ECONOMETRIC REVIEWS, 11(1), 1-71 (1992)

THE ECONOMETRICS OF FEMALE LABOR SUPPLY AND CHILDREN

Alice Nakamura and Masao Nakamura
Faculty of Business
University of Alberta
Edmonton, Alberta
Canada T6G 2R6

Key Words and Phrases: female labor supply; fertility; life cycle models; household decision models; demand analysis; endogeneity; instrumental variables.

JEL classification: C50, D13, J13, J22, J24

ABSTRACT

This is a survey of applied econometric research on the effects of children on female labor supply. Reasons for interest in the topic, and a basic model and terminology, are reviewed. Concerns are raised about the possible endogeneity of child status variables, and about the instrumental variables approach for dealing with this problem. Alternative ways of conceptualizing and estimating child status effects are considered, together with selected empirical evidence. Relevant developments from the household demand literature are summarized. Basic issues of model choice are also discussed.

I. INTRODUCTION

This is a survey of applied econometric research on the effects of children on female labor supply. Other surveys of research on female labor supply have focused on the impacts of economic variables such as wage rates, in keeping with the emphasis in the literature being reviewed.[1] However, there is considerable public policy interest in the interrelationships between children and women's work behavior. Most empirical analyses of female labor supply do contain findings about child status effects. In fact, child status variables usually account for a large share of the explained variation in empirical studies of female labor supply. Also, there is a growing body of research that explicitly addresses the problems of estimating the effects of children on the labor supply of women. It is time to take stock of what has been, or can be, learned from this research.

The questions addressed in this survey are basic:

Why are measures of child status effects needed?

What is meant by child status effects?

How should child status variables be defined?

Can theorized child status effects be appropriately represented in available models of female labor supply?

How should we choose among alternative models, and among differing estimates of child status effects?

The evidence and theories reviewed are found to be helpful in answering these questions, though the answers are far from complete.

Section II summarizes the stated reasons for interest in measuring the effects of children on the work behavior of women. For readers who are not specialists, a typical specification for a static, one-period model of female labor supply is outlined in section III. Also relevant terminology, notation and concepts are introduced. Concerns about the possible endogeneity of child status effects and related econometric issues are addressed in section IV. Alternative estimation methods have been shown to yield quite different estimates of child status responses. Selected aspects of this issue are discussed as well in section IV. Section V outlines two directions that have been pursued in the search for better

[1]Other surveys of research on female labor supply include Killingsworth (1983), Heckman and MaCurdy (1986), Killingsworth and Heckman (1986), Blundell (1990) and Browning (1992).

instrumental child status variables. This material raises questions about the treatment of time, and about the causal nature of child status effects. These issues are discussed and direct and indirect child status effects are defined in section VI.

Sections VII and VIII address the issue of how direct child status effects can be represented in models of female labor supply. Alternative ways of allowing for intertemporal effects are discussed in section IX. The final two sections focus on the problem of choosing among or ranking alternative models all of which must be viewed as approximate. The output space methods recommended in section X are used in section XI for comparing three models of female labor supply incorporating different amounts of information about work behavior in the previous year. Conclusions are summarized in section XII.

II. REASONS FOR INTEREST IN CHILD STATUS EFFECTS

Most of the stated and implied reasons for estimating child status effects on female labor supply can be grouped under three headings:

1. To improve estimates of wage and income elasticities.
2. To improve the ability to forecast the labor supply and earnings of women.
3. To improve basic knowledge about how children affect female labor supply.

Judgments about the appropriateness of various simplifying assumptions and about model performance should depend, in part, on the intended uses for the results. Leamer makes this point with a maps analogy:

> We may take as a theory of the world an enormously detailed globe which identifies every object down to the smallest grain of sand. The complexity of this theory effectively prevents us from using it for any purpose whatsoever. Instead, we simplify it in the form of a set of maps. I use one map to find my way to the subway station, another to select the station at which to depart. The pilot of the airplane uses yet another to navigate from Boston to Washington. Each map is a greatly simplified version of the theory of the world; each is designed for some class of decisions and works relatively poorly for others.
>
> (Leamer, 1978, p. 205)

Taking Leamer's arguments seriously, the main types of reasons for estimating child status effects are briefly discussed in the following subsections.

1. **Estimating Wage and Income Elasticities**

In most economic studies of female labor supply, the stated objective is obtaining estimates of wage and income elasticities. This reflects policy concerns in the taxation and income support areas, as well as the fact that the conceptual origins of labor economics lie largely in the theory of consumer demand.

Child status variables are nuisance variables in studies falling in this first category. They are included to avoid bias problems and to improve the efficiency of the estimates of the wage and income effects. Biases in, or problems of interpreting, the coefficient estimates of the child status variables are important in this context only to the extent that there are implications for the quality of the estimates of the economic responses of prime interest.

2. **Forecasting Female Labor Supply**

Since World War II there have been important changes in the representation of women in the workforce. There have also been changes in fertility rates. These changes have spurred interest in forecasting female labor supply and in investigating the interrelationships between fertility and labor supply behavior. Forecasts are needed for analyses of current and future labor market conditions, as well as for more specialized purposes such as for evaluating proposed tax laws and projecting revenue and benefit flows for public benefit programs like the U.S. Social Security system.[2]

[2]Aggregate time series models have been widely used for forecasting the labor force, but it is difficult to construct aggregate child status variables. Household microsimulation models are another means for generating labor supply forecasts, as in Orcutt, Caldwell and Wertheimer (1976), Blomqvist and McKee (1986), and studies cited in Citro and Hanushek (1991). The microsimulation approach allows the explicit representation of tax and transfer programs affecting individuals. This approach can also incorporate research findings that focus on the roles of women in families, as in the studies of Reimers (1984) and Duleep, Regets and Sanders (1992).

In a private sector context, employers considering women for jobs requiring substantial employer-subsidized training must form judgments about expected future labor supply. Labor supply and earnings forecasts for women are needed in legal cases involving wrongful dismissal, or death or disability due to negligence.[3] There may also be product marketing implications of continuing changes in the labor market participation and earnings of women.[4]

In a forecasting context, the choice of child status variables and an understanding as to what the coefficients of these variables are capturing is vital.

3. **Improving Behavioral Knowledge**

Many of the terms used in econometrics reflect the close relationship between behavioral research and forecasting. For instance, the coefficients of the explanatory variables in econometric models are sometimes referred to as "response coefficients". This is in keeping with the interpretation of each coefficient as the expected change in the dependent variable given a ceteris paribus, one unit change in the associated explanatory variable.

On a conceptual level, constrained utility maximization models embed a treatment-response view of human behavior. In the theory of consumer demand from which economic models of female labor supply have evolved, the treatment and response variables are easily identified and (in principle, at least) observable: the treatment variables are the prices and the income or expenditure variables, and the response variables are the measures of demand. However, in most studies presenting estimates of child status effects on female labor supply, the conceptual meaning of the relevant treatment variable(s) is neither stated nor clear. Consider the coefficient for a variable for the number of children ever born in a study where the dependent variable is labor supply in a given year. Is the

[3]Capozza, Nakamura and Bloss (1989) provide evidence that erroneous estimates of child status effects can lead to systematic inequities in the resulting financial settlements.

[4]Studies of possible marketing implications include Strober and Weinberg (1980), Weinberg and Winer (1983) and Douthitt and Fedyk (1988).

hypothetical child status "treatment" the birth of one more child in that year, or in any previous year? Is the contemplated "treatment" an unplanned "exogenous" child or one wanted and planned? Does the health status (or any other attribute) of the "added" child matter?

One of the primary objectives of this survey is to try to develop a clearer understanding of the relevant child status "treatments" in models of female labor supply, and of how these "treatments" are (or could be) represented in terms of observable variables.[5] We seek to accomplish this through an examination of the basic structure of economic models of female labor supply (section III), through reviewing behavioral hypotheses about how children affect the work behavior of women (sections VI and VII), and through analysis of the definitions and empirical evidence concerning commonly used child status variables (sections IV-V and VIII-IX).

III. A STATIC ONE PERIOD MODEL OF FEMALE LABOR SUPPLY

1. The Basic Conceptual Framework

Most economic models of female labor supply have the same basic form: a conditional household utility function is maximized subject to time and budget constraints. This maximization problem typically yields a decision rule for a woman's labor supply that can be described in terms of a reservation wage function giving the minimum (or shadow) wage for which a woman would be willing to work under various circumstances, and a market wage function for the wage (or the highest expected wage) a woman could command from an employer. This labor supply decision rule is often stated in two parts: (1) a woman is assumed to work in a designated time period (such as a year) if her market wage is greater than or equal to her reservation wage at zero hours of work, and (2) the amount she works, if she is employed, is assumed to be determined so as to equate her reservation wage (usually treated as an increasing function of time spent in market work) and her market wage.

[5]Heckman (1990) provides a fundamental and far reaching discussion of the econometrics of isolating treatment effects.

There are many simplifying assumptions embedded in this conceptual framework. For example, Manser and Brown (1980) and McElroy and Horney (1981) draw attention to the treatment of household utility and household decision making processes. From an empirical perspective, however, this framework has aided thinking about factors that might affect the willingness to work versus job opportunities. Factors that shift the reservation wage function upward (downward) are hypothesized to decrease (increase) labor supply, while factors that shift the market wage function upward (downward) are expected to have the opposite effect.

The similar structure of most economic models of female labor supply implies a shared set of econometric issues and problems, many of which can be discussed within the context of the static, one-period model outlined in this section.

2. Notation

Let G_t denote the amount of a Hicksian composite good with unit price p_t that is consumed by a family in unit period t. Let $h_{i,j,t}$ denote the hours in period t spent by family member i in time use j. For want of better terminology, the adults who form a heterosexual family will be referred to as the female and male partners (denoted by sub f and sub m, respectively).

In this paper, the time uses that will be differentiated are market work (W), family work (F), and "leisure" (L) defined as all uses of time other than market or family work. In many studies, family work is not distinguished from "leisure" on either a conceptual or semantic level. In our notation, the complement of market work (the "leisure" category in most studies) is denoted by NW. Thus, for example, for the female partner $h_{f,NW,t} = h_{f,F,t} + h_{f,L,t}$. When there is no risk of confusion, market work will be referred to simply as "work".

The wage rate family member i can command in the labor market is denoted by $w_{i,t}$, and $y_{i,t}$ stands for the period t earnings of family member i (with $y_{i,t} = w_{i,t}h_{i,W,t}$). Nonlabor family income is denoted by A_t.

For now, all child status variables are included in the vector N_t. Other family and individual attributes that might affect a woman's asking wage are denoted by the vector Z^*.

3. The Maximization Problem

Typically, a family is viewed as maximizing a conditional utility function of the form

$$U = U(G_t,\ h_{f,NW,t};\ y_{m,t},\ A_t,\ N_t,\ Z^*_t). \tag{III-1}$$

The earnings of the male partner ($y_{m,t}$) and nonlabor income (A_t) are often treated econometrically as well as conceptually as predetermined or exogenous variables. This is also the case for the child status variables (N_t) and the variables for other family attributes (Z^*_t).

The time constraint for the female partner is given by

$$h_{f,NW,t} + h_{f,W,t} = H_t, \tag{III-2}$$

where H_t denotes total hours in unit period t.[6] The family budget constraint is

$$p_tG_t = y_{m,t} + A_t + w_{f,t}\ h_{f,W,t}. \tag{III-3}$$

Assuming that $0 < h_{f,NW,t} \le H_t$, or $0 \le h_{f,W,t} < H_t$, the Lagrangean is

$$L = U(G_t,\ h_{f,NW,t};\ y_{m,t},\ A_t,\ N_t,\ Z^*_t) + \lambda_G(y_{m,t} + A_t + w_{f,t}\ h_{f,W,t} - p_tG_t) + \lambda_W h_{f,W,t}, \tag{III-4}$$

where λ_G and λ_W are, respectively, an unconstrained and a nonnegative Lagrange multiplier. The first-order optimality conditions are:

[6]Note that time for personal care (such as sleeping time) is not excluded from H_t; this time is included in leisure time. Note also that H_t does not exclude time spent on children, as, for example, Blundell (1987) does. In this study child care time is included in the family work component of H_t.

$$U_G - \lambda_G\ p_t = 0, \tag{III-5}$$

$$-U_{NW} + \lambda_G\ w_{f,t} + \lambda_W = 0, \tag{III-6}$$

$$y_{m,t} + A_t + w_{f,t}\ h_{f,W,t} - p_tG_t = 0, \text{ and} \tag{III-7}$$

$$\lambda_W\ h_{f,W,t} = 0, \tag{III-8}$$

where U_G and U_{NW} are the partial derivatives of the utility function with respect to purchased goods and services (G_t) and the female partner's nonwork time ($h_{f,NW,t}$). Condition (III-5) implies that $\lambda_G > 0$.

From the first-order condition (III-6), it follows that

$$w_{f,t} = (U_{NW}/\lambda_G) - (\lambda_W/\lambda_G) = w^*(h_{f,W,t}) - (\lambda_W/\lambda_G) \tag{III-9}$$

where $w^*(h_{f,W,t}) = (U_{NW}/\lambda_G)$ is the *reservation wage* at $h_{f,W,t}$ hours of work. Condition (III-9) implies that $w_{f,t}$ is less than or equal to $w^*(h_{f,W,t})$. Also, from (III-8) it can be seen that for positive hours $\lambda_W = 0$. Hence the *condition for work* in period t is

$$w_{f,t} > w^*(h_{f,W,t} = 0). \tag{III-10}$$

The model implies that a woman who works will choose her hours of work so that the following equilibrium condition is satisfied:

$$w_{f,t} = w^*(h_{f,W,t}) \quad \text{when } h_{f,W,t} > 0. \tag{III-11}$$

From how it is defined above, it can be seen that a woman's reservation wage depends on $h_{f,W,t}$, $y_{f,W,t}$ $(= w_{f,t}h_{f,W,t})$, p_t, $y_{m,t}$, A_t, N_t, and Z^*_t when $h_{f,W,t}$ is positive; and on p_t, $y_{m,t}$, A_t, N_t and Z^*_t when $h_{f,W,t}$ is zero.

4. Estimation of Labor Supply and Market Wage Relationships

In applied studies, it is common to use approximations for a woman's *reservation wage function* such as

(III-12)

$$\ln w^*(h_{f,W,t}) = \begin{cases} \beta_0 + \beta_1(y_{m,t}+A_t) + N_t\beta_2 + Z^*{}_t\beta_3 \\ \quad + \beta_4 \ln w_{f,t} + \beta_5 h_{f,W,t} + u^*{}_t & \text{if } h_{f,W,t} > 0 \\ \beta_0 + \beta_1(y_{m,t}+A_t) + N_t\beta_2 + Z^*{}_t\beta_3 + u^*{}_t & \text{if } h_{f,W,t} = 0, \end{cases}$$

where ln denotes the natural logarithm, $u^*{}_t$ is an error term and the β's are the parameters to be estimated. The logarithm of a woman's offered wage is usually expressed as some function of job-related personal characteristics (denoted here by the vector Z_t), and labor market conditions (denoted by E_t). A typical sort of specification is

(III-13) $$\ln w_{f,t} = \alpha_0 + Z_t\alpha_1 + E_t\alpha_2 + u_t,$$

where u_t is an error term and the α's are unknown parameters.

Using (III-12), (III-13), and the decision rule given in (III-10) for work in period t, the *probability of work* in period t is represented as

(III-14) $$P(h_{f,W,t} > 0) = P[\ln w_{f,t} > \ln w^*(0)] = F(\phi_{f,t}),$$

where F denotes the cumulative density function for $(u^*{}_t - u_t)$, and the *index for the probability of work* function is given by

(III-15)
$$\phi_{f,t} = (1/\sigma)[(\alpha_0 - \beta_0) + Z_t\alpha_1 + E_t\alpha_2 - \beta_1(y_{m,t} + A_t) - N_t\beta_2 - Z^*{}_t\beta_3]$$

with σ denoting the standard deviation of $(u^*{}_t - u_t)$. If $u^*{}_t$ and u_t are jointly normally distributed with zero means and constant variances, then F in (III-14) is the cumulative density function for the standard normal distribution, and the coefficients of (III-15) can be estimated by *probit analysis*. Other estimation methods are appropriate given other distributional assumptions.

When condition (III-10) is satisfied so that a woman's hours of work are positive, then condition (III-11) also applies. An expression for a woman's hours of work can be obtained by equating ln w and the right-hand side of the expression given in (III-12) for

$\ln w^*(h_{f,W,t})$ when hours of work are positive, and then solving for $h_{f,W,t}$. The resulting *hours equation* is:

(III-16) $$h_{f,W,t} = (1/\beta_5)[(1 - \beta_4)\ln w_{f,t} - \beta_0 - \beta_1(y_{m,t} + A_t) - N_t\beta_2 - Z^*{}_t\beta_3 - u^*{}_t].$$

The offered wage equation (III-13) is often estimated by regression analysis, with the inverse Mills ratio (the *Heckit selection bias term*), $f(\phi_{f,t})/[1-F(\phi_{f,t})]$, included along with the explanatory variables to account for sample selection due to the condition for work given in (III-10).

The hours equation could also be estimated as specified in (III-16) with the addition of a Heckit selection bias term. Often, however, there is concern that the wage variable in the hours equation is correlated with the equation error term. Two alternative approaches for dealing with this correlation problem are the use of an instrumental wage variable, and elimination of the wage variable in equation (III-16) using equation (III-13). Of course, elimination of the wage variable also eliminates the possibility of directly estimating offered wage effects on hours of work.

An alternative way of allowing for the fact that wage rates and hours of work are only observed for those who work is to estimate an integrated model for both the decision to work and the hours of work using *Tobit analysis*. The estimates of child status effects can be quite different depending on which of these approaches is adopted. In the case of the Heckit approach, separate estimates are obtained for (β_2/σ) in the index for the probability of work (expression (III-15)) and for (β_2/β_5) in the hours equation (III-16), whereas the Tobit estimation approach constrains the elements of β_2 to have the same values in the decision rule for the probability of work and in the relationship determining hours of work.[7]

[7]Relevant references include Heckman (1974, 1979, 1981), Amemiya (1985), Mroz (1987), Nakamura and Nakamura (1989) and Berndt (1990, Ch. 11). The Berndt reference includes well-designed exercises with a data set provided on a diskette and a readable presentation of the econometric methods.

IV. ALTERNATIVE ESTIMATES OF CHILD STATUS EFFECTS

1. Possible Endogeneity of the Child Status Variables

In early cross-sectional studies of female labor supply, child status variables were directly introduced along with the other explanatory variables into relationships for the probability of work and hours of work. In fact, there are many recent, econometrically sophisticated studies in which child status variables are directly introduced. However, T. Paul Schultz and others including Dooley (1982), Moffitt (1984), and Hotz and Miller (1988) have persistently argued that this practice may result in bias problems because of correlations between child status variables and the error terms for the labor supply relationships. Two conceptually distinct sources of this alleged correlation problem are singled out. The first is that families treat the choice of how many children to have and the decision of how much of the female head's time should be devoted to market work as aspects of a joint decision problem. A second suggested source of correlations between the child status variables and the equation error terms are persistent omitted factors that affect both child status and the labor supply of the mother.

Regardless of the source of the alleged correlation problem, the most widely recommended remedy is the use of instrumental child status variables. Auxiliary equations are specified for the child status variables, and predicted values are substituted for the original child status variables in the labor supply relationships.

2. Estimates Produced by Different Econometric Methods

Table 1 shows alternative estimates of the effects of the number of children ever born on maternal labor supply. The results are presented in elasticity form, as originally presented by Schultz (1978). We first discuss the results in the top half of Table 1, and then compare these with the results in the bottom half.

In the top half of Table 1, the results in the first column are *OLS estimates for all women*. In particular, these results are for an hours equation like (III-16) that has been estimated for all women with hours of work set equal to zero for those who did not work in the given year, and with a predicted wage variable obtained using data for women who worked.

TABLE 1

ESTIMATES OF THE ELASTICITY OF THE ANNUAL LABOR SUPPLY OF WIVES WITH RESPECT TO CHILDREN EVER BORN

	Annual hours of work			Probability of work
	All wives		Wives who worked	All wives
	OLS	ML Tobit	OLS	Logit
White, 18-24	-.55	-.76	-.10	-.60
White, 25-29	-1.04	-1.12	-.29	-.80
White, 30-34	-.51	-.62	.06	-.49
White, 35-39	-.47	-.51	-.18	-.33
Instrumental variable for children ever born				
White, 18-24	-.24	-.39	.05	-.37
White, 25-49	-.95	-.90	-.14	-.74
White, 30-34	-.92	-1.13	.07	-.95
White, 35-39	-1.13	-1.41	-.27	-1.06

Source: Table 6, p. 299 in Schultz (1978).

Tobit estimates are presented in column 2. The Tobit approach, which also uses data *for all women*, allows for the limited nature of the dependent variable (hours of work cannot be negative), for the fact that wages cannot be observed for those who do not work, and for the possibility of selection bias problems. The estimates in both columns 1 and 2 imply a strong negative relationship between expected labor supply and the number of children.

The estimates shown in column 3 are for the same linear hours equation for which estimation results are shown in column 1, except that now the *OLS estimates* are computed using data *for only those women who worked*. These estimates correspond to the second stage of the Heckit estimation method discussed above, except that no selection bias term has been included. Results reported by others indicate that ignoring the selection bias term probably has little impact on the hours equation estimates of the child effects. Note that the estimated responses shown in column 3 are small in magnitude compared to the results when data for all women are used (columns 1 and 2).

As already noted, when the error terms for the labor supply model obey normal distributions, then F in (III-14) denotes the normal cumulative density function, and the coefficients of the index for the probability of work (III-15) can be appropriately estimated using probit analysis. These results are often treated as the first stage in the Heckit estimation procedure. When F in (III-14) denotes the logistic rather than the normal cumulative density function, then the coefficients of (III-15) are appropriately estimated using logit analysis. Except in the tails of the distributions, the logistic and the normal density functions are similar; hence probit and logit analysis usually yield similar estimated effects for the explanatory variables.

In column 4 we show Schultz's *logit analysis estimates* of the elasticity for the probability of work with respect to the number of children ever born. A woman's unconditional expected labor supply in a year can be represented as the product of the probability she will work times her expected hours of work if she does work. The elasticity estimates in column 4 only take account of child status effects on the first of these two components, while the estimates in column 3 are measures of the second component. Note that the elasticity estimates in column 4 for the probability of work are almost as large as the estimates in column 1 and in column 2 for the unconditional annual labor supply responses.

A first implication Schultz draws from these results is that child-related variations in labor supply primarily manifest themselves in alterations in the propensity to work. Others have also noted this (see, for example, Bowen and Finegan, 1969, p. 100).

A second implication is that integrated probability-hours estimation procedures, like the Tobit approach, may not be appropriate.

A third implication emerges from comparing the findings of Schultz and others (including ourselves) who have focused on the decision to work and hours of work in an annual context with recent findings by Blank (1988, 1989) and Fuchs (1988, pp. 44-48). Integrated treatments of the weeks and the hours-per-week dimensions of annual hours of work may be inappropriate, and could possibly obscure the effects of children on the annual hours of work of women who are employed.

The elasticity estimates presented in the bottom panel of Table 1 correspond, column by column, to those in the top panel, except that they are computed using an instrumental variable (IV) for the number of children born. The IV coefficient estimates are systematically different from the corresponding figures in the top panel of the table: they are consistently less negative for wives younger than 30, but almost always more negative for older wives. However, the fact that the IV estimates are different does not establish that they are better. The difficulty lies with the near impossibility of finding suitable exogenous variables. With respect to Schultz's own choice of exogenous variables, he writes:

> I will assume here that the wife's residential origins at age 16, her age, and the schooling of both spouses are exogenous determinants of fertility, but that these variables exert no direct role in determining the current labor market behavior of the wife.... It should be repeated, however, that there is no strong theoretical presumption that these identifying restrictions are in some sense 'correct'....
> (Schultz, 1978, p. 291)

Unfortunately, it seems conceivable that a woman's residential origins, her cohort (identified by her age), and her own and her spouse's schooling experiences could have important effects on her tastes for a career, the occupations in which she can find work, her offered wage, and her ideas about the needs of small children for maternal care. Surely these factors could also affect current labor supply. Moreover the designated "exogenous" variables will not reflect any of the truly chance variation in fertility due to factors like contraceptive failure or multiple births, or in the level of care parents provide for children. The problem of correlations between child status variables and the error terms for the labor supply equations could be more, rather than less, serious with an IV approach using these "exogenous" variables.

None of the available exogeneity tests provide a satisfactory means for examining this problem. All rely on *finding* suitable exogenous variables to form an instrumental child status variable. These tests just move the problem of testing for endogeneity one step back. Furthermore the tests do not provide a metric for judging the seriousness of endogeneity problems in terms of the magnitudes of the biases in the coefficient estimates of interest. Another drawback of these tests is that their power drops severely as the explanatory power of the auxiliary equation(s) used in forming the instrumental variable(s) falls. The explanatory power is usually very low for equations for child status variables. These issues are explored more formally in Nakamura and Nakamura (1985c).

Moreover, even if suitable exogenous variables could be found to explain a single child status variable such as the number of children ever born, the limited availability of exogenous variables means that very few dimensions of child status can be dealt with using an IV approach. This is why the only child status variable in Schultz's 1978 study is a linear term for the number of children ever born. Schultz recognizes this shortcoming, and acknowledges that "it may not be the number of children a woman has borne ... that directly affects her current labor supply..." (p. 286). A study such as Lehrer (1991) illustrates the potential need for taking account of multiple dimensions of child status.

V. A SEARCH FOR IMPROVED INSTRUMENTAL CHILD STATUS VARIABLES

In search of better exogenous factors to use in forming instrumental child status variables, researchers have looked to sources of unplanned births and to the technology of producing (and preventing) babies. The Rosenzweig and Wolpin (1980) twins-first study reviewed in subsection V.1 is an example of the first of these lines of inquiry. The Rosenzweig and Schultz (1985) study outlined in subsection V.2 is an example of the second approach.

1. A Twins-First Approach

Rosenzweig and Wolpin suggest using the exogenous variation in the number of children ever born that is afforded by multiple births in order to obtain better estimates of the effects of children on female labor supply. Rosenzweig and Wolpin claim:

> In particular, we show how a natural event, the occurrence of a multiple birth of "twins," can be used as an instrument for exogenous fertility movements. The variable we propose, a twins outcome in the first birth, approximates the social experiment we wish to perform not only in that some families receive an unanticipated child, while others do not, but also in that the treatment and control groups are randomly selected with respect to characteristics that may be related to market participation. It is therefore unnecessary to utilize any information on the determinants of labor supply behavior in order to determine the "true" exogenous fertility effect by this method.
> (Rosenzweig and Wolpin, 1980, pp. 335-336)

The discussion in their study is cast in a life cycle context, but the measure of labor supply used is a dichotomous variable for current work status.

Data are pooled from two national random surveys of women: a 1965 survey conducted by the Office of Population Research and a 1973 survey carried out by the U.S. Department of Health, Education and Welfare. Of 15,000 available observations for women 15-44 years of age, 12,605 were used. The main reason for excluding observations was childlessness. The twins-first methodology requires a data sample for women who have had at least one birth. Of the 12,605 observations, 87 were for women whose first children were twins. (The rarity of multiple births is obviously a drawback of the twins-first methodology.) Rosenzweig and Wolpin find that an "extra" child on the first birth (before age 35) reduces the mother's current probability of labor force participation by 35 percentage points for women 15-24 years of age, and by about 10 percentage points for women 25-34 years of age; but increases the current probability of participation by 6 to 14 percentage points for women 35-44 years of age.

2. A Birth Technology Approach

Rosenzweig and Schultz (1985) seek to obtain better estimates of child status effects by taking more explicit account of the production process for babies. They model the actual number of births to a couple in a given time period as a sum of a time invariant, couple-specific fecundity factor; a random component; an age factor; and fertility control measures. The *fecundity factor* is found to be associated with persistent differences in labor supply. Unexpected births (that is, the *random component* of fertility) are found to affect the timing of a woman's labor supply but to have little or no effect on the total amount of labor supplied in a woman's lifetime. (Recall that Rosenzweig and Wolpin also found evidence of offsetting timing effects for unexpected multiple births.)

Rosenzweig and Schultz estimate a *reproduction technology function* with actual fertility as the dependent variable. The explanatory variables are measures of the type and amount of birth control used, births before 1970, coital frequency, whether the woman smoked, and the woman's age and age squared. However, all of these variables are treated as endogenous except age and age squared. The *exogenous variables* used to "explain" the endogenous variables include age, education, religious background and the husband's income. Denoting the actual birth rate of couple j in period i as n_{ij}, and the predicted birth rate based on the estimated reproduction technology function as n^e_{ij}, the *fecundity factor* is computed as

$$\mu_j = \Sigma^A_{i=1} (n_{ij} - n^e_{ij})/t$$

where t is the number of time periods and the *random component* is given by

$$\varepsilon_{ij} = n_{ij} - n^e_{ij} - \mu_j.$$

Particular attention should be paid to the fact that the couple-specific fecundity factor and the random component, taken together, are determined as the residual between the actual and the predicted birth rates. As a consequence, these factors will collectively pick up any persistent effects on the actual birth rate for a couple (such as tastes for family life) that are not captured by the predicted birth rate and that may be associated with persistent differences in labor supply.

Also, the focus on births in both the Rosenzweig-Wolpin and the Rosenzweig-Schultz studies draws attention to the issue of whether this is really the most relevant child status "treatment" with regard to female labor supply. From both casual observation and family expenditure and time use studies, it is clear that parental choice plays a major role in determining the level of care provided for even "exogenous" children (multiple births, products of contraceptive failure, children acquired through marriage, and so forth). There is also empirical evidence of economies of scale, parity (perhaps learning!) effects, and effects of advance planning (that is, differences depending on whether a birth was planned) that may make labor supply responses to the birth of a twin or an unplanned child atypical of the labor supply responses to other sorts of births.

The problems and issues raised by efforts to construct appropriate instrumental child status variables draw attention again to three of the questions raised at the start of this survey:

What is meant by child status effects?

How should child status variables be defined?

How can theorized child status effects be represented in models of female labor supply?

In seeking answers to these questions, in the next section we take a closer look at the treatment of time in models of female labor supply, and at various behavioral theories about how children affect the work behavior of women.[8]

VI. THE TREATMENT OF TIME, AND DIRECT VERSUS INDIRECT CHILD STATUS EFFECTS

Reviewing the work of others on the effects of children on female labor supply has made us aware of the crucial role which time plays in the definition and modeling of these effects. Subsection VI.1 draws attention to differences between alternative models with respect to the treatment of time and provides terms for discussing these differences. We suggest a time-related breakdown of the effects of children on female labor supply. Two categories of effects are defined: direct and indirect. Indirect effects are discussed in subsection VI.2, and direct effects in subsection VI.3.

1. Planning Horizons and Unit Time Periods

Arguments about the endogeneity of child status variables are usually couched in life cycle terms. Addressing these arguments requires differentiating life cycle models from one-period models of the sort reviewed in section III. Dynamic and static models must also be distinguished.

The *planning horizon* is the entire time period considered in the family labor supply decision problem. *Life cycle models* have planning horizons that stretch over whatever the model builder views as the relevant life cycle, such as the years a couple hope to be together from the time of their marriage. If it is assumed in such a model that there are no bequests and that families cannot pass out of existence with unpaid bills or debts, then the family budget must be balanced over the specified life cycle planning horizon.

[8]Economists have not been alone in wrestling with these issues of how the effects of children on family decision making should be represented. See, for example, Waite and Stolzenberg (1976), Cramer (1980), Oppenheimer (1982), Bagozzi and Van Loo (1988, 1991) and Birg (1991). There is also an important related literature on factors affecting the timing of births and on the interrelationships between the timing of births and female labor supply. See, for example, Easterlin (1987), Ward and Butz (1980), Moffitt (1984), Heckman, Hotz and Walker (1985), Ni Bhrolchain (1986), Calhoun and Espenshade (1988), Spitze (1988), Cigno and Ermisch (1989), Hoem and Hoem (1989), Birg (1991) and Siegers, de Jong-Gierveld and van Imhoff (1991).

The *unit time period* can be thought of like budget years for businesses. Families are often viewed as making long-term plans (stretching over their planning horizons) which consist of subplans for consecutive unit time periods. Sometimes it is also assumed that the family budget must be balanced for each unit period. Alternatively, conditions for intertemporal borrowing and saving can be specified as well as conditions for how a family values resources deferred to or borrowed from other unit time periods.

In models of female labor supply, both the budget constraints and the utility functions are usually specified for families. This is because family members are viewed as pooling and sharing their resources so as to maximize family well-being. However, by virtue of the physical nature of human existence, the time constraints in these models are for the individuals whose labor supply decisions are being analyzed and are specified with respect to the unit time period.

In models of female labor supply, the treatment of time is crucial for the definition and interpretation of the effects of children. It is also crucial for differentiating among one-period, life cycle, static and dynamic models. It would be easier to ascertain how time is treated in specific studies if the terminology for different types of variables were more specific with respect to the presumed time attributes of the variables as well as the way in which the variables are thought to relate to the decision processes of interest. The following nomenclature for variables that might help to meet this need.

All family decision models treat subsets of the full array of decisions that real families make. We will refer to the specified decisions treated in a model or study as a *decision complex*. Variables can be categorized by whether or not they are outcomes of a given decision complex. Those that are we propose to classify as *choice variables*. Those that are not are *exogenous variables*.

Choice variables can be classified by the timing of the relevant decisions. Variables for decisions made prior to the current unit time period, t, could be called *predetermined variables*. All variables can also be classified by when the outcomes of associated decisions or the exogenous circumstances are realized, and hence observed. Depending on whether the point of realization is prior to, during, or after the current unit period t, variables could be referred to, respectively, as *lagged*, *current*, or *future*.

As with current practice for classifying variables, in applying our nomenclature from study to study, the same variables could be designated as endogenous or predetermined or exogenous. For example, consider the current earnings of the male partner. In a model of female labor supply where it is postulated that the male and female partners decide together each unit period how much each should work, the current earnings of the male will be a current endogenous variable. If men are regarded as having careers, with their current earnings primarily reflecting the outcomes of previous human capital investment and lifestyle choices and the customary hours of work and wage profiles for particular occupations, then the current earnings of the male partner would be viewed as a current predetermined variable. If male earnings are viewed as being determined separately from the labor supply of their female partners, then the earnings of the men would be viewed as exogenous.

Likewise, when child status variables are included to reflect fertility outcomes, in annual models they will be predetermined (births follow pregnancies) or exogenous. However, when these variables are included as proxies for current demands on the resources of families, with the level of care for children largely a matter of choice, then the child status variables should be treated as proxies for current endogenous variables.

The terms that have been defined are convenient for distinguishing static versus dynamic, and one-period versus life cycle, models.

Dynamic models are models in which lagged endogenous or predetermined variables are hypothesized to affect the determination of the current endogenous variables, and where the intertemporal linkages are modeled in at least a reduced form sense. In other words, these are models in which *past* decisions about *choice variables* are hypothesized to affect current period decisions. The intertemporal causal chain could be quite complex. For example, future wage expectations could be hypothesized to have affected past schooling choices, with years of schooling affecting current wage offers and current labor supply.

Models where current decisions are not specified to be affected by choices made in previous periods are *static models*.

A model for a unit time period is called a *one-period model*. If the unit time period is the entire life cycle planning horizon, then this is also a *life cycle model*. One-period models for which the unit time period is a shorter period, such as a year, are also still referred to sometimes as life cycle models where the utility function for the unit period is conceptualized as a subfunction of a life cycle utility function. For reasons of data availability, the most common length for the unit time period is a year.

With a unit time period such as a year, we find it useful to think of child status effects as falling into two time-related categories: direct and indirect effects. The *direct effects* are defined to be the current period labor supply responses of women to the time, effort and other resource requirements of having and caring for children. Through their actual and anticipated direct effects on current labor supply, children may also affect factors such as the accumulation of work experience that, in turn, affect the future labor supply behavior of the mothers. These child-related effects on labor supply that operate, over time, via other variables are termed *indirect effects*.

2. Indirect Effects of Children

Actually, a better term for indirect child status effects might be "career implications of alternative lifestyle preferences and expectations". For the most part, these are hypothesized responses to expectations of future labor supply, taking account of expected future family responsibilities. Three sorts of hypothesized indirect effects are distinguished. It is often suggested that these child-status effects are transmitted, in part at least, via the offered wage. Hence the large body of literature on offered wage effects on female labor supply is also selectively reviewed.

Personal and Family Investment Decisions

During their child rearing years, if women devote less time to market work than most men, then the women will accumulate less job experience. Also, formal job-related training is usually obtained at some cost. Mincer and Polachek (1978) argue that, since the returns on these investments are only realized through market work, the less an individual expects to work, the weaker the incentives will be to augment job skills over the life cycle.

In summary, it is argued that most women choose to work less during their child rearing years and that, anticipating this, they invest less in job-related training. Hence most women are expected to have lower wage rates than men with the same levels of formal education because their stocks of job-related human capital are smaller.[9]

Occupational Choice

Polachek (1981) specifies a model of occupational choice in which an individual's stock of human capital "atrophies" during years of home time at a rate determined by the type of human capital. He hypothesizes that women anticipating intermittent work due to family responsibilities will tend to choose occupations for which the required types of human capital have lower atrophy rates. H. Zellner (1975) notes that many types of jobs with steep earnings-experience profiles are thought to have lower starting wages because employees share the costs of specific training in this way. She argues that women who anticipate dropping out of the work force to raise children will not select jobs of this sort because the

[9]Other studies providing discussion or evidence of anticipatory effects of children on human capital investment (and occupational choice) include Becker (1981), L.B. Shaw (1983), Dolton and Makepeace (1987), Simpson (1986), Goldin (1990), K.L. Shaw (1989), Blau and Ferber (1991), and Ermisch and Wright (1991).

expected appreciation in earnings will not provide adequate compensation for the lower starting wages.

The implications of these anticipatory, child-related effects are that women will tend to be more constrained in their occupational choices than men, and will tend to have lower wage rates because of the nature of these constraints.[10]

Employer Effects

If women work intermittently, they will accumulate less job experience than most men. It is hypothesized that rational employers will pay relatively lower wage rates to workers with less experience, and also will be less inclined to make training investments in workers who are expected to supply less labor in future time periods.

According to the theory of statistical discrimination, even women who do not (and may never) have children, and who plan to (and will) work as much in years to come as any man, may still be passed over by employers when it comes to training and other career advancement opportunities. The alleged reason is that these women cannot be distinguished reliably and cheaply enough from other women.[11]

Offered Wage Effects

For the indirect effects discussed above, the relevant child status "treatments" are, at best, loosely related to observable child status variables. These indirect effects of children can be redefined, however, in terms of direct effects of observable factors such as the amount of previous work experience, occupational specialization, and the offered wage. Taking this approach, research on indirect child status effects could proceed in two

[10]For further discussion and empirical evidence concerning these hypothesized effects of children, see Bergman (1986), Corcoran, Duncan and Ponza (1983), Cox (1984), England (1984), Mincer and Ofek (1982), Mincer and Polachek (1978) and Polachek (1981).

[11]See, for example, Phelps (1972) and Bergmann (1986) for discussions of statistical discrimination.

phases. Phase one would involve determining the effects of children on women's offered wage rates and any other observable factors through which indirect child status effects are believed to be transmitted. Phase two would then involve determining the impacts of the(se) observable factors on female labor supply.

There is already a considerable body of research suggesting that the child bearing and rearing activities of women (and perhaps even anticipations of future child bearing and rearing) reduce their offered wage rates. If this is the case, then the indirect effects of children on female labor supply will be negative if there is a positive relationship between women's offered wage rates and their labor supply, and will be positive otherwise.

A woman's expected labor supply can be expressed as the probability she will work in the designated time period times her expected hours of work if she is employed. Models of female labor supply typically imply that a woman will work in a given unit time period if her offered wage exceeds her reservation wage evaluated at zero hours of work (as specified in Condition (III-10)). Therefore, factors that decrease (increase) a woman's offered wage are expected to decrease (increase) the probability that she will engage in market work in the unit period.

The nature of the effect of a wage change on the hours of work of a working woman is hypothesized to depend on the trade-off between income and substitution effects. Two lines of argument suggest that, for many women, negative income effects may be weak in comparison with positive substitution effects.

According to Mincer (1962), women with families divide their time among work at home ($h_{f,F,t}$ in our notation), work in the market ($h_{f,W,t}$), and leisure ($h_{f,L,t}$). If substitution is possible between the female partner's time and market-produced goods and services (like day care), it is argued that the income effect of a given wage change should be weaker for a woman than for an otherwise similar man who divides his time between work in the market and leisure (with $h_{m,F,t} \simeq 0$).

The second line of argument is simpler. Holding hours of work constant, the dollar value of the change in earned income associated with any given wage change will be smaller the shorter the hours of work are. Hence, for women who work less because of child care responsibilities, the income effects associated with any given wage change should be weaker than for men working longer hours.

If the negative income effects of a wage change are weaker for most working women than for men, and if the positive substitution effects are similar for both, then the hours-wage relationship should be more positive for women than for men. Some have argued that the net effects for most working women are positive, and possibly large in magnitude. For example, Blau and Ferber write:

> With respect to the hours decision, empirical evidence indicates that for men the income effect generally tends to offset or even dominate the substitution effect and that they do not decrease, or may even increase, the amount of nonmarket time as their wage rate goes up.... The situation is quite different for women ... empirical studies for the most part find that women's labor supply is strongly positively related to the wage rate.
>
> (Blau and Ferber, 1986, p. 95)

Nevertheless, estimates for married women of the impact of a wage change on hours of work that are small and sometimes negative have been obtained in a number of studies. Using micro data from the 1971 Census of Canada for married women in the five-year age groups of 20-24 through 45-49, Nakamura, Nakamura and Cullen (1979) obtain estimated wage elasticities in the same range reported by other researchers for men. Of course, married women in Canada and the U.S. might differ in their work behavior because of country-specific factors like tax laws. In order to investigate this possibility, Nakamura and Nakamura (1981) present tax-corrected results based on micro data from the 1970 U.S. Census and from the 1971 Census of Canada. Others have suggested that perhaps wives working full-time have wage responses similar to those of men, but that wives working part-time will exhibit large positive elasticities. However, Nakamura and Nakamura (1983, p. 246), obtain similar elasticity estimates for wives working part-time and for those working full-time. Killingsworth (1983, pp. 200-201) suggests the results presented in Nakamura, Nakamura and Cullen (1979) and in Nakamura and Nakamura (1981) may be "aberrant" because work experience is not directly controlled for. This conjecture is explored empirically and dismissed in Nakamura and Nakamura (1985b, pp. 180-190 and pp. 278-293). On the basis of these findings and the findings of others,[12] we conclude that the wage responses of hours of work for working married women are modest, as for men, and possibly negative. If this is true, the indirect effects of children on the hours of work of working women might be expected to be modest also.

This does not mean, of course, that child-related effects on women's earnings are small. If children have negative effects on women's wage rates, this will lower their earnings even if wages have no effect on labor supply. Also, this does not mean that the indirect, offered wage effects of children on the probability (as opposed to the hours) of work are small.[13]

3. Direct Effects of Children

Direct effects of children have been defined as the current period labor supply responses to the time, effort and other resource requirements of having and caring for children. There are two initial sets of considerations in attempting to measure direct child status effects. The first is that the model must include child

[12]Findings similar to ours are reported for the U.S. and Canada by others including Robinson and Tomes (1985), Smith and Stelcner (1988) and Mroz (1987). Also, similar estimates for men and for women have been obtained using data from negative income tax experiments (see Killingsworth, 1983, pp. 398-399, Table 6.2 for a list of these studies).

[13]For a wealth of information on trends in twelve countries in female wage rates, participation and employment rates, hours of work, fertility, and family formation and breakdown, see the special issue of the *Journal of Labor Economics* edited by Layard and Mincer (1985), and Mincer's (1985) introduction. A variety of approaches are pursued for trying to obtain estimates of wage, child-related, and other effects on female participation versus the hours of work for working women. These approaches range from the cohort analyses of Smith and Ward (1985) and Joshi, Layard and Owen (1985) to attempts to include selectivity corrected wage predictions for all women in cross-sectional micro data equations for the probability of work (see also Hartog and Theeuwes, 1986). The problem of finding variables that can be included in Z or E but excluded from Z^* (that is, variables that affect women's market wage rates but not their reservation wage rates) is evident in several of these studies.

status variables that can capture the hypothesized direct effects of children. The second is that indirect child status effects must be adequately controlled for because, otherwise, indirect effects may be captured as well by the child status variables.

The empirical separation of direct from indirect effects is important because indirect effects are "sunk effects" that could not be immediately altered by changes in current child status. For example, suppose a young child starts school. This will free the mother's time during the school days, and this current period child status effect may increase the likelihood of the woman working in the current period: a direct child status effect. Suppose also that the mother has not worked since the birth of this child; hence she has not added to her stock of job-related human capital through work experience during this period. Her job skills may even have "atrophied" during her years out of the labor force. Because of this, the wage offers this woman can command may be quite low, and for this reason she may not be interested in working full time now. The woman's low wage situation, and the wage effects on her labor supply, cannot be altered in the current period through any sort of variation in her child status (including even the death of the child starting school). Hence, if this indirect effect on the woman's labor supply is picked up, in part, by the child status variables, the estimated coefficients of these variables will not properly reflect expected current period variations in labor supply in response to current period variations in child status.

The remainder of this paper focuses on the estimation of direct child status effects.

VII. REPRESENTING DIRECT CHILD STATUS EFFECTS IN MODELS OF FEMALE LABOR SUPPLY

If the level of care parents provide for their children is, in part, a choice variable, the behavioral model introduced in subsection III.1 is inadequate as a conceptual framework for studies focusing on direct child status effects. The deficiencies of this model begin with the utility function.

The utility function specified in subsection III.1 is

(III-1) $$U = U(G_t,\ h_{f,NW,t};\ y_{m,t},\ A_t,\ N_t,\ Z^*_t).$$

DeTray (1973) suggests introducing "child services" into this utility function. He hypothesizes that the flow of child services is a function of two "intermediate products": per child quality (denoted by Q_t) and the current number of children. Surely, though, the satisfactions parents derive from children are also affected by other attributes determined through the birth process such as the age range and spacing of children. Some of these other birth-determined attributes may also affect the per-child costs of producing "child quality". In order to establish a conceptual link between the arguments of the utility function and the cost and time factors entering the family budget and individual time constraints, a child services "production" relationship is defined as

(VII-1) $$C_t = C(N_t,\ Q_t).$$

In (VII-1), N_t is a vector of child attributes (including, but not limited to, number) determined through the birth process and Q_t denotes attributes shaped by how children are raised, such as school performance, emotional stability and health. Also a "production function" for child quality is defined as

(VII-2) $$Q_t = Q(h_{f,F,t},\ G_c;\ y_{m,t},\ A_t,\ N_t,\ Z^*_t,\ Q_{t-1})$$

where $G_{c,t}$ is the portion of the Hicksian consumption good devoted to child rearing and $G_{\bar{c},t}$ is what is used for everything else, and where the lagged child quality vector (Q_{t-1}) is included in recognition of the developmental processes for children's personalities, abilities, and even physical health. We will not follow DeTray in specifying a current production relationship for N_t. Rather (following the example in subsection VI.1) N_t will be treated as predetermined.

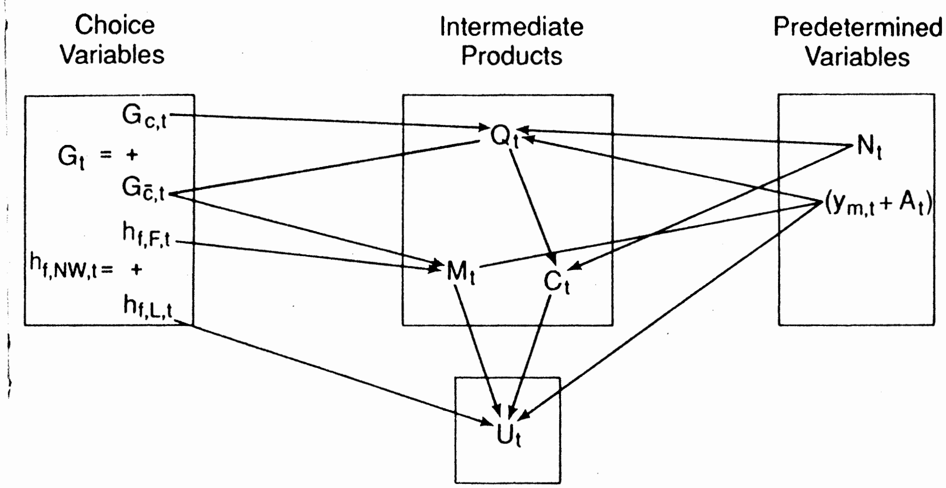

Figure 1. Inputs into Family Utility

Simplistically, we will represent the process whereby the portion of the Hicksian consumption good not devoted to child rearing ($G_{\bar{c},t}$) is transformed into material services (denoted by M_t) as

(VII-3) $$M_t = M(G_{\bar{c},t},\ h_{f,F,t};\ y_{m,t},\ A_t,\ Z^*_t).$$

Now the utility function can be represented as

(VII-4) $$U_t = U(C_t,\ M_t,\ h_{f,L,t};\ y_{m,t},\ A_t,\ N_t,\ Z^*_t,\ Q_{t-1}).$$

The time constraint is still (III-2) with $h_{f,F,t} + h_{f,L,t} = h_{f,NW,t}$, and the family budget constraint is still (III-3) with $G_t = G_{c,k} + G_{\bar{c},t}$.

The causal interactions specified in relationships (VII-1) through (VII-4) are depicted in Figure 1, with the predetermined factors in Z^* and Q_{t-1} ignored for expositional simplicity.[14]

[14]See also Becker and Lewis (1974) on interactions between the number of children and their quality. Functions for household production technologies have a history, as well, in the demand literature. Gorman (1976) generalized commodity-specific adult equivalent scales into linear household technologies. Lewbel (1985, p. 6) points out the advantages of this approach when the production technologies for intermediate goods differ depending on household characteristics.

There are no ready measures for either child or "material" services. However, introduction of these services into the utility function and the conceptual specification of production relationships for of these services draws attention to the proxy roles of the included child status variables in models of female labor supply. It also helps to clarify the types of benefit and cost factors that should be kept in mind in choosing child status variables, and why the number of children ever born is not the most relevant child status attribute with respect to the determination of female labor supply.

VIII. CHILD STATUS VARIABLES

In some of the early micro data studies of female labor supply there is a clear recognition that both time and material investments in children may have important effects on female labor supply. It is also recognized that these effects will surely differ depending on the maturity as well as the number of children. In their classic 1969 study Bowen and Finegan write:

> The presence of children can be expected to influence the labor force participation of a married woman in three ways: (1) by increasing the amount of work to be done in the home; (2) by increasing the family's need for money income; and (3) in the case of older children, by providing a source of assistance with home tasks. These considerations pull in different directions.... The quantitative impact of each of these considerations, and thus the overall net effect of children, should be expected to vary according to both the ages and number of children....
> (Bowen and Finegan, 1969, p. 96)

Bowen and Finegan did not have direct expenditure and time use information with which to quantify or check their a priori notions. Instead they amassed evidence on the net associations between various aspects of the child status and labor supply of married women. Their evidence and conclusions have influenced the types of child status information collected in household surveys, and the choice of child status variables in subsequent studies of female labor supply. Based on micro data from the 1960 U.S. Census, Bowen and Finegan conclude:

TABLE 2

PROBABILITY OF LABOR FORCE PARTICIPATION FOR AN AVERAGE[a] MARRIED WOMAN WITH AT LEAST ONE CHILD UNDER 6

Age of youngest child	Participation rate
Less than one	10.5
One	18.8
Two	21.6
Three	23.4
Four	27.5
Five	28.5

Source: Bowen and Finegan (1969, p. 102).

[a] Average in terms of the values of the other variables (besides the child status variables) included in the multiple regression for which estimated coefficients are shown in Bowen and Finegan (1969, Appendix, Table A-10).

> First, it is the presence of children under 6 which takes precedence over all other aspects of the age distribution of children in determining the mother's labor force status....
> (Bowen and Finegan, 1969, p. 98)

Bowen and Finegan conducted a further analysis limited to married women with at least one child younger than 6. Their results are reproduced in Table 2. A new baby is found to have the biggest negative impact on a mother's probability of work.

Bowen and Finegan also suggested there might be important interaction effects. Other researchers have attempted to empirically explore these hypothesized child-related interaction effects.[15] However, in studies that allow for certain interactions,

[15]Studies allowing for demographic interaction effects or attempting to take account of child costs and other child-related expenditure effects include Hill and Stafford (1980), Strober and Weinberg (1980), McLaughlin (1982), Weinberg and Winer (1983), Hunt and Kiker (1984), Johnson and Pencavel (1984), Juster and Stafford (1985), Salkever (1987), Blau and Robins (1989), Douthitt and Fedyk (1988), Leibowitz, Waite and Witsberger (1988), Nock (1988), Blank (1989), Graham and Beller (1989), Mason and Kuhlthau (1989), Presser (1989), Klerman and Leibowitz (1990), and Trzcinski and Alpert (1991).

other theoretically possible demographic interactions are simply ignored. Blundell and Meghir (1986, p. 55) note the tendency in labor economics to adopt highly restrictive functional forms. In the consumer demand area, researchers have experimented with a wide range of functional forms for introducing demographic effects, to the extreme of allowing all the response parameters for other variables to differ depending on family type.[16]

Nevertheless, degrees of freedom limitations and preferences for relationships that are relatively simple to interpret mean that, in practice, there is a tradeoff between the number of demographic attributes allowed for and the complexity of the specified functional relationships. In the demand analysis literature, researchers have opted for less restrictive functional forms but parsimony in the choice of demographic variables. Often only a single demographic variable, the number of children (or family size), is considered.[17] What is needed is an empirical research strategy for determining *which* dimensions of family child status, and *which* possible interaction effects and higher order terms, have the most important impacts on female labor supply. The need for

[16]The oldest and most common method of allowing for demographic effects in consumer demand studies are the adult equivalent scales introduced by Engel (1895). The practice of allowing for commodity-specific demographic effects goes back to Sydenstricker and King (1921), and has been extended and popularized by, among others, Prais and Houthakker (1955), Gorman (1976), Muellbauer (1977), Pollak and Wales (1981), and Ray (1986).

[17]For example, Pollak and Wales (1981) allow for families with one man, one woman, and 1, 2, or 3+ children; Barnes and Gillingham (1984) distinguish families with 1, 2, or 3+ children; and Rossi (1988) differentiates families by size (1, 2, 3, 4, 5, or 6+ persons) as well as other variables not directly related to child status. Nicol (1989) stratifies his data by the number of children. In a few studies, the potential importance of allowing for ages of children is recognized. For example, Muellbauer (1977) takes account of the numbers of children 0 to 5, and 5 to 16 years of age; and Blundell and Meghir (1986) and Blundell and Walker (1986) take account of the presence and/or the number of children in the age groups of 0 to 4, 5 to 10, and 11 to 18. Deaton and Muellbauer (1980, pp. 191-213) provide a very helpful overview of the treatment of demographic effects in the demand analysis literature.

such a strategy is even more apparent when intertemporal effects are considered.

IX. ALLOWING FOR INTERTEMPORAL EFFECTS

The degrees of freedom problem, and the problem of keeping models sufficiently simple to facilitate understanding, become even more severe when intertemporal effects are considered. Pencavel draws attention to this as a general limitation of the life-cycle approach:

> The empirical implementation of the life-cycle model would appear to require a great volume of data: to understand an individual's labor supply today, the economist needs information on prices and wages throughout the individual's life!
>
> (Pencavel, 1986, p. 46)

Pencavel summarizes the two main approaches economists have devised for dealing with the data requirements of life cycle models:

> One derives from the literature on habit persistence and stock adjustment and specifies the individual's utility function in period t as conditional on the individual's consumption and hours of work in the previous period. The notion that the standards by which individuals gauge their welfare are molded by their prior experiences is, of course, an old one.... Whereas in this specification the lifetime utility function is intertemporally not (strongly) separable, the opposite hypothesis is maintained in the second approach to the individual's life-cycle labor supply problem.
>
> (Pencavel, 1986, p. 46)

Separability is a property of the preference structure. If some of the arguments of a utility function are separable from the rest, this means that the preference structure for those goods and services is independent of the relative quantities consumed of all of the other goods and services. A subutility function can be specified for the goods and services in each separable group, and the subutilities can be combined (additively, when what is called "strong" or "additive" separability is assumed) to give total utility. Moreover the demand for any good or service in a group that is (at least weakly) separable can be expressed as a function of prices for the goods and services in just that group and a given level of expenditure on that group of commodities. Deaton and Muellbauer discuss the role of intertemporal separability in empirical demand analysis studies:

> ... we may wish to assume that preferences are weakly *intertemporally* separable so that goods in each period form a closely related group with only general relations between periods .... the adoption of the assumption allows the expression of commodity demands in each period as a function of total outlay and prices in that period alone in accordance with usual practice.... Note too that there is no other way to justify this; regressing demands on current variables rules out any specific interaction between commodities consumed in different time periods.
> (Deaton and Muellbauer, 1980, pp. 124-125)

However, in labor supply models "total outlay" in the current period depends on labor supply in the current period which, in turn, may depend on wage rates in the current and other time periods. In fact, much of the interest in dynamic models of labor supply stems from interest in possible intertemporal shifting of period-to-period labor supply depending on variations in the offered and reservation wage rates (some of which might be due to changes over the life cycle in child status). Thus even the imposition of strong separability, with exogenously determined market wage rates,[18] does not permit wage effects from other time periods to be entirely ignored. Nevertheless, under some conditions these assumptions allow other-period wage effects to be parsimoniously summarized, as in the model presented by Heckman and MaCurdy (1980).[19]

In justifying labor supply models embedding intertemporal separability assumptions, Killingsworth (1983, p. 221) asserts that: "For many purposes, this assumption of 'intertemporal separability' ... is fairly innocuous." And, in commenting on the practice of

[18]See Killingsworth (1983). K.L. Shaw (1989) estimates intertemporal labor supply elasticities for male workers using a model which allows for endogenous human capital accumulation.

[19]For the development and application in the consumer demand literature of demand functions of this sort see Browning, Deaton and Irish (1985) and Browning (1989).

assuming additive separability in life cycle models of consumption and labor supply, Deaton and Muellbauer (1980, p. 311) write, "in intertemporal choice, we should perhaps not expect very strong specific substitutabilities and complementarities between periods, so that in this context additivity is likely to be more acceptable."

Surely, though, if a couple have developmental goals for their children, then time and material investments in these children in one year must be strongly complementary with investments in other years. Concern for year-to-year stability and consistency in how children are raised comes through even in the conventions for establishing child support levels when parents divorce. The guiding principle is to try to maintain support of children in the manner to which they have been accustomed.

Some of these concerns might be addressed by treating children as "durables". Commenting on the analysis of the demand for durable goods, Diewert writes:

> If rental markets for all consumer durables exist and the consumer merely rents the services provided by a consumer durable, then we may apply the standard model.... But many consumers typically purchase consumer durables rather than rent them. Thus in order to apply the standard theory, our approach will be to decompose the consumer's purchase of a durable into a rental part and an investment in an asset part.
> (Diewert, 1974, pp. 503-504)

Of course, in the case of children, there is no rental market. Nor is there a "used good" market for children, or "free disposal" of excess quantities (except prior to birth).

There are a number of other relevant differences. For instance, in the investment literature costs of a durable good consist of the purchase price (or, more appropriately, the opportunity cost of this outlay) and maintenance and enhancement (or replacement) expenses. For children, the "purchase cost" is the expense of the pregnancy and delivery. The larger, and largely discretionary, costs of children are the maintenance and enhancement costs. The acquisition of a child is a one-time decision; but the choice of how to raise a child is an ongoing sequence of expenditure decisions stretching out over much of the adult lives of the parents. These arguments suggest that the habit persistence models of Pollak (1970) and others in the demand analysis literature, or some variant of the state variable approach of Houthakker and Taylor (1970), might provide a more suitable basis for simplifying assumptions about intertemporal behavior in models of female labor supply than the additive separability approach to this problem.

Habit persistent models including lagged dependent variables as explanatory variables raise both econometric and conceptual problems. The conceptual problems have their origin in the fact that utility functions have traditionally been specified in unconditional form; that is, a preference map is specified without reference to the standard of living to which an individual or family has been accustomed. Because of this, the theoretical implications that are derived are also unconditional in nature, and it is difficult (and perhaps inappropriate) to test these properties in an empirical model incorporating lagged dependent variables as conditioning factors. Among other econometric problems, the coefficients of wage rate and other variables in models containing lagged dependent variables cannot be regarded as unbiased estimates of the corresponding theoretical effects in an unconditional model of the sort that is usually specified as the theoretical context for the empirical work. Deaton and Muellbauer discuss these tradeoffs in the demand analysis area:

> In the face of these difficulties, and given the ambiguity of interpretation of the lagged dependent variable, it might be tempting to revert to the estimation of models ... which relate consumption directly to its ultimate determinants--dispensing with the lagged dependent variables.... However, such a procedure is extremely dangerous ... the inclusion of the lagged variable will largely protect investigators against the most obvious spurious correlations. However, the frequent exclusion of lagged dependent variables on the grounds that their entry causes the relationship to "collapse" is a testimony both to the effectiveness of the cure and to the greater difficulty of finding a theoretically satisfactory relationship that outperforms a simple autoregression.
>
> (Deaton and Muellbauer, 1980, pp. 332-333)

X. CHOOSING AMONG ALTERNATIVE APPROXIMATE MODELS

What is needed is an empirical research strategy for determining *which* dimensions of family child status, and *which* possible interaction effects and higher order terms (including intertemporal effects) have the most important impacts on female labor supply. Accepted model choice procedures, as explained in econometrics texts such as Kmenta (1986), rely on narrowing the list of candidate models by eliminating those with statistically significant estimated properties that violate implications of maintained economic theory. This practice is at odds with efforts to check and modify maintained theories on the basis of empirical findings. Moreover, even without problems of inconsistency, as already noted when estimated models are approximations it is often difficult to establish a correspondence between the estimation results and theoretical expectations.

A second, and increasingly popular, mainstay of model choice practices is the elimination, or modification, of models that fail statistical tests for properties required to insure that the chosen estimation method yields consistent parameter estimates. Yet the relevant question is not the existence, but the *extent*, to which departures from these statistical properties cause deteriorization in the goodness of the predictions or behavioral insights provided by alternative empirical models (versus the performance, say, of "corrected" empirical models). Most specification error tests are not helpful in addressing this question, as discussed in Nakamura and Nakamura (1985c).

There would also be advantages if economic theory and evidence about possible statistical problems could be used to broaden, rather than to limit, the list of candidate models. This orientation adds urgency to the need to develop model choice methods for *ranking* alternative imperfect models. Procedures for evaluating empirical models based on the correspondence between the actual and predicted values of the dependent variable(s) can be used for this purpose, as discussed at length in Nakamura, Nakamura and Duleep (1990) and as is explicitly recognized by Sawa (1978), Leamer (1978), and Vuong

(1989). Output space evaluation methods focus on the predicted versus the actual values of the dependent variable. A better fit, in some specified sense, is taken to be evidence of a better model.

Variants of Pearson's chi-square statistic have been used in some studies for making output space comparisons of models of female labor supply.[20] The first application of this sort that we are aware of in the labor economics literature is due to Heckman (1981). He uses this approach to rank alternative models for the simple binary choice of working versus not working each year. One of the models Heckman considers in this study is explicitly approximate in that it incorporates a proxy explanatory variable. Nakamura and Nakamura (1983, 1985a, 1985b) extend this approach to accommodate models involving multiple discrete choices, continuous dependent variables, and variables that are functions of the outputs of multiple behavioral relationships. Heckman and Walker (1988) use this and other approaches in examining and ranking alternative models of fertility. Several of the models compared in these studies are nonnested.

There is more promise for output space methods in a micro data environment than in the macro time series setting where these sorts of methods first evolved. In a micro data environment, often there are sufficient data that some portion can be used for estimation and the rest can be reserved for immediate out-of-sample testing. The out-of-sample data are not usually as strongly autoregressively related to the in-sample data as with macro time series data. Thus the out-of-sample tests that can be carried out are more informative. Also, as noted in Nakamura and Nakamura (1985b, p. 198), in a micro data (and particularly in a panel data) environment it is often the case that not all of the information in the in-sample data is used in model estimation. In this situation, even appropriate in-sample predictive tests may provide useful insights as to the relative goodness of estimated models.

[20]See Andrews (1988), Bagozzi and Yi (1988), Heckman (1981, 1984), Heckman, Hotz and Walker (1985), Heckman and Walker (1988), Nakamura and Nakamura (1985a, 1985b), and Heckman, Nakamura and Walker (1991).

Output space model evaluation methods have been criticized on the grounds that they focus exclusively on the predictive abilities of models. Applied economists are often more concerned about behavioral responses. Yet there is a relationship between predictive ability and the ability of a model to properly capture responses to the included explanatory variables. An estimated model (including the assumed and estimated properties of the error term) should be able to reproduce key features of the observed distribution of the dependent variable conditional on observed in-sample and out-of-sample values for the explanatory variables. If it cannot, then at the very least there is probably some difficulty with the specification of the properties of the error term, which may mean that the standard tests of significance are inappropriate.

Comparisons between the predicted and actual distributions of variables are inherently multidimensional. In analyses of female labor supply, there may also be interest in comparisons for several dimensions of the employment and earnings behavior of women. The limitations of single statistic methods of model evaluation and choice are acknowledged by even some of those who have developed these methods. There is an older tradition in econometrics of multidimensional output space evaluation methods that are descriptive and exploratory in nature. The spirit of this approach was to develop a better understanding of aspects of alternative models that might matter for particular applications, and to determine where research should be directed in order to improve the usefulness of particular models. These are different, though related, objectives from general investigations of specification problems or attempts to determine the range of alternative models that yield essentially the same inferences (a Leamer-style investigation of fragility).

However, multidimensional evidence, by its nature, does not yield a simple decision rule for accepting or rejecting, or for ranking, alternative empirical models. In our book, *The Second Paycheck*, we argue for a courtroom-style approach to the problem of choosing among, or ranking, alternative empirical models on the basis of multidimensional evaluation evidence:

> Our behavioral conclusions all rest on the accumulation of circumstantial evidence.... Moreover, the issue of when the weight of accumulated evidence is sufficient to warrant a particular conclusion is treated as a matter of judgment. In a courtroom proceeding, eyewitness reports, expert testimony and various sorts of circumstantial evidence may all be brought before the court, but it is the ultimate responsibility of a judge or jury to weigh this evidence and reach a verdict. In a study ... in which there is uncertainty about the proper specification of the functional forms of the behavioral relationships, about the distributions of the disturbance terms, and so forth, we do not believe that better conclusions will necessarily be reached by avoiding the degree of arbitrariness inherent in judgmental decision making by appeals to mathematical statistics predicated on assumptions that cannot be checked.
>
> (Nakamura and Nakamura, 1985b, p. 365)

We feel that prior information about the aspects of an approximation that are most crucial for a particular application, or concerning the interpretation of the empirical findings, should enter into the "court-room" consideration even if it is not obvious how this information can be summarized in the form of prior distributions or a mathematically specified loss function. We also feel that the potential for exploring model outputs from different perspectives should be exploited as fully as possible. Indices and other evidence of model fit should be examined for interesting subsets of the data as well as for the entire data set, out-of-sample as well as in-sample, at different levels of aggregation, and for meaningful combinations of model outputs (such as annual earnings defined as the product of the hourly wage rate times annual hours of work) as well as for individual dependent variables (such as annual hours of work), with the intended uses of the model guiding this investigation. This is the approach adopted, for example, in Nakamura and Nakamura (1985b) and in Heckman and Walker (1988).

In his Foreword to *The Second Paycheck*, Heckman (1985) explains his views on a courtroom approach to model choice in a labor economics context:

> The approach pursued in many recent studies of labor supply has been to arrive at final, empirical specifications for a single demographic group by means of a battery of "t" and "F" tests on the coefficients of candidate variables. The problem of pretest bias [the multiple tests problem] is conveniently ignored. Only rarely ... do analysts ask how well fitted micro relationships explain other aspects of labor supply such as the aggregate time series movement....
>
> This book does not adopt the conventional "t" ratio methodology. The authors estimate the same models for a variety of age, marital status, and sex groups and look for commonalities in findings across groups. They look for consistency in the impact of explanatory variables on different dimensions of labor supply. Models are simulated both within samples and out of samples.... The simulation format has the additional feature of spelling out the implications of complex models that are not obvious from reported coefficient estimates. The rigorous body of tests proposed and implemented by the authors of this book ... sets a new, high standard that will be followed by all serious scholars of the subject.
>
> (Heckman, 1985, pp. xi-xii)

IX. AN OUTPUT SPACE APPRAISAL OF THREE MODELS

In this section, descriptive output space methods are used to compare and evaluate three models of female labor supply incorporating different amounts of information about work in the previous year.

1. The Models

In all three of the models considered, probit relationships describe the probability of work, and linear regression relationships are used for the log wage rates and annual hours of work for women who do work. In addition to a fairly standard set of control variables, the probit indices for the probability of work and the hours equations include a "baby dummy" set equal to 1 if a woman has a new baby, and a "young child dummy" set equal to 1 if a woman's youngest child is younger than 6 but was not born in the current year.[21] Heckit selection bias terms are included in both the wage and hours equations.

What we term the *Standard Model* incorporates no information about work behavior in the previous year (like the model in III.1). The *Dummy Model* is the same as the Standard Model, except that the equations also include a dummy variable set equal to 1 if a woman worked in the previous year, and set equal to 0 otherwise. Finally the *Inertia Model* consists of the same equations as the Standard Model, with these equations estimated separately for women who did not and for women who did work in the previous year. Also lagged wage rate and hours of work variables are included in the Inertia Model equations for women who worked in the previous year. The specifications for these models, estimation details, summary statistics for all variables, and comparative estimation and simulation results are given in Nakamura and Nakamura (1985b). All three of these models were estimated using pooled data from the Michigan Panel Study of Income Dynamics (PSID) for married women aged 21-46 for the years of 1971-1978.

Table 3 summarizes what the three models predict about the probability that a 25-year-old married woman will continue working given that in the previous year she worked 2,200 hours at a wage of $4.00 per hour (in 1967 dollars). The top number in each group of three is for the Standard Model, the next is for the Dummy Model, and the bottom one is for the Inertia Model. It can be seen that the predicted negative impacts of a new baby on the probability a woman will work another year decline dramatically the more information a model includes about work behavior in the previous year.

From Table 4 we see that the Inertia Model predicts large differences in the probability that a woman will work depending on

[21]The baby dummy equals 1 if the number of children in the family aged 0 to 17 has increased by 1 since the previous year, and if the youngest child is 23 months of age or less. Otherwise this variable equals 0. The young child dummy equals 1 if the youngest child in the family is less than 6 and if the baby dummy equals 0. Otherwise this variable equals 0.

TABLE 3

PROBABILITY OF WORK ESTIMATES FOR A 25-YEAR-OLD WIFE WITH THE DESIGNATED CHARACTERISTICS[a], BASED ON THE STANDARD, DUMMY AND INERTIA[b] MODELS

	No children	One new baby
18 years of schooling		
Husband's income[c]		
$4,000	.94[d]	.82
	.95	.90
	.97	.92
$7,900	.90	.75
	.94	.88
	.97	.92
12.6 years of schooling		
Husband's income		
$4,000	.79	.57
	.90	.80
	.96	.88
$7,900	.72	.49
	.87	.77
	.96	.88

[a] In addition to the characteristics specified above, it is assumed that the woman was also married in the previous year, that she is not black, that she did not receive Aid to Families with Dependent Children, and that the federal unemployment rate is 3.4 percent and the value of the national wage index (an index of compensation per worker hour with 1977=100) is 91.9.

[b] For the Inertia Model we also assume the woman worked 2,198 hours in the previous year for an hourly wage of $3.95 in 1967 dollars. These are the average hours of work and wage rate reported in Nakamura and Nakamura (1985b, Table 2.7.19, p. 57 and Table 2.7.22, p. 58) for men aged 21-46.

[c] This variable is measured as before-tax earnings in 1967 dollars. The mean value of this variable reported in Nakamura and Nakamura (1985b, p.53, Table 2.7.13) for working wives is $7,900.

[d] The top number in each group of three is computed using the Standard Model, the next is for the Dummy Model, and the bottom one is for the Inertia Model.

TABLE 4

INERTIA MODEL PREDICTIONS OF THE PROBABILITY OF CONTINUING TO WORK FOR A 25-YEAR-OLD WIFE WITH 12.6 YEARS OF SCHOOLING, A HUSBAND WHO EARNS $7,900, A NEW BABY, AND WHO WORKED THE DESIGNATED HOURS IN THE PREVIOUS YEAR AT THE DESIGNATED WAGE RATE[a]

Hours of work in previous year	Wage rate in previous year: $2.24	$3.95
600	.53	.54
1400	.73	.74
2200	.87	.88

[a] All other characteristics of the woman are as specified in footnote a of Table 3.

how many hours she worked in the previous year. This is so despite the fact that all of the variables customarily included in models of female labor supply were held constant in calculating these predicted probabilities.

2. Output Space Evaluation Results

The estimated models have very different behavioral implications, but which one is the best model?

In Table 5 we show the actual and simulated distributions for years of part-time and of full-time work over the 7-year in-sample period. The first number in each group of four gives the *observed proportion* of women with the designated numbers of years of full-time (over 1,400 hours) and part-time (1 to 1,400 hours) work over the in-sample simulation period. The next three numbers are the *predicted proportions* for the Standard, Dummy and Inertia Models, respectively. The pseudo chi-square values in the bottom row of this table summarize the goodness of fit of the predicted distributions to the observed distribution, with smaller values

TABLE 5

ACTUAL AND SIMULATED PROPORTIONS OF WOMEN 21-46 YEARS OF AGE AND MARRIED IN 1971 BY NUMBERS OF YEARS OUT OF 7 OF FULL-TIME (OVER 1,400 HOURS) AND PART-TIME (1 TO 1,400 HOURS) WORK

Number of years with over 1,400 hours	Number of years with 1 to 1,400 hours: 0	1-3	4-6	7
0	.16[a] .00 .16 .15	.18 .14 .13 .18	.07 .05 .02 .08	.05 .00 .00 .01
1-3	.01 .06 .03 .03	.12 .54 .32 .24	.12 .12 .19 .13	
4-6	.00 .02 .01 .01	.18 .05 .14 .14		
7	.10 .00 .00 .02			

Model	Standard	Dummy	Inertia
Pseudo chi-square statistic	877	251	106

Note: Computations are based on 424 observations.
Source: Nakamura and Nakamura (1985b, p. 409, Table B.24).

[a] The top number is the actual proportion, the next is the simulated proportion using the Standard Model, the next is for the Dummy Model, and the bottom number is for the Inertia Model.

TABLE 6

ACTUAL AND SIMULATED PROPORTIONS OF WOMEN 21-46 YEARS OF AGE IN 1971 BY EARNED INCOME CUMULATED OVER THE 7-YEAR PERIOD OF 1971-1977

Earned income over 7 years (1967 dollars)	Actual	Standard	Dummy	Inertia
Under $0	.17	.00	.16	.15
Less than $10,000	.40	.44	.29	.44
$10,000 - 19,000	.20	.42	.29	.18
$20,000 - 29,999	.12	.10	.15	.12
$30,000 - 39,999	.07	.02	.08	.06[a]
$40,000 - 59,999	.05	.01	.03	.05
$60,000 - 79,999	.00	.00	.00	.00
Over $79,999	.00	.00	.00	.00
Pseudo chi-square statistic		207	37	4
Number of observations			424	

Source: Nakamura and Nakamura (1985b, p. 415, Table B.31).

[a] Values below the line across the table were combined in computing the pseudo chi-square values, as is also done in Tables 7 and 8.

of this statistic indicating a better fit.[22] According to this criterion the Dummy and Inertia Models perform far better than the Standard Model, with the Inertia Model performing best of all.

A more comprehensive picture of model performance can be obtained by examining the actual and predicted distributions of the earned incomes of women cumulated over the 7-year simulation period. The predicted earned income for each woman in each year was set

[22]This is the usual chi-square goodness-of-fit statistic except that the roles of the observed and expected (the simulated) frequencies have been reversed to insure nonzero values for the denominators. See Nakamura and Nakamura (1985a, 1985b).

equal to zero if she was not predicted to work, and was set equal to the product of her predicted wage rate and hours of work if she was found to work. The Dummy and Inertia Model predictions were computed using the predicted, rather than the observed, work behavior in the previous year except in the initial year when actual values for the previous year were used. Summing the predicted annual earnings for each woman over the 7-year simulation period should magnify most types of systematic errors. From Table 6 we see that, according to this criterion, the Dummy and Inertia Models far outperform the Standard Model, with the Inertia Model again performing best of all.

Predictions of the impacts of children on the probability of work and hours of work are of key interest in this survey. Thus in Table 7 we examine the abilities of the Dummy and Inertia Models to capture observed differences in the cumulative income distributions for women who did versus women who did not have a baby or young child over the course of the simulation period. Both models reflect the basic differences in shape for the actual income distributions, with the Inertia Model performing somewhat better than the Dummy Model. (The Standard Model is omitted because of its poor performance in other more basic respects.)

Finally in Table 8 we show out-of-sample simulation results for the cumulative income distribution for a different group of married women over the period of 1981-1982. Again the Dummy and Inertia Models are found to far outperform the Standard Model, with the Inertia Model performing best of all.

The evidence on the continuity of female labor supply that is summarized in this subsection is taken from Nakamura and Nakamura (1985b). This evidence suggests that there is a great deal of year-to-year continuity in female labor supply, that this continuity cannot be explained by persistent variables like years of schooling, and that this continuity is likely to be spuriously accounted for by the child status variables if it is not adequately controlled for. This evidence also provides examples of at least some of the sorts of output space analyses that could be useful in ranking alternative approximate models of female labor supply.

TABLE 7

ACTUAL AND SIMULATED PROPORTIONS OF WOMEN 21-46 YEARS OF AGE AND MARRIED IN 1971 WHO HAD, OR WHO DID NOT HAVE, A BABY OR YOUNG CHILD IN THE 7-YEAR PERIOD, BY EARNED INCOME CUMULATED OVER 1971-1977

Earned income over 7 years (1967 dollars)	Baby or young child			No baby or young child		
	Actual	Dummy	Inertia	Actual	Dummy	Inertia
$0	.16	.18	.16	.18	.14	.14
Under $10,000	.46	.34	.51	.30	.20	.32
$10,000 - 19,000	.20	.29	.14	.19	.30	.23
$20,000 - 29,999	.09	.11	.10	.17	.20	.14
$30,000 - 39,999	.05	.06	.05	.10	.10	.08
$40,000 - 59,999	.04	.01	.05	.05	.06	.06
$60,000 - 79,999	.00	.00	.00	.01	.00	.01
Over $79,999	.00	.00	.00	.00	.00	.00
Pseudo chi-square statistic		26	7		19	5
Number of observations		255			169	

Source: Nakamura and Nakamura (1985b, Tables D.9 and D.10, p. 444).

3. Confirming Evidence from Census Data

Census data are usually viewed as cross-sectional. Nevertheless the 1980 U.S. Census provides information about employment status in the 1980 Census Reference Week as well as about weeks of work in 1979. Thus Census data can also be used to examine the importance of accounting for labor supply in the previous year. The large sample sizes that are possible with census data are useful for exploring different dimensions of how children affect the work behavior of their mothers.

Observations from the 1980 U.S. Census were selected for 20-45-year-old, white, married women with and without children. Observations were also selected for 20-45-year-old, white, married

TABLE 8

ACTUAL AND SIMULATED PROPORTIONS OF WOMEN 21-46 YEARS OF AGE AND MARRIED IN 1981 BY EARNED INCOME CUMULATED OVER THE 2-YEAR PERIOD OF 1981-1982

Earned income over 2 years (1967 dollars)	Actual	Standard	Dummy	Inertia
$0	.22	.11	.22	.21
Under $1,000	.13	.07	.05	.09
$1,000 - 2,499	.11	.18	.11	.13
$2,500 - 4,999	.15	.27	.21	.19
$5,000 - 7,499	.13	.18	.17	.15
$7,500 - 9,999	.11	.09	.10	.09
$10,000 - 14,999	.10	.08	.09	.08
$15,000 - 19,999	.03	.01	.04	.03
$20,000 - 29,999	.01	.01	.01	.01
Over $29,999	.00	.00	.00	.00
Pseudo chi-square statistic		448	155	65
Number of observations			1724	

Source: Nakamura and Nakamura (1985b, p. 474, Table E.3).

men with children, since the behavior of married, prime aged men is a common standard of reference in labor market analyses. The observations were grouped by education (less than high school, high school graduate, more than high school) and weeks of work in the previous year (0, 1-26, 27-47, 48+). The observations for the women were further grouped by the number of children ever born (0, 1, 2, 3+) and the age of the youngest child for those with children ever born (under 1, 1-3, 4-6, 7-11, 12+). For each group, what is shown are the proportions of those who were employed in the Census Reference Week, with the cell counts given in parentheses.

TABLE 9

1980 REFERENCE WEEK EMPLOYMENT RATES FOR MEN AND WOMEN WITH WITH LESS THAN A HIGH SCHOOL EDUCATION, GROUPED BY WEEKS OF WORK IN 1979 AND PARITY

Number of children ever born	Weeks of work in 1979: All	0	1-26	27-47	48+
White, married men 20-45 with children					
1+	.82 (316)[a]	.08 (12)	.44 (32)	.76 (67)	.94 (205)
White, married women 20-45 with no children					
0	.52 (548)	.08 (207)	.53 (83)	.75 (73)	.92 (185)
White, married women 20-45 with children, all 12 or older					
1	.51 (109)	.00 (40)	.38 (13)	.60 (10)	.98 (46)
2	.56 (260)	.05 (92)	.53 (32)	.77 (35)	.97 (101)
3+	.48 (552)	.06 (249)	.48 (64)	.87 (60)	.95 (179)

[a] Sample sizes are shown in parentheses in Tables 9-15.

TABLE 10

1980 REFERENCE WEEK EMPLOYMENT RATES FOR MEN AND WOMEN WITH A HIGH SCHOOL EDUCATION, GROUPED BY WEEKS OF WORK IN 1979 AND PARITY

Number of children ever born	Weeks of work in 1979: All	0	1-26	27-47	48+
White, married men 20-45 with children					
1+	.92 (765)	.36 (11)	.79 (34)	.75 (119)	.98 (601)
White, married women 20-45 with no children					
0	.74 (2032)	.11 (323)	.55 (240)	.75 (280)	.94 (1189)
White, married women 20-45 with children, all 12 or older					
1	.64 (298)	.04 (92)	.62 (29)	.97 (30)	.95 (147)
2	.66 (743)	.07 (198)	.48 (73)	.80 (109)	.99 (363)
3+	.65 (970)	.09 (287)	.60 (113)	.88 (144)	.97 (426)

TABLE 11

1980 REFERENCE WEEK EMPLOYMENT RATES FOR MEN AND WOMEN WITH MORE THAN A HIGH SCHOOL EDUCATION, GROUPED BY WEEKS OF WORK IN 1979 AND PARITY

Number of children ever born	Weeks of work in 1979: All	0	1-26	27-47	48+
White, married men 20-45 with children					
1+	.96 (1264)	.62 (16)	.64 (53)	.90 (144)	.99 (1051)
White, married women 20-45 with no children					
0	.82 (2990)	.16 (266)	.57 (377)	.85 (673)	.96 (1674)
White, married women 20-45 with children, all 12 or older					
1	.73 (169)	.09 (32)	.33 (18)	.89 (36)	.99 (83)
2	.70 (457)	.09 (99)	.51 (63)	.96 (117)	.95 (178)
3+	.69 (502)	.09 (122)	.54 (61)	.93 (115)	.95 (204)

In Tables 9-11, employment rates can be compared for men, childless women, and mothers with a youngest child at least 12 years of age. For each of these groups, it can be seen that the probability of work in the Reference Week rises steeply as the number of weeks worked in the previous year increases. For the men classified by weeks of work in 1979, there is a tendency for the probability of work in the Reference Week to rise with increases in the level of schooling. There is a similar tendency for the women. There are no clear-cut parity-related patterns in the results.

TABLE 12

1980 REFERENCE WEEK EMPLOYMENT RATES FOR WHITE, 20-45, MARRIED WOMEN WITH CHILDREN GROUPED BY AGE OF THE YOUNGEST CHILD, PARITY AND EDUCATION: THOSE WITH NO WEEKS OF WORK IN 1979

Number of children ever born	Age of youngest child: under 1	1-3	4-6	7-11
Less than high school				
1	.06 (33)	.04 (82	.13 (44)	.06 (46)
2	.07 (72)	.05 (149	.05 (76)	.06 (115)
3+	.03 (98)	.05 (213	.07 (177)	.04 (254)
High school				
1	.04 (68)	.08 (281)	.10 (116)	.10 (100)
2	.05 (179)	.06 (462)	.07 (255)	.06 (264)
3+	.04 (129)	.06 (394)	.06 (265)	.04 (369)
More than high school				
1	.04 (57)	.04 (262)	.03 (71)	.04 (47)
2	.04 (170)	.05 (386)	.12 (184)	.09 (195)
3+	.05 (127)	.04 (263)	.07 (131)	.12 (184)

TABLE 13

1980 REFERENCE WEEK EMPLOYMENT RATES FOR WHITE, 20-45, MARRIED WOMEN WITH CHILDREN GROUPED BY AGE OF THE YOUNGEST CHILD, PARITY AND EDUCATION: THOSE WITH 1-26 WEEKS OF WORK IN 1979

Number of children ever born	Age of youngest child			
	under 1	1-3	4-6	7-11
Less than high school				
1	.15	.36	.28	.65
	(13)	(44)	(18)	(17)
2	.33	.30	.37	.42
	(33)	(44)	(32)	(40)
3+	.19	.44	.38	.42
	(16)	(48)	(48)	(83)
High school				
1	.22	.45	.44	.56
	(102)	(147)	(59)	(62)
2	.28	.52	.47	.63
	(67)	(140)	(93)	(122)
3+	.26	.44	.50	.60
	(57)	(86)	(76)	(139)
More than high school				
1	.29	.52	.64	.61
	(154)	(167)	(39)	(41)
2	.28	.53	.58	.66
	(69)	(145)	(79)	(125)
3+	.33	.52	.54	.58
	(33)	(79)	(61)	(97)

TABLE 14

1980 REFERENCE WEEK EMPLOYMENT RATES FOR WHITE, 20-45, MARRIED WOMEN WITH CHILDREN GROUPED BY AGE OF THE YOUNGEST CHILD, PARITY AND EDUCATION: THOSE WITH 27-47 WEEKS OF WORK IN 1979

Number of children ever born	Age of youngest child			
	under 1	1-3	4-6	7-11
Less than high school				
1	.75	.77	.67	.79
	(8)	(17)	(12)	(14)
2	.78	.67	.61	.74
	(9)	(30)	(28)	(35)
3+	.80	.77	.77	.85
	(10)	(31)	(22)	(52)
High school				
1	.49	.58	.72	.80
	(79)	(71)	(39)	(45)
2	.51	.75	.87	.78
	(55)	(57)	(62)	(142)
3+	.74	.70	.75	.79
	(23)	(40)	(44)	(160)
More than high school				
1	.45	.87	.76	.94
	(132)	(143)	(55)	(52)
2	.69	.83	.78	.92
	(62)	(102)	(87)	(139)
3+	.77	.81	.78	.92
	(26)	(53)	(36)	(114)

TABLE 15

1980 REFERENCE WEEK EMPLOYMENT RATES FOR WHITE, 20-45, MARRIED WOMEN WITH CHILDREN GROUPED BY AGE OF THE YOUNGEST CHILD, PARITY AND EDUCATION: THOSE WITH 48+ WEEKS OF WORK IN 1979

Number of children ever born	Age of youngest child			
	under 1	1-3	4-6	7-11
Less than high school				
1	.75	.94	.83	.86
	(8)	(17)	(23)	(37)
2	.80	.83	.93	.97
	(5)	(30)	(42)	(63)
3+	.77	.95	.95	.94
	(13)	(37)	(58)	(137)
High school				
1	.69	.93	.95	.96
	(85)	(144)	(123)	(155)
2	.89	.95	.96	.94
	(35)	(146)	(169)	(320)
3+	.72	.93	.96	.93
	(32)	(88)	(130)	(363)
More than high school				
1	.79	.96	.96	.98
	(109)	(160)	(84)	(120)
2	.80	.96	.95	.96
	(51)	(140)	(120)	(218)
3+	.82	.97	.96	.96
	(27)	(60)	(93)	(193)

Tables 12-15 give figures for mothers with a youngest child less than 12. The patterns are generally similar to those seen in Tables 9-11 for mothers with no children younger than 12. However, there does seem to be a tendency for the probability of work in the Reference Week to rise with increases in the age of the youngest child, with the biggest increases in moving from the "under 1" to the "1-3" age-of-youngest-child groups for women who worked in the previous year.

The results presented in Tables 9-15 are supportive of the Inertia Model introduced in the previous subsection. Labor supply behavior is found to be persistent in the sense that women who work (do not work) in one year are likely to continue working (not working) in the next. In fact, women who worked 48 or more weeks in the previous year are found to be almost as likely to continue working as men, regardless of the ages or numbers of their children, except for those women who have a new baby.

4. Supporting Findings in the Literature

The evidence in subsections IX.2 and IX.3 on the continuity of female labor supply is supported by findings in a number of studies drawing on other data sources. These studies include Heckman and Willis (1979), Mott and Shapiro (1983), L.B. Shaw (1983), and Even (1987). Based on archival micro data for the early 1900's, Goldin finds that continuity of individual labor supply even characterizes historical female work behavior in the United States:

> A more surprising and rather significant finding is that married women who were in the labor force, even when participation rates were low, had substantial attachment to their jobs and to paid work in general. Data on the life cycle of female employment reveal considerably more continuity of employment than has been presumed. Work experience among those currently employed, therefore, was high, not low. Women have been depicted as working when young and single, then for a brief period after marriage, and again later in their lives. But for most cohorts, this characterization is simply inaccurate. Married women were either in the labor force or out; the majority did not engage in much intermittent labor force activity.
>
> (Goldin, 1990, pp. 12-13)

XI. CONCLUSIONS

The following conclusions that emerge from this survey are a mixture of findings and suggested directions for further research:

1. It is important to recognize when child status variables are included in models of female labor supply simply to improve the quality of the estimates of other responses (such as wage elasticities) versus when care is taken in modeling and measuring the impacts of children on women's work behavior. With studies falling into both categories, judgment should be exercised in attributing causal significance to the estimated coefficients for child status variables.

2. Having a baby, or even a young child, does seem to reduce the probability of a woman being employed, though this effect is less dramatic than indicated in many earlier studies. The evidence concerning child status effects on annual hours of work for women who work in the year is unclear, but this may be due to a failure to allow for differences in the effects on weeks of work per year versus hours of work per week.

3. Indirect child status responses (effects over time of existing or anticipated children on other variables that affect labor supply) are hypothesized to be transmitted primarily via the wage offers women can command from employers. But wage effects on the *hours of work* for working women have been shown to be small, as for men. This does not rule out the possibility that there are important indirect child status effects on the *probability of work* for women.

4. Instrumental variables approaches to estimating child status effects are based on assertions that a second set of variables, that are correlated with the child status variables, are uncorrelated with the unobservable factors that are the presumed cause of the correlation problem. These assertions cannot be tested.

5. It is not feasible to develop instrumental child status variables for several different dimensions of child status.

6. Interest in using instrumental child status variables has had the unintended consequence of inhibiting investigations of which dimensions of child status are most important for characterizing the direct effects of children on female labor supply. This is because it is difficult to construct multiple instrumental child status variables. Moreover there is no convincing evidence that the estimates of child status responses obtained using instrumental variables methods are better than the estimates obtained in studies that simply ignore the possibility of correlations between child status variables and omitted factors such as tastes for work and job search costs. Other approaches for dealing with these correlation problems should be explored such as conditioning on past labor supply behavior.

7. Efforts to develop improved ways of measuring child status effects have been hampered by confusion about the types of child status "treatments" for which response estimates are needed. Parental discretion concerning the nature and amount of care to be provided per child probably has at least as large an impact on female labor supply as the number, and even the ages, of children in a family.

8. Efforts to estimate direct child status effects must control for the demonstrated temporal persistence of female work behavior. If this persistence cannot be accounted for with observable variables viewed as exogenous to the labor supply decision problem, other approaches such as the inclusion of lagged labor supply variables should be explored. In order to evaluate the success of these approaches, methods need to be developed for evaluating and ranking alternative estimated models all of which must be viewed as approximations.

ACKNOWLEDGEMENTS

This research was supported by the Social Sciences and Humanities Research Council of Canada, the U.S. Donner Foundation (through the Conference on the Economic Well-Being of Women and Children), the U.S. National Institute of Dental Research (through the Cornell CORSIM project headed by Steve Caldwell), and the Faculty of Business at the University of Alberta.

The authors are particularly indebted to Harriet Duleep and John Pencavel for painstaking comments on earlier versions of this paper, and to Victor Fuchs, Jim Heckman, Jan Kmenta, Guy Orcutt, and Chris Sims for comments and encouragement on earlier efforts without which the present paper would never have been written. The authors also wish to thank the Editor, Esfandiar Maasoumi, two anonymous referees, conference and workshop participants, and Richard Blundell, Martin Browning, Martin Dooley, Greg J. Duncan, Paula England, John Ermich, Evelyn Lehrer, T. Paul Schultz and James Walker for helpful comments and discussions on the present paper. We are grateful to Louise Hebert for perseverance in typing.

REFERENCES

Amemiya, T., Advanced Econometrics, Harvard Univ. Press, 1985.

Andrews, D.W.K., "Chi-Square Diagnostic Tests for Econometric Models: Theory," Econometrica 56, 1988, 1419-1453.

Bagozzi, R.P. and Van Loo, M.F., "An Investigation of the Relationship Between Work and Family Size Decisions Over Time," Multivariate Behavioral Research 23, 1988, 3-34.

Bagozzi, R.P. and Van Loo, M.F., "A Purposeful Behavior Theory of Work and Family Size Decisions," in Siegers, de Jong-Gierveld and van Imhoff, 1991, 101-129.

Bagozzi, R.P. and Yi, Y., "On the Evaluation of Structural Equation Models," J. of the Academy of Marketing Science 16, 1988, 74-94.

Barnes, R. and Gillingham, R., "Demographic Effects in Demand Analysis: Estimation of the Quadratic Expenditure System Using Microdata," Rev. of Economics and Statistics 66, 1984, 591-601.

Becker, G.S., A Treatise on the Family, Harvard Univ. Press, 1981.

Becker, G.S. and Lewis, H.G., "Interaction Between Quantity and Quality of Children," in T.W. Schultz (Ed.), Economics of the Family, Univ. of Chicago Press, 1974, 81-90.

Bergmann, B., The Economic Emergence of Women, Basic Books, 1986.

Berndt, E.R., The Practice of Econometrics: Classic and Contemporary, Addison-Wesley, 1990.

Birg, H., "A Biographic/Demographic Analysis of the Relationship between Fertility and Occupational Activity of Women and Married Couples," in Siegers, de Jong-Gierveld and van Imhoff, 1991, 133-157.

Blank, R.M., "Simultaneously Modeling the Supply of Weeks and Hours of Work among Female Household Heads," J. of Labor Economics 6(2), 1988, 177-204.

Blank, R.M., "The Role of Part-Time Work in Women's Labor Market Choices Over Time," American Economic Rev. 79(2), 1989, 295-299.

Blau, D.M. and Robins, P.K., "Fertility, Employment, and Child-Care Costs," Demography 26(2), 1989, 287-299.

Blau, F.D. and Ferber, M.A., The Economics of Women, Men and Work, Prentice-Hall, 1986.

Blau, F.D. and Ferber, M.A., "Career Plans and the Expectations of Young Women and Men: The Earnings Gap and Labor Force Participation," J. of Human Resources, 1991, forthcoming.

Blomqvist, A. and McKee, "Eliminating the 'married exemption' in the Canadian Income Tax," Canadian J. of Economics 19, 1986, 309-318.

Blundell, R., "Econometric Approaches to the Specification of Life-Cycle Labour Supply and Commodity Demand Behavior," Econometric Rev. 6, 1987, 103-165.

Blundell, R., "Evaluating Structural Microeconometric Models of Labour Supply," invited lecture at the Sixth World Congress of the Econometric Society, Barcelona, August, 1990.

Blundell, R. and Meghir, C., "Selection Criteria for a Microeconometric Model of Labour Supply," J. of Applied Econometrics 1, 1986, 55-80.

Blundell, R. and Walker, I., "A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data," Rev. of Economic Studies 68, 1986, 539-558.

Bowen, W.G. and Finegan, T.A., The Economics of Labor Force Participation, Princeton Univ. Press, 1969.

Browning, M., "The Intertemporal Allocation of Expenditure on Non-Durables, Services, and Durables," Canadian J. of Economics 22, 1989, 22-36.

Browning, M., "Children and Household Economic Behavior," J. of Economic Literature, 1992, forthcoming.

Browning, M.J., Deaton, A. and Irish, M., "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," Econometrica 53, 1985, 503-543.

Calhoun, A.C. and Espenshade, T.J., "Childbearing and Wives' Foregone Earnings," Population Studies 42, 1988, 5-37.

Capozza, D.R., Nakamura, A. and Bloss, G., "Work History in Female Earnings Loss," J. of Forensic Economics, 1989, 55-68.

Cigno, A. and Ermisch, J.F., "A Microeconomic Analysis of the Timing of Births," European Economic Rev. 33, 1989, 737-760.

Citro, C.F. and Hanushek, E.A. (Eds.) Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling Volume I Review and Recommendations, National Academy Press, 1991.

Corcoran, M., Duncan, G.J. and Ponza, M., "Longitudinal Analysis of White Women's Wages," J. of Human Resources 18, 1983, 497-520.

Cox, D., "Panel Estimates of the Effects of Career Interruptions on the Earnings of Women," Economic Inquiry 22, 1984, 386-403.

Cramer, J.C., "Fertility and Female Employment: Problems of Causal Direction," American Sociological Rev. 45, 1980, 167-190.

Deaton, A. and Muellbauer, J., Economics and Consumer Behavior, Cambridge Univ. Press, 1980.

DeTray, D.N., "Child Quality and the Demand for Children," J. of Political Economy 82(2, part II), 1973, S70-S95.

Diewert, W.E., "Intertemporal Consumer Theory and the Demand for Durables," Econometrica 42, 1974, 497-516.

Dolton, P.J. and Makepeace, G.H., "Marital Status, Child Rearing and Earnings Differentials in the Graduate Labour Market," Economic J. 97, 1987, 897-922.

Dooley, M.D., "Labor Supply and Fertility of Married Women: An Analysis with Grouped and Individual Data from the U.S. Census," J. of Human Resources 17, 1982, 499-532.

Douthitt, R.A. and Fedyk, J.M., "The Influence of Children on Family Life Cycle Spending Behavior: Theory and Applications," J. of Consumer Affairs 22(2), 1988, 220-248.

Duleep, H.O., Regets M., and Sanders, S., The Economic Assimilation of Asian Immigrant Families, W.E. Upjohn Inst. for Employment Research, 1992, forthcoming.

Easterlin, R.A., Birth and Fortune, Univ. of Chicago Press, 1987.

Engle, E., "Die Lebenskosten Belgisher Arbeiter-Familien Fruher and Jetzt," International Statistical Inst. Bulletin 9, 1895, 1-74.

England, P., "Wage Appreciation and Depreciation: a Test of Neoclassical Economic Explanations of Occupational Sex Segregation," Social Forces 62, 1984, 726-749.

Ermisch, J.F. and Wright, R.E., "Wage Offers and Full-Time and Part-Time Employment by British Women," J. of Human Resources, 1991, forthcoming.

Even, W.E., "Career Interruptions Following Childbirth," J. of Labor Economics 5, 1987, 255-277.

Fuchs, V.R., Women's Quest for Economic Equality, Harvard Univ. Press, 1988.

Goldin, C., Understanding the Gender Gap: An Economic History of American Women, Oxford Univ. Press, 1990.

Gorman, W.M., "Tricks with Utility Functions," in M.J. Artis and A.R. Nobay (Eds.), Essays in Economic Analysis: Proceedings of the 1975 AUTE Conference, Sheffield, Cambridge Univ. Press, 1976.

Graham, J.W. and Beller, A.H., "The Effect of Child Support Payments on the Labor Supply of Female Family Heads," J. of Human Resources 24(4), 1989, 664-688.

Hartog, J. and Theeuwes, J., "Participation and Hours of Work: Two Stages in the Life-cycle of Married Women," European Economic Rev. 30, 1986, 833-857.

Heckman, J.J., "Shadow Prices, Market Wages and Labor Supply," Econometrica 42, 1974, 679-694.

Heckman, J.J., "The Sample Selection Bias as a Specification Error," Econometrica 47, 1979, 153-162.

Heckman, J.J., "Heterogeneity and State Dependence," in S. Rosen (Ed.), Studies in Labor Markets, Chicago: Univ. of Chicago Press, 1981, 91-139.

Heckman, J.J., "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata," Econometrica 52, 1984, 1543-1547.

Heckman, J.J., "Foreword," in Nakamura and Nakamura, 1985b, xi-xii.

Heckman, J.J., "Alternative Approaches To the Evaluation of Social Programs: Econometric and Experimental Methods," Barcelona Lecture, Sixth World Congress of the Econometric Society, Barcelona, August, 1990.

Heckman, J.J., Hotz, V.J. and Walker, J., "New Evidence on the Timing and Spacing of Births," American Economic Rev. 75, 1985, 179-184.

Heckman, J.J. and MaCurdy, T.E., "A Life Cycle Model of Female Labour Supply," Rev. of Economic Studies 47, 1980, 47-74.

Heckman, J.J. and MaCurdy, T.E., "Labor Econometrics," in Z. Griliches and M. Intriligator (Eds.), Handbook of Econometrics, Vol. 3, Elsevier Science, 1986, 1917-1977.

Heckman, J.J., Nakamura, A. and Walker, J.R., "Choosing Among Competing Estimates of the Effects of Children on Female Labor Supply," working paper, 1991.

Heckman, J.J. and Walker, J.R., "Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite Data," in C.C. Clogg (Ed.), Sociological Methodology, American Sociological Assoc., 1988.

Heckman, J.J. and Willis, R.J., "Reply to Mincer and Ofek," J. of Political Economy 87(1), 1979, 203-211.

Hill, C.R. and Stafford, F.P., "Parental Care of Children: Time Diary Estimates of Quantity, Predictability, and Variety," J. of Human Resources 15, 1980, 219-239.

Hoem, B. and Hoem, J.M., "The Impact of Women's Employment on Second and Third Births in Modern Sweden," Pop. Studies 43, 1989, 47-67.

Hotz, V.J. and Miller, R.A., "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," Econometrica 56, 1988, 91-118.

Houthakker, H.S. and Taylor, L.D., Consumer Demand in the United States, 2nd edition, Harvard Univ. Press, 1970.

Hunt, J.C. and Kiker, B.F., "Parental Time Devoted to Children in Two- and One-Wage-Earner Families," Economics of Education Rev. 3(1), 1984, 75-83.

Johnson, T.R. and Pencavel, J.H., "Dynamic Hours of Work Functions for Husbands, Wives, and Single Females," Econometrica 52, 1984, 363-389.

Joshi, H.E., Layard, R. and Owen, S.J., "Why Are More Women Working in Britain?" in Layard and Mincer, 1985, S147-S176.

Juster, F.T. and Stafford, F.P. (Eds.), Time, Goods, and Well-Being, Univ. of Michigan, 1985.

Killingsworth, M.R., Labor Supply, Cambridge Univ. Press, 1983.

Killingsworth, M.R. and Heckman, J.J., "Female Labor Supply: A Survey," in O. Ashenfelter and R. Layard (Eds.), Handbook of Labor Economics, Vol. 1, Elsevier Science, 1986, 103-204.

Klerman, J.A. and Leibowitz, A., "Women's Labor Market Mobility: Evidence from the NLS," Amer. Econ. Rev. 80(2), 1990, 284-288.

Kmenta, J., Elements of Econometrics, 2nd Edition, MacMillan, 1986.

Layard, R. and Mincer, J. (Eds.), Trends in Women's Work, Education, and Family Building, special issue of J. of Labor Economics 3 (1, part 2), 1985.

Leamer, E.E., Specification Searches: Ad Hoc Inference with Nonexperimental Data, John Wiley & Sons, 1978.

Lehrer, E.L., "The Impact of Children on Married Women's Labor Supply: Black-White Differentials Revisited," J. of Human Resources, 1991, forthcoming.

Leibowitz, A., Waite, L.J. and Witsberger, C., "Child Care for Preschoolers: Differences by Child's Age," Demography 25(2), 1988, 205-220.

Lewbel, A., "A Unified Approach to Incorporating Demographic or Other Effects into Demand Systems," Rev. of Econ. Studies, 1985, 1-18.

Manser, M. and Brown, M., "Marriage and Household Decision-Making: A Bargaining Analysis," International Econ. Rev. 21, 1980, 31-44.

Mason, K.O. and Kuhlthau, K., "Determinants of Child Care Ideals among Mothers of Preschool-Aged Children," J. of Marriage and the Family 51, 1989, 593-603.

McElroy, M.B. and Horney, M.J., "Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand," International Economic Rev. 22(2), 1981, 333-349.

McLaughlin, S.D., "Differential Patterns of Female Labor-Force Participation Surrounding the First Birth," J. of Marriage and Family, 1982, 407-420.

Mincer, J., "Labor Force Participation of Married Women: a Study of Labor Supply," in Aspects of Labor Economics, National Bureau of Economic Research, Princeton Univ. Press, 1962, 63-97.

Mincer, J., "Intercountry Comparisons of Labor Force Trends and of Related Developments," in Layard and Mincer, 1985, S1-S32.

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

Mincer, J. and Polachek, S., "Women's Earnings Reexamined," J. of Human Resources 13, 1978, 118-134.

Moffitt, R., "Profiles of Fertility, Labour Supply and Wages of Married Women: A Complete Life-Cycle Model," Rev. of Economic Studies 51, 1984, 263-278.

Mott, F.L. and Shapiro, D., "Complementarity of Work and Fertility among Young American Mothers," Pop. Studies 37, 1983, 239-252.

Mroz, T.A., "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica, 1987.

Muellbauer, J., "Testing the Barten Model of Household Composition Effects and the Cost of Children," Economic J. 87, 1977, 460-487.

Nakamura, A. and Nakamura, M., "A Comparison of the Labor Force Behavior of Married Women in the United States and Canada, with Special Attention to the Impact of Income Taxes," Econometrica 49, 1981, 451-489.

Nakamura, A. and Nakamura, M., "Part-Time and Full-Time Work Behavior of Married Women: A Model with a Doubly Truncated Dependent Variable," Canadian J. of Economics 16, 1983, 229-257.

Nakamura, A. and Nakamura, M., "Dynamic Models of the Labor Force Behavior of Women Which Can Be Estimated Using Limited Amounts of Past Information," J. of Econometrics 27, 1985a, 273-298.

Nakamura, A. and Nakamura, M., The Second Paycheck: A Socioeconomic Analysis of Earnings, Academic Press, 1985b.

Nakamura, A. and Nakamura, M., "On the Performance of Tests by Wu and by Hausman for Detecting the Ordinary Least Squares Bias Problem," J. of Econometrics 29, 1985c, 213-227.

Nakamura, A. and Nakamura, M., "Selection Bias: More Than a Female Phenomenon," in B. Raj (Ed.), Advances in Econometrics and Modelling, Kluwer Academic, 1989, 143-158.

Nakamura, M., Nakamura, A. and Cullen, D., "Job Opportunities, the Offered Wage, and the Labor Supply of Married Women," American Economic Rev. 69, 1979, 787-805.

Nakamura, A., Nakamura, M. and Duleep, H., "Alternative Approaches to Model Choice," J. of Economic Behavior and Organization 14(1), 1990, 97-125.

Ni Bhrolchain, M., "Women's Paid Work and the Timing of Births: Longitudinal Evidence," European J. of Population 2, 1986, 43-70.

Nicol, C.J., "Estimating a Third-Order Translog Using Canadian Cross-Sectional Micro-Data," Canadian J. of Economics 22(3), 1989, 543-560.

Nock, S.L., "Time With Children: The Impact of Couples' Work-Time Commitments," Social Forces 67(1), 1988, 59-85.

Oppenheimer, V.K., Work and the Family: A Study in Social Demography, Academic Press, 1982.

Orcutt, G., Caldwell, S. and Wertheimer, R., Policy Exploration Through Microanalytic Simulation, The Urban Institute, 1976.

Pencavel, J., "Labor Supply of Men: A Survey," in O. Ashenfelter and R. Layard (Eds.), Handbook of Labor Economics, Vol. 1, North-Holland, 1986, 3-102.

Phelps, E.S., "The Statistical Theory of Racism and Sexism," American Economic Rev. 62, 1972, 659-661.

Polachek, S., "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure," Rev. of Economics and Statistics 58, 1981, 60-69.

Pollak, R.A., "Habit Formation and Dynamic Demand Functions," J. of Political Economy 78, 1970, 60-78.

Pollak, R.A. and Wales, T.J., "Demographic Variables in Demand Analysis," Econometrica 49, 1981, 1533-1551.

Prais, S.J. and Houthakker, H.S., The Analysis of Family Budgets, Cambridge Univ. Press, 1955.

Presser, H.B., "Can We Make Time for Children? The Economy, Work Schudules, and Child Care," Demography 26(4), 1989, 523-543.

Ray, R., "Demographic Variables and Equivalence Scales in a Flexible Demand System," Applied Economics 18, 1986, 265-278.

Reimers, C.W., "Sources of the Family Income Differentials among Hispanics, Blacks, and White Non-Hispanics," American J. of Sociology 89(4), 1984.

Robinson, C. and Tomes, N., "More on the Labour Supply of Canadian Women," Canadian J. of Economics 18, 1985, 156-163.

Rosenzweig, M.R. and Schultz, T.P., "The Demand for and Supply of Births: Fertility and Its Life Cycle Consequences," American Economic Rev. 75, 1985, 992-1015.

Rosenzweig, M.R. and Wolpin, K.I., "Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models," J. of Political Economy 88, 1980, 328-348.

Rossi, N., "Budget Share Demographic Translation and the Aggregate Almost Ideal Demand System," European Economic Rev. 31, 1988, 1301-1318.

Salkever, D.S., "Children's Health Problems and Maternal Work Status," J. of Human Resources 17, 1987, 97-109.

Sawa, T., "Information Criteria for Discriminating Among Alternative Regression Models," Econometrica 46(6), 1978, 1273-1291.

Schultz, T.P., "The Influence of Fertility on Labor Supply of Married Women: Simultaneous Equation Estimates," in R. Ehrenberg (Ed.), Research in Labor Econ., Vol. 2, JAL Press, 1978, 273-351.

Shaw, K.L., "Life-Cycle Labor Supply with Human Capital Accumulations," International Economic Rev. 30, 1989, 431-456.

Shaw, L.B., "Causes of Irregular Employment Patterns," in L.B. Shaw (Ed.), Unplanned Careers: The Working Lives of Middle-Aged Women, Lexington Books, 1983, 45-59.

Siegers, J.J., de Jong-Gierveld, J. and van Imhoff, E. (Eds), Female Labour Market Behaviour and Fertility, Springer-Verlag, 1991.

Simpson, W., "Analysis of Part-time Pay in Canada," Canadian J. of Economics 19, 1986, 798-807.

Smith, J.B. and Stelcner, M., "Labour Supply of Married Women in Canada, 1980," Canadian J. of Economics 21, 1988, 857-870.

Smith, J.P. and Ward, M.P., "Time-Series Growth in the Female Labor Force," in Layard and Mincer, 1985, S59-S90.

Spitze, D., "Women's Employment and Family Relations: a Review," J. of Marriage and the Family 50, 1988, 595-618.

Strober, M.H. and Weinberg, C.B., "Strategies Used by Working and Nonworking Wives to Reduce Time Pressures," J. of Consumer Research 6, 1980, 338-348.

Sydenstricker, E. and King, W.I., "The Measurement of the Relative Economic Status of Families," Quarterly Publication of the American Statistical Assoc. 17, 1921, 842-857.

Trzcinski, E. and Alpert, W.T., "Strategies and Costs of Handling Work During Leave," J. of Management Issues, 1991, forthcoming.

Vuong, Q.H., "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," Econometrica 57, 1989, 307-333.

Waite, L.J. and Stolzenberg, R.M., "Intended Childbearing and Labor Force Participation of Young Women: Insights from Nonrecursive Models," American Sociological Rev. 41, 1976, 235-252.

Ward, M. and Butz, W., "Completed Fertility and Its Timing," J. of Political Economy 88, 1980, 917-940.

Weinberg, C.B. and Winer, R.S., "Working Wives and Major Family Expenditures: Replication and Extension," J. of Consumer Research 10, 1983, 259-263.

Zellner, H., "The Determinants of Occupational Segregation," in C. Lloyd (Ed.), Sex, Discrimination, and the Division of Labor, Columbia Univ. Press, 1975, 125-145.

ECONOMETRIC REVIEWS, 11(1), 73-77 (1992)

Comments on The Econometrics of Female Labor Supply and Children by Alice and Masao Nakamura

The reader of this survey will be struck by the important part the analysis of female labour supply continues to play in empirical microeconomics. Although it has been the focus of microeconometric research for some twenty years it continues to raise issues which are at the frontier of econometrics and yet it also remains at the centre of many policy debates. The interaction with demographic choices has largely been left in the sidelines and the authors must be congratulated for bringing these issues back into the foreground so lucidly.

Given such a sustained amount of econometric research it is, at first, worrying that empirical results on female labour supply differ so widely and definitive results have yet to emerge. However, taking the evidence in this survey with that reported in Mroz(1987), we can now begin to attribute different empirical results to the different assumptions utilised in specific empirical studies. This is a highly constructive approach and I will adopt it to frame my comments.

There are four broad areas on which I wish to comment. The first relates to the specification child status variables. The second concerns the modelling of state-dependence through the direct inclusion of lagged labour supply behaviour. The third area concerns nonparametric identification and the use of instrumental variables in censored models. Finally I wish to discuss briefly the authors' approach to model evaluation and selection.

To address the first issue I found it most productive to think in terms of a dynamic optimisation model. Although, as the authors note, many labour supply models are primarily "static" it would seem inappropriate to analyse the interactions between labour supply and children in anything but an intertemporal model. The sequential nature of the standard intertemporal model can be used to show under what conditions certain conditional, even "static" models are appropriate and, when they are not, what biases to structural parameters are likely to occur. Moreover, it can show which structural parameters can be identified by particular model specifications. This is important since more structural information is often required for some analysis, welfare measurement for example, than is needed for, say, labour supply forecasts.

I take as an illustration the additively separable joint utility maximising model which, at each period s, may be represented by:

$$\max\left\{U_s = \sum_{t=s}^{L} F(v(w_t,\mu_t;z_t),Z_t) + \theta(Z_t)\right\}$$

where w_t is the period t marginal wage, μ_t is period t virtual income and z_t is the demographic profile (henceforth "child status") as it directly affects current period decisions while Z is the complete life-time profile of children as seen in period t as affecting life-time utility. The function $v_t(.)$ is the usual period t indirect utility function conditional on z_t. The standard labour supply function (III-16), expressing labour supply (and participation) as a function of w_t, μ_t and *conditioning* on children z_t, is completely summarised by indirect utility $v_t(.)$. All the remaining parameters in F reflect intertemporal substitution and those in θ the pure effects of children on life-time utility. This decomposition is valid whether or not child status variables are endogenous for current labour supply.

An intertemporal substitution model can identify F(.) but only a model of demographic choice (endogenous fertility) can recover θ. This simple intertemporal specification (which rules out state dependence) can be used to justify the standard "static" labour supply model in a dynamic setting if μ_t is chosen so as to fully reflect asset changes (see Blundell(1987)) - the usual other income measure $y_t + A_t$ in (III-4) will not do. Indeed, given that the life-cycle model suggests that assets will change in a way that depends on children, the standard other income measure is almost sure to lead to a mispecification of the child status parameters. If μ_t is defined incorrectly or differently then it is also likely that the estimated parameters on children will differ and this in a predictable way.

This sequential nature of the life-cycle problem allows a sequential decomposition of child status effects. To evaluate the effect of child status on labour supply given asset choices (or savings), the standard cross-section labour supply model (with μ_t appropriately defined) will do, to predict the evolution of labour supply responses over time allowing savings to adapt, the full intertemporal parameter set is required and panel data is all but essential. To make welfare comparisons necessary to measure the constant utility welfare costs in terms of lost earnings that children induce, θ itself is needed. Depending on the objective a different set of parameters become sufficient and require specific data and econometric techniques. Of course μ_t may be endogenous and whether or not it is treated as such will very likely change the interpretation and the empirical estimates of the child status parameters. It should be pointed out however, that a correct choice of μ_t justifies a life-cycle model and this extends importantly to the fixed costs framework of Cogan (1981).

This simple intertemporal specification is invalidated under most forms of state dependence. History enters this standard additively separable model only through the last period's asset levels in μ_t. Once individuals realise that history influences current decisions in any other direct way, then a structural model will predict that future outcomes (as well as past decisions) will enter the labour supply function. Although the authors find clear prior and empirical support for lagged behaviour entering current decisions they do not formulate such a model systematically. This, I will argue, has certain drawbacks.

There are essentially three distinct forms of state-dependence. Each breaks the intertemporal separability assumption underlying U above and invalidates (III-16). The most common is via habit persistence in which lagged leisure enters current period utility. This appears to be the form underlying the authors alternative inertia and dummy models of section IX although these are not derived from a dynamic optimising model. To my mind the following two alternatives appear more likely. The first of these breaks separability through lagged labour supply behaviour (experience) entering the wage equation. This is commonly observed for women (see Mroz, 1987) and yet its implications for labour supply are rarely acknowledged (see Shaw, 1989 for an exception). Employment today has a payoff for tomorrow's wage and may reduce the chance that a women, once in employment, would leave temporarily to care for children. A similar implication arises in search models which comprise the second alternative. Here a different layoff and arrival rate ensure current employment status directly affects tomorrow's choice set. In this model a women may choose to remain in employment while caring for young children even though preferences alone would induce her to quit temporarily. Since the authors use a labour force participation model with and without lagged behaviour as a comparison for evaluation of the "standard" framework (Tables 6, 7 and 8) such a discussion of state-dependence is relevant. Although their inertia and dummy models may provide a useful test of the standard model they do not provide structural alternatives and should be interpreted, at best, as reduced forms. Each of the alternative state-dependent models have quit different implications for behaviour over the business-cycle and life-cycle as well as different policy prescriptions.

The authors point out the established inconsistency with empirical evidence of the restrictive Tobit assumption in the standard static labour supply model. For women with children, fixed costs make it particularly unreasonable, however the search costs model also invalidates such an assumption. If the Tobit framework is almost surely incorrect the selectivity model seems more appropriate. However, this requires stronger identification conditions or a parametric assumption on the joint distribution of hours and participation. To identify participation (non-parametrically) the participation index should include an additional exogenous (weakly) regressor. Choosing identification restrictions is a serious and tenuous consideration especially once the possibility of endogeneity for the child status variables is acknowledged. The authors spend some time on this issue. They compare OLS results with IV results where child status variables are instrumented. Even if the instruments used are potentially invalid, where the estimates differ I can place little faith in the OLS results. In principle a minimum of four (sets of) valid exclusion restrictions among weakly exogenous variables are required for identification. One for participation (child care costs or local labour market conditions), one for the wage (education), one

for other income (husband's education or beginning of period asset income) and one for child status (?). The latter is the most difficult. The twins first approach can be thought of as a within groups IV transformation that sweeps out correlation with permanent effects. Shultz uses residential origin but, as the authors point out, this is potentially endogenous and may induce spurious inferences. In my view the fertility literature itself would seem the best place to start. To model the potential simultaneity of labour supply and recent births taking past births as weakly exogenous would seem a reasonable starting point. This would clearly require panel data and would relate *changes* in child status to changes in labour supply parameters across just indentified models. It is unlikely, in this case, that a unique most preferred model can be found, or that within sample information alone and standard statistical tests will be sufficient to choose between models.

Richard Blundell, University College London and Institute for Fiscal Studies

ECONOMETRIC REVIEWS, 11(1), 79-83 (1992)

Comments on The Econometrics of Female Labor Supply and Children

by Alice and Masao Nakamura

In the past three decades many estimates of the effects of children on labour supply have been published. Some strong relationships have been found but, as the authors of this survey point out, this literature has, with some notable exceptions, focused on the estimation of wage and income elasticities. This survey provides a much needed corrective in the form of an extended and insightful discussion of the estimation of child status 'effects' on labour supply.

1. The authors pose five questions to be addressed in this survey. Among the more important in our view (and, we believe, that of the authors) is "What is meant by child status effects?" In section III-3 they ask whether such effects can be interpreted in the traditional "treatment-reponse" framework and state that a development of a clear understanding of this issue is one of the primary objectives of this survey.

In section VI-2 the authors provide a straightforward interpretation of the indirect effects of child status as the influence of child rearing activities on subsequent offered wages as transmitted through occupational choice, human capital investment and employer perceptions.

The authors provisionally define the direct effects of children (page 19) as the "current period labour supply responses to the time, effort and other resource requirements (our emphasis) of having and caring for children". However, the general drift of the argument throughout the paper is that this is an inappropriate way to think of child status effects. This is the case even for the case of unexpected births, e.g. the "twins first" data of Rosenzweig and Wolpin. In particular, the very useful figure 1 suggests that we cannot view child status effects as being "treatments" with a corresponding time allocation "response". This point (which has been made forcibly by, amongst others, T. Paul Schultz) is related to the discretion that parents have over their time allocation to children. The Nakamuras implicitly conclude that the "treatment-response" interpretation is not warranted. At this point, however, the paper turns to other issues. We hope and trust that the authors will return to this important point in future work.

Our reading of the Nakamuras' figure 1 is that the estimated parameters of the joint distribution of labour supply and child status, conditional on other observed variables, reflect many unobserved factors including tastes, home technology, the price of various child care services, whether births are planned or not and expectations concerning future labour supply. Current data availability makes a convincing resolution of these various factors impractical.

In this paper and elsewhere, the Nakamuras have argued for the importance of labour force continuity in estimating child status effects. They have drawn our attention to the fact that much of the increase in the labor supply of mothers reflects an increase in the proportion of women who choose a life cycle path of continuous and strong labour force attachment. In particular, they find the child status variables lose much of their predictive power in the presence of controls for lagged labour supply. The Nakamuras readily acknowledge that the inclusion of lagged labour supply likely does not provide consistent estimates of structural models. They state, however, that the relevant problem is one of choosing amongst alternative approximate models given the data demands of the life cycle model.

The Nakamuras argue for a 'courtroom strategy' for ranking alternative approximate models. An example is provided in an instructive extract from their 1985 book which compares the predictive power of three models: the standard model, the dummy model and the inertia model. This paper's excellent discussion of endogenous fertility and the difficulty of finding appropriate instruments suggests a fourth variant which excludes all child status and lagged labour supply variables from the right hand side. We label this the "purist" model while acknowledging that the potential endogeneity of such 'standard' regressors as schooling and husband's income is highly suggestive of remaining impurities.

The final section of the paper contains two examples of a 'courtroom' strategy. In the first example, the standard, dummy and inertia models are compared in terms of their ability to predict cumulative labour supply and earnings using in-sample evidence from a seven year stretch of PSID data. The second is a comparison of estimates obtained with the PSID data and with 1980 U.S. Census data. This test for stability across data sets is one which the Nakamuras have employed in past work. It is an approach that is to be highly recommended.

A finding that the inertia model is much more stable over time than estimates of the purist or standard models would help to focus future research and data collection. For example, it would focus attention on the interaction between career plans and fertility plans made early in the life cycle.

2. We have an immediate problem with (III-1) which specifies the (conditional) utility function for goods (G_t) and female non-work time ($h_{f,NW,t}$). This conditions on nonlabor income (A_t) and the male's income ($y_{m,t}$), as well as children (N_t) and other family attributes (Z_t^*). The inclusion of the two incomes as conditioning goods for preferences has no basis in utility theory. To be sure, we might want to condition on the male's hours ($h_{m,NW,t}$) to allow for some complementary or substitutability with the female's hours or consumption but this is a long way from conditioning on earnings. Of course, earnings do enter the <u>indirect</u> utility function (since they enter the budget constraint (III-3)); consequently the first order conditions given in (III-5) to (III-8) are correct. This inappropriate conditioning is repeated in may places after (for example, in (VII-2)) and could lead to confusion.

There is an important corollary to these remarks: the formulation in (III-5) to (III-8) does assume that preferences over consumption and female non-work time are separable from male hours. This may be acceptable but it at least deserves explicit statement. This tends to get lost when the utility function incorrectly conditions on incomes.

3. Another issue arises from the conditioning in (III-1). The authors remark after (III-1) that the conditioning variables are treated as 'predetermined or exogenous variables'. This suggests that exogeneity and conditioning are somehow synonymous but this is not the case. Whilst we can choose to condition on current exogenous variables we can also condition on current endogenous variables and we can obviously choose not to condition on particular exogenous variables. As an example, in Browning and Meghir (1991) preferences for commodity demands are conditioned on the labor force variables of the male and female in the household. Associated with such a representation is a commodity demand system that depends on commodity prices, the total expenditure on commodities and the labor supplies. The possible endogeneity of the latter to the demand system does not affect their status as conditioning variables. Thus we feel that the sentence following (III-1) is potentially misleading and certainly redundant.

4. The arguments that the Nakamuras have made regarding the importance of conditioning on past labor supply are well taken. Not so clear are the sources of this dependence. It is a slippery road to impute them to habit persistence effects as in section 9. The problem with this justification is that if past actions affect the present (other than through the budget constraint) then rational agents will take account of the effect of current actions on future preferences. Thus we need to model not only the dependence of current actions on past actions but also on expectations about the future environment.

To illustrate, suppose that, as suggested in section 9, (time and money) expenditures on children really are complementary over time (or auto-complementary as Browning (1991) terms it). Then an increase in future (discounted) price and wages will reduce current levels of 'investment' (time and money) expenditures on children. Indeed auto-complementary goods are like habits; we reduce consumption now if future prices rise.

On the other hand, we might question whether (time and money) expenditures on children really are auto-complementary as suggested. If, for example, parents are forced to spend less than they would like in a particular period then auto-complementary implies that they will voluntarily reduce

expenditures on children in all future periods (and even reduce expenditures before the enforced drop if the latter is anticipated). It seems at least as plausible to us that expenditures on children are auto-substitutes. As an example, this implies that if expenditures are forced to be low in any period then expenditures in adjacent periods will be increased to compensate. A particularly interesting example of this arises for schooling. If the quality of this rises (exogenously for parents) then auto-complementary implies that parents of pre-schoolers will raise their (time and money) expenditures on these children whereas auto-substitutibility implies that they will decrease such expenditures. Of course, auto- (or want) independence implies that a rise in the quality of schooling will have no effect on child (investment) expenditures.

Martin Browning and Martin Dooley, McMaster University

References

Browning, Martin (1991), "A Simple Non-Additive Preference Structure for Models of Household Behaviour Over Time", Journal of Political Economy.

Browning, Martin and Costas Meghir (1991), "The Effects of Male and Female Labour Supply on Commodity Demands", Econometrica.

ECONOMETRIC REVIEWS, 11(1), 85-87 (1992)

Comments on The Econometrics of Female Labor Supply and Children by Alice and Masao Nakamura.

In a longer time perspective it is reasonable to regard labor supply and "child production" as being simultaneously determined. This means that there is an endogeneity problem if we want to assess the effect of children on labor supply. Alice and Masao Nakamura have written a useful review of estimation problems arising because of this endogeneity. Alternative solutions to many of the problems are discussed. Most of the proposed methods do not model the joint decision, but concentrate on how to take account of the endogeneity of child status variables when estimating labor supply functions. It would have been more exciting, and without doubt much more difficult, with more emphasis on how to structurally model the joint decision of child production and labor supply. To do this a life cycle model would be needed.

A large part of the survey deals with labor supply in a static context where the child status variables are predetermined. This is a sensible simplification. However, starting out with, and probing deeper into, a structural life cycle model might have given insights about how to model the decisions within a period, where the child status variables are given. Even within the static framework I would have liked more emphasis on how to estimate structural relationships. The fact that a woman has children affects both the utility function and the budget constraint. A clear separation of effects that work via the utility function versus the budget constraint would be useful. We can in principle observe budget constraints directly, but preferences only indirectly. My view is that we therefore always should model budget constraints as carefully as possible. Details of the budget constraint inadvertantly excluded will be intertwined with the estimated preference parameters. A change in individuals' budget constraints will then change the "preference" parameters, i.e. the parameter estimates will be unstable.

There are many problems if we want to construct a structural model as many of the costs of bringing up children are hard to measure. However, some important costs are , in principle, fairly easy to measure. Children must be taken care of, either by a parent or somebody else. In some countries day care is available in the form of au pair girls, in other countries via day care centres, which in some countries are heavily subsidized. The effect of children on female labor supply is presumably heavily dependent on the cost of good quality day care. This dependence is via the budget constraint. If the cost of children's day care is not taken care of in the buget constraint, the effect of this cost will instead be intertwined with the preference parameters. A change in ,say, the subsidies to the day care system would change these parameters; that is, the parameter estimates would be unstable.

Nakamura and Nakamura do not discuss at all how to take nonlinear taxes into account. Many of the estimation methods which are consistent if budget constraints are linear are inconsistent if the budget constraint is nonlinear. I am not sure if this negligence of taxes should be interpreted as an implicit hypothesis that the direct effect of children will be estimated without severe bias, even if the coefficients for the economic variables are inconsistently estimated. Apriori my guess would be that this hypothesis is false and that different ways to account for taxes would result in different estimates of the child effect. However, results in Blomquist and Hansson–Brusewitz (1990) indicate that the estimates might be robust to the choice of taking care of taxes. In that paper maximum likelihood methods and information of the complete budget constraint are used to estimate female labor supply functions. One of the explanatory variables, NC1, measures the number of children aged 0–5, another variable , NC2, measure the number of children aged 6–19. If the labor supply function is assumed to be linear in all variables the two child coefficients are estimated as –281 and –172 respectively. On my behalf Hansson–Brusewitz has also estimated a linear supply function, using individuals' before tax budget constraints, i.e. taxes are not accounted for. (This supply function is not included in the published article.) The constant, the wage and income coefficients are estimated quite differently as compared to when taxes are accounted for. However, the coefficients for the two child variables are almost identical to those obtained when we do take account of taxes. This is certainly scanty evidence and the robustness of estimated child effects to different ways to account for taxes should be studied further, using other functional forms.

Another interesting question is how sensitive estimated child effects are to the functional form of the labor supply function. The results in Blomquist and Hansson–Brusewitz (1990) indicate that the estimated child effects might be sensitive to how the wage is included in the labor supply function.

N. Sören Blomquist, Uppsala Universitet

Reference

N. S. Blomquist and U. Hansson–Brusewitz. (1990) "The Effect of Taxes on Male and Female Labor Supply in Sweden." **Journal of Human Resources** 25, 317–357.

ECONOMETRIC REVIEWS, 11(1), 89-91 (1992)

Comments on *The Econometrics of Female Labor Supply and Children* by Alice and Masao Nakamura

The topic addressed by Alice and Masao Nakamura is very wide and has generated several strands of literature. I learned a lot by reading their survey, which I will not attempt to summarize here. I will rather comment on aspects which in my opinion deserve more attention than they received in this survey.

I shall begin with considerations on model specification and then go on to econometric problems, although the two points are intimately linked in the choice of a research strategy.

All the reasons for interest in child status effects listed in Section II should lead a researcher to try to separate the impact of these effects on preferences from their impact on constraints.

Considering preferences, the first issue to be dealt with, if only for the sake of reference, is the structure of household decision-making. It may be useful to test the collective rationality assumption at the household level, see Chiappori (1986). Having admitted the validity of a household utility function to rationalize household choices, the researcher might still want to test certain commonly adopted separability assumptions. For example, equation (III-1) embodies the assumption of separability of consumption and female 'non-work' time from male 'non-work' time. A further disturbing aspect of equation (III-1) is that it mixes elements of the budget constraint with other conditioning variables, but this has no impact on the subsequent analysis in the paper. Empirical evidence on separability of male 'non-work' time from other arguments of the household utility function is not clear-cut: Blundell and Walker (1986) do not reject it, but Kooreman and Kapteyn (1987) and Browning and Meghir (1989) do. Altug and Miller (1990) reject separability of male non-work time and consumption from female non-work time. The study of Kooreman and Kapteyn distinguishes eight different types of time use for two potential earners and this framework seems attractive when studying the impact of child effects on female labour supply. At any rate the consideration of a household production function is attractive, if only to guide functional form specification and help interpret the estimated coefficients (see Blundell and Walker, 1984, and Browning et al., 1985).

However, every responsible father will be depressed by looking at equation (VII-2), from which he can gather that, as far as the 'quality' of his children is concerned, he might just as well just do anything with his non-work time. Again this points to the usefulness of disaggregating 'non-work' time in several time uses: equation (VII-2) extended to a vector of female time uses would be acceptable if all fathers had identical vectors of time uses (disregarding differences in efficiency of course).

Turning to constraints it is surprising that the authors say so little about taxes and benefits, since these may cause major problems for the identification of the preference parameters. Thus, Blomquist (1988) shows that the recurring empirical result of a wage coefficient close to zero for men, leading to the hasty conclusion that taxes may have only a negligible impact on their labour supply, may well result from an improper treatment of the tax system when taxes do indeed affect labour supply. Moffitt (1988) discusses the difficulty of disentangling child effects on preferences from their effects on the budget constraint through child-related benefits and marginal taxes.

Hours restrictions should be considered as well. Specifying the model in terms of yearly hours rather than weekly hours solves only some of the problems here. The availability of part-time jobs may have an important impact on the labour market participation of mothers. Neglecting this aspect will cause bias in the preference parameters (see for instance the model of Ilmakunnas and Pudney, 1990). Fixed costs of work are mentioned in the survey, but they are absent from all models discussed, since the reservation wage is equated with the virtual wage at 0 hours. Again proper account of these fixed costs may be important for the identification of the impact of child effects on preferences. Moreover, an institutional aspect like the availability of day care may take a wide share of the explanation of the participation of women with young children. It seems to me an act of faith to think that all these aspects are properly taken care of by the comparison of a reservation wage and a market wage.

In view of the general difficulty of disentangling the impact of child effects on constraints from their impact on preferences, it may prove useful to conduct studies along the lines of Nakamura and Nakamura (1981 a). They estimate the same model for Canada and the U.S., but the idea I have in mind would be to maintain the assumption of identical preferences across the border as an identifying assumption.

Turning now to econometric aspects, what I have to say is almost entirely implied by my previous discussion of a modelling strategy. While I agree with the authors on the limitations of exogeneity tests due to the difficulty of finding appropriate instruments, and while I respect their emphasis on output space model evaluation methods, I still think that they go too far in their distrust of other methods of statistical inference. Even with the critique of Leamer (1978) concerning specification searches in mind, the usual specification tests based on M-estimation or on GMM-estimation are a useful investigation tool which may lead to a better understanding of the shortcomings of a particular model. Examples of this are given by Chesher (1990) and Laisney et al. (1991). While a sequence of alterning test and estimation stages based on a single data set is problematic from a statistical point of view, it would seem unreasonable to disregard the wealth of insights than can be gained by thoroughly testing a *given* specification. To paraphrase the authors of this survey, I insist that such tests should be admitted in the court-room.

François Laisney
GREMAQ, Université de Toulouse I, and
Institut für VWL und Statistik, Universität Mannheim

Additional references

Altug, S. and R.A. Miller (1990): 'Household Choices in Equilibrium', *Econometrica*, 58, 543-570.

Blomquist, N.S. (1988): 'Nonlinear Taxes and Labour Supply', *European Economic Review*, 32, 1213-1226.

Blundell, R.W. and I. Walker (1984): 'A Household Production Specification of Demographic Variables in Demand Analysis', *Economic Journal*, 94, 59-68.

Browning, M. and C. Meghir (1989): 'Testing for the Separability of Commodity demands from Male and Female Labour supply', *UCL discussion paper* 89-05.

Chesher, A. (1990): 'Optimal Scrapping, Vehicle Operating Costs and the Estimation of the Benefits to Highway Improvements', Chapter 14 in Florens et al. (eds.), *Microeconometrics: Surveys and Applications*, 368-393. Oxford: Basil Blackwell.

Chiappori, P.A. (1988): 'Rational Household Labour Supply', *Econometrica*, 56, 63-89.

Ilmakunnas, S. and S. Pudney (1990): 'A Model of Female Labour Supply in the Presence of Hours Restrictions', *Journal of Public Economics*, 41, 183-210.

Kooreman, P. and A. Kapteyn (1987): 'A Disaggregated Analysis of the Allocation of Time within the Household', *Journal of Political Economy*, 95, 223-249.

Laisney, F., M. Lechner and S. Strøm (1991): 'Lessons from Specification Tests for a Labour Supply Model', *Annales d'Économie et de Statistique*, 21/22, 193-217.

Moffitt, R. (1988): 'Demographic Behavior and the Welfare State: Econometric Issues in the Identification of the Effects of Tax and Transfer Programs', paper presented at the Conference of the European Society for Population Economics, June, University of Mannheim.

ECONOMETRIC REVIEWS, 11(1), 93-96 (1992)

REPLY

We are grateful to the reviewers for their thoughtful comments.

Blundell lists three sources of state dependence. We wonder why he feels the habit persistence source is least important, but are glad it has been pointed out that readers could conclude (incorrectly) that we feel the other sources are unimportant.

Note that the Inertia Model wage equations include lagged labor supply variables. One reason for this is that "Longer hours of work mean more hours of accumulated experience on the job ..." (Nakamura and Nakamura, 1985b, p. 25). This source of state dependence is dealt with as an "indirect" child status effect in our survey. State dependence due to employment status-related differences in the arrival of wage offers also falls under the "indirect" heading but is neglected in our survey paper. Chapter 7 of our 1985 book discusses this issue, but a better reference is Devine and Kiefer (1991).

Browning and Dooley clarify the important point that conditioning variables need not be predetermined or exogenous. The Browning and Meghir paper they cite provides a cogent example. This point is developed from a different perspective for microeconomic research and microsimulation modeling in Orcutt (1987).

The new results Blomquist reports, on the robustness of estimated child status effects to tax corrections, are interesting. The models and data used in Nakamura, Nakamura and Cullen (1979) and in the Canadian portion of Nakamura and Nakamura (1981) are identical. Coefficient differences are due solely to the correction for federal and provincial taxes in the 1981 paper. In accordance with the new Hansson-Brusewitz and Blomquist results, the child status coefficients are very similar in our 1979 and 1981 papers.

We apologize to François Laisney and fathers everywhere for ignoring their parental contributions! (Alice Nakamura also wishes to acknowledge Masao's efforts in this respect.) Laisney's insistence on an open door at the courthouse is welcome too. We had only intended to remind readers of the tenuous foundations of inference in a world of approximate models, and to argue against summary elimination of models that fail specification tests.

THE QUEST FOR STRUCTURAL RELATIONS

The reviewers suggested that our research is not structural. We feel it *is* structural in the original sense of this term.

As Epstein (1987) notes, the term "structural" has its econometric origins in the writings of Tinbergen, and reflects his optimism that economists would soon be able to provide reliable predictions of the effects of macroeconomic stabilization policies.

In the form in which the structural approach was codified in the early econometrics texts, Tinbergen's original goal of predicting the effects of real policy actions was clearly evident. Consider the following excerpts from Christ's classic text:

> It is natural to call the autonomous equations "structural", because it is natural to think of the economy as a "structure" whose component parts are the various groups or sectors each of which corresponds to an autonomous equation.
> (Christ, 1966, p. 21)

> ... a *structure* is a set of autonomous relationships sufficient to determine uniquely the *conditional probability distributions* of the endogenous variables, given the values of the exogenous variables.
> (Christ, 1966, p. 153)

> The kind of structural change that is likely to invalidate the parameters of the old reduced form is likely to stem from a change in one or a few of these structural relationships, and to leave other structural relationships unaltered. But *all* the reduced-form relationships are likely to be altered by any single structural change, because typically each is dependent upon every structural relationship. We describe this situation by saying that structural relationships have a high degree of *autonomy* ... and that reduced-form equations have only a low degree of autonomy....
> (Christ, 1966, p. 247)

A BEHAVIORAL RESEARCH STRATEGY

In our research on labor supply behavior, we have demonstrated a persistent concern for finding relationships with high degrees of autonomy: relationships stable in different institutional environments (the U.S. and Canada), over time, and for different demographic groups. We have sought definitions for, and ways of isolating, factors that could be altered by government policies (the industrial mix, opportunities for women to find employment in specific industries and occupations, and the after-tax wage rate in our 1981 and 1983 studies; and our search for ways of characterizing child status effects in our recent work). We have devoted effort to exploring the fit between the predicted and actual conditional distributions of the dependent variables for our models, given the values of selected explanatory variables.

Modern structural proponents do not accept research efforts such as ours as structural because the parameters of the hypothesized underlying preference structure and budget constraint cannot be recovered. They push for models that are causally complete at the level of the individual and the family.

Perhaps relabelling our efforts as "behavioral research" might help us escape the "reduced form" connotations of mindless forecasting exercises. The term "behavioral" highlights our objective: the prediction of behavioral responses to identifiable actions or to observable changes in circumstances.

The cornerstone of the approach is the identification of the actions or circumstances of interest. This is why we ask: "What is meant by child status effects?" A second principle is to develop models in which the variables for the specified actions or circumstances are explicitly included. For example, if what is needed are estimates of the labor supply responses of women to the birth of their children, then the model must include a "birth" variable.

Principle three is that the actions or circumstances of interest should be represented together with any important real life contextual conditions. For example, every mother of a new baby either was or was not working in the year prior to the birth. We are grateful to Browning and Dooley for pointing out that a reader could surmise that "The Nakamuras implicitly conclude that the 'treatment-response' interpretation is not warranted." This is not the case. Rather we feel that the child status effects *and the relevant contexts for these effects* need to be more thoughtfully and fully defined, with attention to what factors are observable (and hence can be conditioned on) in the situations in which the estimation results are to be utilized.

The final principle is to check for spurious associations, data permitting, by using the estimated model to forecast the behavioral responses of interest in multiple settings. For example, if there is concern that tax laws should have been accounted for, the goodness of the predictions can be compared for different U.S. states with different state tax laws, or for Canadian versus U.S. women.

Alice Nakamura and Masao Nakamura

ADDITIONAL REFERENCES

Devine, T.J. and Kiefer, N.M., Empirical Labor Economics: The Search Approach, Oxford Univ. Press, 1991.

Epstein, R.J., A History of Econometrics, Elsevier Science, 1987.

Orcutt, G.H., "Joint Conditional Probability Functions for Modeling National Economies," in M.L. King and D.E.A. Giles (Eds.), Specification Analysis in the Linear Model, Routledge and Kegan Paul, 1987.