimator exists.

<u>Proposition 3</u>

Define $\theta_3^{(N)}$ by (A.20) and Σ_2 by (A.21). Then $N^{1/2}(\theta_3^{(N)} - \theta_0)$ converges in distribution to a normal with mean 0 and covariance matrix Σ_2.

4. AN EMPIRICAL APPLICATION

The Data

The data comes from the 1986 Family-Individual File of the Michigan Panel Study of Income Dynamics (PSID) (Waves I through XIX); aside from some exceptions noted in Appendix B, they refer to the period 1967 through 1985. The main advantage of working with the Family-Individual File is that it contains a separate record for each member of all households (in the survey in any given year), and consequently one can more easily track the behavior of married women as well as the behavior of unmarried women who are or were heads of their own households. Our subsample consists of women who either currently belong or have just left families which responded to the questionnaire in 1986. Appendix B describes how all the variables used in our study were constructed. It also describes in detail the selection criteria that led to the effective subsample used in our study. The characteristics of our subsample are displayed in Table I.

A Parameterization

The application we estimate assumes there is one skill factor, which represents the wage of a standardized unit of labor, and that lags of up to 3 annual periods may be important for determining current utility and wages. Thus $\nu = 1$ and $\rho = 3$. In particular

(4.1) $$u_{0nt} = (x'_{nt}, y'_n)B_0 d_{nt} + (1-d_{nt})\varepsilon_{0nt} + d_{nt}\varepsilon_{1nt}$$

$$u_{1nt} = (x'_{nt}, y'_n)B_1 l_{nt} + \Sigma_{s=0}^{3}\delta_s l_{nt} l_{n,t-s} + \alpha_1 l_{nt} l_{nt-1}^2$$

$$u_{2nt} = \exp(x_{nt}B_2 + \varepsilon_{2nt})c_{nt}^{\zeta}/\zeta$$

where ε_{0nt} and ε_{1nt} are distributed as Type 1 extreme value with location parameter 0 across (n,t), while $E(\varepsilon_{2nt}) = 0$ and $E(\varepsilon_{2nt} y_{nt}) = 0$ for the set of instruments y_{nt}. With regards the wage rate (2.4), we assume:

$$(4.2) \qquad w_{nt} = w_t \left[x_{nt} B + \Sigma_{s=1}^{4} \gamma_s l_{n,t-s} + \gamma d_{n,t-1} \right]$$

The parameter vector $B = (B^{(0)}, \ldots, B^{(8)})'$ represent coefficients on a constant 1, schooling E_n, a race dummy W_n (which takes on 1 if the person is white and 0 otherwise), locational dummies for the northeast NE_n, northcentral regions NC_n, age A_{nt}, age squared A_{nt}^2, and an age-education interaction $A_{nt} E_n$. Finally ψ_t is assumed to be multivariate lognormal with mean μ and variance σ^2.

5. EMPIRICAL RESULTS

This section reports our empirical results. First the wage equation was estimated to assess the empirical importance of learning by doing in market work. Then we estimated u_{2nt}, the marginal utility of consumption for each household n at different points in time t, and hence obtain the nonparametric estimates of η_n. The last two parts estimated the labor supply participation and Euler equations, with and without aggregate shocks.

Wages

A convenient way to obtain estimates of $(B, \gamma, \gamma_1, \ldots, \gamma_4)$, the parameters determine the wage function (2.4), and the standardized wage rates $(w_1, \ldots, w_T)$, is to first regress x_{nt} on w_{nt} for each $t \in \{1, \ldots, T\}$, and then derive a minimum distance (MD) estimator from the T ordinary least squares (OLS) parameter vector estimates. Assuming the measurement error is independently and identically distributed across (n, t) coordinant pairs running OLS on:

$$(5.1)\qquad z_{nt} \equiv (1, E_n, W_n, NE_n, NC_n, A_{nt}, A^2_{nt}, A_{nt}E_n, l_{n,t-1}, \ldots, l_{n,t-4}, d_{n,t-1})'$$

for each $t \in \{1,\ldots,T\}$ produces asymptotically normal estimators of the reduced form parameter vector

$$(5.2)\qquad \pi_t = w_t(B', \gamma_1, \ldots, \gamma_4, \gamma)'.$$

Let $(\tilde{\pi}'_1, \ldots, \tilde{\pi}_T)'$ denote the OLS estimates. The structural parameters were then found by minimizing

$$(5.3)\qquad \sum_{t=1}^{T}\left[\tilde{\pi}'_t - w_t(B', \gamma_1, \ldots, \gamma_4, \gamma)\right]\Omega^{-1}_{1t}\left[\tilde{\pi}_t - w_t(B', \gamma_1, \ldots, \gamma_4, \gamma)'\right]$$

with respect to $(B', \gamma_1, \ldots, \gamma_4, \gamma, w_1, \ldots w_T)'$, where:

$$(5.4)\qquad \Omega_{1t} = \left[\sum_{n=1}^{N_t}(w_{nt} - z_{nt}\tilde{\pi}_t)^2\right]\left[\sum_{n=1}^{N_t} z_{nt}z'_{nt}\right]^{-1}$$

Appealing to Chamberlain (1982, p.22), the resulting structural estimator is $N^{1/2}$ and asymptotically normal with covariance

$$(5.5)\qquad [\partial\pi'(\theta_1)/\partial\theta]' \cdot \Omega^{-1}_1\ [\partial\pi(\theta_1)/\partial\theta]$$

where Ω_1 is block diagonal with Ω_{1t} in the t^{th} position. Under the assumption that measurement error in wages is not correlated over time, one can show this estimator achieves the lowest covariance matrix within the GMM class; see Newey and West (1987).

The importance of both general human capital and aggregate shocks is evident from Table II, which reports our findings on the wage equation. The first 11 columns and 19 rows of each table are the OLS results for each year; the last row and column show the minimum distance estimates of the standardized wages and the parameters characterizing the wage equation. Thus, by (5.2), multiplying the s^{th} element of last column by the t^{th} element in the last row yields the restricted estimate of component s in π_t, whose unrestricted estimate is the (s,t) component of the table (viewing the estimates as a matrix). Although few of the OLS estimates are significant, most of structural coefficients are significantly different from 0. The signs of the coefficients are plausible; wages increase with education and are quadratic in age. Previous work experience raises current wages, more recent experience having the largest impact. Finally the overidentifying restrictions implied by the minimum distance estimator cannot be rejected, but the null hypothesis that standardized wages do not fluctuate over time is strongly rejected.

Marginal Utility of Wealth

A nonparametric estimator for (a linear transformation of) the marginal utility of wealth was obtained in two stages from data on household demographics and their food consumption. First we estimated the parameters characterizing preferences over consumption from the first order condition (2.9). Then the residuals obtained from these results were used as the dependent variable in a nonparametric regression on the permanent characteristics which we assume determine wealth, and consequently its marginal utility.

Following Altug and Miller (1990), the first stage is based on differencing the logged first order condition for consumption, stacking the resulting set of orthogonality conditions for each t, and forming sample moments over n. Letting Δ stand for the first difference operator, it follows from (2.9) and (4.1) that:

$$\text{(5.6)} \qquad \Delta\varepsilon_{2nt} = (1-\zeta)\Delta\ell n(c_{nt}) - \Delta x'_{nt}B_2 + \Delta\ell n(\lambda_{ct})$$

For an R dimensional vector of instruments z_{nt} assumed to satisfy the orthogonality conditions $E(\varepsilon_{nt}z_{nt}) = 0$, a GMM procedure was used to estimate the identified parameters:

$$\text{(5.7)} \qquad b \equiv \left((1-\zeta)^{-1}B_2',\ (1-\zeta)^{-1}\Delta\ell n(\lambda_{01}),\ldots,(1-\zeta)^{-1}\Delta\ell n(\lambda_{0T})\right)'$$

Respectively define the vector the T-1 dimensional vector Y_n, the Q × (T-1) matrix X_n and the square (T-1) matrix W_n as:

$$\text{(5.8)} \qquad Y_n \equiv (\Delta\ell n(c_{n1}),\ldots,\Delta\ell n(c_{n,t-1}))'$$

$$X_n \equiv \begin{pmatrix} \Delta x_{n1},\ldots,\Delta x_{n,T-1} \\ D_1,\ldots,D_{T-1} \end{pmatrix}$$

$$W_n \equiv E[(Y_n - X_n b)(Y_n - X_n b)' \,|X_n]$$

It is now straightforward to show that a GMM estimator achieving the lowest covariance within this class is:

(5.9) $$b_{(N)} = \left[N^{-1}\sum_{n=1}^{N}(X_n\tilde{W}_n^{-1}X_n)\right]^{-1}\left[N^{-1}\sum_{n=1}^{N}X_n\tilde{W}_n^{-1}Y_n\right]$$

where $\tilde{W}_n$ is any consistent estimator of W_n. The covariance of b_N is $E\left\{\left[X_nW_n^{-1}X_n'\right]\right\}^{-1}$. This application further assumed ε_{2nt} is homoskedastic; hence a consistent estimator of W_n, which is constant across n in this case, can be obtained from an average the outer of product of the redisuals from regressing Y_n on X_n. However relaxing the homoskedasticity assumption is a straightforward exercise; see Robinson (1987).

Table 3 reports our findings from the first stage. The table shows food consumption increses with family size; children consume less than adults, but over the lifecycle it is concave. All the coefficients characterizing these effects are highly significant, as are the regional dummy variables which capture the effects of transportation costs and the climatic conditions. Contingent claims prices deviate significantly from what a perfect foresight world with a constant interest rate would predict; the test statistic for the null hypothesis that $\Delta\ell n(\lambda_t) = \Delta\ell n(\lambda_{t+1})$ for $t \in \{4,5,\ldots,19\}$ is 558, yet under the null it would be distributed χ^2 with 11 d.f.

The main interest in the first stage is, of course, as input to the second. With this in mind we offer two brief remarks. First the fact that ζ, the concavity parameter measuring the degree of relative risk aversion, is unidentified, does not hinder identification of the nonseparability parameters for leisure. Consequently the estimation approach is robust to what one assumes about ζ; this is an attractive feature of the model given the difficulty researchers have in pinning it down. Second, the high level of significance achieved by the socioeconomic characteristics suggests that $\ell n(c_{nt})$ itself is not a reasonable proxy for ϕ_{nt}.

The next stage formed $\phi_{nt}^{(N)}$, and $N^{1/2}$ consistent estimator of:

$$(5.10) \qquad \phi_{nt} \equiv (\eta_n - \lambda_{01} + \varepsilon_{nt})/(1-\zeta)$$

by setting:

$$(5.11) \qquad \phi_{nt}^{(N)} \equiv x'_{nt}\left[(1-\zeta)^{-1}B_2\right]^{(N)} - \ell n(c_{nt})$$

Substituting $\phi_{nt}^{(N)}$ for ϕ_n into (3.14) then generated a consistent estimate of $(1-\zeta)^{-1}(\eta_n + \lambda_{01})$ for each $n \in \{1,\ldots,N\}$. The components of y_n, the vector used to measure proximity in wealth, consisted of race, plus a number of characteristics associated with their achievements by age 25 (including years of schooling and two locational dummies), and by age 30. (This included whether they had been married, the number of children they had given birth to, the age distribution of those children, and whether their household owned the house they lived in.) The rate of convergence of $\phi_n^{(N)}$ is less than $N^{1/2}$; therefore using b_N, rather than its limit b_0, to construct $\phi_n^{(N)}$ has no asymptotic consequences.

Nonseparable Preferences and Participation Costs

The procedures developed in Section 3 are now applied to estimate B_0, the parameter characterizing participation costs (δ_1, $\delta_1,\ldots,\delta_4$, which measures the effects of previous labor supply choices on current utility, B_1, which shows the immediate effects of the current decision, and (μ, σ), the parameters determining the price shocks of realized contingent claims.

The parameterization in (4.1) and (4.2) implies the other expressions required to apply our estimator are straightforward to calculate. From Hotz and Miller (1990), the Type 1 extreme value assumption implies:

(5.12) $q(p) = \ell n[p/(1-p)]$

The assumption of quadratic preferences implies $u_{nt}^{(k,s)} = 0$ for all $s \geq k$ and $k \in \{0,1\}$. Substituting these expressions into (3.8) and (3.9), we obtain:

(5.13)
$$g_{1nt}(\theta, p_n, \eta_n) = w_{nt}\eta_n + x_{nt}B_1 + \sum_{s=0}^{3}\delta_s 1_{n,t-s} + \alpha_1 1_{nt-1}^2$$

$$+ \sum_{s=1}^{4}\beta^s\left\{\left[p_{nt}^{(1,s)}q_{nt}^{(1,s)} + q_{nt}^{(1,s)}\right]\left(\frac{f_{1nt}^{(s)\prime}}{f_{1nt}^{(s)}} - \frac{f_{0nt}^{(s)\prime}}{f_{0nt}^{(s)}}\right)p_{nt}^{(1,s)}\right\}$$

$$g_{2nt}(\theta, p_n\eta_n) = \left[\eta_n w_{nt} + (x_{nt}', y_n')B_1\right]1_{nt} - \sum_{s=0}^{3}1_{nt}1_{n,t-s} + \sigma_\varepsilon^2 + \alpha_1 1_{nt}1_{nt-1}^2$$

$$+ \sum_{s=1}^{4}\left[p_{nt}^{(1,s)}q_{nt}^{(1,s)} - p_{nt}^{(0,s)}q_{nt}^{(0,s)}\right] - q_{nt}^{(0,0)}$$

where:

(5.14)
$$E_t[\varepsilon_{1nt}|h_{nt} = h_{nt}^*] - E_t\varepsilon_{0nt}|h_{nt} = 0]$$

$$= 2\gamma p_{nt}^{(0,0)} - \gamma - p_{nt}\ell n(p_{nt}^{(0,0)}) + (1 - p_{nt}^{(0,0)})\ell n(1 - p_{nt}^{(0,0)})$$

and γ is Euler's constant (≈0.576).

For the purposes of comparison we first estimated a model without aggregate effects. This specialization is of independent interest extending, as it does, the empirical work of Eckstein and Wolpin (1988) to environments where hours worked is a continuous variable. The absence of common shocks allows us to simplify the orthogonality conditions in two ways. First, orthogonality conditions can be formed directly from sample realizations of (2.8) without jeopardizing $N^{1/2}$ consistency. Consequently, g_{1nt}, defined in (5.13), is replaced with:

$$(5.15) \qquad g_{0nt} = (x'_{nt}, y'_{n})B_{1} + \sum_{s=0}^{3} \delta_{s} l_{n,t-s} + \delta d_{n,t-1} + \alpha_{1} l^{2}_{nt-1}$$

$$+ \sum_{s=1}^{4} \beta^{s} \delta_{s} l_{n,t+s} + \beta\delta d_{n,t+1} - \eta_{n}\left(w_{nt} + \sum_{s=1}^{4} \gamma_{s} w^{s}\right)$$

The second difference is that, as in Hotz and Miller (1989), the nonparametrically estimated conditional choice probabilities used in g_{2nt}, can be formed directly from sample without first simulating hypothetical futures for each data point.

Table 4 reports the no aggregate shock case. The instruments we used for this case included current values of children less than 6 years old, older children, age, age squared, age times education, house plus rental value plus lagged values of household income, female labor supply and real wages. It is worth noting that since g_{nt} is linear in the reduced form parameters listed in the second column, a closed form solution to the unconstrained estimator exists. As indicated by the J_{N} and d.f. statistics, the overidentifying orthogonality conditions are rejected at the 0.1 but not the 0.05 level. More troublesome is the observation that, although many of the reduced form parameters are highly significant, different estimators of the same structural parameter (such as B_{1} which appears in both the participation and Euler equations) or simple transformations of the same parameter (such as β, β^{2}, β^{3} and β^{4}) seem incompatible.

This conjecture is verified by the test statistic for the constrained estimates, which strongly rejects the no aggregate shock specification. The estimates themselves are nevertheless plausible. The estimated subjective discount factor lies between 0 and 1 but is significantly different from both these numbers, which suggests expectations over the future are being modelled in a reasonable way; young children are complimentary with nonmarket time while older children are not, a common finding in the literature on female labor supply; finally there is evidence that preferences are nonseparable, but the coefficients on successive lags switch sign. No evidence for human capital accumulation on the job is found here: although the coefficients are positive but declining with lag length (as we found in the wage equation), the null hypothesis that all 4 coefficients are 0 cannot be rejected at the .05 level.

Finally we reestimated the participation and Euler equations, now incorporating aggregate shocks into the analysis as prescribed by Section 3. An attractive computational feature of this parameterization is that given σ^2, the variance of the aggregate shock, the remaining parameter estimates have a closed form solution. Therefore the criterion function can be concentrated in all parameters bar σ^2, thus reducing the minimization algorithm to numerically searching over the positive real line.

Three main findings emerged. First, the overall specification is not rejected; the J_N test statistic is 6.7 which under null hypothesis is distributed χ^2 with 29 d.f. Second, the criterion function rises to 17.5 when σ^2 is restricted to 0, implying that its estimated value of 4.8 is highly significant. Third, all the other coefficients are insignificant even at the 10 percent level (which explains why their estimated values are not reported here); in particular we find no evidence against the hypothesis that preferences over female labor supply are additively separable over time.

Overall our empirical findings suggest that while labor market experience increases wages, the role of nonseparabilities in a female's preferences for leisure taken at different times seems limited. Restated in the language of household production functions, the latter result implies there is little investment value from extra experience in nonmarket activities beyond that acquired by full time female workers. On the other hand, a spurious reversal occurs if aggregate fluctuations are ignored; on the job training in market work seems inconsequential, while nonseparabilities in preferences over leisure become significant. This ambiguity is resolved by noting that ignoring aggregate shocks produces biased estimates; the overidentifying restrictions of the econometric framework are rejected only when aggregate shocks are ignored, and furthermore the estimated variance of the aggregate shock process is itself significant.

REFERENCES

Altug, S. and R. Miller (1990), "Household Choices in Equilibrium," Econometrica, 58, 543-570.

Becker, G. (1971), The Economics of Discrimination, 2nd ed., Chicago, University of Chicago Press.

———— (1975), Human Capital, 2nd ed., Chicago, University of Chicago Press.

Ben Porath, Y. (1967), "The Production of Human Capital and the Life Cycle of Earnings," Journal of Political Economy 75, 352-365.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," Journal of Econometrics, 18, 5-46.

———— (1984), "Panel Data" Chapter 22 in Handbook of Econometrics 2, Zvi Griliches and Michael Intriligator, eds., North Holland: Amsterdam and New York.

Corcoran, M. (1979), "Work Experience, Labor Force Withdrawals, and Women's Wages: Empirical Results Using the 1976 Panel of Income Dynamics," Chapter 13 in Women in the Labor Market, C. B. Lloyd, E. S. Andrews, and C. L. Gilroy, eds., New York: Columbia University Press, 216-245.

Corcoran, M. and G. Duncan (1979), "Work History, Labor Force Attachment, and Earnings Differences Between Races and Sexes," Journal of Human Resources 14, 3-20.

Eckstein, Z. and K. Wolpin (1989), "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience," Review of Economic Studies 56, 375-390.

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica 50, 1029-1054.

Hansen, L. and K. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," Econometrica 50, 1269-1286.

Heckman, J. and T. MaCurdy (1980), "A Life Cycle Model of Female Labor Supply,"Review of Economic Studies 47, 47-74.

———— (1982), "Corrigendum on a Life Cycle Model of Female Labour Supply," Review of Economic Studies 49, 659-660.

Hotz, V., F. Kydland and G. Sedlacek (1988) "Intertemporal Preferences and Labor Supply," Econometrica 56, 335-360.

Hotz, V. and R. Miller (1989), "Conditional Choice Probabilities and the Estimation of Dynamic Models," GSIA 88-89-10.

Killingsworth, M. (1983), Labor Supply. Cambridge: Cambridge University Press.

Lazear, E. (1979), "Male-Female Wage Differentials: Has the Government Had Any Effect," Chapter 19 in Women in the Labor Market, C. B. Lloyd, E. S. Andrews, and C. L. Gilroy, eds., New York: Columbia University Press, 331-351.

Lucas, R. (1978), "Asset Prices in an Exchange Economy," Econometrica 46, 1429- 1445.

MaCurdy, T. (1981), "An Empirical Model of Labor Supply in a Life-Cycle Setting," Journal of Political Economy 89, 1059-1085.

Miller, R. (1984), "Job Matching and Occupational Choice," Journal of Political Economy 92, 1086-1120.

Mincer, J. and H. Ofek (1982), "Interrupted Work Careers," Depreciation and Restoration of Human Capital," Journal of Human Resources 17, 3-24.

Mincer, J. and S. Polachek (1974), "Family Investments in Human Capital: Earnings of Women," Journal of Political Economy 82, S76-S108.

Mroz, T. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica 55, 765-799.

Newey, W. and K. West (1987), "Hypothesis Testing with Efficient Methods of Moments Estimation," International Economic Review, 28, 777-787.

Novales, A. (1990), "Solving Nonlinear Rational Expectations Models: A Stochastic Equilibrium Model of Interest Rates," Econometrica 58, 93-112.

Oaxaca, R. (1973), "Male-Female Wage Differentials in Urban Labor Markets," International Economic Review 14, 693-709.

Polachek, S. (1975), "Differences in expected Post-School Investment as a Determinant of Market Wage Differentials," International Economic Review 16, 451-470.

Polachek, S. (1979), "Occupational Segregation Among Women: Theory, Evidence and A Prognosis," Chapter 9 in Women in the Labor Market, C. B. Lloyd, E. S. Andrews, and G. L. Gilroy, eds., New York: Columbia University Press, 137-157.

Polachek, S. (1981), "Occupational Self-Selection: A Human Capital approach to Sex Differences in Occupational Structure," Review of Economics and Statistics, 60-69.

Powell, J., J. Stock and T. Stoker (1989), "Semiparametric Estimation of Index Coefficients," Econometrica 57, 1403-1430.

Robinson, P. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," Econometrica, 55, 875-892.

Rosen, S. (1972), "Learning and Experience in the Labor Market," Journal of Hüman Resources 7, 327-342.

Sandell, S. and D. Shapiro (1978), "The Theory of Human Capital and the Earnings of Women: A Reexamination of the Evidence," Journal of Human Resources 13, 103-17.

Shaw, K. (1989), "Life-Cycle Labor Supply with Human Capital Accumulation," International Economic Review 30, 431-456.

Weiss, Y. and R. Gronau (1981), "Expected Interruptions in Labor Force Participation and Sex-Related Differences in Earnings Growth," Review of Economic Studies XLVIII, 607-619.

Wolpin, K. (1984), "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," Journal of Political Economy 92, 852-874.

APPENDIX A

Proof of Proposition 2

To prove $\theta_2^{(N)}$ is consistent, we first note the nonparametric estimators for η_n and $p_n^{(N)}$ are uniformly consistent for $n \in \{1,\ldots\}$. Since $g_n(\theta,p)$ is continuous in (θ,p) it follows that $g_n(\theta,p_n^{(N)})$ converges to $g_n(\theta,p_n^{(N)})$ almost surely for all $\theta \in \Theta$. Noting $\theta_1^{(N)}$ is a set of first order conditions defining an optimization estimator, it follows from Theorem 2.1 of Hansen (1982, p. 1035), for example, that $\theta_1^{(N)}$ converges to θ_0 almost surely. Therefore $\theta_2^{(N)}$ does too.

We preface the proof of asymptotic normality with explicit definitions for the nonparametric estimators of p_n, which leads us to a more precise definition of $\theta_1^{(N)}$. Recall:

$$\text{(A.1)} \qquad p_n^{(N)} \equiv \left\{p_{nt}^{(0,s,N)},\ p_{nt}^{(1,s,N)},\ f_{0nt}^{(s,N)},\ f_{1nt}^{(s,N)},\ f_{0nt}^{(s,N)\prime},\ f_{1nt}^{(s,N)\prime}\right\}_{s=1}^{\rho}$$

for each $n \in N$. In turn, the $6\rho N$ components depend upon the nonparametric estimator for η_n, as well as the structural parameters θ. To facilitate the exposition of the proof, we restate our definitions of the kernel estimators for the conditional choice probabilities, as quotients (of density weighted regression function estimators and estimators of their respective probability density functions). Recall ν is the dimension of $(x_{nt}, \underset{\sim}{1}_{nt}, \lambda_t)$. Accordingly given any $(s,k) \in \{1,\ldots,\rho\} \times \{0,1\}$, let

$$\text{(A.2)} \qquad \rho_{rn}^{(N)}(\theta,\underset{\sim}{\eta}) = (N-1)^{-1}\sum_{m=1,\, m\neq n}^{N} \delta_{pN}^{-\upsilon} d_{mt}\, J\left(\frac{x_{nt}^{(s)} - x_{mt}}{\delta_{pN}},\ \frac{\underset{\sim}{1}_{nt}^{(k,s)} - \underset{\sim}{1}_{mt}}{\delta_{pN}},\ \frac{\eta_n \lambda_{nt}^{(s)}(\theta)*(1,y_n) - \eta_m(1,y_m)}{\delta_{pN}}\right)$$

for

$$0 < r = \rho k + s \leq 2\rho.$$

Also define for:

$$2\rho < r = \rho(2 + k) + s \leq 4\rho$$

the density estimator:

(A.3) $$\rho_{rn}^{(N)}(\theta, \underset{\sim}{\eta}) = (N-1)^{-1}\sum_{m=1, m\neq n}^{N} \delta_{pN}^{-\upsilon} J\left(\frac{x_{nt}^{(s)} - x_{mt}}{\delta_{pN}}, \frac{\underset{\sim}{l}_{nt}^{(k,s)} - \underset{\sim}{l}_{mt}}{\delta_{pN}}, \frac{\eta_n \lambda_{nt}^{(s)}(\theta)*(1, y_n) - \eta_m(1, y_m)}{\delta_{pN}}\right)$$

Given $\underset{\sim}{\eta}^{(N)}$ and $\theta^{(N)}$ (estimates of $\underset{\sim}{\eta}$ and θ respectively), (A.2) and (A.3) yield a kernel estimator of p_{nt} for each (k,s,n,t), namely:

(A.4) $$p_{nt}^{(k,s,N)} = \rho_{s+\rho k,n}^{(N)}\left(\theta^{(N)}, \underset{\sim}{\eta}^{(N)}\right) / \rho_{s+\rho(2+k),n}^{(N)}\left(\theta^{(N)}, \underset{\sim}{\eta}^{(N)}\right)$$

The kernel estimator of the probability density function for $f_{knt}^{(s)}$ is defined in a similar manner. Let the variable $d_n^{(s)} \in \{0,1\}$ indicate whether a person has participated in the last s periods or not. That is:

(A.5) $$d_n^{(s)} = \Pi_{r=1}^{s-1}(1 - d_{n,t-r})$$

We note the support of $f_{knt}^{(s)}$ excludes all women for whom $[1 - k(1 - d_{nt})]d_n^{(s)} = 0$. Accordingly an estimator for $f_{knt}^{(s)}$ is defined as:

$$\text{(A.6)} \qquad f_{knt}^{(s,N)}(\theta,\underset{\sim}{\eta}) = (N-1)^{-1}\sum_{m=1,m\neq n}^{N} \delta_{pN}^{-\nu}[1 - k(1 - d_{mt})]d_m^{(s)}$$

$$J\left(\frac{x_{nt}^{(s)} - x_{mt}}{\delta_{pN}}, \frac{\underset{\sim}{1}_{nt}^{(k,s)} - 1_{mt}}{\delta_{pN}}, \frac{\eta_n\lambda_{nt}^{(s)}(\theta)^*(1,y_n) - \eta_m(1,y_m)}{\delta_{pN}}\right)$$

(where the dependence on $(\theta,\underset{\sim}{\eta})$ is now made explicit) while the estimator of its derivative with respect to 1_{nt}, denoted $f_{knt}^{(s,N)\prime}(\theta,\underset{\sim}{\eta})$, is here taken as the derivative of $f_{knt}^{(s,N)}(\theta,\underset{\sim}{\eta})$ (the estimator itself). Then for notational convenience we set:

$$\text{(A.7)} \qquad \rho_{rn}^{(N)}(\theta,\underset{\sim}{\eta}) = \begin{cases} f_{knt}^{(s,N)}(\theta,\underset{\sim}{\eta}) & 4\rho < r = \rho(4 + k) + s \le 6\rho \\ \\ f_{knt}^{(s,N)\prime}(\theta,\underset{\sim}{\eta}) & 6\rho < r = \rho(6+k) + s \le 8\rho \end{cases}$$

By construction, $p_n^{(N)}$ is clearly a mapping from $\rho_n^{(N)}$, the 8ρ dimensional vector $\rho_n^{(N)} \equiv (\rho_{1n}^{(N)},\ldots,\rho_{8\rho,n}^{(N)})$. The components of $\eta_n^{(N)}$ are similarly defined. The kernel estimator for η_n, denoted $\eta_n^{(N)}$ and defined by (3.16) may be written of the quotient of $\eta_{1n}^{(N)}/\eta_{2n}^{(N)}$, where:

$$\text{(A.8)} \qquad \eta_{1n}^{(N)} = (N-1)^{-1}\sum_{m=1,m\neq n}^{N}\left(\delta_N^{(\eta)}\right)^{-\nu}\phi_m J\left(\frac{y_m - y_n}{\delta_N^{(\eta)}}\right)$$

$$\eta_{2n}^{(N)} = (N-1)^{-1}\sum_{m=1,m\neq n}^{N}\left(\delta_N^{(\eta)}\right)^{-\nu} J\left(\frac{y_m - y_n}{\delta_N^{(\eta)}}\right)$$

Let $\underset{\sim}{\eta}_n^{(N)} \equiv \left(\eta_{1n}^{(N)}, \eta_{2n}^{(N)}\right)$ and define $\underset{\sim}{\eta}_n$ as its pointwise limit. The moment conditions (3.10) and (3.11) can now be restated in terms of $(\theta, \underset{\sim}{\eta}_n, \rho_n(\theta, \eta_n))$ rather than (θ, η_n, p_n) by defining:

$$\text{(A.9)} \qquad h_{1n}\left(\theta, \underset{\sim}{\eta}_n^{(N)}, \rho_n^{(N)}\left(\theta, \underset{\sim}{\eta}_n^{(N)}\right)\right) = g_{1n}\left(\theta, \eta_n^{(N)}, p_n(\theta, \eta_n^{(N)})\right)$$

$$h_{2n}\left(\theta, \underset{\sim}{\eta}_n^{(N)}, \rho_n^{(N)}\left(\theta, \underset{\sim}{\eta}_n^{(N)}\right)\right) = g_{2n}\left(\theta, \eta_n^{(N)}, p_n(\theta, \eta_n^{(N)})\right)$$

We set $h_n \equiv (h_{1n}, h_{2n})'$ and define the estimator $\theta_2^{(N)}$ for θ_0 by restating (3.12) so that:

$$\text{(A.10)} \qquad N^{-1} A_N \sum_{n=1}^{N} y_{nt} \otimes h_n\left(\theta, \underset{\sim}{\eta}_n^{(N)}, \rho_n^{(N)}, (\theta, \underset{\sim}{\eta}_n^{(N)})\right) = 0$$

To prove $\theta_2^{(N)}$ is asymptotically normal, we first investigate another (hypothetical) estimator, denoted $\theta^{(N)}$, and show $N^{1/2}\theta^{(N)}$ asymptotically normal. Then we establish the difference between $\theta^{(N)}$ and $\theta_2^{(N)}$ is $o_p(N^{-1/2})$. Hence, by result (x)(d) in Rao (1973, p. 122) $N^{1/2}\theta_2^{(N)}$ also converges to a normal random variable.

Consider the class of hypothetical estimators $\theta^{(N)}$ defined by the equations:

(A.11) $$- N^{-1/2}A\sum_{n=1}^{N} y_{nt} \otimes h_n\left(\theta_0, \underset{\sim}{\eta}_n, \rho_n(\theta_0, \eta_n)\right)$$

$$= N^{-1/2}A\sum_{n=1}^{N} y_{nt} \otimes \left\{\left(\frac{\partial h_n}{\partial \theta} + \frac{\partial h_n}{\partial \rho_n}\frac{\partial \rho_n}{\partial \theta}\right)(\theta^{(N)} - \theta_0) + \left(\frac{\partial h_n}{\partial \underset{\sim}{\eta}_n} + \frac{\partial h_n}{\partial \rho_n}\frac{\partial \rho_n}{\partial \eta_n}\right)\left(\underset{\sim}{\eta}_n^{(N)} - \underset{\sim}{\eta}_n\right) + \frac{\partial h_n}{\partial \rho_n}(\rho_n^{(N)} - \rho_n)\right\}$$

where all the derivatives are evaluated at their true values $(\theta_0, \underset{\sim}{\eta}_n, \rho_n(\theta_0, \eta_n))$. By a central limit theorem, the left side of (A.11) is asymptotically normal. Assuming

$$AN^{-1}\sum_{n=1}^{N} y_{nt} \otimes \left(\frac{\partial h_n}{\partial \theta} + \frac{\partial h_n}{\partial \rho_n}\frac{\partial \rho_n}{\partial \theta}\right)$$

is invertible, the asymptotic normality of $N^{1/2}(\theta^{(N)} - \theta_0)$ follows by showing

(A.12) $$AN^{-1/2}\sum_{n=1}^{N} y_{nt} \otimes \left(\frac{\partial h_n}{\partial \underset{\sim}{\eta}_n} + \frac{\partial h_n}{\partial \rho_n}\frac{\partial \rho_n}{\partial \eta_n}\right)\left(\underset{\sim}{\eta}_n^{(N)} - \underset{\sim}{\eta}_n\right)$$

and

(A.13) $$AN^{-1/2}\sum_{n=1}^{N} y_{nt} \otimes \frac{\partial h_n}{\partial \rho_n}\left(\rho_n^{(N)} - \rho_n\right)$$

are both asymptotically normal too. The components of (A.12) and (A.13) can be treated the same way. Expanding the first components in (A.13), for example, we obtain

(A.14) $$N^{-1/2}\sum_{n=1}^{N} y_{nt} \otimes \frac{\partial h_n}{\partial \rho_{1n}}\left(\rho_{1n}^{(N)} - \rho_{1n}\right)$$

$$= N^{-1/2}(N-1)^{-1}\sum_{n=1}^{N}\sum_{\substack{m=1\\ m\neq n}}^{N} y_{nt} \otimes \frac{\partial h_n}{\partial \rho_{1n}}\left[\delta_{pN}^{-0} d_{mt} J\left(\frac{x_{nt}^{(1)} - x_{mt}}{\delta_p^{(N)}},\ \frac{\underset{\sim}{1}_{nt}^{(0,s)} - \underset{\sim}{1}_{mt}}{\delta_p^{(N)}},\right.\right.$$

$$\left.\left.\frac{\eta_n \lambda_{nt}^{(1)}(\theta_0)^*(1,y_n) - \eta_m(1,y_m)}{\delta_p^{(N)}}\right)\right]$$

$$= N^{-1/2}(N-1)^{-1}\sum_{n=1}^{N}\sum_{m=n+1}^{N-1} \nu_{mn}^{(N)}$$

where the symmetric kernel $\nu_{mn}^{(N)}$ is defined as:

$$\nu_{mn}^{(N)} \equiv A_N \delta_N^{-k} \sum_{s=1}^{\rho}\left[y_{nt} \otimes \frac{\partial g_n}{\partial \tilde{\rho}_{ns}} d_m J\left((x_m - x_n^{(s)})/\delta_p^{(N)}\right) + z_m \otimes \frac{\partial g_n}{\partial \tilde{\rho}_{ns}} d_m J\left((x_m - x_n^{(s)})/\delta_p^{(N)}\right)\right]$$

$x_m \equiv (x_{mt},\ \underset{\sim}{1}_{mt},\ \eta_m y_m)$. Let $\|\nu_{mn}^{(N)}\|^2 \equiv \nu_{mn}^{(N)\prime}\nu_{mn}^{(N)}$, and suppose

(A.15) $$E\left[\|\nu_{mn}^{(N)}\|^2\right] = o(N)$$

Then by Lemma 3.1 of Powell, Stock and Stoker (1989, p. 1410)

$$N^{1/2}\left[N^{-1}(N-1)^{-1}\sum_{n=1}^{N}\sum_{m=n+1}^{N}\nu_{mn}^{(N)}\right]$$

$$= N^{1/2}\left[N^{-1}\sum_{n=1}^{N}E(\nu_{mn}|n) - E(\nu_{mn})\right] + o_p(1)$$

Because the second line converges to a normal random variable, it follows that (A.14) is asymptotically normal if (A.15) is true. But:

$$E\left[\|\nu_{mn}^{(N)}\|^2\right]$$

$$= O(1)\delta_N^{-2k}E\left\{\left\|\sum_{s=1}^{\rho}\left[z_n \otimes \frac{\partial g_n}{\partial\tilde{\rho}_{ns}} p_m J_N(x_m - x_n^{(s)}) + z_m \otimes \frac{\partial g_m}{\partial\tilde{\rho}_{ms}} p_n J_N(x_n - x_m^{(s)})\right]\right\|^2\right\}$$

$$= O(1)\delta_N^{-2k}E\left\{\left\|\sum_{s=1}^{\rho}\left[J_N(x_m - x_n^{(s)}) + J_N(x_n - x_m^{(s)})\right]\right\|^2\right\}$$

$$= O(1)\delta_N^{-2k}E\left\{\|J_N(x_m - x_n^{(s)})\|^2\right\}$$

$$= O(1)\delta_N^{-2k} \int\|J(u)\|^2 \, dF(u)$$

$$= o(N)$$

The second line follows from the fact that α_N is $O_p(1)$ (converging to a matrix of full rank); the second line uses the fact that z_n and $\partial g_n/\partial\tilde{\rho}_{ns}$ are both $O_p(1)$; the third line repeatedly exploits the Cauchy Schwartz inequality, and uses the fact that x_m is independent of $x_n^{(s)}$ for all (m,n,s); the fourth line undertakes the change in variables $u = (x_m - x_n)/\delta_N$ for x_m; finally δ_N^{-k} is $o(N)$ by assumption.

Repeating the same argument for the other components of (A.12) and (A.13) the asymptotic normality of $\theta^{(N)}$ follows.

The last step of the proof shows:

$$o_p(N^{-1/2}) = \theta^{(N)} - \theta_2^{(N)}$$

Taking a Taylor expansion about the defining equation for $\theta_2^{(N)}$:

$$\text{(A.15)} \qquad - N^{-1/2} A_N \sum_{n=1}^{N} y_{nt} \otimes h_n\left(\theta_0, \eta_n, \rho_n(\theta_0, \eta_n)\right)$$

$$= N^{-1/2} A_N \sum_{n=1}^{N} y_{nt} \otimes \left(\frac{\partial \tilde{h}_n}{\partial \theta} + \frac{\partial \tilde{h}_n}{\partial \rho_n} \frac{\partial \tilde{\rho}_n}{\partial \theta}\right) \left(\theta_1^{(N)} - \theta_0\right)$$

$$+ N^{-1/2} A_N \sum_{n=1}^{N} y_{nt} \otimes \left(\frac{\partial \tilde{h}_n}{\partial \eta_n} + \frac{\partial \tilde{h}_n}{\partial \rho_n} \frac{\partial \rho_n}{\partial \eta_n}\right) \left(\eta_n^{(N)} - \tilde{\eta}_n\right)$$

$$+ N^{-1/2} A_N \sum_{n=1}^{N} y_{nt} \otimes \frac{\partial \tilde{h}_n}{\partial \rho_n} \left(\rho_n^{(N)} - \rho_n\right)$$

where each of the derivatives are evaluated at points in (arbitrarily small uniform) neighborhoods of the two values. (This is signified by a tilde superscript.) Considering each of the expressions above, we see that:

$$(A.16)\qquad N^{-1/2}A_N\sum_{n=1}^{N} y_{nt} \otimes h_n\big(\theta_0,\eta_n,\rho_n(\theta_0,\eta_n)\big)$$

$$= N^{-1/2}A\sum_{n=1}^{N} y_{nt} \otimes h_n\big(\theta_0,\eta_n,\rho_n(\theta_0,\eta_n)\big) + (A_N - A)O_p(1)$$

$$= N^{-1/2}A\sum_{n=1}^{N} y_{nt} \otimes h_n\big(\theta_0,\eta_n,\rho_n(\theta_0,\eta_n)\big) + o_p(1)$$

The second line follows from Lemma 4.5 of White (1984, p. 63) and the third from the convergence of A_N to A. Turning to the last expression on the right side of (A.15):

$$(A.17)\qquad N^{-1/2}A_N\sum_{n=1}^{N} \tilde{z}_{nt} \otimes \frac{\partial\tilde{h}_n}{\partial\rho_n}\big(\rho_n^{(N)} - \rho_n\big)$$

$$= N^{-1/2}A\sum_{n=1}^{N} \tilde{z}_{nt} \otimes \frac{\partial h_n}{\partial\rho_n}\big(\rho_n^{(N)} - \rho_n\big)$$

$$+ N^{-1/2}\sum_{n=1}^{N}\left(A_N\tilde{z}_{nt} \otimes \frac{\partial\tilde{h}}{\partial\rho_n} - A\tilde{z}_{nt} \otimes \frac{\partial h}{\partial\rho_n}\right)\big(\rho_n^{(N)} - \rho_n\big)$$

$$= N^{-1/2}A\sum_{n=1}^{N} z_{nt} \otimes \frac{\partial h_n}{\partial\rho_n}\big(\rho_n^{(N)} - \rho_n\big) + O_p(N^{-1/2})$$

Using a similar argument:

$$\text{(A.18)} \qquad N^{-1/2}A_N\sum_{n=1}^{N}\tilde{z}_{nt} \otimes \left(\frac{\partial\tilde{h}_{nt}}{\partial\eta_n} + \frac{\partial\tilde{h}_n}{\partial\rho_n}\frac{\partial\rho_n}{\partial\eta_n}\right)\left(\eta_n^{(N)} - \eta_n\right) + o_p(1)$$

$$= N^{-1/2}A\sum_{n=1}^{N}\tilde{z}_{nt} \otimes \left(\frac{\partial h_{nt}}{\partial\eta_n} + \frac{\partial h_{nt}}{\partial\rho_n}\frac{\partial\rho_n}{\partial\eta_n}\right)\left(\eta_n^{(N)} - \eta_n\right)$$

Substituting (A.11), (A.17) and (A.18) into (A.15), it immediately follows that:

$$\text{(A.19)} \qquad N^{-1/2}A\sum_{n=1}^{N}\tilde{z}_{nt} \otimes \left(\frac{\partial h_n}{\partial\theta} + \frac{\partial h_n}{\partial\rho_n}\frac{\partial\rho_n}{\partial\theta}\right)\left(\theta^{(N)} - \theta_2^{(N)}\right) = o_p(N^{-1/2})$$

Multiplying both sides of (A.30) by the inverse of

$$AN^{-1}\sum_{n=1}^{N}\tilde{z}_{nt} \otimes \left(\frac{\partial h_n}{\partial\theta} + \frac{\partial h_n}{\partial\rho_n}\frac{\partial\rho_n}{\partial\theta}\right)$$

(which exists with arbitrarily high probability for sufficiently large N) the desired result is obtained. ‖

<u>Proposition 3</u>

Define a fixed number of estimators $\theta_{2g}^{(N)}$ for $g \in \{1,\ldots,G\}$, which differ by the kernel bandwidth used in the nonparametric incidental parameter estimation. Then $\theta_3^{(N)}$ is defined as a certain linear combination of the original G estimators. By Proposition 2 all G + 1 estimators are consistent and, by the arguments given above, $N^{1/2}(\theta_{2g}^{(N)} - \theta_0)$ is asymptotically normal for each $g \in \{1,\ldots,G\}$. This implies $N^{1/2}(\theta_3^{(N)} - \theta_0)$ is also asymptotically normal, because it is a linear combination of asymptotically normal random variables. Therefore we only have to demonstrate $N^{1/2}(\theta_3^{(N)} - \theta_0)$ is asymptotically unbiased, and exhibit its covariance matrix. Accordingly, consider the following linear combination of estimators, defining $\theta_3^{(N)}$ as

$$\text{(A.20)} \qquad \theta_3^{(N)} = \left(\theta_{2G}^{(N)} - \sum_{g=1}^{G-1} c_g \theta_{2g}^{(N)}\right) \Big/ \left(1 - \sum_{g=1}^{G-1} c_g\right)$$

where:

(i) $G = (k+4)/2$ if k is an even number and $G = (k+3/2$, for k is an odd number; k is the cardinality of the arguments in the nonparametric estimators.

(ii) $\theta_{2g}^{(N)}$ for $g = 1,\ldots,G-1$, is an estimator formed in the manner described above, where the bandwidth used in the kernel function is $h_{Ng} = \psi_g h_{NG}$, where $\psi_1,\ldots,\psi_{G-1}$ are distinct but otherwise arbitrary positive constants; and

(iii) $c_1,\ldots,c_{G-1}$ are a set of weights given by:

$$\begin{bmatrix} c_1 \\ \vdots \\ c_{G-1} \end{bmatrix} = \begin{bmatrix} \psi_1 & ,\ldots, & \psi_{G-1} \\ \vdots & & \vdots \\ \psi_1^{G-1} & ,\ldots, & \psi_{G-1}^{G-1} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

It now follows directly from Hotz and Miller (1990, Appendix B) that $\theta_3^{(N)}$ is $N^{1/2}$ consistent, and the random variable $N^{1/2}(\theta_3^{(N)} - \theta_0)$ is asymptotically normal, with mean and covariance matrix $\Sigma_2^{(N)}$, defined

$$\text{(A.21)} \qquad \Sigma_2^{(N)} = \Gamma^{-1\prime} E\left[(\pi_n + \nu_n)(\pi_n + \nu_n)'\right]\Gamma^{-1}$$

where π_n is the vector of orthogonality conditions for n, ν_n is the nonparametric correction for the incidental parameters.

APPENDIX B

Our subsample is selected as follows. There are 25,236 individuals included in the nineteen-year Family-Individual Respondents File of the PSID. According to the definition used in the PSID, an individual is denoted a *main family nonresponse* in a given year if both the individual and his or her family are lost to the study in that year. Alternatively, an individual may be a *mover-out nonresponse* is he or she has left a family that is still included in the study in a given year. The individual may subsequently become response if he or she moves into a panel family or becomes a splitoff by forming a new panel family or household. Mover-out nonrespondents have some nonzero individual data in the year that they became nonresponse because they were part of a panel family in the year preceding the one when they became nonresponse. The nineteen-year Family-Individual Respondents File contains data on individuals (and families) that were respondents as of the 1986 interviewing year as well as individuals who became mover-out nonrespondents in that year. In our selection, we did not distinguish between respondents and individuals who had become mover-out nonrespondents during a given year.

We initially selected a sample of women who were in either of one of the above categories as of 1986 by setting the individual-level variables "Relationship to Head" to head or wife, "Sex of the Individual" to female and the "Why Nonresponse" variable to the zero category, which denotes individuals who were still a member of a panel family. Since individuals who had become nonrespondents as of 1986 either because they and their families were lost to the study or they were mover-out nonrespondents in years prior to the 1986 interviewing year are not included in the nineteen Family-Individual Respondents File, the number of individuals included in our subsample increases with time.

Based on this initial selection, the total number of women in each year for the years 1967 to 1985 is 2474, 2592, 2761, 2912, 3079, 3260, 3445, 3619, 3815, 3973, 4130, 4363, 4597, 4793, 4987, 5153, 5358, 5652 and 5900, respectively. However, our effective sample was reduced further due to the existence of missing data or inconsistent observations. The selection of our effective sample can be motivated by the nature of the variables available in the PSID.

Our measures of annual hours and average hourly earnings are identical to the PSID variables of the same names. In the PSID data-tapes, average hourly earnings for both husbands and wives are defined from the ratio of total labor income to total annual hours of work. We encountered cases (due to reporting or coding errors) for which annual hours were positive but average hourly earning zero or vice versa. There is also an issue about the way average hourly earnings was coded in 1968 versus the remaining survey years. The number of person-years lost due to this coding error was 980. In 1968, 9's were coded instead of 0's when the head or wife did not work for money and therefore had no hourly earnings. In the remaining years, average hourly earnings above 99.99 dollars were coded as 99.99 dollars. The number of person-years lost due to this criterion was 40.

We obtained our measure of food expenditures for a given year 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 that year. We then measured consumption expenditures for year t by taking 0.25 of the value of this variable for year $t - 1$ and 0.75 of its value for 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 food 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 in the 1968 survey year is 999 dollars 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 dollars. We lost 452 person-years due to the truncation of the different consumption variables.

Our empirical study also uses variables describing various demographic characteristics of the women in our sample. First, we obtained the age of each woman from the individual variables located in the latter part of the data records of the Family-Individual File. For this variable, a value of 99 indicates missing data. We lost 74 person-years due to missing values in the age variable.

There are no separate individual variables describing the race of the individual or the region where they are currently residing. Hence, variables from the family portion of the data record must be used for this purpose. We defined the region variable to be the geographical region 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 1-4 correspond to the regions Northeast, North Central, South, West. For 1970 and 1971, the values 5 and 6 denote the regions Alaska and Hawaii and foreign country, respectively. After 1971, value of 9 indicates missing data but no person-years were lost due to missing data for this variable.

Third, we used the family variable "Race of the Household head" to measure the race variable in our study. There is a family variable that records information about the race of the wife but this variable was included in the PSID only for the interviewing years 1985 and 1986. Defining the race variable in our empirical study as the race of the household head should not create much measurement error because the women in our subsample are either household heads themselves or wives of such heads. For the interviewing years

1968-1970, the values of 1 to 3 denote white, black, Puerto Rican or Mexican, respectively 7 denotes other (including Oriental, Philippino) and 9 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, this variable was coded such that values of 1 -4 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 and a value of 9 denotes missing data. We lost 200 person-years due to missing data in this variable.

We also used the family variables that indicate the educational attainment level of the household head or wife to measure the education variable. We did this because the variable "Completed Education" recorded in the individual part of the data record does not apply if the individual is a household head or wife. However, one difficulty in using the family level education variables is that if the individual was a wife of a PSID household head for the interviewing years 1969, 1970 or 1971, there is no information about her education attainment level because questions regarding the wife's completed education level were not asked for those years. A second difficulty is that the variables denoting the head's and wife's completed education level are not strictly comparable across the different waves of the PSID. Since 1975, information pertaining to advanced (graduate or professional) degrees as well as that pertaining to additional nonacademic training have been coded for this variable. Another noncomparability problem is that the question regarding difficulty in reading or writing was omitted from the coding of this variable after 1984. For both the head and wife, the coding of this variable is as follows: 1: 0-5 grades, 2: 6-8 grades, 3: 9-11 grades, 4: 12 grades, and no further training 5: 12 grades plus nonacademic training, 6: College but no degree, 7: College BA but no advanced degree and 8: College and advanced or professional degree. For both the head's and wife's education variable, a value of 9 denotes missing data. Our effective subsample reflects a loss of 2282 person-years due to missing data for the education variable.

The marital status of a woman in our subsample was determined from the marital status of the head. This variable was coded differently for the interviewing year 1968, on the one hand, and the remaining years on the other. For 1968, the values 1 through 5 denote the categories married, single, widowed, divorced and separated, respectively, 8 denotes married but spouse absent and 9 missing data. After 1968, the sixth category is dropped.

The number of individuals in a 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 other years, missing data were assigned. The second variable, which indicates the total number of children under 18 in the family regardless of their relationship to the head, was truncated above at the value of 9 for the interviewing years 1968 to 1971. After 1975, this variable denotes the actual number of children within the family unit.

We constructed some additional variables that were used as instruments.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 the variable of the above variables showing the value of rent paid and rental value of free housing for a household. Finally, household income was measured from the PSID variable total family money income, which included taxable income of head and wife, total transfers of head and wife, taxable income of others in the family units, and their total transfer payments.

The issue of truncation from above also arises for the variables used to construct measures of the above variables. However, we did not eliminate any observations or person-years due to the existence of such upper limits because the fact that some of the variables used as instruments were truncated from above for certain years does not invalidate the use of these instruments.

We used two different deflators to convert such nominal quantities as average hourly earnings, household income, etc. to real. First, we defined the (spot) price of food consumption to be the numeraire good at t in the theoretical framework of Section 2. We accordingly measured real food consumption expenditures and real wages as the ratio of the nominal consumption expenditures and wages and the annual implicit price deflator for food consumption expenditures published in Table 7.12 of the National Income and Product Accounts. (See the U.S. Department of Commerce, Bureau of Economic Analysis publication *Business Statistics 1986*, a supplement to the *Survey of Current Business*. On the other hand, we deflated variables such as the nominal value of home ownership or nominal family income by the implicit price deflator for total personal consumption expenditures.

We also constructed variables that show the age distribution of children within the family. For the interviewing years 1975 to 1986, we were able to obtain the number of children in the family between the ages of 1-2, 3-5, and 6-13 from family-level variables which show the total number of children in these age groups who were currently in the family unit. For the years 1968 to 1974, we constructed a series showing the number of children in the age categories less than 1, 1-2, 3-5, and 6-13 years by using the birth dates of the (eight) children raised by the wife. Since this variable is recorded in 1976, it allows us to go back to 1968. Finally, for the years 1975 to 1986, we interpreted an increase in the number of children in the family unit across two consecutive years as a birth or equivalently, an increase in the number of, of children in the family unit across two consecutive years as a birth or equivalently, an increase in the number of children less than one year old.

TABLE I

	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
(1) Number of Individuals in Effective Sample	2439	2549	2704	2840	2982	3127	3288	3449	3628	3778	3922	4135	4347	4506	4682	4835	5014	5235	5444
(2) Demographic Characteristics																			
(i) Average Age	39.2	39.4	39.1	39.3	39.3	39.4	39.4	39.5	39.6	39.9	40.1	40.1	40.1	40.2	40.5	40.9	41.2	41.3	41.6
(ii) Average Number of Children	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.3	1.2	1.2	1.2	1.1	1.1	1.1	1.1	1.0
(iii) Average Number in Family	4.2	4.1	3.9	3.8	3.8	3.6	3.5	3.4	3.4	3.3	3.3	3.2	3.1	3.1	3.0	3.0	3.0	3.0	3.0
(iv) Proportion Married	0.74	0.74	0.72	0.72	0.72	0.72	0.71	0.70	0.70	0.69	0.69	0.68	0.68	0.66	0.66	0.65	0.65	0.65	0.64
(v) Proportion White	0.66	0.66	0.65	0.65	0.65	0.65	0.65	0.64	0.64	0.64	0.64	0.63	0.63	0.662	0.662	0.62	0.62	0.63	0.63
(vi) Proportion Black	0.33	0.32	0.33	0.33	0.32	0.33	0.33	0.33	0.33	0.33	0.34	0.34	0.35	0.35	0.35	0.35	0.35	0.35	0.35
(vii) Average Education Level	3.8	---	---	---	4.0	4.0	4.0	4.1	4.1	4.1	4.2	4.2	4.3	4.3	4.3	4.4	4.4	4.5	4.5
(viii) Proportion from Northeast	0.17	0.17	0.17	0.17	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.15	0.15	0.15	0.15
(ix) Proportion from Northcentral	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.24	0.24	0.24	0.24	0.24
(x) Proportion from West	0.42	0.42	0.41	0.42	0.42	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.44	0.44	0.44	0.44	0.45	0.45	0.45
(3) Hours, Earnings and Participation																			
(i) Proportion with Positive Annual Hours	0.55	0.58	0.62	0.62	0.61	0.59	0.60	0.61	0.61	0.61	0.62	0.64	0.65	0.65	0.64	0.64	0.64	0.67	0.66
(ii) Average Annual Hours for Workers	1313	1323	1303	1316	1334	1357	1375	1361	1349	1377	1421	1444	1428	1450	1465	1463	1473	1526	1561
(iii) Average Hourly Earnings for Workers	1.72	5.52	5.90	6.02	6.30	6.40	6.66	6.61	6.23	6.39	6.64	6.65	6.83	6.57	6.64	6.45	6.78	6.58	6.86
(4) Food Consumption, Income and Housing																			
(i) Average Annual Value of Food Consumption at Home	1558.8	3994.4	4021.6	3913.8	3641.0	----	4341.1	4037.5	3682.8	3340.0	3421.3	3552.3	3369.4	3259.4	3093.0	2908.5	2880.9	2947.5	2796.3
(ii) Average Annual Value of Amount Eating Out	----	756.7	781.4	778.1	767.0	----	940.0	718.7	715.8	762.2	812.8	881.7	822.3	787.4	750.2	730.0	800.0	841.0	817.5
(iii) Average Annual Value of Food Stamps	178.0	959.3	1608.1	1649.6	1494.4	----	1662.1	1315.6	1196.5	1140.0	1157.7	1102.5	1227.0	1175.2	1242.0	1147.0	1246.0	1344.0	1253.5
(iv) Average Annual Income	20270	21863	22519	22786	23235	23814	24495	23986	23230	24206	24530	25045	25198	25173	24327	23920	24307	25210	25538
(v) Average House Value of Homeowners	44970	42837	44447	44519	45494	47576	48748	48630	49446	52983	56775	61675	63367	62125	59388	56935	57715	58737	60046
(vi) Average Annual Rent of Renters	2224	2257	2323	2438	2404	2485	2482	2426	2438	2496	2475	2527	2516	2492	2517	2535	2609	2710	2791
(vii) Average Annual Value of Free Housing	1644	1432	1388	1520	1637	1714	1965	1769	1986	2178	2337	2656	2692	2545	2359	2411	2613	2672	2531
(viii) Proportion of Homeowners	0.53	0.54	0.54	0.55	0.56	0.57	0.58	0.57	0.57	0.58	0.59	0.59	0.59	0.58	0.57	0.57	0.57	0.57	0.57
(ix) Proportion of Renters	0.42	0.42	0.42	0.41	0.38	0.38	0.38	0.38	0.38	0.37	0.37	0.37	0.37	0.37	0.38	0.38	0.38	0.38	0.38
(x) Proportion Receiving Food	0.12	0.07	0.07	0.12	0.12	----	0.13	0.15	0.16	0.15	0.14	0.14	0.17	0.17	0.16	0.17	0.16	0.14	0.14

TABLE 11

Estimates of the Wage Equation

Variable	t=9	t=10	t=11	t=12	t=13	t=14	t=15	t=16	t=17	t=18	t=19	Estimates
1) Constant	-7,5090	-12,3505	-6,3007	-6,0809	-8,1243	-6,2800	-6,4236	-3,8097	-5,4222	-6,2684	-9,6467	-6,3103
	(2,3265)	(3,3411)	(2,5849)	(2,3556)	(3,0011)	(1,9876)	(1,6416)	(1,5046)	(2,3944)	(1,8985)	(2,2330)	(1,0033)
2) Education	1,2490	2,234	1,004	0,9862	1,4454	1,2000	0,9562	0,7756	1,1682	1,1170	1,1588	1,0700
	(0,3611)	(0,500)	(0,4080)	(0,3721)	(0,4826)	(0,3226)	(0,2782)	(0,2502)	(0,3772)	(0,2856)	(0,3322)	(0,1605)
3) Race Dummy: White	3,5815	3,3008	-0,8431	-2,7543	1,0812	-0,2162	-0,2700	-1,5978	-1,2952	-1,9830	1,6305	-0,3858
	(1,8900)	(2,7984)	(2,2264)	(2,0135)	(2,5774)	(1,6691)	(1,3976)	(1,3037)	(2,0647)	(1,7678)	(1,9545)	(0,8394)
Region Dummies:												
4) Northeast	0,2377	0,8851	0,9850	0,5257	0,3080	1,4300	0,0646	0,2354	0,4106	0,5200	0,8703	0,3384
	(0,3507)	(0,4649)	(0,4031)	(0,3986)	(0,5034)	(0,3109)	(0,2823)	(0,2710)	(0,3579)	(0,2925)	(0,3304)	(0,1596)
5) Northcentral	0,7049	0,5784	-0,3693	0,6550	-0,0041	0,0567	0,0018	-0,3584	-0,05700	-0,3259	0,3573	0,0387
	(0,3067)	(0,4275)	(0,3614)	(0,3448)	(0,4149)	(0,2718)	(0,2311)	(0,2316)	(0,2991)	(0,2630)	(0,2815)	(0,1360)
6) Age	1,1878	2,5511	1,6371	2,0092	1,9372	1,4562	1,6242	1,1028	1,2880	2,0955	2,1885	1,6214
	(0,6450)	(0,9104)	(0,7220)	(0,6789)	(0,8476)	(0,5369)	(0,4321)	(0,4127)	(0,5680)	(0,4760)	(0,5389)	(0,2645)
7) Age ** 2	-0,0643	-0,1682	-0,1089	-0,1710	-0,1440	-0,1208	-0,1222	-0,0844	-0,0848	-0,1907	-0,1580	-0,1267
	(0,0564)	(0,0785)	(0,0628)	(0,0596)	(0,0741)	(0,0470)	(0,0380)	(0,0352)	(0,0470)	(0,0387)	(0,0441)	(0,0225)
8) Age * Education	-0,0056	-0,0153	-0,0137	-0,0166	-0,0103	-0,0124	-0,0122	-0,0094	-0,0131	-0,0113	-0,0073	-0,0106
	(0,0054)	(0,0073)	(0,0059)	(0,0058)	(0,0070)	(0,0044)	(0,0038)	(0,0036)	(0,0047)	(0,0041)	(0,0043)	(0,0221)
Lagged Hours:												
9) Hours (-1)	3,1885	2,4722	1,6041	1,5950	1,8663	2,0708	3,1700	2,5272	2,4018	2,2242	2,0489	2,3486
	(-0,5291)	(0,8397)	(0,6141)	(0,6327)	(0,7371)	(0,4289)	(0,4259)	(0,3987)	(0,5646)	(0,4433)	(0,5033)	(0,2605)
10) Hours (-2)	-0,1242	1,3841	0,7964	1,3624	2,0100	1,2788	-0,0500	1,1707	0,9361	0,6138	1,1033	0,8104
	(0,5446)	(0,8328)	(0,6280)	(0,6955)	(0,7949)	(0,4826)	(0,4821)	(0,4165)	(0,6417)	(0,5153)	(0,5511)	(0,2610)
11) Hours (-3)	-0,0382	-0,7411	0,8310	-1,1095	-1,1790	0,1864	0,2968	-0,5412	0,1387	0,5552	0,9598	0,0485
	(0,5939)	(0,7728)	(0,6625)	(0,6635)	(0,8455)	(0,5076)	(0,4468)	(0,4191)	(0,5775)	(0,5072)	(0,5625)	(0,2580)
12) Hours (-4)	0,4362	-0,0375	0,2725	1,3206	0,6749	-0,9546	0,8801	1,0768	1,2378	-0,1412	0,0040	0,4408
	(0,5411)	(0,7380)	(0,6506)	(0,6481)	(0,7774)	(0,5214)	(0,4545)	(0,4260)	(0,5538)	(0,5100)	(0,5717)	(0,2558)
13) Hours (-5)	-0,1841	-0,0791	-0,8444	1,9176	0,1360	1,4044	-0,2336	0,1213	-0,2702	0,4519	0,0258	0,2097
	(0,5958)	(0,7280)	(0,6133)	(0,6428)	(0,7505)	(0,5021)	(0,4576)	(0,4288)	(0,5754)	(0,4953)	(0,5570)	(0,2542)
14) Hours (-6)	1,3025	0,3150	0,4966	-0,9786	-0,6985	0,6479	1,6250	-0,5535	0,8612	1,2655	-0,2412	0,4514
	(0,5763)	(0,8068)	(0,6187)	(0,6172)	(0,7052)	(0,5232)	(0,4505)	(0,4672)	(0,5969)	(0,5187)	(0,5611)	(0,2599)
15) Hours (-7)	-0,2903	-0,2405	-0,9462	0,2608	0,9128	-0,7398	-0,2116	0,9506	-1,4192	-1,2318	0,7209	-0,1471
	(0,6426)	(0,8694)	(0,7401)	(0,6385)	(0,6862)	(0,4731)	(0,4510)	(0,4518)	(0,5816)	(0,4877)	(0,5444)	(0,2575)
16) Hours (-8)	-0,6944	0,0017	1,2150	-0,4600	0,2140	0,4524	-0,5448	-0,1229	1,1831	0,2035	0,1427	0,0762
	(0,5110)	(0,7217)	(0,6000)	(0,5548)	(0,6142)	(0,3662)	(0,3666)	(0,3811)	(0,4722)	(0,4026)	(0,4244)	(0,2105)
Lagged Participation:												
17) Ind (-1)	2,7296	2,1968	3,3078	2,5484	2,1200	2,3007	2,1136	2,2350	2,0366	1,7854	2,6095	2,2583
	(0,3714)	(0,5329)	(0,4417)	(0,4611)	(0,5488)	(0,3505)	(0,3173)	(0,3053)	(0,4167)	(0,3227)	(0,3918)	(0,1996)
18) White Dummy * Education	-0,4095	-0,5065	0,3461	0,5888	-0,0727	-0,0811	0,2514	0,2583	0,1962	0,2179	0,0710	0,1166
	(0,2652)	(0,3950)	(0,3338)	(0,3017)	(0,4011)	(0,2598)	(0,2198)	(0,2028)	(0,3102)	(0,2519)	(0,2725)	(0,1259)
19) White Dummy * Age	-0,0404	-0,0390	-0,0083	0,0248	-0,0212	0,0200	0,0005	0,0081	0,0034	0,0189	-0,0247	0,0004
	(0,0300)	(0,4520)	(0,0355)	(0,0312)	(0,0367)	(0,0246)	(0,0200)	(0,0200)	(0,0266)	(0,0251)	(0,0268)	(0,01220)
20) $\ln(w_t)$ (MOE)	--	0,0787	0,0275	0,0898	0,0527	-0,0052	0,0084	-0,0256	0,0332	0,0040	0,0401	J_N : 97,12
		(0,0751)	(0,0659)	(0,0656)	(0,0717)	(0,0547)	(0,0525)	(0,0510)	(0,0568)	(0,0527)	(0,0556)	Degrees of Freedom: 180
Number of Observations	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1034	1079	1107	1156	1237	1274	

TABLE III

Estimates of the FOC for Consumption

Variable	Parameter	Estimate	Aggregate Price	Estimate
Δ (Number in Family)	$B_{20}/(1-\zeta)$	0.0157 (0.0037)	$\ell n(\lambda_4/\lambda_3)/(1-\zeta)$	0.0473 (0.0231)
Δ (Number of Children Less than 6)	$B_{21}/(1-\zeta)$	0.0391 (0.0028)	$\ell n(\lambda_5/\lambda_4)/(1-\zeta)$	-0.0135 (0.0080)
Δ (Number of Children Between 6 and 14)	$B_{22}/(1-\zeta)$	0.0057 (0.0014)	$\ell n(\lambda_9/\lambda_8)/(1-\zeta)$	-0.0700 (0.0173)
Δ (Age squared)	$B_{23}/(1-\zeta)$	-0.0742 (0.0024)	$\ell n(\lambda_{10}/\lambda_9)/(1-\zeta)$	-0.1800 (0.01200)
Δ Region Dummies:			$\ell n(\lambda_{11}/\lambda_{10})/(1-\zeta)$	-0.0540 (0.0102)
Northeast	$B_{24}/(1-\zeta)$	-0.1884 (0.0188)	$\ell n(\lambda_{12}/\lambda_{11})/(1-\zeta)$	0.0320 (0.0100)
Northcentral	$B_{25}/(1-\zeta)$	-0.1191 (0.0153)	$\ell n(\lambda_{13}/\lambda_{12})/(1-\zeta)$	-0.0604 (0.0100)
South	$B_{26}/(1-\zeta)$	-0.0772 (0.0114)	$\ell n(\lambda_{14}/\lambda_{13})/(1-\zeta)$	-0.0926 (0.0100)
			$\ell n(\lambda_{15}/\lambda_{14})/(1-\zeta)$	-0.0814 (0.0100)
			$\ell n(\lambda_{16}/\lambda_{15})/(1-\zeta)$	-0.1117 (0.0090)
			$\ell n(\lambda_{17}/\lambda_{16})/(1-\zeta)$	-0.0412 (0.0084)
			$\ell n(\lambda_{18}/\lambda_{17})/(1-\zeta)$	0.0247 (0.0084)
			$\ell n(\lambda_{19}/\lambda_{18})/(1-\zeta)$	-0.0303 (0.0100)

TABLE IV

Estimates of the Euler and Participation Equations
Without Aggregate Shocks

Variable	Parameter	Unconstrained	Constrained
(i) Participation Equation			
l_{nt}	B_{10}	-58.8584	-20.4394
		(0.3618)	(0.0375)
$nk_{1nt} \cdot l_{nt}$	B_{11}	-24.9590	-8.3730
		(0.4435)	(0.0306)
$nk_{2nt} \cdot l_{nt}$	B_{12}	26.3147	10.5746
		(0.3287)	(0.0410)
$l_{nt}^2 - var(\varepsilon_{nt})$	δ_0	0.0236	0.0900
		(0.0004)	(0.0001)
$l_{nt} \cdot l_{nt-1}$	δ_1	-0.6024	-0.5314
		(0.0100)	(0.0011)
$l_{nt} \cdot l_{nt-2}$	δ_2	0.0239	0.0437
		(0.0002)	(0.00003)
$l_{nt} \cdot l_{nt-3}$	δ_3	0.0078	0.01307
		(0.0002)	(0.00003)
$l_{nt} \cdot l_{nt-1}^2$	α_1	0.0003	0.0002
		(0.000005)	(5.6E-7)
$p_{nt}^{(1,1)}q_{nt}^{(1,1)} - p_{nt}^{(0,1)}q_{nt}^{(0,1)}$	β	-149.5400	0.7588
		(4.7440)	(0.0006)
$p_{nt}^{(1,2)}q_{nt}^{(1,2)} - p_{nt}^{(0,2)}q_{nt}^{(0,2)}$	β^2	46.5319	--
		(3.0484)	
$p_{nt}^{(1,3)}q_{nt}^{(1,3)} - p_{nt}^{(0,3)}q_{nt}^{(0,3)}$	β^3	183.9014	--
		(3.5187)	
$p_{nt}^{(1,4)}q_{nt}^{(1,4)} - p_{nt}^{(0,4)}q_{nt}^{(0,4)}$	β^4	-7.9742	--
		(0.2909)	

(ii) Euler Equation

Constant	B_{10}	7.2871	--
		(0.4332)	
nk_{1nt}	B_{11}	12.0878	--
		(0.2330)	
nk_{2nt}	B_{12}	3.8934	--
		(0.0580)	
l_{nt}	δ_0	4.2765	--
		(0.0348)	
l_{nt-1}	δ_1	0.1284	--
		(0.0022)	
l_{nt-2}	δ_2	-0.0061	--
		(0.0001)	
l_{nt-3}	δ_3	-0.0022	--
		(0.00004)	
l_{nt+1}	$\beta\delta_1$	35.5131	--
		(0.4357)	
l_{nt+2}	$\beta^2\delta_2$	32.0452	--
		(0.5311)	
l_{nt+3}	$\beta^3\delta_3$	18.6238	--
		(0.2403)	
$\eta_n l_{nt+1}$	$\beta\gamma_1$	-10.7277	30.8362
		(0.1491)	(38.3389)
$\eta_n l_{nt+2}$	$\beta\gamma_2$	-12.5382	4.0722
		(0.1984)	(21.6639)
$\eta_n l_{nt+3}$	$\beta^3\gamma_3$	-6.1184	0.7905
		(0.0858)	(39.4428)
$\eta_n l_{nt+4}$	$\beta^4\gamma_4$	-0.7822	0.2216
		(0.0101)	(124.0000)
l^2_{nt-1}	α_1	-0.00003	--
		(0.00001)	
$l_{nt+1} \cdot l_{nt}$	$\beta\alpha_1$	-0.0052	--
		(0.0001)	
J_N		8.7255	568.27
Degrees of Freedom		4	19