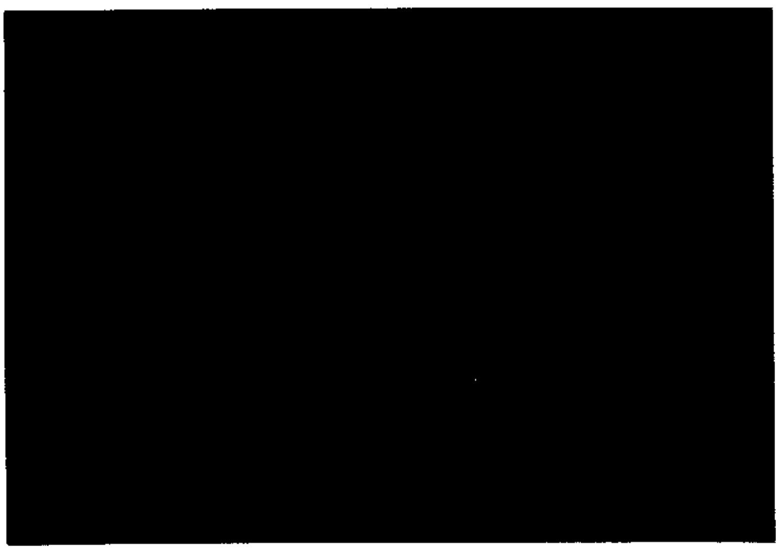
National Longitudinal Surveys

U.S. Department of Labor
Bureau of Labor Statistics

Discussion Paper

The National Longitudinal Surveys (NLS) program supports many studies designed to increase understanding of the labor market and methods of data collection. The *Discussion Papers* series incorporates some of the research coming out of the program. Many of the papers in this series have been delivered as final reports commissioned as part of an extramural grant program, under which grants are awarded through a competitive process. The series is intended to circulate the findings to interested readers within and outside the Bureau of Labor Statistics.

Persons interested in obtaining a copy of any *Discussion Paper*, other NLS reports, or general information about the NLS program and surveys should contact the Bureau of Labor Statistics, Office of Economic Research, Washington, DC 20212-0001 (202-606-7405).



NLS Discussion Papers

Dynamic Models of the Joint Determination of Labor Supply and Family Structure

Saul Schwartz
Robert Hutchens
George Jakubson

June 1991

Report: NLS 92-3

Dynamic Models of the Joint Determination of Labor Supply and Family Structure

Saul Schwartz
Carleton University and Tufts University

Robert Hutchens
Cornell University

George Jakubson
Cornell University

June 1991

This project was funded by the U.S. Department of Labor, Bureau of Labor Statistics under grant number E-9-J-8-0090. The authors wish to thank the participants in seminar at BLS, Tufts University, Carleton University, and Cornell University. In addition, some of this research was conducted at the Cornell National Supercomputer Facility, Center for Theory and Simulation in Science and Engineering, which is funded in part by the National Science Foundation, New State, and IBM Corporation. Opinions stated in this paper do not necessarily represent the official position or policy of the U.S. Department of Labor.

Dynamic Models of the Joint Determination of Labor Supply and Family Structure

Saul Schwartz
Carleton University and Tufts University

Robert Hutchens
Cornell University

George Jakubson
Cornell University

This project was funded by the U.S. Department of Labor, Bureau of Labor Statistics under Grant Number E-9-J-8-0090. Opinions stated in this document do not necessarily represent the official position or policy of the U.S. Department of Labor.

We would like to thank a number of individuals for their assistance during this project. They include Dr. Michael Pergamit, our project officer at the Bureau of Labor Statistics (BLS), as well as participants in seminars at BLS, Tufts University, Carleton University, and Cornell University. In addition, some of this research was conducted at the Cornell National Supercomputer Facility, Center for Theory and Simulation in Science and Engineering, which is funded in part by the National Science Foundation, New York State, and IBM Corporation.

Dynamic Models of the Joint Determination of
Labor Supply and Family Structure

TABLES OF CONTENTS

DYNAMIC AND SIMULTANEOUS MODELS OF THE JOINT DETERMINATION OF LABOR SUPPLY AND FAMILY STRUCTURE

Executive Summary

The past twenty-five years have seen the emergence of a number of important longitudinal data sets. Foremost among these is the set of surveys known, collectively, as the National Longitudinal Surveys (NLS). The availability of nationally-representative, longitudinal data has spawned a variety of econometric methods designed to study the economic behavior of individuals over time. These include hazard rate analysis, event history studies and techniques for pooling time-series and cross-sectional data.

This report deals with another econometric model developed to exploit longitudinal data - dynamic stochastic discrete choice models (Eckstein and Wolpin, 1989). We use data from the National Longitudinal Survey - Youth Cohort (NLS-Y) to explore a dynamic discrete choice model of the labor force participation and marital status of young mothers. The theory underlying such a model is quite appealing. Expectations about the future are allowed to influence current decisions in an explicit utility-maximization framework. In that sense, our model is a <u>structural</u> one.

The econometric estimation of our model requires the solution to a recursive dynamic programming problem and the maximization of a multi-period, multi-state likelihood function.

The programs required to estimate the parameters of our model are available from the authors upon request (see Appendix C).

Because the dynamic discrete choice model is relatively new and somewhat complicated, our work (and this report) moves from simple models to more complex models. In the first part of Chapter 1, we estimate relatively simple models of labor force participation and marriage using standard discrete-choice techniques. Then, we exploit the longitudinal nature of the NLS-Y by adding lagged values of the two dependent variables to the simple models. The most complicated model in Chapter 1 is a two-equation simultaneous probit model of labor force participation and marriage.

Aside from the development and implementation of our dynamic discrete choice model using the NLS-Y, several other interesting results arise from of our work on this project:

(1) we found no evidence of any interdependence between marital status and labor force participation. In particular, in the simultaneous probit model estimated in Chapter 1, current labor force participation did not affect current marital status nor did current marital status affect current labor force participation. Furthermore, we could not reject the hypothesis that the covariance between the error terms of the equations representing labor force participation and marital status was zero;

(2) adding lagged dependent variables as explanatory variables to the models estimated in Chapter 1 indicated that there is a higher-than-expected correlation between past status and current status. This leads to the conclusion that, except for unobserved factors, the determinants of the "initial condition" - the marital status/labor force participation prevailing at the time the woman first had a child - seems to persist over time. For example, the most important determinant of whether a woman worked in 1985 was whether she worked in 1984. By contrast, demographic variables pale in significance beside lagged dependent variables. For example, once we account for past participation in the Aid

to Families with Dependent Children (AFDC) program, the race and ethnicity of the young mother becomes irrelevant to her labor force participation decision. While there is substantial change in labor force participation and marital status - 50% of the young mothers in the sample change one or the other before 1985 - these changes seem to be the result of factors that we cannot observe. These conclusions are buttressed by similar results from the dynamic model of Chapter 2;

(3) the dynamic stochastic discrete choice model did <u>not</u> lead to any results that were substantially different that the results obtained from the simpler Chapter 1 models. Past values of the "state" variable were quite important while current demographic characteristics were relatively unimportant.

Our model posits rational decision-making by these young mothers. Since this assumption is not directly tested in our model, our results may be affected if the assumption is incorrect. Others researchers, however, have found this model more useful than standard models in other contexts (see Eckstein and Wolpin, 1989, for a review of this literature).

The constraints imposed by the computational burden of the estimation forced us to keep our model quite simple. The similarity of results across dynamic and static models may indicate only that simplicity. While the model may be too simple to capture behavior adequately, it is a step in the right direction. If there is to be progress in modeling labor force participation, we believe that a structural approach is absolutely essential.

DYNAMIC AND SIMULTANEOUS MODELS OF THE JOINT DETERMINATION OF LABOR FORCE SUPPLY AND FAMILY STRUCTURE

Introduction

The past twenty-five years have seen the emergence of a number of important longitudinal data sets. Foremost among these is the set of surveys known, collectively, as the National Longitudinal Surveys (NLS). The availability of nationally-representative, longitudinal data has spawned a variety of econometric methods designed to study the economic behavior of individuals over time. These include hazard rate analysis, event history studies and techniques for pooling time-series and cross-sectional data.

This report deals with another econometric model developed to exploit longitudinal data - dynamic stochastic discrete choice models (Eckstein and Wolpin, 1989). We use data from the National Longitudinal Survey - Youth Cohort (NLS-Y) to explore a dynamic discrete choice model of the labor force participation and marital status of young mothers. The woman's choices are discrete because in any time period, she is either part of the labor force or she is not; in any time period, she is either married or she is not. Furthermore, these models are stochastic in the sense that observably identical individuals may not behave in identical ways because of factors that are unobservable to the researcher.

The advantages of such a dynamic model are best understood when contrasted with a static model. A static model explains the labor force participation and marital status decisions in terms of current variables (such as current wages, current numbers of children and current age). In contrast, a dynamic model such as ours explains labor force participation in terms of both current variables and the expected values of future variables. For example, a dynamic model would incorporates the idea that today's decision to participate in the labor force will affect future levels of income, children, and schooling. Moreover, one's expectations about these potential consequences feed back into today's decision to participate in the labor force. Thus, a dynamic model has the advantage of yielding a more realistic picture of actual behavior.

Because of the explicit utility-maximization that is its theoretical base, we think of our model as a structural one. An alternative to structural estimation, an alternative explored in Chapter 1 of this report, is the estimation of a model that "approximates" the reduced form of the structural model. Under this alternative strategy, one implicitly solves the dynamic structural model for its reduced form, in which the endogenous variables are a function of current and past realizations of the exogenous variables. Although the explicit reduced form solution to a structural dynamic model is usually nonlinear and extremely complex, it is always possible to take a Taylor expansion to obtain a linear approximation of this reduced form. In the reduced form, each endogenous variable is a function of a linear combination of coefficients and exogenous variables as well as an error term. The coefficients are then the object of estimation.

The principal advantage to the "approximation" approach is that it is less restrictive. As such, it may provide estimates of how a large variety of exogenous variables affect endogenous variables. The structural approach, implemented in Chapter 2, involves using an iterative maximization routine to solve a system of nonlinear equations in each time period. This complexity limits the range of explanatory variables that can be incorporated into the analysis. The "approximation" approach is computationally simpler, requiring less programming and computer time.

The structural approach has, however, other advantages. First, and perhaps most important, estimation is focused on utility functions and constraints. In contrast to the "approximation" approach, the assumptions underlying the estimation are explicit. Second, the structural approach can provide more precise parameter estimates and stronger (more restrictive) tests of the theory. Wolpin (1984) argues quite forcefully that the structural model, if correctly specified, implies restrictions that permit more precise inference and a more parsimonious representation of complex relationships. But, he goes on to say, if the model is incorrectly specified, all statistical inferences may be contaminated, regardless of the offending assumptions (p. 854).

Our work involves specifying <u>four</u> possible choices for each woman - two labor force participation "states" (in the labor force and not in the labor force) as well as two marital statuses (married or not married). The estimation of the parameters of such a model involves complicated and computer-intensive maximum-likelihood techniques. The relevant programs, which we have used to estimate models of up to six "states" were developed by George Jakubson and are available upon request.

Working with longitudinal data also requires considerable effort to ensure that survey responses are consistent over time. This is especially true of the NLS-Y codes for the inter-relationships among the individuals who move in and out of families over time. We spent a great deal of time "cleaning" the data as part of this project and as part of another, related project (Hutchens, Jakubson and Schwartz, 1990b). The result of those efforts is a relatively "clean" set of data on the family structure of NLS-Y female respondents (see Appendix A). We were ably assisted in that effort by Angela Mikalauskas.

Since our focus is primarily methodological, we do not spend a great deal of space reviewing the vast literature on the determinants of labor force participation or on the smaller literature concerning the determinants of marital status. For reviews of that literature, see Johnson and Skinner (1986), McElroy (1985), Gonul (1989), and Killingsworth (1980).

Because dynamic discrete choice models are relatively new and somewhat complicated, our work (and this report) begins with simple "approximation" models and then moves to the more complex "structural" model. In the first part of Chapter 1, we estimate relatively simple models of labor force participation and marriage using standard discrete-choice techniques. Then, we exploit the longitudinal nature of the NLS-Y by adding lagged values of the two dependent variables to the simple models. The most complicated model in Chapter 1 is a two-equation simultaneous probit model of labor force participation and marriage.

Chapter 2 contains the theoretical and empirical versions of our structural dynamic model. The chapter begins by laying out the utility maximization assumptions that underlie the later empirical work. After describing the statistical issues involved in estimating the dynamic programming model implied by the theory, we present our empirical parameter estimates. A short summary concludes the report.

We would like to thank a number of individuals for their assistance during this project. They include Dr. Michael Pergamit, our project officer at the Bureau of Labor Statistics (BLS), as well as participants in seminars at BLS, Tufts University, Carleton University, and Cornell University. In addition, some of this research was conducted at the Cornell National Supercomputer Facility, Center for Theory and Simulation in Science and Engineering, which is funded in part by the National Science Foundation, New York State, and IBM Corporation.

CHAPTER 1

Cross-sectional Models of Labor Force Participation and Marital Status

In this chapter, we begin our examination of the relationship between marital status and labor force participation. We start by selecting a sample of women with children. We then estimate a series of single-equation models of the two dependent variables. In the next part of the chapter, we estimate the structural parameters of a cross-sectional two-equation system in which one equation represents the marriage decision and the other equation represents the labor force participation decision. This bivariate simultaneous equations model enables us to estimate, for these young mothers, the impact of labor force participation on marriage and the effect of marriage on labor force participation.

We exploit the time-series nature of the NLS-Y data by introducing past values of the dependent variables into our econometric models. Adding this dynamic element to the model allows us to account for the impact of past decisions on current status. For example, we can answer questions such as "Does last year's marital status affect this year's labor force participation decision?"

The AFDC program (which is an important source of financial support for these women) enters the model in that past AFDC participation, treated as an exogenous variable, is allowed to affect both current labor force participation and current marital status.

We estimate three sets of models in this chapter. The first set are essentially single-equation models of marriage and labor force participation. In that context, we look at how past marital status and labor force participation affect current marital status and labor force participation.

The second set of models are bivariate simultaneous equations models without any dynamism - only current values of the variables appear. The major difference between this model and a standard simultaneous equations model is that both dependent variables - marital status and labor force participation - are dichotomous. Thus our model is a "simultaneous probit" model. The third set of models combines the first two by introducing lagged dependent variables into the simultaneous probit models.

I. Single-Equation Models

In this section, we estimate two single-equation, reduced-form models of labor force participation and marital status, respectively. As noted in the introduction, these are linear "approximations" to the reduced form of a structural model, described in Chapter 2.

The basic facts about marital status and labor force participation in our sample are straightforward. The women in the sample were all 14-21 years of age in 1979, when the NLS-Y began. In 1985, the year to which these estimates apply, the women were aged 20-27. The sample consists of 2,221 women with children; of these, about 47% were working in 1985 and 65% were married (or living with a "partner"). Of the married women, almost 49% were working while 38% of the unmarried women were working. The patterns are quite different across racial and ethnic groups. Among the 715 black women in the sample, almost 62% are unmarried. By contrast, of the 1,506 women who are not black, only 23% were unmarried.

The models we estimate are cross-sectional, dichotomous probit models. These single-equation reduced-form models are too simple, in theoretical terms, to capture some potentially important links between the two decisions. For example, the two decisions are assumed to be made independently. Despite their simplicity, these models form a baseline from which we can assess the gain to be realized from more complicated models.

We assume that labor force participation depends, in general, on a woman's comparison of her reservation wage to the market wage available to her. Market wages depend, in turn, on previous labor force experience; past labor force participation should thus increase the probability of current labor force participation, holding other variables constant. Market wages also depend on educational attainment, represented here by the "highest grade completed" by the woman. Reservation wages depend on the woman's tastes for leisure as well as her other obligations. Foremost among those obligations for our sample is childcare. The more children a woman has, and the younger the children are, the less likely a woman is to work.

A controversial and still unsettled question is the extent to which past AFDC participation affects current labor force participation. If AFDC participation, in and of itself, reduces a woman's inclination or ability to work, then past AFDC participation should decrease the current probability of labor force participation.

To an unknown extent, past marital status may influence current labor force participation. As in the case of past AFDC participation, the causal links between past marital status and current labor force participation are unclear.

Theoretical explanations for labor force participation are considerably more developed than theoretical explanations for marriage. In particular, the literature on labor force participation leads to clear implications for the specification of an econometric model. The theoretical rationale for marriage depends on the productivity gains potentially available to both parties. Marriages break down whenever those gains are insufficient (or whenever the gains are divided in such a way that one party is worse off than they would be outside the marriage). But since the productivity gains are unobservable, as is the distribution of the gains between partners, empirical analysts must be satisfied with an econometric model that asserts the importance of a number of observed variables (such as education and labor force experience) in terms of their potential contribution to marital "productivity."

Race is clearly an important correlate of the probability of marriage, simply because of the observation, noted above, that a much smaller proportion of black mothers are married.

Table 1.1 contains parameter estimates from two simple reduced-form cross-sectional models of marital status and labor force participation. Variable definitions are shown in Appendix Table A1. As we would expect, the number and ages of children are correlated both with being married and with working. Having children under 3 years of age substantially increases the probability that a woman is married and substantially lowers the probability that she works. To a lesser extent, this is also true of children between the ages of 3 and 6. Once a woman's children are of school age, however, they affect neither the probability of working nor the probability of being married. Black women are much less likely to be married, even in the multivariate context. Hispanic women are also somewhat less likely to be married than white women.

Turning to the determinants of labor force participation for women with children, we find, as expected that women with more years of education are significantly more likely to work than those with fewer years of education. Black women are somewhat less likely to be working than white women, while Hispanic women are somewhat more likely to working than white women. Those who live in the South and those who live in SMSAs are more likely to be working than those who do not. The AFDC system does not have a significant impact on either marriage or working.

As a method of describing the characteristics of women who are married or who are working, these simple cross-sectional models are quite informative. They illustrate the importance of the number and ages of children as correlates of both marriage and working and they point to race and ethnicity as two other important correlates.

But as a way of understanding the underlying relationship between marriage and working, these models are too simple because that relationship, if any, is not made explicit. For example, one might think that unmarried women were more likely to work than married women since they lack the economic support that might be provided by a husband or partner. But looking at the bivariate relationship between marriage and working, one might come to the opposite conclusion. After all, as noted above, more married women were working than unmarried women. But this observation may be explained by the fact that unmarried women with children are eligible for economic support from the AFDC program, support that might enable them to stay out of the labor market and in their homes with their children.

Our immediate goal is to see if there is a relationship between marriage and work (and AFDC participation). We expect that there are unobserved factors affecting both the decision to work and the decision to marry, so we do not include "working" as an explanatory variable in the marriage equation, nor do we include "married" as an explanatory variable in the equation for labor force participation. These variables are endogenous and their inclusion on the right-hand side of the simple probit models would lead to inconsistent parameter estimates, estimates that would compromise our assessment of the relationship between marriage and work. Moreover, because almost all women receiving AFDC are unmarried and because the decision to participate in the AFDC program involves the same unobserved factors as play a role in the marriage and work decisions, we cannot include AFDC status as an explanatory variable in either equation.

Using linear models of this type, we have two ways of trying to disentangle the relationship between marriage and working. The first is to exploit the time-series nature of the NLS-Y by including lagged values of the dependent variables as explanatory variables in the equations for current marital status and current labor force participation. Furthermore, we can include lagged AFDC participation as a way of accounting for the availability of financial support for unmarried women. The second method is to use simultaneous equations techniques to test the hypothesis that unobserved factors affect both the labor force participation and the marital status decisions. The following two sections attempt those two extensions of the simple single-equation reduced-form models.

II. Single-Equation Models with Lagged Dependent Variables

The results of our inclusion of lagged dependent variables in our model are shown in Table 1.2. Lagged AFDC participation has a dramatic impact on both marriage and work. Those who received AFDC in 1984 were very much less likely to be married in

1985. This is not surprising since most AFDC recipients are unmarried. And, since relatively few women become married over a one year period, being unmarried in 1984 is a good "proxy" for not being married in 1985.

In the marital status equation, women who worked in 1984 were not significantly more or less likely to be married in 1985. The inclusion of lagged AFDC and lagged labor force participation does not change the numerical magnitude of the other significant coefficients, reported in Table 1.1. Those with young children remain more likely to be married; blacks and Hispanics remain less likely to be married.

In the labor force participation equation, the lagged AFDC variable is large and its standard error is small. Those on AFDC in 1984 were quite unlikely to be working in 1985. That is, very few of these mothers leave AFDC and enter the labor force in any one year. The coefficient on lagged marital status is significantly different from zero but very small. The lack of importance of lagged marital status in the labor force participation equation and of lagged labor force participation in the marital status equation is an early indication of the seeming independence of those two decisions. That independence is a theme that runs through the entirety of this report.

Interestingly, when we condition on lagged AFDC, race is no longer important. Put differently, when past AFDC participation is not accounted for, as in Table 1.1, it would seem that blacks are less likely to work than whites. AFDC participants are less likely to work than non-AFDC participants and the seeming correlation between race and working is caused by the correlation between race and AFDC participation. The inclusion of the lagged dependent variables also reduces the numerical magnitude of the regional variables, SOUTH and SMSA. In and of themselves, these variables seem not to affect the probability of working. Instead, they appear to affect the probability of receiving AFDC and determine the probability of working only through that indirect channel.

The importance of including lagged variables is not only to estimate coefficients more accurately in cross-sectional models. As can be seen above, marriage and work do not adjust instantaneously. In that sense, "where you are depends on where you've been". Marital status in 1985 and labor force participation in 1985 are greatly influenced by past marital status and past labor force participation.

That notion is an integral part of the structural model in Chapter 2, where we make explicit the links between past decisions and current states. To the extent that current variables are correlated with past dependent variables, their impact will be overstated in cross-sectional models. The example of how including past AFDC participation eliminates the correlation between race and labor force participation illustrates this idea.

Table 1.3 shows the effect of including a complete set of lagged dependent variables, variables going back over the 1979-

1985 NLS-Y survey period. Six years of lagged AFDC participation (1979-1984) as well as corresponding years of lagged labor force participation and lagged marital status are included along with the same set of current exogenous variables as appear in Table 1.1 and 1.2.

The inclusion of a more complete history for the young woman further reduces the impact of current variables. For example, with that history in place, only the presence of very young children (less than three years old) affects the probability of working. Women who have a history of working continue to work unless they have very young children.

Similarly, once "history" has been held constant, the presence of school-age children actually increases the probability of working. This is entirely plausible; having older children (compared to otherwise similar women) means than a woman is "farther along" in the life-cycle and is returning to work.

More importantly for our purposes, the inclusion of lagged marital status in the marriage equation and the inclusion of lagged labor force participation in the current labor force participation equation points up how strongly current status depends on past status. The single most important determinant of whether a woman worked in 1985 was whether she worked in 1984. By far the most important determinant of whether a woman was married in 1985 was whether she was married in 1984.

The policy implication here is that the immediate impact on marriage and labor force of almost any microeconomic policy is likely to be quite small. But any effects that policy does have will continue reverberate into the future.

In cross-sectional models, correcting for simultaneous equations bias when trying to ascertain the impact of, say, labor force participation on the probability of being married is very important. Part of the task is accomplished by including lagged dependent variables in the model and excluding the current dependent variables. This is because part of the impact of current labor force participation on current marital status is really the impact of past labor force participation (which is being picked up the current value of labor force participation).

III. Simultaneous Probit Models

In this section, we estimate simultaneous equations models of marital status and labor force participation. In light of our effort (in Chapter 2) to construct a structural model of discrete decisions, these simultaneous equations models need some explanation.

As discussed in the introduction, the model in Chapter 2 is a structural model in the sense that estimation is predicated on the maximization of utility functions subject to constraints. We think of all of the models estimated in this Chapter as linear approximations to a "structural model" of the sort laid out in Chapter 2. The models in the last section are also "reduced-form" models in the sense that only exogenous and lagged endogenous variables appear on the right-hand side.

But what about the models in this section, where we estimate simultaneous equations models in which current endogenous variables can appear on the right-hand side. In standard textbook discussions, such models are portrayed as "structural" but we reserve that term for models such as that presented in Chapter 2.

Econometrically, a simultaneous equations model simply imposes constraints on the reduced form parameters and then tests those restrictions. For example, the coefficient on an exogenous variable might be constrained to be zero (to have no impact) in one equation while it is allowed to be nonzero in another. Rather than thinking of the simultaneous equations models as the true "structure" of the joint decisions, we view them simply as a way of testing a set of plausible constraints on the reduced-form parameters. That is, our simultaneous equations model is another "reduced form" model.

While that view of such models seems simple, the actual estimation process is not. Both labor force participation and marital status are dichotomous variables, so standard simultaneous equations techniques must be modified in order to estimate the parameters of the models. We now discuss some of the econometric issues that arise in making those modifications.

The unifying principle of most models with limited dependent variables is the notion that while we, as researchers, might be able to observe only a limited number of values for a dependent variable (such as working or not working, or being married or unmarried), these observed values have been generated by continuous, unobserved latent underlying variables.

In our case, there are two such unobserved latent variables. Let A^* be an unobserved, continuous index of a young mother's desire to be married. Let L^* be another unobserved, continuous index of a woman's propensity to participate in the labor force.

As noted earlier, we have a fairly clear theory about how L^* "works" - if a woman is offered a wage higher than her reservation wage (which is a function of her marginal productivity at home) then she works; if her wage offer is less than her reservation wage she stays home. L^* is a function of the difference between the woman's market wage and her reservation wage. This difference is unobserved since reservation wages are unobserved.

Suppose that both A^* and L^* are both functions of a set of variables X. At this point, suppose X might include current and lagged exogenous variables as well as current and lagged endogenous variables. The coefficients on X for A^* and L^* will be denoted β_A and β_L, respectively. So,

(1) $A^* = X'\beta_A - \epsilon_A$

(2) $L^* = X'\beta_L - \epsilon_L$

where ϵ_A and ϵ_L are unobserved serially uncorrelated errors whose distribution is bivariate normal. The variances of ϵ_A and ϵ_L are unidentified and set to unity while their covariance is a parameter to be estimated.

While the woman is assumed to know both A^* and L^*, the researchers can observe only whether or not the woman is married and whether or not the woman participates in the labor force. Let the variables observed by the researcher be A and L where:

(3) $A = 1$ if $A^* > 0$; $A = 0$ otherwise;

(4) $L = 1$ if $L^* > 0$; $L = 0$ otherwise.

The models in the last section (Tables 1.1-1.3) treated the A^* and L^* variables in isolation using the NLS-Y variables defining A and L. The probability that $A = 1$ (that the NLS-Y respondent was married) was assumed to be a function of an X vector that included only current exogenous and lagged endogenous variables.

We now want to expand the scope of our model in order to explain <u>joint</u> decisions concerning not only marital status but also labor force participation. Thus we are concerned with the effect of current labor force participation on current marital status and with the effect of current marital status on current labor force participation. In addition, we want to allow the unobserved factors that influence labor force participation and marital status (that is, the error terms in equations (1) and (2)) to be correlated. This correlation, if any, would imply that there are common unobserved factors influencing both decisions.

These additional considerations add up to a hypothesis that decisions about marital status and labor force participation are made simultaneously. In terms of specifying an econometric model, that hypothesis implies that when we include "labor force participation" in X, its parameter estimate in equation (1) will be significantly different from zero. When we include "marital status" in X in the estimation of equation (2), its coefficient should also be significant. The way in which "labor force participation" and "marital status" enter equations (1) and (2), however, makes a difference in estimating the models.

One specification is that it is the qualitative variable - for example, whether or not one is in the labor force - that influences marital status, rather than the continuous underlying variable.

Another plausible specification is that the latent variable is the important determining factor; that, for example, the value of L^* is important in determining the value of A^*. A variant on this second model is perhaps most plausible. In that variant, the observed qualitative variables depends on the current latent variable and the lagged qualitative variables. That is, in making current decisions about labor force participation, a woman considers the current value of her propensity to be married (the current value of the latent variable) but only the observed value of past marital status. In other words, last year's marital status has a 0-1 impact but last year's propensity to be married (last year's value of the latent variable) is now forgotten or irrelevant.

We label these two models "A" and "B." Not only is each model specified differently, but each model must be estimated differently.

Specification of Model A

In Model A, using the definitions of Equations (1)-(4):

(5) $A^* = X'\beta_A + \gamma L - \epsilon_A$

(6) $L^* = X'\beta_L + \lambda A - \epsilon_L$

where X and the two corresponding vectors of parameters have been redefined to exclude A and L, the observed qualitative variables for marital status and labor force participation, respectively.

Unfortunately, as specified in Equations (5) and (6), this model is underidentified, regardless of the exclusions that might be imposed on the X vector. The reason for the underidentification is that the model is logically inconsistent (Schmidt, 1982). The problem can be seen as follows.

Following Maddala (1983), suppose the vectors β_A and β_L are all zero and that ϵ_A and ϵ_L are independent, normal variates; the argument holds even when these assumptions are not made but the point will be clearer if we make them. With these assumptions,

(7) $\Pr(A=1, L=1) = \Pr(A^*>0, L^*>0) = \Pr(\gamma - \epsilon_A > 0, \lambda - \epsilon_L > 0)$
$= \Pr(\epsilon_A < \gamma, \epsilon_L < \lambda) = F_A(\gamma) * F_L(\lambda)]$

(8) $\Pr(A=1, L=0) = \Pr(A^*>0, L^*<0) = \Pr(- \epsilon_A > 0, \lambda - \epsilon_L < 0)$
$= \Pr(\epsilon_A < 0, \epsilon_L > \lambda) = F_A(0) * [1 - F_L(\lambda)]$

(9) $\Pr(A=0, L=1) = \Pr(A^*<0, L^*>0) = \Pr(\gamma - \epsilon_A < 0, - \epsilon_L > 0)$
$= \Pr(\epsilon_A > \gamma, \epsilon_L < 0) = [1 - F_A(\gamma)] * F_L(0)$

(10) $\Pr(A=0, L=0) = \Pr(A^*<0, L^*<0) = \Pr(- \epsilon_A < 0, - \epsilon_L < 0)$
$= \Pr(\epsilon_A > 0, \epsilon_L > 0) = [1 - F_A(0)]*[1 - F_L(0)]$

The four probabilities in equations (7)-(10) must add up to unity since they represent the only four possibilities for any given woman. The sum of the four probabilities is:

$$(11) \quad 1 + F_A(0)*F_L(0) - F_A(\gamma)*F_L(0) - F_A(0)*F_L(\lambda) + F_A(\gamma)*F_L(\lambda)$$

Equation (11) will equal unity if γ or λ or both are equal to zero but not otherwise.

In general, a latent variable cannot be a function of its observed indicator in a single equation model and, in a two-equation model, only one of the 0-1 observed dependent variables can appear on the right hand side.

Thus, looking back to equations (5) and (6), either observed marital status (A) cannot be in the labor force participation equation or observed labor force participation (L) cannot be in the marital status equation. The constraint is imposed by the econometrics of the models and not by any economic reasoning. To make the model both econometrically estimable and economically plausible, we have to make an assumption about which decision "comes first".

For example, we could assume that the marital status decision is made first, as a function of only age, education, race, region of residence and the AFDC parameters and <u>not</u> as a function of labor force participation. Then, the labor force participation decision could be made as a function of the same demographic variables plus observed marital status.

We begin by estimating two versions of equation (5) and (6), denoting them as Model A1 and Model A2. In both cases, the vector

X contains the same list of current exogenous and lagged endogenous variables defined in Appendix Table A1 and used in the simple reduced-form equation models. The difference between models A1 and A2 is that one constrains the parameter γ to be zero (Model A1) and the other constrains the parameter λ to be zero (Model A2).

Specification of Model B

Model B is a cross-sectional simultaneous probit model of the 1985 living arrangements and labor force participation of young women with children. The dependent variables are the 0-1 labor force participation status and the 0-1 living arrangement status of a sample of women with children drawn from the National Longitudinal Survey - Youth cohort.

The "strength" of each woman's decision - represented by the amount by which A^* exceeds zero - is irrelevant in Model A. A woman whose labor force participation decision is "easy" (because her market wage is much higher than her reservation wage) is no more or less likely to be in the labor force than a woman for whom the decision to work was marginal (in the sense that her reservation wage is close to her market wage).

In Model B, we assume that it is not the 0-1 labor participation decision that is relevant but that it is rather the "strength" of that decision that is important. "Strength" is captured by the values of the unobserved latent variables, A^* and L^*. Algebraically,

(12) $$A^* = X'\beta_A + \gamma' L^* - \epsilon_A$$

(13) $L^* = X'\beta_L + \lambda'A^* - \epsilon_L$

While this model seems only slightly different from Model A, as represented by equations (5) and (6), it does not have the (econometric) problem of logical inconsistency.

Models A and B were both estimated using the LISREL based estimation package known as LISCOMP (Muthen, 1988). Using methods-of-moments type estimators, LISCOMP provides a flexible environment for estimating the parameters of latent variable models.

Tables 1.4 and 1.5 show the results of our estimation of those simultaneous (and dynamic) models of marriage and labor force participation.

The major result is easily stated. There does not seem to be any simultaneous equations bias to be corrected. Current labor force participation and current marital status seem to be independent. Furthermore, there seems to be no correlation between the current error terms of the two equations. This result echoes the lack of importance of lagged labor force participation and lagged marital status in the single equation models (see p.16). Tables 1.4 and 1.5 make this quite clear by putting three equations side-by-side for marriage and labor force participation, respectively.

In column (1) of each table is a reduced-form probit model of the dependent variable, with only current exogenous variables and lagged values of the dependent variable. Column (2) shows the equation from either Model A1 or Model A2 in which the current value of one dependent variable appears as an independent variable in the equation for the other dependent variable. For example, column (2) of Table 1.4 allows observed current labor force participation to affect current marital status. Finally, column (3) of each table is one of the two equations from Model B, the fully simultaneous model. In that model, the latent labor force participation variable is allowed to affect current marital status and the latent marital status variable is allowed to affect current labor force participation.

The thrust of both Tables 1.4 and 1.5 is that there is little or no evidence of any simultaneity between labor force participation and marriage, once the "history" of labor force participation and marriage is included in the models.

The simple reduced-form probit coefficients are essentially unchanged when we allow for a nonzero covariance between the error terms of the two equations and when we include current labor force participation in the marital status equation and vice-versa. This is true regardless of which method we use to introduce the simultaneity - using the current observed value of LFP or marital status (Models A1 or A2) or using the current latent LFP or marital status (Model B).

In no case is the estimated error covariance significantly different from zero; even the point estimates are quite small. Furthermore, the coefficients on LFP in the marital status equation and on marital status in the LFP equation are also very small in magnitude and not significantly different from zero.

Table 1.1

Single-Equation Cross-sectional Models of
Marital Status and Labor Force Participation
Women Aged 20-27, with Children, in 1985

Dependent Variables: LFP85 = 1 if respondent is working in 1985; 0 otherwise

MARRY285 = 1 if respondent is married or lives with a "partner"; 0 otherwise

Coefficient Estimates (Standard Errors)

Independent Variables	Marital Status (1)		Labor Force Participation (2)	
AGE	0.17	(.32)	0.48	(.30)
AGESQ/100	-0.18	(.66)	-0.92	(.62)
EDUC	-0.17	(.09)	0.26	(.10)
EDUCSQ/100	0.99	(.44)	-0.42	(.45)
BLACK	-1.17	(.07)	-0.25	(.07)
HISPANIC	-0.25	(.08)	0.15	(.08)
SOUTH	0.17	(.09)	0.29	(.08)
SMSA	0.02	(.07)	0.16	(.06)
KIDS2185	0.40	(.05)	-0.46	(.05)
KIDS2285	0.19	(.05)	-0.21	(.05)
KIDS2385	-0.02	(.05)	-0.07	(.05)
AFDCG85	-0.11	(.37)	0.10	(.37)
AFDCW85	*	*	*	*
Constant	-2.16	(3.9)	-8.49	(3.7)
Sample Size	2,221		2,221	
Mean of Dep. Variable	0.47		0.65	
-2 log likelihood	629.7		253.1	

A "*" indicates that the coefficient estimate was less than 0.005.

Table 1.2

Models of Labor Force
Participation and Marriage
with Lagged Dependent Variables
Women with Children, Aged 20-27, in 1985

Dependent Variables: LFP85 = 1 if respondent is working in 1985;
0 otherwise

MARRY285 = 1 if respondent is married or lives with a "partner";
0 otherwise

Coefficient Estimates (Standard Errors)

Independent Variables	Marital Status (1)		Labor Force Participation (2)	
LFP84	-0.09	(.07)	-	
MARRY284	-		-0.13	(.07)
AFDC84	-1.28	(.08)	-0.96	(.08)
AGE	-0.02	(.34)	0.36	(.31)
AGESQ/100	0.15	(.69)	-0.71	(.64)
EDUC	-0.11	(.10)	0.26	(.10)
EDUCSQ/100	0.58	(.46)	-0.53	(.44)
BLACK	-0.94	(.07)	-0.06	(.07)
HISPANIC	-0.30	(.09)	0.14	(.08)
SOUTH	-0.02	(.09)	0.17	(.09)
SMSA	0.00	(.07)	0.16	(.06)
KIDS2185	0.46	(.06)	-0.45	(.06)
KIDS2285	0.28	(.06)	-0.16	(.05)
KIDS2385	0.07	(.06)	-0.05	(.05)
AFDCG85	0.29	(.31)	0.35	(.29)
AFDCW85/100	0.04	(.30)	-0.60	(.27)
Constant	0.55	(4.1)	-6.51	(3.8)
Sample Size	2,221		2,221	
Mean of Dep. Variable	0.47		0.65	
-2 log likelihood	728.4		433.8	

Table 1.3

Models of Labor Force Participation and Marital Status with Lagged Dependent Variables Women with Children, Aged 20-27, in 1985

Independent Variables	Coefficient Estimates (Standard Errors) Marital Status (1)		Labor Force Participation (2)	
LFP84	0.12	(.09)	0.90	(.07)
LFP83	-0.05	(.10)	0.31	(.08)
LFP82	-0.11	(.10)	0.23	(.08)
LFP81	0.09	(.09)	0.19	(.08)
LFP80	-0.18	(.09)	0.16	(.08)
LFP79	0.09	(.09)	0.15	(.07)
MARRY284	1.80	(.10)	-0.13	(.10)
MARRY283	0.42	(.11)	0.05	(.10)
MARRY282	-0.03	(.12)	0.00	(.10)
MARRY281	-0.00	(.12)	0.16	(.10)
MARRY280	-0.07	(.12)	0.03	(.10)
MARRY279	0.19	(.12)	0.08	(.10)
AFDC84	-0.68	(.12)	-0.61	(.11)
AFDC83	0.00	(.14)	0.00	(.12)
AFDC82	0.00	(.14)	0.29	(.12)
AFDC81	-0.22	(.14)	0.13	(.12)
AFDC80	0.24	(.14)	-0.08	(.13)
AFDC79	0.07	(.15)	-0.12	(.14)
AGE	-0.52	(.42)	-0.28	(.34)
AGESQ/100	1.04	(.85)	0.46	(.70)
EDUC	-0.16	(.13)	0.16	(.10)
EDUCSQ/100	0.86	(.59)	-0.35	(.46)
BLACK	-0.41	(.10)	0.14	(.08)
HISPANIC	-0.18	(.11)	0.26	(.09)
SOUTH	0.03	(.11)	0.10	(.09)
SMSA	0.10	(.09)	0.14	(.07)
KIDS2185	0.24	(.07)	-0.33	(.06)
KIDS2285	0.09	(.07)	-0.04	(.06)
KIDS2385	0.08	(.07)	0.10	(.06)
Constant	6.23	(5.1)	1.90	(4.2)
Sample Size	2,221		2,221	
Mean of Dep. Variable	0.47		0.65	
-2 log likelihood	1486.4		869.2	

Table 1.4

Simultaneous and Dynamic Models
of Marital Status
Women with Children, Aged 20-27, in 1985

MARITAL STATUS

Independent Variables	Coefficient Estimates (Standard Errors)					
	Reduced-Form Probit		Model A2		Model B	
	(1)		(2)		(3)	
LFP85	-		0.12	(.19)	-0.01	(.06)
MARRY284	1.79	(.10)	1.80	(.10)	1.76	(.10)
MARRY283	0.41	(.11)	0.42	(.12)	0.38	(.12)
MARRY282	-0.01	(.12)	-0.02	(.13)	0.06	(.13)
MARRY281	0.00	(.12)	0.00	(.12)	0.01	(.12)
MARRY280	-0.07	(.12)	-0.06	(.12)	-0.08	(.12)
MARRY279	0.19	(.12)	0.18	(.12)	0.19	(.12)
AFDC84	-0.70	(.12)	-0.67	(.13)	-0.78	(.13)
AFDC83	0.01	(.13)	0.02	(.14)	0.10	(.14)
AFDC82	0.03	(.14)	0.02	(.15)	0.06	(.15)
AFDC81	-0.21	(.14)	-0.21	(.15)	-0.25	(.15)
AFDC80	0.24	(.14)	0.24	(.14)	0.31	(.14)
AFDC79	0.07	(.15)	0.08	(.17)	0.04	(.17)
AGE	-0.55	(.41)	-0.58	(.42)	-0.46	(.42)
AGESQ/100	1.08	(.84)	1.2	(.9)	0.9	(.9)
EDUC	-0.16	(.13)	-0.17	(.16)	-0.18	(.16)
EDUCSQ/100	0.86	(.58)	0.9	(.8)	0.9	(.7)
BLACK	-0.40	(.10)	-0.40	(.10)	-0.38	(.10)
HISPANIC	-0.18	(.11)	-0.19	(.11)	-0.24	(.11)
SOUTH	0.04	.11)	0.03	(.12)	0.08	(.12)
SMSA	0.10	(.09)	0.09	(.09)	0.14	(.09)
KIDS2185	0.23	(.07)	0.25	(.08)	0.22	(.08)
KIDS2285	0.10	(.07)	0.11	(.07)	0.08	(.08)
KIDS2385	0.08	(.07)	0.08	(.07)	0.09	(.07)
Error Covariance	-		0.01	(.09)	-0.06	(.08)
Constant	6.61	(5.0)	-7.04	(5.2)	5.64	(5.2)
Sample Size	2,221		2,221		2,221	
Mean of Dep. Variable	0.47		0.47		0.47	

Table 1.5

Simultaneous and Dynamic Models
of Labor Force Participation
Women with Children, Aged 20-27, in 1985

LABOR FORCE PARTICIPATION	Coefficient Estimates (Standard Errors)					
	Reduced-Form Probit		Model A1		Model B	
Independent Variables	(1)		(2)		(3)	
MARRY285	-		-0.04	(.13)	-0.01	(.04)
LFP84	0.91	(.07)	0.91	(.07)	0.91	(.07)
LFP83	0.31	(.08)	0.31	(.08)	0.32	(.08)
LFP82	0.22	(.08)	0.22	(.08)	0.25	(.08)
LFP81	0.18	(.08)	0.18	(.08)	0.18	(.08)
LFP80	0.15	(.07)	0.15	(.08)	0.14	(.08)
LFP79	0.15	(.07)	0.15	(.07)	0.15	(.07)
AFDC84	-0.61	(.10)	-0.63	(.10)	-0.65	(.11)
AFDC83	0.01	(.12)	0.00	(.11)	0.07	(.12)
AFDC82	0.28	(.12)	0.28	(.12)	0.34	(.12)
AFDC81	0.11	(.12)	0.11	(.12)	0.10	(.13)
AFDC80	-0.10	(.12)	-0.10	(.13)	-0.11	(.13)
AFDC79	-0.15	(.13)	-0.15	(.14)	-0.23	(.14)
AGE	-0.20	(.34)	-0.20	(.35)	-0.10	(.35)
AGESQ/100	0.32	(.70)	0.3	(.7)	0.1	(.7)
EDUC	0.16	(.10)	0.16	(.11)	0.25	(.11)
EDUCSQ/100	-0.41	(.46)	-0.4	(.5)	-0.8	(.5)
BLACK	0.01	(.08)	0.09	(.09)	0.11	(.09)
HISPANIC	0.24	(.09)	0.24	(.09)	0.24	(.09)
SOUTH	0.10	(.09)	0.10	(.10)	0.13	(.10)
SMSA	0.14	(.07)	0.14	(.07)	0.14	(.07)
KIDS2185	-0.33	(.06)	-0.33	(.07)	-0.38	(.07)
KIDS2285	0.00	(.06)	0.01	(.06)	-0.03	(.06)
KIDS2385	0.11	(.06)	0.11	(.06)	0.14	(.07)
Error Covariance	-		0.01	(.09)	-0.06	(.08)
Constant	0.68	(4.2)	-0.74	(4.2)	-0.97	(5.2)
Sample Size	2,221		2,221		2,221	
Mean of Dep. Variable	0.65		0.65		0.65	

Chapter 2

A Dynamic Stochastic Discrete Choice Model of Labor Force Participation and Marital Status

In Chapter 1, we estimated the parameters of one-period static models of labor force participation and marital status. Here, we estimate the dynamic four-alternative version of the same decisions.

As before, the relevant theoretical model refers to a young mother who chooses among four alternative states, defined by whether the women is married and whether she participates in the labor force. The four alternatives states are:

(1) married and in the labor force;
(2) married and not in the labor force;
(3) not married and in the labor force; and
(4) not married and not in the labor force.

Our model is explicitly dynamic. In choosing alternative i, the woman not only considers her utility in that alternative today, but also the utility she can expect to obtain in the future.

The model begins with a rational young woman with a time-invariant utility function and accurate forecasts of her expected utility (that is, forecast errors have zero mean). She exercises choice among the four alternative states, recognizing that today's decisions may have long-term effects. For example, choosing not to work in any period may reduce her future income (inside and outside of marriage). If she has a high discount rate, however, such future consequences may carry little weight in her decision-making.

The work reported in this chapter should be viewed as an exploration of multiple-alternative discrete choice models. While our model focuses on four alternatives, it can be expanded to any number of alternatives.[1]

In theoretical models of this type, current actions affect future decisions in two different ways. First, current actions might affect the returns to future actions or the constraints faced by the decision-maker in the future. Second, the decision-maker <u>recognizes</u> that fact and takes account of the probable future effects of current actions when deciding on the current action. Thus the object of maximization in the current period no longer involves only the utility function in the current period. Instead it is a "value function" which incorporates the current utility function and also the discounted expected value of next period's utility function. The decision-maker calculated the "discounted expected value" conditional on what she knows in the current period (the current period's information set).

There are two basic approaches to building an estimating model that respects the above discussion. If the dependent variable is continuous, the "Euler equations" approach is appropriate.[2] But the Euler equation approach is not applicable when, as in our model, the action to be explained is discrete rather than continuous. Instead, one must enumerate all the possible actions and evaluate the value function for each. This entails evaluating, for each possible action in the <u>current</u>

period, the expected future return of each possible <u>future</u> action. A rational decision-maker then takes that action which maximizes the value function. This approach is typically termed "dynamic programming."

There are then two approaches to implementing a dynamic programming model. In some cases the environment is stationary and one can show that the value function takes the same form over time (for example, some of the simple job search models take this form). In such cases one can work with the same function in each period.

In our problem the environment is not stationary so we cannot use this approach. For example, the number of children changes over time. Hence we must use a solution technique known as "backwards recursion." We first pick a terminal date, say T. Given her position at date T-1, the woman then faces a static optimization problem. We can thus characterize the optimal decision at T as a function of the values of the state variables in period T-1 along with any other exogenous variables at date T. Since T is the terminal period, no expectations of future events need to be calculated.

Now, at T-1, we calculate the expected value of making alternative decisions, conditioning on the values of the state variables at T-2. These are expected values since they include the (discounted) expected value of the period T value function, for the different possible period T-1 decisions. We must calculate the expected value associated with all possible decisions at date T-1. We then move backwards through time.

In general, at time t we calculate the expected values associated with the possible decisions which can be made at t, conditioning on the levels of the state variables at time t-1, and calculating the expected values at time t+1, conditioning on the choices made at t.[3]

I. Theoretical Model

Consider a young woman with at least one child and a time-invariant utility function. In period t, that woman chooses among our four possible marital status/labor force participation alternatives. In making her decision, she considers both the utility available in each alternative ($U_0(t)$, $U_1(t)$, $U_2(t)$ and $U_3(t)$), and the utility she can expect in the future given her chosen alternative in period t. It is this last element that distinguishes this model from the static model presented in the last chapter.

Introducing notation, let $d_i(t) = 1$ if alternative i is chosen at time t and $d_i(t) = 0$ otherwise, where $i = 1,...,4$. Alternatives are mutually exclusive; that is, $\Sigma\, d_i(t) = 1$.

We assume that $U_i(t)$ is a linear function of a vector of exogenous variables that are the same for all alternatives ($X(t)$), and a vector of dummy variables indicating the woman's alternative in period t-1. Thus,

(1) $U_i(t) = \beta_i X(t) + \alpha_i D(t-1) + u_i(t) + \epsilon(t)$, $i=0,...,3$; $t = 1,T$.

where:

$U_i(t)$ - the woman's utility in alternative i in period t;

$X(t)$ – a vector of exogenous characteristics (mostly demographic) of the woman in period t;

$D(t-1)$ – $\{d_0(t-1), d_1(t-1), d_2(t-1), d_3(t-1)\}$ is a 4 variable vector indicating the alternative chosen in period t-1;

$\epsilon(t)$ – a normally distributed random error that is uncorrelated with $X(t)$, $D(t-1)$, and $\epsilon(t')$;

$u_i(t)$ – an error term that is drawn from an extreme value distribution of the form, $F(u_i)=\exp\{-\exp\{-u_i/\tau\}\}$. It is pure white noise - $E\{u_i(t,j), u_i'(t',j')\}=0$, $i \neq i'$, $j \neq j'$, $t \neq t'$. Moreover, it is uncorrelated with $D(t-1)$, $X(t)$, and $\epsilon(t)$.

Note that the error term, $\epsilon(t)$ is not subscripted - it does not depend upon the alternative chosen in period t. β_i and α_i are vectors of parameters to be estimated. Also note that since D(t-1) enters into the U(t) function, past choices influence today's utility, and today's choices influence future utilities.

The woman's objective at any time t = 0,1,...,T, is to maximize,

$$(2)\quad E\left\{ \sum_{j=t}^{T} \rho^{j-t} \sum_{i=0}^{3} U_i(j) d_i(j) \,|\, \Omega(t) \right\}$$

where,

ρ is the woman's discount factor, and

$\Omega(t)$ is her information set at time t.

The woman maximizes (2) by choosing the optimal sequence of control variables for all future periods. Thus, she chooses the optimal $d_i(j)$, $i = 0,\ldots,3$; $j = t, t+1,\ldots, T$.

This problem can be solved through backward sequential solution of Bellman's equation (Bellman, 1957). In particular, let the value of choosing alternative i at time t be written,

(3) $L_iV(\Omega(t)) = U_i(t) + \rho E\{V(\Omega(t+1))|d_i(t) = 1\}$, $t = 1,\ldots,T-1$,
where $V(\Omega(t+1)) = \text{Max}_i\{L_iV\ \Omega(t+1)\}$.

Thus, $E\{V(\Omega(t+1))|d_i(t) = 1\}$ is the maximum expected value of utility in period t+1 given that the individual has chosen alternative i in period t. In period T, the value of choosing alternative i is simply,

(4) $L_iV(\Omega(T)) = U_i(T)$

As demonstrated below, the solution for $L_iV(\Omega(t))$ is obtained by substituting recursively from T.

II. Analytic Forms

To estimate this model one needs analytic forms for (3) and (4) as well as an expression for the probability of choosing alternative i. To that end it is simplest to rewrite the value of choosing alternative i at time t as,

(5) $L_iV(\Omega(t)) = L_iV(t)^* + \epsilon(t) + u_i(t)$,

where

$u_i(t)$ is the i.i.d. extreme value error,

$\epsilon(t)$ is the normally distributed random error,

$L_iV(t)^* = \beta_iX(t) + \alpha_iD(t-1) + \rho E\{V(\Omega(t+1))|d_i(t)=1\}$, for $t = 1,\ldots,T-1$, and

$L_iV(T)^* = \beta_iX(T) + \alpha_iD(T-1)$.

The term, $L_iV(t)^*$, is obtained by substituting (1) into equations (3) and (4).

Since $u_i(t)$ is distributed i.i.d. extreme value, the probability that the woman chooses alternative i can be written as a logit. To see this, let $P(i,t|D(t-1))$ be the probability that the woman chooses alternative i in period t conditional on the alternative chosen in the previous period. Then,

$$
\begin{aligned}
P(i,t|D(t-1)) &= \text{Prob}\{L_iV(\Omega(t)) > L_jV(\Omega(t)) \text{ for } j \neq i\} \\
&= \text{Prob}\{L_iV(t)^* + \epsilon(t) + u_i(t) > \\
&\qquad L_jV(t)^* + \epsilon(t) + u_j(t) \text{ for } j \neq i\} \\
&= \text{Prob}\{ u_j(t) - u_i(t) < L_iV(t)^* - L_jV(t)^*\}
\end{aligned}
$$

Note that since $\epsilon(t)$ is identical for all alternatives, it drops out of the last line of the above expression. Since $u_i(t)$ is distributed i.i.d. extreme value,

$$(6) \quad P(i,t|D(t-1)) = \exp\{L_iV(t)^*\}/ \sum_{j=0}^{3} \exp\{L_jV(t)^*\}$$

To compute this probability, one must first compute $L_iV(t)^*$. And from equation (5), that requires information on $E\{V(\Omega(t+1))|d_i(t) = 1\}$. The extreme value distribution of $u_i(t)$ implies an expression for $E\{V(\Omega(t+1))|d_i(t) = 1\}$. From Berkovec and Stern (1988), p. 8,

$$(7) \quad E\{V(\Omega(t+1))|d_i(t) = 1\} = \tau\{\gamma + K + \ln (\sum_{j=0}^{3} \exp\{D_jV(t+1)^*/\tau\})\},$$

where,

γ is Euler's constant (= .5772); and

K is a constant equal to the expected value of $\ln(4\exp\{\epsilon(t)\})$.

A solution is obtained by computing $L_iV(t)^*$, $i = 0,\ldots,3$ for the last period (period T), and then using equation (7) to compute $L_iV(t)^*$ for the next to the last period, T-1. Continuing this backward recursion, one obtains values of $L_iV(t)^*$ for all time periods. And given the values of $L_i V(t)*$, one can compute the probabilities in equation (6) for all time periods.

III. Estimation of the Dynamic Model

The goal of estimation is to use data on the exogenous variables (X(t) and D(t-1)) and the endogenous variables, D(t) to estimate the parameters τ, ρ, β_i, α_i, i = 0,...,3. To that end, note that the likelihood that the woman chooses alternative i in period t is,

$$(8) \quad \prod_{i=0}^{3} P(i,t|D(t-1))^{d_i(t)}$$

Generalizing slightly, the likelihood that she chooses the sequence of control variables $d_i(j)$, i = 0,...,3; j = 1,..,T is,

$$(9) \quad \prod_{t=1}^{T} \prod_{i=0}^{3} P(i,t|D(t-1))^{d_i(t)}$$

To estimate the model, we first form a sample likelihood function by taking the product of these individual likelihood functions, and then use a maximization routine with numerical derivatives to find the parameters, τ, ρ, β_i, and α_i, i = 0,...,3 which maximize the likelihood function.

Several of our data handling procedures must be discussed, however, before we describe the actual estimation process. First, we have assumed, from the onset, that a woman without children is very different than a woman with children, so we have excluded women without children from the analysis. For women with children, we begin the problem when the woman first has a child.

Specifically, we start by determining the first year in which the woman has a child of her own in the household. The woman's status in the year prior to that year is then her "initial condition."[4] Since the NLS-Y respondents had children at different times, the number of observed statuses (the $d_i(t)$) will vary across individuals; different women will contribute different numbers of decisions to the statistical problem.

Having assigned each woman an initial state, we then determine her status in each year, along with the vector of explanatory variables corresponding to each year.

We delete any observation for which any of the data we require is missing.[5] We are left with a sample of 1,983 young women. The upper panel of Table 2.1 shows how many of the respondents were in each of the four statuses in each year. Also shown, for each year prior to 1985, is the number of women who do not yet have a child and who are not yet included in the likelihood function.

The lower panel of Table 2.1 displays the distribution of these women by the number of decision periods. The sample is weighted (but not heavily weighted) in favor of those with a greater number of decision periods. This is partly because two groups of women contribute six decision periods, those with a child already present in 1979 and those whose first child appears on the record in 1980.

Table 2.2 displays means and standard deviations for the demographic variables used here. Since different women contribute different numbers of decisions, Table 2.2 displays the means of these variables for the whole sample (upper panel) and for the sample which is "active" during each year (lower panel).

The key part of our maximization of the likelihood function (shown in equation (9)) is the backwards recursion that takes place for a given woman.[6] There are two types of parameters there: (1) the status-specific coefficients on explanatory demographic variables (the β_i in the theoretical discussion above) and; (2) the status-specific coefficients on the woman's status in the previous period (the α_i in the theoretical discussion above).

The scale parameter of the extreme value distribution, τ, is not identifiable and is normalized to unity. This underidentification is common in models of discrete choice. For example, the probit model normalizes its variance to unity also. The discount factor, ρ, is identifiable in principle. We found it impossible, however, to identify this parameter.[7]

For each sample observation, we begin with the terminal date and calculates the value function, $L_iV(T)$, associated with each possible choice, as a function of the previous period's state. As shown in the discussion after equation (3), this value is a function only of terminal period demographic variables, X(T), and the woman's status in the last period, D(T-1).

Having calculated $L_iV(T)$, we then move to the previous time period. For periods other than T, the maximization routine must first calculate, for each status, the expected value of utility in the next period for each possible choice. That is, the routine must calculate the values of $E\{V(\Omega(t+1))\}$ for the periods other than T.

Using the values of $E\{V(\Omega(t+1))\}$, we can then calculate the value for the current period, $L_iV(t)$. We continue this process until we have exhausted all the decision points for this observation. Using the calculated values of $L_iV(t)$, for all four statuses, we then calculate the choice probabilities in equation (7). These probabilities represent the contribution of each observation to the log likelihood.

IV. Results from the Dynamic Model

Table 2.3 shows our estimates for the parameters α_i and β_i in equation (1); we show the absolute value of the asymptotic normal statistic, for the null hypothesis that the coefficient is zero, below each parameter estimate.

The estimates of β_i are presented as a matrix in the upper panel of Table 2.3. The rows of this matrix indicate the characteristic under consideration while the columns indicate the status whose utility function is being estimated. Each coefficient is an estimate of the effect of individual demographic characteristics on the utility of being in any one of the four labor force participation/marital status categories. The estimates of α_i are presented as a matrix in the lower panel of Table 2.3. The rows of the matrix indicate the previous year's status while the columns again indicate the status whose utility function is being estimated. Each coefficient indicates the effect of last period's status on the utility of being in any one of the four labor force participation/marital status categories this period.

As an example of the interpretation of the estimates of β_i, the negative coefficients on the variable BLACK in the first two columns of Table 2.3 indicate that black mothers gain less utility from being married than comparable white mothers, regardless of labor force participation. The positive coefficients on the variable BLACK in third and fourth columns of Table 2.3 indicate that black mothers gain more utility from not being married than white mothers, regardless of labor force participation. By contrast, the across-the-board positive coefficients on the variable HISPANIC indicate that Hispanic mothers receive have higher utility than white mothers in all four statuses, ceteris paribus. None of the coefficient estimates, however, allow us to reject the hypothesis that the coefficients are zero in the population.

Our estimates of β_i, the coefficients on the demographic variables, are uniformly insignificant. Looking at the algebraic sign of the coefficient estimates, we see that age

(AGE/10) has a small positive effect on the utility of being in the labor force. Number of children (KIDS1) has a positive impact on the utility of all statuses except being unmarried and in the labor force. The coefficients on race/ethnicity were discussed above.

As an example of the interpretation of the estimates of α_i, note the large and statistically significant coefficient (2.27) in the first column and first row of Table 2.3. This coefficient indicates the high utility associated with being married and in the labor force for those women who were also married and in the labor force in the previous period.

In the context of a dynamic programming model, the estimates for α_i reflect the value of remaining in the same status as in the previous period. If one made the optimal choice in the previous period, then the only reason to change status in the current period is the arrival of new information, either in terms of the disturbance or in terms of one of the explanatory variables. Therefore we expect the diagonal elements to be positive, or at least not negative and significant. This pattern is strongly supported by the coefficient estimates in Table 2.3; all of the coefficients on the diagonal of the α_i matrix are large, positive and statistically significant.

The off-diagonal elements, which represent the change in mean utility from changing status, should be negative. The argument is the same. If the previous decision were optimal, then the mean change in utility from the change in status should be negative. Changes do occur, but only in response to new unexpected information, represented here by the disturbance term.

As indicated in Table 2.3, though, some of the off-diagonal elements of the table are both positive and statistically significant. Entering the labor force increases utility. The change in utility when a women moves from being married and out of the labor force to being married and in the labor force is 1.44 with a normal statistic of 3.09. For unmarried women moving into the labor force, the relevant coefficient is 1.25 (2.68).

Conclusion

In this report, we have estimated a dynamic stochastic model of the labor force participation and marital status decisions of young mothers. In implementing that model empirically, we used data from the on-going National Longitudinal Survey - Youth Cohort.

The major advantage of such a model is theoretical. It incorporates the appealing notion that young mothers think about the future in making decisions today. The model uses an explicit utility-maximization framework, in contrast to less "structural" models as have been more commonly used.

Empirically, the model we use estimates the parameters of a four-state model. The same programs, however, can be used to estimate the parameters of larger models; we report its use in a six-state model in Appendix B to this report.

In order to assess the usefulness of the dynamic model, we have estimated a series of models, of increasing complexity. In this particular context, there does not seem to be much gain in using more complicated cross-sectional models. In particular, the earlier models, discussed in Chapter 1, indicate that the labor force participation and marital status decisions are independent of each other. These indications first appear in cross-sectional models using data for 1985.

The cross-sectional models also suggested that once past values of labor force participation and marital status are included in the analysis, demographic variables (such as race, ethnicity, age and education) are relatively unimportant in determining current labor force participation and marital status. This cross-sectional conclusion appears again in the dynamic model, which uses data from all years.

There are two ways to view that result. One is that there is little to be gained from using the dynamic model because the same conclusion can be drawn from the simpler model. The other view is that the dynamic model is working properly <u>because</u> it leads to the same conclusion as the simple model.

Our view is that the dynamic model is the theoretically appropriate model in this context. The lack of appreciable "gain" (in the form of more precise and plausible parameter estimates) should not impede its adoption.

The constraints imposed by the computational burden of the estimation forced us to keep our dynamic model quite simple. The similarity of results across dynamic and static models may indicate only that simplicity. While the model may be too simple to capture behavior adequately, it is a step in the right direction. If there is to be progress in modeling labor force participation, we believe that a structural approach is absolutely essential.

Table 2.1

Descriptive Statistics for the Sample of
Young Mothers Used in the Dynamic Model

A. Classification of Women with Children by Year, by Marital Status and Labor Force Participation

1985 Status	Number of Women in Each Category					
	1980	1981	1982	1983	1984	1985
Married						
In Labor Force	128	210	256	355	460	561
Not in Labor Force	248	331	434	503	535	571
Not Married						
In Labor Force	116	152	192	247	323	377
Not in Labor Force	230	322	412	447	456	474
Without Children	1261	968	689	431	209	0
Total	1983	1983	1983	1983	1983	1983

1985 Status	Percentage of Women in Each Category					
	1980	1981	1982	1983	1984	1985
Married						
In Labor Force	6.5	10.6	12.9	17.9	23.2	28.3
Not in Labor Force	12.5	16.7	21.9	25.4	27.0	28.8
Not Married						
In Labor Force	5.8	7.7	9.7	12.5	16.3	19.0
Not in Labor Force	11.6	16.2	20.8	22.5	23.0	23.9
Without Children	63.6	48.8	34.7	21.7	10.5	0
Total	100.0	100.0	100.0	100.0	100.0	100.0

B. The Distribution of Women with Children By Number of Available Decision Periods in Dynamic Model

	Number of Periods						
	1	2	3	4	5	6	Total
Number of Women	209	222	258	279	293	722	1,983
Percentage	10.5	11.2	13.0	14.1	14.8	36.4	100.0

Table 2.2

Means and Standard Deviations for Independent Variables in the Dynamic Programming Model Women with Children in 1985

Independent Variable	Mean	Standard Deviation
AGE	24.45	2.21
BLACK	0.32	0.47
HISPANIC	0.18	0.38
KIDS1	1.55	0.78

Sample Size = 1,983

Means for Independent Variables in the Dynamic Programming Model "Active" Decision Makers, by Year

Independent Variable	Year 1980	1981	1982	1983	1984	1985
AGE	20.57	21.17	21.93	22.68	21.55	24.45
BLACK	0.38	0.36	0.34	0.33	0.33	0.32
HISPANIC	0.14	0.16	0.17	0.18	0.18	0.18
KIDS1	1.08	1.18	1.28	1.38	1.46	1.55
Sample Size	722	1015	1294	1552	1774	1983

Table 2.3

Coefficient Estimates for a Four State Dynamic Programming Model of Marriage and Labor Force Participation

Status-Specific Coefficients on Demographic Variables, β_i
(absolute value of asymptotic normal statistic)

	Married In LF	Married Not in LF	Not Married In LF	Not Married Not in LF
CONSTANT	-.398 (.610)	.869 (1.342)	-.847 (1.293)	.772 (1.189)
BLACK	-.025 (.050)	-.332 (.662)	.316 (.630)	.437 (.871)
HISPANIC	.147 (.294)	.078 (.156)	.026 (.052)	.145 (.288)
AGE/10	.441 (.869)	-.363 (.716)	.783 (1.541)	-.453 (.894)
KIDS1	.083 (.165)	.397 (.793)	-.315 (.629)	.234 (.467)

Status-Specific Coefficients on Past Status Variables, α_i
(absolute value of asymptotic normal statistic)

	Current Status			
Previous Status	Married In LF	Married Not in LF	Not Married In LF	Not Married Not in LF
Married In LF	2.269 (4.872)	1.420 (3.041)	.124 (.262)	-.471 (.980)
Married Not in LF	1.444 (3.093)	2.634 (5.672)	-.447 (.941)	.734 (1.563)
Not Married In LF	.392 (.833)	.114 (.242)	2.127 (4.567)	1.580 (3.377)
Not Married Not in LF	-.599 (1.263)	.606 (1.296)	1.254 (2.685)	2.833 (6.099)

Value of Log Likelihood Function: -7209.10038

Appendix A

Variable Definitions and Data Preparation Issues

This Appendix begin with the definition of the variables appearing in the body of our report. The definitions appear in Table A1. The remainder of the Appendix discusses, in substantially greater detail, some of the problems in using the NLS-Y for time-series analysis of decisions concerning family structure.

The Problems in Defining Family Structure over Time in the NLS-Y

In order to make our results comparable to those of earlier work done on Current Population Survey (CPS) cross-sections, we decided to construct CPS-type marital status and living arrangement definitions, such as "primary family", "subfamily" and "unrelated individuals." A description of the available variables in the NLS-Y documentation suggested that these living arrangement definitions were feasible and would require fairly straightforward manipulations of the data. Unfortunately, we encountered numerous problems in the construction of our marital status and living arrangement measures because of inaccuracies in the documentation or miscodings in the data themselves. The latter problem diminishes in the later years of the survey, but is particularly prevalent during the early years of the survey (1979-1981).

Table A1

Variable Definitions for Models of Labor Force Participation and Marital Status

LFPxx	Labor force participation variable defined as 1 if the respondent is in the labor force in year xx and 0 otherwise.
MARRY1xx	Marital Status variable defined as 1 if respondent is married in year xx and 0 otherwise.
MARRY2xx	Marital Status variable defined as 1 if respondent is married or living with a "partner" in year xx and 0 otherwise.
AFDCxx	Welfare participation variable defined as 1 if the respondent received income from AFDC in year xx and 0 otherwise.
AGE(SQ)	The respondent's age in years, measured continuously from birth. AGESQ is AGE squared.
EDUC(SQ)	The highest grade completed by the respondent as of the date of interview in 1985. EDUCSQ is EDUC squared.
BLACK	Takes the value 1 if the respondent reports her race as black; 0 otherwise.
HISPANIC	Takes the value 1 if the respondent reports her ethnicity as Hispanic; 0 otherwise.
SOUTH	Takes the value 1 if the respondent's residence is in the South in 1985; 0 otherwise.
SMSA	Takes the value 1 if the respondent's residence is in an SMSA in 1985; 0 otherwise.
KIDS2185	The number of children (own, adopted or partners) of age 0, 1 or 2 years in 1985.
KIDS2285	The number of children (own, adopted or partners) of age 3, 4 or 5 years in 1985.
KIDS2385	The number of children (own, adopted or partners) 6 years old or more in 1985.
KIDS1	The total number of children, between 0 and 3 years of age, present in the respondent's household.

Table A1

Variable Definitions for Models of Labor Force Participation and Marital Status

(Continued)

AFDCG	The relevant 1985 AFDC maximum payment, for the respondents geographic state and family size.
AFDCW	The estimated difference, in 1985, between AFDC payments for a household head and a subfamily head.

The NLS-Y survey gathers information on all individuals (to a maximum of 15) who live in the same household as the respondent and classifies household members into families. The information collected includes each household member's sex, age, relationship to the respondent. In addition, the NLS-Y documentation indicates that the first individual in the household record is the household head. Taken together, this information should have been sufficient to construct definitions of living arrangement measures that are consistent with the CPS.

After some data manipulation, however, it became clear that there were serious inconsistencies in the data. First, the individual who appears in the first position of the household record cannot be reliably declared as the household head. This was later confirmed by the NLS-Y data archivists at Ohio State. Household head information was consistently collected in 1979, by means of a separate survey question. In subsequent survey years, however, the interviewer became responsible for correctly placing the household head in the first position of the household record data. Unfortunately, this approach has proved to be unreliable. Some attempt to use mortgage information to identify the household head was made, but this also proved to be unsuccessful. This inability to identify the household head has limited the extent to which the living arrangement measures created from the NLS-Y parallel the CPS definitions.

Second, each household member is assigned a family unit number which identifies the family to which the household member belongs. Theoretically, the family unit number could then be used to determine the number of families within a dwelling unit as well as identifying members within a family. Individuals are considered to be members of the respondent's family if they are related by blood or marriage. Unrelated individuals, including cohabitation partners, should not be coded as members of the respondent's family. This information, however, was found to be fairly inconsistent. For example, unrelated individuals were often given the same family unit number as the respondent suggesting a single family unit in the household. Yet, a respondent living with siblings or other relatives did not share the same family unit number suggesting multiple families within the dwelling unit. These inconsistencies were sufficiently common that any systematic use of the family unit number was abandoned.

Given the problems associated with identifying the household head and using the family unit number to unravel multiple family households and their members, it became necessary to base the living arrangement measures solely on the "relationship to youth" codes. This task was further complicated by the fact that individuals may appear in any order within the fifteen household records for a single year. For example, there may be information in positions one, three and six of the household record, with no information in any other positions. Further, the positions with data are not consistent from year to year. In one year the respondent may be in position three and the spouse in position one, yet the following year the respondent is in position one and the spouse in position two, with no change in overall family composition. In addition, the creation of marital status and living arrangement measures was further complicated by the need to allow for partners as well as spouses.

A respondent can declare an individual as a spouse even if the marital union is not legally binding. That is, a respondent can be legally married or simply regard the individual with whom they are cohabiting as a spouse. A partner, on the other hand, is an individual of the opposite sex who lives with the respondent as a cohabitant and is identified as such by the respondent.

Essentially, for each year, it was necessary to loop through all fifteen household records and classify any individual who resided in the household into the relevant categories of living arrangement. This included whether or not the respondent was living with parent(s) or parent(s)-in-law; living with relatives over 18; living with nonrelatives; living with a spouse or partner; living with own, step or adopted children; living with partner's children. Given the complexity of the task at hand, and cognizant of the apparent limitations of the data themselves, other variables were used to cross-check the living arrangement measures which had been created using the "relationship to youth" codes.

One of the relationship to youth codes specifically categorizes an individual within the household as being the respondent's "partner." There were numerous cases, however, where an interviewer check question indicated that the respondent was currently living with an individual of the opposite sex as a partner but the "relationship to youth" code revealed no partner living in the household. To resolve this inconsistency, it was necessary to look more closely at individuals in the household who were coded as nonrelatives of the respondent. To check if an individual coded as a nonrelative was really a partner, one of two routes was taken. The first systematically looked at nonrelatives when there was no partner or spouse in the household. Specifically, if there was only one individual in the household who was coded as a nonrelative, and was an adult male, then that individual was reclassified as a "partner." The second route consisted of dumping data records and hand-coding the relevant variables for that observation when inconsistencies were found between interviewer checks and the relationship to youth codes. Hand-coding of observations will be described more fully in a subsequent section of this appendix.

Having identified a "partner," we made an attempt to determine if any of the children in the household who were coded as nonrelatives could be reclassified as the partner's children. If an expanded definition of being married includes partners along with spouses, then the partner's children should be classified as part of the respondent's family. This required some data manipulation because the relationship codes offer no clue as to the parenthood or guardian relationship of nonrelative children to the respondent or other members of the household. An individual was defined as a partner's child if all of the following conditions held:

(a) The partner's family unit number was different from the respondent's. If the partner and the respondent shared the same family unit number, it was assumed that all children relevant to their family would have been coded as the respondent's own, step or adopted children;

(b) When the respondent and partner had different family unit numbers, the individual's family unit number had to be the same as the partner's. That is, any potential child of the partner should be coded as belonging to the partner's family;

(c) In the "relationship to youth" code, the individual was coded as being a nonrelative. Any potential child of the partner should have no family relationship to the respondent;

(d) The individual was under 18 years of age;

(e) The partner was at least 16 years older than his potential child.

There are serious limitations with the approach used to identify the partner's children which must be pointed out. First, this is at best an educated guess of which children in the household, who are not related to the respondent, could conceivably be the partner's children. The relationship codes are simply not sufficiently detailed to be able to determine the identity of the partner's children without error. Second, it was necessary to use the family unit numbers in this endeavor and the limitations of those numbers have already been described. We hope that the criteria used were sufficiently stringent that the probability of error in classifying partner's children was minimized.

Capturing and Correcting Data Errors

After repeated iterations of doing consistency checks and printing out inconsistent records, we were able to program many of the corrections. However, for a subset of observations this proved to be impossible. We therefore recoded these observations manually after examining the records closely.

The inconsistencies and errors appear to occur most frequently in households with large groupings of individuals where the possibility of shared living arrangements with family members and/or nonrelatives was the highest. Also, many of the inconsistencies were related to difficulties in correctly identifying the respondent's partner. To simply delete these records from the sample would have resulted in disproportionately dropping those cases in which the respondent was in a shared living arrangement or cases where a partner was present in the household. Yet, these were exactly the cases of primary interest to the analysis.

A number of data checks were used to validate some of the living arrangement measures created. One of the data checks used initially was the recorded household record type. Three versions of household records are used by the NLS-Y. Version A is used if the respondent is living with parent(s) or parent(s)-in-law. In this case, the household interview, which collects information about the occupants of the household, is conducted with one of the parents. Version B is used if the respondent is living in a temporary dwelling unit such as a sorority, fraternity, dormitory or military quarters. These respondents are not considered to be within the sample relevant to this analysis and were dropped at the beginning of the analysis. Version C is used if the respondent is living in their own dwelling unit or is the head of a family unit. Attempts to compare household record type against the created measure of whether the respondent is living with parent(s) or parent(s)-in-law based on the relationship to youth codes proved to be futile. The NLS-Y allowed interviewers to use household record version C even when the respondent was under 18 and living with parent(s) or parent(s)-in-law if the interviewer ascertained that contacting the parent would be awkward or there was reason to suspect the parent would not consent to the interview. Other exceptions are based on the respondent's age (either younger or older than 18) and whether they have lived continuously with parent(s) or parent(s) in-law. These exceptions made it impossible to use this variable as a check against whether the respondent was sharing the household with parent(s) (in-law) based on the relationship to youth codes.

Three specific checks of the constructed living arrangement measures were made. They were concerned with the correct identification of spouses and partners, and accurately distinguishing partners from nonrelatives. The first two checks were constructed from NLS-Y interviewer check questions. Specifically, they ask "is the respondent married and the spouse listed on the household record" and "does the respondent live with an adult nonrelative of the opposite sex." After 1981, the latter question becomes more specific and asks "is the respondent currently living as a partner with an opposite sex adult." The answers to both these questions were compared to constructed variables concerned with whether the respondent had a spouse or partner based on the "relationship to youth" codes. The last check was concerned with flagging any respondent which reported multiple spouses or partners or both a spouse and a partner in the household.

When an inconsistency was found, data from multiple years was printed. Specifically, data from the year in which the inconsistency was found (year t), as well as data from the previous (year t-1) and subsequent (t+1) years was printed. If the inconsistency occurred in the first year of the data survey (1979), however, the two subsequent years (t+1, t+2) were printed. While some inconsistencies could have been resolved from a single year's data, others could only be resolved by observing the age and sex composition of the household in past or future years. A total of 210 records were examined and corrected.

Some attempt has been made to construct general categories of errors found when looking at the printed records for the years t, t-1 and t+1 (or t, t+1 and t+2 when the inconsistency occurred in 1979). It should be noted that corrections were only made to constructed variables in order to maintain the integrity of the original data set. What follows is a discussion of each type of error in descending order of frequency.

First and most frequently, the interviewer check indicated that the respondent was living with an adult of the opposite sex as a partner. Yet, according to the "relationship to youth" codes, no individual was coded as a partner. Multiple nonrelatives lived in the household, however. This particular scenario led to four sub-categories of problems and solutions.

(a) Amongst the male nonrelative(s) in the household, no single individual could be discerned as being the respondent's partner even after comparing the sex and age composition of the household in year t with years t-1 and t+1. In these cases (frequency=52), the respondent was recoded as being single;

(b) The respondent's partner could be discerned from the male nonrelative(s) in the household after comparing the sex and age composition of the household in year t with years t-1 and t+1. In these cases (frequency=21), the respondent was recoded as living with a partner. In a similar case, the interviewer check indicated there was no opposite sex adult living with the respondent. Yet, a single individual was coded as being the respondent's partner from the "relationship to youth" codes. In this case (frequency=1), the respondent was recoded as living with a partner;

(c) All the nonrelatives in the household shared the same family unit number, suggesting they formed a single family that was unrelated to the respondent. It was very difficult to discern, however, if one of the male family members was the respondent's partner. As a result, the respondent was recoded as being single in these cases (frequency=10);

(d) All of the nonrelatives in the household were females. In these cases (frequency=7), the respondent was recoded as being single.

Second, the marital status of the respondent (single, married with spouse present or living with partner) as determined from the "relationship to youth" codes was inconsistent with one or both of the interviewer checks. Sub-categories of this problem are discussed below. In general, discrepancies were resolved by ignoring the interviewer checks and classifying marital status based on who resided in the household, as determined from the relationship codes.

(a) The interviewer checks indicated that both the respondent's spouse and partner were present in the household. According to the relationship codes, however, only a spouse resided in the household. In these cases (frequency=20), the respondent was recoded as married, spouse present;

(b) The interviewer checks indicated that both the respondent's spouse and partner were present in the household. Furthermore, the relationship codes found either a spouse and partner or a spouse and male nonrelative residing in the household with the respondent. In these cases (frequency=12), the respondent was coded as married, spouse present. It was assumed that the male nonrelatives in these cases were not living as partners with the respondent. In addition, any individual coded as a partner was viewed as a miscode and subsequently counted as a male nonrelative. This was done because it was difficult to imagine a household where the respondent was living with a spouse and a live-in companion of the opposite sex simultaneously. The spousal relationship took precedence over the partner relationship because the spousal relationship has generally been less difficult to discern in this data set;

(c) The interviewer check indicated the respondent was living with an opposite sex adult as a partner. No partner or male nonrelative was identified, however, from the relationship codes. In these cases (frequency=17), the respondent was recoded as being single;

(d) The interviewer check indicated the respondent was married, with spouse present, but according to the relationship codes, no spouse was present in the household. In these cases (frequency=11), the respondent was recoded as being single;

(e) The interviewer codes indicated the respondent had a spouse but no partner. Yet, the relationship codes revealed a partner but no spouse. In these cases (frequency=6), the respondent was recoded as living with a partner.

Third, and most disturbing, were errors in the relationship codes found through various discrepancies in one or more of the three data checks outlined earlier. A total of 42 relationship coding errors were found. All could be corrected using information on the sex and age of household members as well as the composition of the household in years t, t-1 and t+1. Some examples of the types of miscodes which occurred include:

(a) The relationship codes revealed multiple spouses, where extra spouses were determined to be a child, sister, brother, cousin or other relative of the respondent;

(b) The respondent's spouse or partner was erroneously miscoded as some other relative (for example, as a sister, brother, father, mother, daughter-in-law or foster child). The most striking example of this type of error were individuals coded as the respondent's sister (relationship code 7) who were also males. Looking at information from years t-1 and t+1, these individuals were subsequently recoded as the respondent's spouse (relationship code 1). It is obvious that relationship codes 7 and 1 were transposed while being transcribed from the original interview sheets;

(c) The last grouping contains miscellaneous coding errors such as a nonrelative miscoded as a foster child; a partner's child miscoded as a partner; a spouse miscoded as another respondent; daughters miscoded as sisters; brothers and sisters miscoded as partners and other-in-laws; partners miscoded as boarders.

Fourth, the relationship codes reveal the respondent was living with more than one partner, where some or all of these partners were really nonrelatives. In these cases (frequency=7), information from years t-1 and t+1 as well as sex and age information from year t was used to try and discern a true partner. Those determined not to be the respondent's partner were recoded as nonrelatives.

Lastly, in a few cases (frequency=2) the relationship codes were either completely missing or so badly miscoded that the entire observation was set to missing for that year. In other cases (frequency=2), only some of the relationship codes were missing and it was possible to reconstruct the composition of the household based on sex, age and household composition information in years t, t-1 and t+1.

While many inconsistencies and coding errors were corrected, it should be noted that the final data set may still contain errors. Of particular concern is any undetected "relationship to youth" coding errors. These would affect cross-sectional analysis as well as transition rates for marital status and living arrangements. One concern is mitigated, however, by the knowledge that the majority of the coding errors which were found in the 210 cases examined manually occurred in the first three years of the data survey.

Appendix B

Additional Models of Labor Force Participation and Marital Status

This appendix present two additional models of labor force participation and marital status. The first is a model of the "initial conditions" for the women in our sample. Since the models presented in the text suggest that a young mother's labor force participation and marital status tend to remain constant, except for random factors, it is of some interest to examine the demographic correlates of those initial conditions. The second model illustrates the geferal applicability of the multi-state, multi-period dynamic model developed by George Jakubson for this project. In that second model, we estimate a <u>six</u> state labor force participation and marital status model. Labor force participation remains as a 0-1 variable but "marital status" can now take on three values - married, unmarried and heading one's own household and unmarried and living with relatives.

<u>A Model of "Initial Conditions"</u>

The results presented in the text highlight the importance of looking at where a woman has been in order to describe where she is now. Lagged marital status is the most important determinant of current marital status; lagged labor force participation is the most important determinant of current labor force participation. Last year's AFDC participation has important negative impacts on current marital status and current labor force participation.

In one sense, this emphasis on lagged dependent variables simply pushes the problem back a few steps. If race affects labor force participation only because black women are more likely to participate in the AFDC program, then why are black women more likely to participate in the AFDC program?

Though these questions are not amenable to statistical analysis, this section addresses the question "what are the correlates of the initial conditions?." The dynamic programming model presented in Chapter 2 deals with four marital status/labor force participation states, so we restrict our attention to those four states here as well.

The dependent variable can take four values:

(1) married and in the labor force;
(2) married and not in the labor force;
(3) not married and in the labor force; and
(4) not married and not in the labor force.

The variable is defined at the time when an NLS-Y respondent first reports having a child of her own in her household. For example, if the woman first reports having a child in 1982, then our dependent variable and all other variables in the model are given their 1982 values. If the woman first reports having a child in 1984, then all variables take on their 1984 values.

By defining the variables in this way, we are trying to look at the correlates of a marital status/labor force participation variable, <u>at the time a woman first has a child</u>.

Consider the following tabulation of the four labor force participation/marital status states in the woman's initial condition as compared with the same four states in 1985.

Initial Condition	Labor Participation/Marital Status in 1985 1	2	3	4	Total
1. Married, in LF	251	125	36	20	432
2. Married, Not in LF	142	259	62	125	494
3. Not Married, in LF	92	62	155	88	397
4. Not Married, Not in LF	76	125	140	319	660
Total	561	571	377	474	1983

Of the 1,983 women who report having a child in their household between 1979 and 1985, inclusive, almost 50% remain in the status that they "started" in.

Our initial conditions model, reported in Table B1, considers the marital status/labor force participation variable as a function of only age, race, ethnicity and the number of young children present. The number of independent variables is limited to correspond to the variables included in the dynamic model in Chapter 2.

The coefficients in the first column enable us to compare the probability of being married and in the labor force to the probability of not being married and not in the labor force. For example, the negative coefficient on BLACK is large (in absolute value), and statistically significant, indicating that being black significantly reduces the probability of being married (at the time when the first child "appears") as compared to being unmarried and not in the labor force. This is not surprising since only 14% of the married women were black as compared to almost 60% of the unmarried women. The same story holds for

Hispanic women as well, though the coefficient is considerably smaller.

The coefficients in the second column compare the probability of being married and not in the labor force to the probability of being not married and not in the labor force (at the time of when the respondent's first child was born). Here again, black and Hispanic women are more likely, compared to white women, to be unmarried and not in the labor force than to be <u>married</u> and not in the labor force.

Age plays a powerful role here, especially considering the limited age range of the NLS-Y respondents. As indicated by the positive and statistically significant coefficients on AGE in Table B1, older women are considerably less likely to be unmarried and out of the labor force. Greater numbers of children under three lowers the probability of being in the labor force.

This overall picture drawn by these "initial conditions" models is not particularly surprising. Roughly put, if a woman is in an economically healthy position before she has a child - in the labor force or married or both, then she is in an economically healthy position <u>after</u> she had a child. The thrust of our other modelling efforts is to show that once an initial condition is established it tends to be perpetuated. Thus our story seems to be that the eventual economic health of women with children is established early on and then tends to persist over time.

<u>A Six-State Model of Labor Force Participation and Marital Status</u>

The results of extending our four-state dynamic model to a six states are shown in Table B2. These results are quite similar to the results of the four state model; we show them largely to indicate the practicality of extending the dynamic model to multiple states. The six states are:[8]

1 = Married, in labor force (MLF)
2 = Subfamily Head, in labor force (SHLF)
3 = Household Head, in labor force (HHLF)
4 = Married, not in labor force (MNLF)
5 = Subfamily Head, not in labor force (SHNLF)
6 = Household Head, not in labor force (HHNLF)

As was true in the four state model, the demographic variables have virtually no significant impact on the utility of being in any one of the six states. The parameter estimates for β_i (Panel 1 of Table B2) are uniformly insignificant. This parallels a similar result from the four-state model.

When we turn to the estimates of the α_i vector, we see again that the "previous status" has the most value. The coefficients on the diagonal (in Panel 2 of Table B2) are the ones that have large normal statistics (indicating a significant difference from zero). The only exception to this is that being a subfamily head and not in the labor force seems to be of value in making the transition to being a subfamily head and in the labor force (the coefficient estimate is 2.4 with a normal statistic of 3.9) and vice-versa (coefficient estimate of 2.3 and normal statistic of 3.8).

Our interpretation of the results of the six state model is that, aside from random factors that are unobserved by the researcher, only past status plays an important role in determining current status.

Table B1

Four-valued Logit Model of Marital Status and AFDC Participation

Initial Conditions Model

NLS-Y Respondents When the First Report Having a Child in Their Household

Frequency Count of Dependent Variable

(1) Married and in the labor force	432
(2) Married and not in the labor force	494
(3) Not married and in the labor force	397
(4) Not married and not in the labor force	660
Total Sample Size	1983

	Married and In the Labor Force		Married and Not in the Labor Force		Not Married and in the Labor Force	
	(1)		(2)		(3)	
BLACK	-2.27	(12.3)	-2.22	(13.1)	-0.64	(4.4)
HISPANIC	-0.78	(3.9)	-0.19	(1.1)	-0.17	(0.9)
AGE	0.60	(15.6)	0.33	(9.4)	0.28	(8.0)
KIDS1	-0.36	(2.3)	0.21	(1.6)	-0.50	(3.4)
Constant	-11.58	(14.8)	-6.09	(8.8)	-5.47	(8.0

Table B2

Coefficient Estimates for a Four State
Dynamic Programming Model of Marriage
and Labor Force Participation

Panel 1. Status-Specific Coefficients on Demographic Variables, β_i

Dependent Variable:

1 = Married, in labor force (MLF)

2 = Subfamily Head, in labor force (SHLF)

3 = Household Head, in labor force (HHLF)

4 = Married, not in labor force (MNLF)

5 = Subfamily Head, not in labor force (SHNLF)

6 = Household Head, not in labor force (HHNLF)

Status-Specific Coefficients on Demographic Variables, β_i
(absolute value of asymptotic normal statistic)

Status	Constant	Black	Hispanic	Age/10	KIDS1
1=MLF	-1.4	-0.5	0.2	0.7	0.2
	(.7)	(1.2)	(0.5)	(1.3)	(0.5)
2=SHLF	-0.2	0.3	-0.1	0.3	-0.2
	(0.2)	(0.7)	(0.1)	(0.6)	(0.4)
3=HHLF	0.7	0.2	0.3	-0.1	0.2
	(0.5)	(0.4)	(0.4)	(0.2)	(0.4)
4=MNLF	1.2	-0.5	0.0	-0.2	0.6
	(0.9)	(1.8)	(0.0)	(0.3)	(1.5)
5=SHNLF	1.9	0.6	-0.1	-0.7	-0.1
	(1.4)	(1.2)	(0.2)	(1.3)	(0.2)
6=HHNLF	-1.5	0.5	0.2	0.6	-0.2
	(0.6)	(1.0)	(0.4)	(1.0)	(0.4)

Table B2
(continued)

Coefficient Estimates for a Four State Dynamic Programming Model of Marriage and Labor Force Participation

Panel 2. Status-Specific Coefficients on Past Status, α_i

Dependent Variable:

1 = Married, in labor force (MLF)

2 = Subfamily Head, in labor force (SHLF)

3 = Household Head, in labor force (HHLF)

4 = Married, not in labor force (MNLF)

5 = Subfamily Head, not in labor force (SHNLF)

6 = Household Head, not in labor force (HHNLF)

Status-Specific Coefficients on Past Status Variables, α_i
(absolute value of asymptotic normal statistic)

	Current Status					
Previous Status	1=MLF	2=SHLF	3=HHLF	4=MNLF	5=SHNLF	6=HHNLF
1 = MLF	3.5 (1.2)	-0.3 (0.3)	0.2 (0.2)	1.3 (2.5)	-0.2 (0.2)	-8.7 (0.4)
2 = SHLF	1.7 (0.6)	3.6 (6.6)	1.8 (2.0)	0.6 (1.0)	2.4 (3.9)	1.4 (0.4)
3 = HHLF	2.1 (0.7)	1.2 (1.3)	2.6 (4.2)	0.5 (0.7)	0.9 (0.9)	2.6 (0.7)
4 = MNLF	2.9 (1.0)	-0.8 (0.9)	-0.5 (0.6)	2.7 (5.5)	0.7 (1.0)	2.0 (0.6)
5 = SHNLF	1.3 (0.5)	2.3 (3.8)	0.8 (1.1)	1.0 (1.8)	3.5 (6.9)	2.9 (0.8)
6 = HHNLF	-7.0 (0.4)	-0.2 (0.2)	1.7 (2.6)	0.9 (1.5)	0.5 (0.5)	4.1 (1.2)

Value of Log Likelihood Function: -760.8

Appendix C

FORTRAN Programs for Dynamic Model

The inclusion of the actual FORTRAN programs used to estimate the parameters of our dynamic would add approximately 40 pages to this report. For that reason, we do not include them here.

However, those programs can be obtained by sending a blank, formatted IBM-compatible diskette to:

Professor George Jakubson
Ives Hall
ILR-Cornell
Cornell University
Ithaca, New York 14853-3901

The programs can also be obtained electronically by sending e-mail to AK5J at CORNELLA.BITNET.

Bibliography

Bellman, R., Dynamic Programming, Princeton, NJ: Princeton University Press, 1957.

Berkovec, James and Steven Stern, "Job Exit Behavior of Older Men," Jefferson Center for Political Economy Discussion Paper #169, University of Virginia, February, 1988.

Eckstein, Zvi and Kenneth I. Wolpin, "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models," Journal of Human Resources, Fall, 1989, p.562-598.

Gonul, Fusun, "Dynamic Labor Force Participation Decisions by Males in the Presence of Layoff and Uncertain Job Offers," Journal of Human Resources, Spring, 1989, p.195-220.

Hotz, V. Joseph and Robert A. Miller, "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," Econometrica, January, 1998, p.91-118.

Hutchens, Jakubson and Schwartz, Preliminary Final Report on Grant #E-9-J-8-0090, August, 1990a.

Hutchens, Jakubson and Schwartz, Final Report to the Department of Health and Human Services, 1990b

Johnson, William R and Jonathan Skinner, "Labor Supply and Marital Separation," American Economic Review, June, 1986

Killingsworth, Mark, Labor Supply, Cambridge: Cambridge University Press, 1983.

Maddala, G.S., Limited-Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press, 1983.

McElroy, Marjorie, "The Joint Determination of Household Membership and Market Work: The Case of Young Men," Journal of Labor Economics, 1985

Muthen, Bengt O., LISCOMP: Analysis of Linear Structural Equations with a Comprehensive Measurement Model, Mooresville, IN: Scientific Software, 1988.

Schmidt, Peter, "Constraints on the Parameters in Simultaneous Tobit and Probit Models," in Manski, Charles and Daniel McFadden, Structural Analysis of Discrete Data: With Econometric Applications. Cambridge, MA: MIT Press, 1982.

Wolpin, Kenneth I., "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," Journal of Political Economy, 1984, p.852-874.

Endnotes

1. In work reported in Appendix B, we show the results of our estimation of the parameters of a six state model.

2. There are two general methods for solving a dynamic programming problem. When the control variable is continuous and there are no censoring or truncation issues, it is straightforward to make use of Bellman's Equation to solve the problem. There are two sets of first order conditions to maximize the value function. The first set essentially mean that there are no within period possibilities for utility increasing reallocations. These are the same first order conditions that arise in a static problem. The second, the "Euler Equations," mean that there are no between period possibilities for utility increasing reallocations, that is, no arbitrage possibilities across time periods.

   The model is typically closed with a rational expectations assumption, so that the forecast errors (for the next period state variables) have zero mean. This allows one to combine the two sets of first order conditions to define a set of equations for the forecast errors. Since these have zero mean (by assumption), this provides a tractable method for specifying an estimating model. There are many examples of this approach in the literature.

   Alternatively, under some conditions the value function defines a contraction mapping. In these cases one can literally compute the value function by iterating the contraction mapping to convergence. The contraction can then be built into the computation of the likelihood of a given sample.

   Unfortunately neither of these two approaches are available to us, because the choices with which we are dealing (e.g., the marriage choice) are intrinsically discrete. Because of the discreteness, the value function is not differentiable with respect to the choice variable, ruling out the Euler Equation approach. And our environment is not stationary because, for example, the number of children varies over time, so we cannot make use of the contraction mapping approach.

3. An important issue here is the choice of terminal date T. In a stationary environment this is not as difficult, so long as there is discounting. As the terminal date T is moved farther into the future, the contribution to current period value of the expected future events grows smaller and smaller. By

pushing T far enough away, any errors made by ignoring time periods later than T become negligible. One can then implement the backwards recursion, since the stationarity assumption implies that the same decision making environment exists at all time periods.

If the environment is not stationary the problem is harder. For example, if there are exogenous variables which affect utility which change over time in a manner which is not completely predictable, then one cannot simulate the decision making environment in periods for which one does not have data. Therefore, the terminal date T cannot be pushed arbitrarily far into the future, but rather must be the latest date for which data are available.

4. When we observe a woman who already has a child in 1979, the first year of the data, we cannot determine her "initial condition" because we do not have the data. We therefore use 1979 as the initial condition for these women and use her decisions from 1980 to 1985 in the estimation.

   In the context of our model, there is no harm in doing this. The disturbances in the model are independent and identically distributed across women, choices, and time periods. The previous period state variables, D(t-1), and the current period exogenous variables, X(t), characterize the decision making environment, so that while we do not see all the choices made by a woman who had a child with her in 1979, those which we do see are made in the same way as those for the women for whom we observe the first appearance of a child.

5. Thus any woman who was not interviewed in each year is deleted. In principle, it is possible to deal with "holes" in the record by integrating over all the possibilities in the missing year(s). The probability weighted values would then be used in the calculations for future years. This approach is, unfortunately, computationally intractable.

6. To maximize the likelihood function in equation (9), we utilize the hill-climbing routine GQOPT (written and maintained by Professor Richard Quandt, Department of Economics, Princeton University) because of its flexibility. Within the package are a number of different algorithms: Davidson-Fletcher Powell (DFP), quadratic hill-climbing (GRADX), a simplex search, a conjugate gradient method, and others. This flexibility is important because the log likelihood function is difficult to maximize. We found it necessary to start from many different places to ensure that we found the values of the parameters for which the function attains its maximum. Different algorithms performed well or poorly in different regions of the parameter space. The

FORTRAN code we use to maximize the likelihood function appears in an appendix.

7. When we free up that parameter, we have serious convergence problems. A grid search over the concentrated log likelihood function (concentrating out all the other parameters) shows that the likelihood value is very insensitive to the value of the discount factor. In the results below, then, the discount factor has been fixed at 0.9.

8. Strictly speaking, our classifications here are not identical to the Census definition of "subfamily" as it appears in the Current Population Survey. Using the NLS-Y, we have no uniformly reliable way to know who owns the dwelling unit or whose name is on the lease. Hence, unlike those working with the CPS, we cannot distinguish between the following two situations:

    1. Respondent and her children live in her parent's home.

    2. Respondent's parents live in the respondent's home with the respondent and her children.

    In the first case, the respondent is a subfamily head. In the second case, the respondent's parents form the subfamily. Since our sample is young, we suspect that the vast majority of shared living arrangements are of the first type and we therefore use the term "subfamily" head.