

ECONOMIC GROWTH CENTER

YALE UNIVERSITY

P.O. Box 1987, Yale Station  
27 Hillhouse Avenue  
New Haven, Connecticut 06520

CENTER DISCUSSION PAPER NO. 455

TWO ESTIMATION METHODS FOR A LIFETIME  
DISCRETE CHOICE MODEL UNDER UNCERTAINTY

Zvi Eckstein  
Tel Aviv University and Yale University

and

Kenneth I. Wolpin  
Ohio State University

May 1984

Notes: We wish to thank Don Andrews for very helpful discussions and gratefully acknowledge Paul McGuire's most valuable research assistance. This work was supported by NSF Grant No. SES 8207594 and funds granted to the Foerder Institute for Economic Research by the NUR MOSHE Fund.

Center Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Discussion Papers should be cleared with the author to protect the tentative character of these papers.

## 1. INTRODUCTION

The development of the techniques of analyses for discrete choice models has proceeded rapidly over the last decade. Original work by McFadden (1976) and others concentrated on static choice frameworks and were developed and applied in the context of cross-section data. More recent work by Heckman (1981) focuses on general dynamic discrete choice models in the context of panel data. In this paper we attempt to provide a firmer theoretical base for this recent literature by specifying the explicit dynamic stochastic optimization problem that underlies the decision rules which are the starting point for the analysis by Heckman. Moreover, we offer two methods of estimating the fundamental taste and constraint parameters of the optimization problem and provide an application.

A general dynamic discrete choice model for panel data is carefully described by Heckman (1981) as follows:

$$Y_t^i = Z_t^i \beta + \gamma_t(L) m_{t-1}^i + \sum_{j=1}^{\infty} \lambda_{t-j}^i \pi_{t-j}^j m_{t-j}^i + G(L) Y_{t-1}^i + \epsilon_t^i$$

$$i = 1, \dots, I, \quad t = 1, \dots, T,$$

where  $m_t^i$  is the discrete choice variable for person  $i$  ( $i = 1, \dots, I$ ) for his lifetime period  $t$  ( $t = 1, \dots, T$ ) such that

$$m_t^i = \begin{cases} 1 & \text{if } Y_t^i \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

the initial conditions of  $m^i$  and  $y^i$  are assumed to be fixed outside the model,  $\gamma_t(L) = \gamma_t^0 + \gamma_t^1 L + \dots + \gamma_t^k L^k$ ,  $G(L) = G^0 + G^1 L + \dots + G^q L^q$  and  $\epsilon_t$  is a normally distributed error with mean zero. The distribution of the vector  $\epsilon^i = (\epsilon_1^i, \dots, \epsilon_T^i)'$  is fully characterized by the assumption

$$\epsilon^i \sim N(0, \Sigma)$$

where  $\Sigma$  is a  $T \times T$  positive definite covariance matrix.

The  $\epsilon^i$ 's are independent across people and components of the vector  $Z_t^i$  are independent of  $\epsilon^i$ .  $Z_t^i$  is a vector of exogenous variables. This is the most general linear discrete choice model that is described in the literature.

In this model  $Y_t^i$  represents the difference between the lifetime utility of a person at time  $t$  given an action ( $m_t^i = 1$ ) is taken and the lifetime utility of the person given the action ( $m_t^i = 0$ ) is not taken, under the assumption that the decisions in the future are optimal. These types of problems include labor force participation (Heckman and Willis, 1977), fertility (Heckman and Willis, 1975) and purchase of a durable (McFadden, 1976). In each application of this general model a set of specific issues are raised both of a practical and more abstract nature. For example: what is the lifetime optimization problem that is behind this intuitively appealing model? What is the lag length for each term in the model and what do these lags represent? What are the underlying sources of the stochastic term, of state dependency and of heterogeneity?

The stochastic model does predict the change in the probability of  $m_s^i$ ,  $\forall s$ , given a change in the exogenous variables at time  $t$  and/or a

change in the parameters in the model. However, it is impossible to interpret the particular change in terms of a policy experiment unless the individual's lifetime problem is carefully described (Marschak, 1953, Lucas, 1976).

One way of providing at least a partial answer to the above questions is to consider the explicit dynamic stochastic optimization problem that the person is supposed to solve. This paper formulates explicitly a lifetime optimization problem that a person faces at each period  $t$  and suggests two methods of estimating the underlying parameters of the preferences and the constraints. Some of the issues that are raised above are solved immediately by the assumptions made about the preferences and the constraints. Given the important empirical issues of state dependency and heterogeneity we emphasize these dynamic stochastic aspects in the context of the three behavioral examples, fertility, schooling, and labor force participation. Some other issues, such as the length of the endogenous lags, are not solved except by computational limits, but they can be interpreted in terms of the economic model.

Estimation of our model's parameters depends on the ability of the econometrician to find an algorithm that can be used to calculate the probabilities of  $m_t^i$  ( $t = 1, \dots, T$ ) conditional on past realizations, and which is consistent with optimization. In this way the likelihood function of each sequence of  $m$ 's can be computed as a product of conditional probabilities. Heckman's statistical model is formulated in a way that enables a straightforward calculation of the conditional probabilities that form the likelihood function, but that formulation is not necessarily consistent with any optimization problem of content. The focus of our work is

on building a bridge between the individual optimization problem and a decision rule that for the econometrician can be stated as a conditional probability for the discrete choice at each  $t$ .

We provide two ways of calculating these conditional probabilities. The first is based on the way Heckman (1981) motivates his model, that is, given optimal decisions in the future the probability of  $m_t^i = 1$  can be calculated from the difference between the lifetime utility level of a person where  $m_t^i = 1$  and the lifetime utility where  $m_t^i = 0$ . We refer to this method as "Full Solution Method", since it requires that we solve completely the dynamic optimization problem of each agent at each time in order to calculate the likelihood function. The computational burden of this method is obvious.

The second method is based on the fact that the optimal individual program for  $m_t^i$  should satisfy certain first order conditions. Therefore, we call this estimation procedure the "Necessary Condition Method". Here we demonstrate a method for calculating the conditional probabilities without fully characterizing the future decisions. We assume, however, that these decisions are made optimally and are predicted optimally by the person (rational expectations).

Since the first method has been used recently (Wolpin, 1982) for estimating a fertility model using Malaysian data, we compare the estimation methods by a Monte Carlo experiment conducted on the Malaysian fertility model. Using this Monte Carlo data on fertility we estimate the parameters of the model with the necessary condition method.

The remaining parts of the paper are organized as follows: In Section 2 we present the model and the three examples. In Section 3 the two estimation

methods are discussed. The results of the Monte Carlo experiment are presented in Section 4 and some remarks are given in Section 5.

## 2. THE MODEL

In this section we describe a class of estimable dynamic models of behavior in which the individual makes a discrete (zero-one) decision in each life cycle period and where the cumulated values of previous choices may affect current welfare and/or costs. An individual is assumed to choose a lifetime contingency plan for the sequence  $\{m_t^i\}_{t=0}^T$ , where  $i$  refers to the individual and  $t$  to the life cycle period, so as to maximize

$$(1) \quad V_0 = E_0^i \sum_{t=0}^T \beta^t U(M_t^i, C_t^i, a_t^i)$$

subject to:

$$(2) \quad M_t^i = M_{t-1}^i + m_t^i - d_t^i$$

$$(3) \quad m_t^i \in \{0, 1\}, (m_t^i, d_t^i) \in \{(0, 0), (1, 0), (1, 1)\}$$

$$(4) \quad C_t^i = NY^i(M_{t-1}^i, m_t^i, h_t^i, t)$$

$$(5) \quad M_{-1}^i \text{ given}$$

In (1)  $\beta$  is the discount factor,  $0 < \beta < 1$ , and preferences are a function of the stock of the control variable  $(M_t^i)$ , and a composite consumption good  $(C_t^i)$ , with  $a_t^i$  a vector of exogenous preference shifters

that may vary across individuals and over the life cycle of the same individual. The stock of the control variable ( $M_t^i$ ) evolves according to (2) with  $m_t^i$ , the discrete decision variable, and  $d_t^i$  a discrete exogenous variable realized after the decision on  $m_t^i$  has been made; the set of feasible values of  $(m_t^i, d_t^i)$  is given in (3). The budget constraint is described in (4);  $NY(\cdot)$  is net income, i.e., income less expenditures on  $m_t^i$ . Net income may depend upon the stock of the control variable, the value of the current decision variable, life cycle period ( $t$ ), and a set of exogenous individual and/or life cycle characteristics ( $h_t^i$ ). The initial value of the stock is given and is non-stochastic.  $E(\cdot)$  is the expectations operator and  $E_t^i(\cdot) = E(\cdot \mid I_t^i)$  where  $I_t^i$  is the information set of individual  $i$  at life cycle stage  $t^1$ .

Dynamics are incorporated into the model both through the utility function and through the net income function. It would, of course, be more general to permit each prior period discrete choice to enter current utility and net income rather than the stock, but that complete generality does not appear tractable to estimate. A feature of the model, due to computational tractability, that is not particularly satisfactory is that we consider only one decision variable. Most econometric applications do so. However, in doing this we are forced to assume that the individual is unable to transfer physical resources between periods. Introducing savings requires solving for an additional decision variable and creates further dynamic interactions.<sup>2</sup> We view this as an important limitation and as a challenge for future work.

Both  $U(\cdot)$  and  $NY(\cdot)$  are assumed to be continuous and differentiable in the decision variables  $(m_t^i, c_t^i)$  at all points on the real line

even though  $m_t^i$  is dichotomous.<sup>3</sup> Existence of a maximum for (1) is guaranteed by the discrete nature of the decision variable, but uniqueness requires some additional regularity conditions on the preference and net income functions. Uniqueness is guaranteed if, upon substituting (2) into (4) and the result into (1) the function

$$E_t \sum_{j=t}^T \beta^j U^i(M_{j-1}^i + m_j^i - d_j^i, NY(M_{j-1}^i, m_j^i, h_j^i, j), a_j^i)$$

is strictly concave in  $m_t^i$  for all  $t = 1, \dots, T$ .

To motivate the above structure, we now provide three examples which have been of major interest to researchers. The reader can doubtless provide many others.

### 1. A Fertility Model with Exogenous Infant Mortality

Equations (1)-(5) correspond to a model of fertility choice over the life cycle under the following definitions:

$M_t^i =$  the stock of surviving children at the end of period  $t$

$m_t^i =$  unity if a child is born at  $t$ , and is zero otherwise

$d_t^i =$  unity if a child born at  $t$  dies at  $t$  with some probability  $\pi$ , and is zero otherwise.

$$NY_t^i = Y_t^i - e_{1t}m_t^i - e_{2t}(m_t^i - d_t^i), \text{ where}$$



$Y_t^i =$  exogenous household income (possibly random),

$e_{it} =$  a fixed cost of child bearing at period  $t$ ,

$e_{2t} =$  a maintenance cost of a child during its first period of life that arises only if the child survives.

Thus, at any period  $t$ , given its stock of surviving children, the household decides whether or not to augment its stock of children by one, based upon future survival prospects, future (uncertain) income, and future child costs. This model has been estimated by Wolpin (1982) and is capable of generating child spacing and in distributing children over different life cycle stages (timing). It is also capable of generating alternative replacement patterns, i.e., reactions to child deaths.

## 2. A Model of Schooling Attainment

The theoretical analysis of schooling attainment has, since Ben-Porath (1967), been conducted in a life cycle framework as a component of a complete human capital accumulation model. Empirical implementation of human capital models have been concerned mostly with estimation of parameters of the human capital production (cost) function (Haley (1976), Brown (1977), Heckman (1976)) using earnings data. School achievement models have not generally been implemented with as careful a connection to underlying theory, though the Wallace and Ihnen (1975) simulation model and the work by Orazem (1982) are exceptions.

The discrete choice dynamic programming framework presented above provides a natural setting within which to consider the schooling choice. To see this, define the variables as follows:

- $M_t^i$  = the stock of schooling (schooling attainment) at the end of period  $t$   
 $m_t^i$  = unity if the individual attends school during period  $t$ , zero otherwise  
 $d_t^i$  = unity if the individual fails to complete the school period due to unforeseen factors, e.g., exogenous illness.  
 $NY_t^i = Y(M_{t-1}^i, m_t^i, t, h_t^i) - e_t m_t^i$  where  
 $e_t$  = the direct cost of a period of schooling at time  $t$   
 $Y_t^i$  = gross income during period  $t$

The gross income generating function  $Y(\cdot)$  depends positively on the stock of schooling and negatively on current school attendance (foregone earnings). It is also permitted to have an age gradient as well as to be influenced by individual and/or calendar time characteristics ( $h_t^i$ ) some of which may be viewed as random by the individual.

Given that  $M_t$  enters into (1), this specification captures both investment and consumption components of schooling choice. The model can clearly generate alternative schooling patterns. In particular, individuals will optimally accumulate schooling as rapidly as possible given either a large enough positive return to schooling at low initial) levels ( $\frac{\Delta Y}{\Delta M_{t-1}}$  is large at  $M_{t-1}$  small), or a large enough consumption value of schooling. The individual in making a current schooling choice considers

the current attainment level, and current and (anticipated) future direct and foregone earnings costs.

### 3. A Model of Labor Force Participation and Wage Determination With Endogenous Experience

Economic models of labor force participation arise naturally from labor supply models in which hours are freely variable. (Heckman and Willis, 1977). With fixed hours, labor force participation models closely resemble job search models, although with costless wage offers during periods of employment. However, the latter models are based on income maximization. The preceding framework with the following definitions can be interpreted either as a labor force participation model or as a job search model. Unlike previous examples, we assume that it is current participation rather than the stock of past leisure that enters preferences. Thus, we modify (1) to:

$$(6) \quad U(m_t^i, C_t^i, a_t^i)$$

where  $\frac{\Delta U}{\Delta m_t^i} < 0$ . In addition,

$$\begin{aligned} M_t^i &= \text{the number of years of labor force experience} \\ m_t^i &= \text{unity if the individual participates in period } t, \text{ zero otherwise} \\ d_t^i &= \text{unity if the individual is (exogenously) laid off, zero otherwise} \\ NY_t^i &= Y(M_{t-1}^i, m_t^i - d_t^i, h_t^i, t) - e_t m_t^i \end{aligned}$$

where income  $Y(\cdot)$  (or the wage rate given fixed hours) is increasing in experience  $(M_{t-1}^i)$  and in current participation  $(m_t^i - d_t^i)$ . There is a fixed cost of work  $e_t$  (Cogan (1980)).  $d_t$  may be viewed by the individual as a random (exogenous) variable. The participation or job acceptance decision at any period depends, therefore, on the stock of accumulated experience, on (expected) future income or wage rate determinants, and on (expected) future layoff propensities.

We have presented these examples to illustrate the applicability of the basic model. It is interesting to note that each of the models contains some form of state dependence (Heckman (1981)) in that current decisions are affected by past states (decisions) in a structural sense. There are major simplifications in each example, but extensions are best explored in the context of the specific problem. It is probably unnecessary to point out that few sequential decision-making models have been estimated directly from theoretical foundations.

### 3. ESTIMATION STRATEGY

The objective of the empirical work is to estimate the parameters that determine the individual choice of  $\{m_t^i\}_{t=0}^T$ . These parameters consist of the discount factor ( $\beta$ ), the parameters of the preference function ( $U(\cdot)$ ) and of the income generating function ( $Y(\cdot)$ ), and the parameters of the distributions of the random variables that affect preferences and income. Estimation obviously requires choosing a particular parameterization of the preference and income functions.

One strategy involves choosing a specification that enables the researcher to solve for the complete characterization of the optimal decision under uncertainty. Assuming that some variables are observed by the individual but not by the researcher (e.g., the preference shifter  $a_t^i$ ) allows for an error in predicting the individual decision. In this way, maximum likelihood estimation can be integrated into the solution of the optimization problem. This procedure corresponds to a full solution estimation method.

Alternatively, as we show below, it is possible to specify a set of necessary conditions which must be satisfied at the maximum of problem (1). Although the decision variable is not continuous, the derivative of the objective function evaluated at an appropriate point can be used to form a set of inequality restrictions that determine the discrete choice. We demonstrate a method for transforming these restrictions into probability statements about any arbitrary sequence of decisions, which leads naturally to a maximum likelihood approach.

### 3.1. The Full Solution Method (Wolpin, 1982)

The full solution method utilizes Bellman's (1957) principle. We first solve for the last period decision and work backwards to the initial period. Using Bellman's equation, at any period  $t$ , the expected utility from the choice  $m_t = 1$  is given by (we omit the index  $i$ ):

$$\begin{aligned}
 (7) \quad E_t(LU_t | m_t = 1, M_{t-1}) &= E_t(U_t | m_t = 1, M_{t-1}) \\
 &+ \beta E_t \max[E_{t+1}(LU_{t+1} | m_{t+1} = 1, m_t = 1, M_{t-1}), \\
 &E_{t+1}(LU_{t+1} | m_{t+1} = 0, m_t = 1, M_{t-1})]
 \end{aligned}$$

Similarly, the expected utility for the opposite choice is

$$\begin{aligned}
 (8) \quad E_t(LU_t | m_t = 0, M_{t-1}) &= E_t(U_t | m_t = 0, M_{t-1}) \\
 &+ \beta E_t \max[E_{t+1}(LU_{t+1} | m_{t+1} = 1, m_t = 0, M_{t-1}), \\
 &E_{t+1}(LU_{t+1} | m_{t+1} = 0, m_t = 0, M_{t-1})]
 \end{aligned}$$

where

$$(9) \quad LU_t = \sum_{j=t}^T \beta^j U(M_j, C_j, a_j)$$

i.e.  $LU_t$  is lifetime utility at  $t$  subject to the conditions (2), (3) and (4) and  $M_{t-1}$  is given.  $t$  is determined by the difference between (7) and (8), namely:

$$\begin{aligned}
 (10) \quad m_t = 1 \text{ iff } J_t &= E_t(LU_t | m_t = 1, M_{t-1}) - E_t(LU_t | m_t = 0, M_{t-1}) \geq 0 \\
 m_t &= 0 \text{ otherwise}
 \end{aligned}$$

In general, calculating (7) and (8) for  $t = 1, \dots, T$  is, even numerically, an intractable task. An enormous simplification is achieved if the utility function is assumed to be quadratic and the constraint (7) is linear (see equations (24) and (25) below). In this case all quadratic terms in random variables in  $J_t$  (10) vanish, so that knowledge of conditional means alone is required. Even more important, if the random preference parameter ( $a_t$ ) is in the linear terms in the utility function (see equation (24)) it turns out that it is additively separable and monotonically increasing in  $a_t$ . Since  $a_t$  is a real number, there always exists an  $a_t^*$  such that  $J_t = 0$ . The importance of this result for estimation will become apparent.

Estimation proceeds in the following manner. For each  $t$  (for a given individual) one can find the unique value  $a_t^*$  for which  $J_t = 0$ , i.e., for which the individual is exactly indifferent between  $m_t = 1$  and  $m_t = 0$ . Given a distribution for  $a_t$ , the probability of the event occurring is

$$(11) \quad \Pr(m_t = 1 | M_{t-1}) = \Pr(a_t > a_t^*)$$

The joint probability of any given sequence of events

$\bar{m}_{t,k} = (\bar{m}_t, \bar{m}_{t+1}, \dots, \bar{m}_{t+k})$  is

$$(12) \quad \Pr(\bar{m}_{t,k} | M_{t-1}) = \Pr(\bar{m}_{t+k} | M_{t+k-1}) \Pr(\bar{m}_{t+1-1} | M_{t+k-2}) \dots \Pr(\bar{m}_t | M_{t-1})$$

and the sample likelihood function for any set of individuals is the product of the probabilities of sequences such as (12) over the individuals. Notice that for each set of parameters, a new set of  $a_t^*$ 's are found by solving

the dynamic programming problem; each evaluation of the likelihood function requires resolving the dynamic program. Optimization must precede numerically since decision rules are not analytic even in the linear-quadratic case.

### 3.2. The Necessary Conditions Method

Any viable alternative to the full solution estimation method should mitigate some deficiencies of that method, namely (1) the necessity for simplified structures in order to permit economical numerical solution, and (2) the large computational burden of even the simplest of models. The necessary condition approach formulated in this section meets both of these criteria, although not without cost.

To demonstrate this estimation method it is useful first to define the "desired" stock as

$$(13) \quad \bar{M}_t = M_t + d_t = M_{t-1} + m_t$$

which implies that  $M_t = \bar{M}_t - \bar{M}_{t-1} + d_{t-1}$ . Substituting (2)-(4) into (1) the problem can now be written as (ignoring the  $i$  superscript)

$$(14) \quad \text{Max } V_0 = E_0 \sum_{t=0}^T \beta^t U(\bar{M}_t - d_t, NY(\bar{M}_{t-1} - d_{t-1}, \bar{M}_t - \bar{M}_{t-1} + d_{t-1}, h_t, t), a_t)$$

by choice of  $\{\bar{M}_t\}_{t=0}^T$ .

Differentiating (14) with respect to  $\bar{M}_t$  at any arbitrary life cycle period  $t$  yields:



$$(15) \quad \frac{\partial V_t}{\partial M_t} = E_t \{ \beta^t [U_1^t + U_2^t NY_2^t] + \beta^{t+1} [U_2^{t+1} (NY_1^{t+1} + NY_2^{t+1})] \}$$

where  $U_j^k$  and  $NY_j^k$  are the partial derivative of the  $j^{\text{th}}$  element of the utility function and net income functions respectively, at age  $k$ . Let

$$(16) \quad Z_t = \beta^t [U_1^t + U_2^t NY_2^t] + \beta^{t+1} [NY_1^{t+1} - NY_2^{t+1}]$$

$$= Z(m_t, m_{t+1}, M_{t-1}, h_t, h_{t+1}, d_t, d_{t+1}, t, \beta)$$

and define the random variable  $\phi_t$  as

$$(17) \quad \phi_t \equiv E_t(Z_t) - Z_t$$

from which it is clear that

$$(18) \quad E_t \phi_t = E \phi_t = 0.$$

Using (17), it is possible to write (15) as

$$(19) \quad \frac{\partial V_t}{\partial M_t} = E_t(Z_t) = Z_t + \phi_t$$

Now, if the  $V$  function is symmetric, it should be clear that an individual will wish to augment the stock, i.e., to choose  $m_t = 1$ , if and only if (19) is positive when evaluated at  $m_t = 1/2$  and all other variables are evaluated at their actual realizations.<sup>5</sup> Thus,

$$(20) \quad m_t = 1 \quad \text{iff} \quad Z_t(m_t = 1/2) + \phi_t \geq 0$$

$$m_t = 0 \quad \text{iff} \quad Z_t(m_t = 1/2) + \phi_t < 0$$

where as noted,  $Z_t(m_t = 1/2) = Z_t(m_t = 1/2, m_{t+1}, M_{t-1}, \dots)$ .

One can ask about the exact distribution of  $\phi_t$  given the distributions of  $h_t$  and  $a_t$ . This distribution is not easy to find and there is no need, in fact, to explicitly calculate it. It is assumed that  $h_t$  and  $a_t$  come from the same distribution over all individuals and in all life-cycle periods. The distributions of  $h_t$  and  $a_t$  are independent over all individuals and periods. Therefore,  $\phi_t$  is also independent over individuals and time. Using a particular distribution for  $\phi_t$  we can write the following probability statement for the choice  $m_t = 1$ :

$$(21) \quad \begin{aligned} \Pr(m_t = 1 | m_{t+1}, M_{t-1}, \text{other exogenous variables}) \\ = \Pr(Z_t(m_t = 1/2) \geq -\phi_t) \quad t = 0, \dots, T \end{aligned}$$

The inequality sign in (2) is reversed for the  $m_t = 0$  probability statement. Note that at  $T$ ,  $m_{T+1} = 0$  (terminal condition) so that we can calculate from (21)

$$(22) \quad \Pr(m_T = j | M_{T-1}, \text{other exogenous variables}) \quad j = 0, 1.$$

For any given observed sequence of choices for an individual,  $\bar{m} = (\bar{m}_0, \bar{m}_1, \dots, \bar{m}_T)$ , the probability of observing that sequence may be written as

$$(23) \quad \Pr(\bar{m}_0, \dots, \bar{m}_T) = \Pr(\bar{m}_T | M_{T-1}) \Pr(\bar{m}_{T-1} | M_{T-2}) \dots \Pr(\bar{m}_0 | M_{-1})$$

where we ignore other exogenous variables and where we use the fact that in the current model only the cumulative value of past decisions are relevant for the current choice.<sup>6</sup> Unfortunately, our necessary conditions lead to probability statements that, except for the last period, are different from those required to form the probability statement given in (23). For all periods other than  $T$ , the probabilities we directly derive are conditional on the one-period ahead choice, i.e.,  $m_{t+1}$  for the  $t^{\text{th}}$  period choice. We prove in Appendix A, however, that (23) can be derived from (21) and (22) which are themselves calculable. The sample likelihood function can thus be formed as products of probabilities of choice sequences over individuals.

### 3.3. A Comparison of the Two Methods

The two models differ with respect to the restrictions that are imposed on the conditional probabilities that are calculated for each  $m_t$ . In the full solution method the dynamic programming problem is completely solved such that there is no error involved due to future decisions. The only source of uncertainty in calculating the probability of  $m_t$  is due to the unobserved preference shifter at time  $t$ ,  $a_t$ . Hence, all the restrictions of the theory are imposed in calculating the likelihood value of observing the

particular sample. Therefore, it is the most efficient method of estimation.

As long as the model is correctly specified, the maximum likelihood estimates should be consistent for both the full solution and the necessary condition approaches. However, the second method is not as efficient as the first. Using the necessary conditions method we do not calculate optimal future decisions. In fact we calculate only their conditional probabilities and we do not use all the information from the theory. The uncertainty that is included in the forecast error of the future endogenous variables depends upon the parameters of the model and the distributions of the exogenous variables. Hence, the likelihood value of the second method should be higher than the likelihood value of the full solution, if everything else is the same. However, we cannot map this difference into a statistical test, since there is no clear way to specify the asymptotic likelihood ratio test. Hence, there is no way to make a formal comparison or to measure the closeness of the estimated parameters using the two methods.

The full solution method does not, however, easily admit to extensions. Since it is necessary to solve for all conditional expectations, nonlinear functions create extreme computational difficulties. The necessary condition approach avoids that problem by exploiting the information from the first-order conditions and the rational expectation assumption that is implicitly imbedded into the solution of the dynamic programming problem. In addition, as we will see in the example presented below, the full solution approach is several times more computationally burdensome.

#### 4. A MONTE CARLO ESTIMATION USING THE NECESSARY CONDITION METHOD

In order to compare the two methods we performed a Monte Carlo experiment with the necessary condition method, using as a basis the fertility model discussed in section two. That model was chosen as it has been estimated by Wolpin (1982) using the full solution method on Malaysian data on fertility and child mortality, and so we have some knowledge about the properties of that model. That data (1976 Malaysian Family Life History) is described in Wolpin (1982) and is used in this exercise as well.

As noted in the previous section, numerical solution of the dynamic programming problem is greatly simplified in the linear-quadratic case. We therefore use the following functional forms:

$$(24) \quad U(M_t^i, C_t^i, a_t^i) = (\alpha_1 + a_t^i)M_t^i - \alpha_2 M_t^{i2} + \beta_1 C_t^i - \beta_2 C_t^{i2} + \gamma_1 C_t^i M_t^i + \gamma_2 S^i M_t^i$$

where in addition to the previously defined terms (see section 2.1),  $S^i$  is the schooling level of the mother and

$$(25) \quad Y_t^i = C_t^i + e_{1t}m_t^i + e_2(m_t^i - d_t^i)$$

Following Wolpin (1982) we assume that  $e_{1t}$  (the fixed cost of a birth) has the time profile:

$$(26) \quad e_{it} = e_1^0 + e_1^1 t + e_1^2 t^2 + e_{11} d_1 + e_{12} d_2$$

where  $d_1$  and  $d_2$  are equal to unity if it is the first and second period in the life cycle respectively and zero otherwise.<sup>7</sup> The income generating function (for the husband alone as the wife is assumed not to work) is

$$(27) \quad \text{Ln} y_t^i = b_0^i + b_1^i t + b_2^i t^2 + v_t^i$$

where  $E_t v_t^i = E v_t^i = 0$  for all  $i$  and  $t$ ; it is estimated for each household in the sample.<sup>8</sup> The exogenous survival probability is assumed to be related only to calendar time and is given by the logistic formulation

$$(28) \quad \log \frac{\pi_t^i}{1-\pi_t^i} = \eta_0^i + \eta_1^i t + \eta_2^i t^2 + u_t^i$$

with  $E_t u_t^i = E u_t^i = 0$  for all  $i$  and  $t$ ; it is estimated from time-series observations on infant mortality for each of the eleven states in Malaysia.

Households are assumed to know the parameter vector  $(b_0, b_1, b_2, \eta_0, \eta_1, \eta_2, \alpha_1, \beta_1, \beta_2, \gamma_1, \gamma_2, e_1^0, e_1^1, e_{11}, e_{12}, \beta, e_2, a_t)$  but do not know the current draws (at  $t$ ) on  $v_t^i, d_t^i, u_t^i$  nor the future draws on  $a_t^i$ . Given the assumption that  $v_t^i$  and  $u_t^i$  are i.i.d., the household revises its decision at each period  $t$  on the basis only of the mortality outcome  $d_t^i$  and the random preference parameter  $a_t^i$ . The researcher observes the state variable  $M_{t-1}^i$  at  $t$  as does the household, but does not observe  $a_t^i$ . Except for  $a_t^i$  the researcher would predict the fertility decision without error. As discussed in the previous section, the full solution estimation

method proceeds by finding critical values for  $a_t^i$  at each  $t$  which makes the household indifferent between  $m_t^i = 1$  and  $m_t^i = 0$  and which, given a distribution for  $a_t^i$ , can be used to formulate the likelihood of observing any particular sequence of choices.

The data on the exogenous variables used in the Monte Carlo experiment was obtained from actual Malaysian data for 188 women and their husbands. Income and survival probability data came directly from the sample data (i.e., the  $b$ 's and  $n$ 's). Fertility outcomes were generated by the dynamic programming solution in the following manner. For a particular set of parameters, taken to be the "true" parameter values (those actually used are shown in Table 1), in each period and for each household an i.i.d. random draw for  $a_t^i$  was obtained from a standard normal density. If, given  $a_t^i$ , it was optimal to have a child at  $t$  as a result of the dynamic optimization, an infant death was randomly generated using the sample death probability. This determined the number of surviving children entering the subsequent period. In this way, we generated a sequence of births (and deaths) for each household over the number of periods each household was actually observed in the Malaysian sample. Together with the life cycle income and survival probability forecasts, the fertility and mortality outcomes comprised the data available to the researcher. The  $a_t^i$  are, of course, observed only by the household (at  $t$ ).

We then used this data to estimate the parameters with the necessary condition approach of the previous section. With the linear-quadratic structure, equation (15) becomes

$$(29) \frac{\partial V_t}{\partial M_t} = E_t \{ \theta_0 + \theta_1 m_t + \theta_2 m_{t+1} + \theta_3 M_{t-1} + \theta_4 d_t + \theta_5 d_{t+1} + \theta_6 Y_t + \theta_7 Y_{t+1} + \theta_8 S \}$$

where the  $\theta$ 's are defined in terms of the fundamental parameters as

$$\theta_0 = \alpha_1 - \beta_1(e_2 + e_{1t})(1-\beta)$$

$$\theta_1 = -\alpha_2 + (e_2 + e_{1t})(-2\gamma_1 + \beta\gamma_1 - \beta_2(e_2 + e_{1t}))$$

$$\theta_2 = \beta(e_2 + e_{1t})(e_2 + e_{1t} + \gamma_1)$$

$$\theta_3 = -\alpha_2 - \gamma_1(e_2 + e_{1t})(1-\beta)$$

$$(30) \quad \theta_4 = \alpha_2 + \gamma_1 e_2 + (e_2 + e_{1t})(\beta_2 e_2 + \gamma_1(1-\beta))$$

$$\theta_5 = (e_2 + e_{1t})\beta(e_2\beta_2 + \gamma_1)$$

$$\theta_6 = \gamma_1 + \beta_2(e_2 + e_{1t})$$

$$\theta_7 = -\beta\beta_2(e_2 + e_{1t})$$

$$\theta_8 = \gamma_2$$

Equation (19) now becomes



$$(31) \quad \frac{\partial V_t}{\partial M_t} = \theta_0 + \theta_1 m_t + \theta_2 m_{t+1} + \theta_3 M_{t-1} + \theta_4 E_t d_t + \theta_5 E_t d_{t+1} + \theta_6 E_t Y_t \\ + \theta_7 E_t Y_{t+1} + \theta_8 S + \theta_2 \phi_t = Z_t + \phi_t$$

with  $\phi_t$  given by  $\theta_2 \phi_t$  corresponding to the term in equation (21).

Table 1 shows the "true" parameter values (column 1) and the estimated parameters using the necessary conditions estimation approach. A single experiment amounts to choosing an  $a_t^i$  for each household over its life cycle. There are 3086 household periods in all. We performed two such experiments for the given "true" parameter values. They are reported separately in columns 2 and 3. Ideally, one would like to perform many more such experiments at alternative sets of "true" values, but the computational burden of such an exercise is prohibitive.

It is difficult to obtain a summary measure of the "closeness" of the approximate approach to the full solution method. Although the former model is nested in the latter, the actual restrictions imposed are not apparent. Restrictions that are automatically taken into account in the full dynamic programming solution are not used in the necessary condition solution. Thus, although the  $\ln$  likelihood value in the full solution is -1920.3 which is substantially higher (in absolute value) than the  $\ln$  likelihood values reported in Table 1, we do not know the number of restrictions and so cannot perform the usual likelihood ratio test.<sup>9</sup>

It is evident, however, from Table 1 that the discount factor is not particularly robust to the estimation method; in the first "experiment" its estimated value is outside any reasonable range. Eyeballing the differences

suggests that the necessary condition approach is possibly quite inaccurate. In both experiments parameters are often orders of magnitude different than the true values. On the other hand, we can evaluate the results independently of the value of the true parameters. Given the data, that by assumption have been generated by optimal dynamic programming, one can ask whether the estimated parameters fit these data well. The answer is that the results are mixed. In the first estimation (column 2))  $\beta_2$  has the wrong sign and the discount factor  $\beta$  is negative. In the second estimation (column (3))  $\beta_2$  and  $e_2$  have the wrong sign. However, the model under the necessary condition method is being much better than a pure chance model since the latter ln likelihood value of -2059.5 (see Wolpin (1982)). Hence, the hypothesis that all the parameters (besides  $\alpha_1$ ) are zero is rejected by any level of significance.<sup>10</sup>

-----  
Table 1 to be Inserted here  
-----

#### CONCLUDING REMARK

In this paper we have suggested a way of formulating a general estimable dynamic discrete choice model. Due to the computational limits of the approach that explicitly specify the individual choice problem, we developed new estimation methods. The full solution method has already proved to be successful in estimating a complicated dynamic fertility model (Wolpin (1982)). Here we have tried to evaluate an alternative method which was

designed to accommodate more complicated dynamic discrete choice models. The estimation results are not very encouraging and the burden of computation has been reduced only by  $1/3$  to  $1/2$ . At this stage of our research we suggest using the full solution method but would encourage the interested researchers to continue the search for estimable models that can accommodate more complex behavioral assumptions.

TABLE 1

Maximum Likelihood Estimates of the Fertility Model:  
The Necessary Condition Method

	(1) True Values	(2)	(3)
$\alpha_1$	$3.432 \times 10^{-2}$	$9.905 \times 10^{-1}$	$3.606 \times 10^{-1}$
$\alpha_2$	$2.939 \times 10^{-1}$	$2.277 \times 10^{-5}$	$3.390 \times 10^{-6}$
$\beta_1$	$6.162 \times 10^{-5}$	$1.178 \times 10^{-5}$	$1.145 \times 10^{-4}$
$\beta_2$	$1.074 \times 10^{-16}$	$-1.376 \times 10^{14}$	$-1.031 \times 10^{15}$
$\gamma_1$	$2.421 \times 10^{-7}$	$4.607 \times 10^{-7}$	$1.64 \times 10^{-6}$
$\gamma_2$	$-5.419 \times 10^{-3}$	$-6.788 \times 10^{-2}$	$-4.050 \times 10^{-2}$
$e_2$	$1.474 \times 10^3$	$5.200 \times 10^3$	$-3.554 \times 10^3$
$e_1^0$	$-1.947 \times 10^3$	$5.391 \times 10^3$	$5.831 \times 10^3$
$e_1^1$	$-2.861 \times 10^2$	$5.926 \times 10^2$	$-1.915 \times 10^2$
$e_1^2$	$6.417 \times 10^1$	$2.978 \times 10^1$	$6.982 \times 10^1$
$e_{11}$	$1.769 \times 10^4$	$1.023 \times 10^6$	$1.834 \times 10^5$
$e_{12}$	$8.093 \times 10^3$	$1.533 \times 10^4$	$1.062 \times 10^4$
$\delta$	$9.215 \times 10^{-1}$	$-2.895$	$5.011 \times 10^{-1}$
$\ln L$		-1829.7	-1837.0

APPENDIX A

In this appendix we demonstrate a method to transform the conditional probability statements derived from the necessary conditions (see (21) and (22)) into the probability statements that are used to form the likelihood function.

We need to derive  $\Pr(m_{t-1} = j \mid M_{t-2} = 0, \dots, T-2)$  for  $j = 1, 0$ , which we will denote by  $\psi_{t-1}$  and  $1-\psi_{t-1}$  respectively. To do so, first write

$$(A.1) \quad \Pr(m_{t-1} = 1 \mid m_t = j, M_{t-2}) = \Pr(m_{t-1} \mid m_t = j, M_{t-2}).$$

$$\begin{aligned} & [\Pr(m_t = j \mid m_{t-1} = 0, M_{t-2}) \Pr(m_{t-1} = 0 \mid M_{t-2}) \\ & + \Pr(m_t = j \mid m_{t-1} = 1, M_{t-2}) \Pr(m_{t-1} = 1 \mid M_{t-2})] \end{aligned}$$

If we sum A.1 over  $j = 1, 0$ , we get the marginal probability for  $m_{t-1}$  in the following form

$$\begin{aligned} (A.2) \quad \psi_{t-1} &= \Pr(m_{t-1} = 1 \mid m_t = 1, M_{t-2}) [\Pr(m_t = 1 \mid m_{t-1} = 0, \\ & \quad M_{t-2})(1-\psi_{t-1}) + \Pr(m_t = 1 \mid m_{t-1} = 1, M_{t-2})\psi_{t-1}] \\ &+ \Pr(m_{t-1} = 1 \mid m_t = 0, M_{t-2}) [\Pr(m_t = 0 \mid m_{t-1} = 0, M_{t-2})(1-\psi_{t-1}) \\ & \quad + \Pr(m_t = 0 \mid m_{t-1} = 1, M_{t-2})\psi_{t-1}] \end{aligned}$$

Solving for  $\psi_{t-1}$ , we get

$$\begin{aligned}
 (A.3) \quad \psi_{t-1} = & \Pr(m_{t-1} = 1 \mid m_t = 1, M_{t-2}) \Pr(m_{t-1} = 1, m_{t-1} = 0 \mid M_{t-2}) \\
 & + \Pr(m_{t-1} = 1 \mid m_t = 0, M_{t-2}) \Pr(m_t = 0 \mid m_{t-1} = 0, M_{t-2}) \\
 & 1 + \Pr(m_{t-1} = 1 \mid m_t = 1, M_{t-2}) [\Pr(m_t = 1 \mid m_{t-1} = 0, M_{t-2}) - \\
 & - \Pr(m_t = 1 \mid m_{t-1} = 1, M_{t-2})] \\
 & + \Pr(m_{t-1} = 1 \mid m_t = 0, M_{t-2}) [\Pr(m_t = 0 \mid m_{t-1} = 0, M_{t-2}) - \\
 & - \Pr(m_t = 0 \mid m_{t-1} = 1, M_{t-2})]
 \end{aligned}$$

Now, consider working backwards from  $T$ .  $\psi_T$  is known directly from the terminal necessary condition (22). From  $\psi_T$ , one can find

$$\begin{aligned}
 (A.4) \quad \Pr(m_T = 1 \mid m_{T-1}, M_{T-2} = x) = & \Pr(m_T = 1 \mid m_{T-1} = 1, d_{T-1} = 1, M_{T-2} = x) \\
 & \Pr(d_{T-1} = 1) \\
 & + \Pr(m_T = 1 \mid m_{T-1} = 1, d_{T-1} = 0, M_{T-2} = x) \Pr(d_{T-1} = 0) \\
 = & \Pr(m_T = 1 \mid M_{T-1} = x) \Pr(d_{T-1} = 1) + \Pr(m_T = 1 \mid M_{T-1} = x+1)
 \end{aligned}$$

and similarly for  $\Pr(m_T = 0 \mid m_{T-1} = 0, M_{T-2})$ ,  $\Pr(m_T = 1 \mid m_{T-1} = 0, M_{T-2})$  and  $\Pr(m_T = 0 \mid m_{T-1} = 1, M_{T-2})$ . Thus, one can find  $\psi_{T-1}$  from A.3 given A.4. Given  $\psi_{T-1}$  one can follow the same procedure as in (A.4) to find  $\psi_{T-2}$  from (A.3) etc.

APPENDIX B

The following table contains some information on the Monte Carlo experiment. The starting values are 10 percent below the true values. The cost information is given below the table.

	True Values	Starting Values	Final Values
$\alpha_1$	$3.4 \times 10^{-1}$	$3.1 \times 10^{-1}$	$3.1 \times 10^{-1}$
$\alpha_2$	$2.9 \times 10^{-1}$	$3.3 \times 10^{-1}$	$3.2 \times 10^{-1}$
$\beta_1$	$6.1 \times 10^{-5}$	$5.5 \times 10^{-5}$	$5.9 \times 10^{-5}$
$\beta_2$	$1.1 \times 10^{-16}$	$.9 \times 10^{-16}$	$.98 \times 10^{-17}$
$\gamma_1$	$2.4 \times 10^{-1}$	$2.1 \times 10^{-1}$	$2.3 \times 10^{-1}$
$e_2$	1467	1600	1880
$\delta$	.92	.84	.92
$\gamma_2$	$-5.4 \times 10^{-3}$	$-5.9 \times 10^{-3}$	$-5.8 \times 10^{-3}$
$e_1^0$	1942	1700	1726
$e_1^1$	-285	-310	-322
$e_1^2$	63.7	58.0	64.5
$e_{11}$	17610	19300	21000
$e_{12}$	8062	7200	7816
$\ln$	-1912.5	-1950.2	-1904.1

L = Likelihood

Number of Iterations = 19

Convergence criterion  $\Delta \ln L = 10^{-10}$

.55 min per evaluation 25-30 evaluations per iteration

## FOOTNOTES

1. Here  $t$  refers to life cycle period (age) rather than calendar time per se, although for different cohorts  $t$  will also correspond to different calendar time. We ignore here any calendar time aspect that affects the individual decision.
2. Note that the no savings assumption is equivalent to assuming a utility function that is linear in consumption which is common to some dynamic estimate model, such as in search models (e.g., Miller, (1983) and Heckman's statistical model that we present in the Introduction).
3. Actually  $U(\cdot)$  and  $NY(\cdot)$  need be differentiable only at one point in the  $(0, 1)$  interval of the  $m_t^i$  variable, as will be discussed later.
4. Strictly speaking, this is only true if  $a_t$  is i.i.d. or if it follows a permanent-transitory scheme. For a more detailed discussion of these issues see Wolpin (1982).
5. Note that in (16) we ignore the unobservable taste elements  $a_t$  and  $a_{t+1}$ . Those elements, if they exist, cannot enter except additively in  $Z_t$ , in order to preserve the validity of the method.
6. That is,  $\Pr(\bar{m}_T, m_{T-1}, \dots, m_0) = \Pr(m_{T-1})$  for the special case we are considering, and similarly for other periods.
7. A period is taken to be eighteen months and the first decision period is age fifteen.



8. The assumption that  $v_t^i$  is serially uncorrelated greatly simplifies the dynamic programming solution since households do not have to update their forecasts each period.
9. The number of restrictions is likely to be large since there is on average sixteen periods per woman and separate restrictions for each period. Twice the difference in the likelihood is 181.2 and the null hypotheses that the parameters in column 1 are the same as those in 2 or 3 would be accepted at the standard significance level, only if the number of restrictions exceeded 150, which seems unlikely.
10. Additional information on the cost of estimating the model using the necessary condition method is given in Appendix B.

# REFERENCES

- Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings, Journal of Political Economy, Aug. 1967.
- Brown, Charles, "A Model of Optimal Human-Capital Accumulation and the Wages of Young High School Graduates," Journal of Political Economy, April, 1976.
- Cogan, John, "Fixed Costs and Labor Supply," Econometrica, July 1981.
- Haley, William, "Estimation of Earnings Profiles from Optimal Human Capital Accumulation," Econometrica, 1976.
- Heckman, J.James, "Statistical Models for Discrete Panel Dats," in Structural Analysis of Discrete Data (eds.) C.F.Marshki and D.McFadden. Cambridge Mass: MIT Press, 1981.
- Heckman, James, J., "A Life Cycle Model of Earnings, Learning, and Consumption," Journal of Political Economy, August, 1976.
- Heckman, James, J. and Robert J.Willis, "Estimation of a Stochastic Model of Reproduction: An Econometric Approach," in Household Production and Consumption. ed. N.Terleckyj, National Bureau of Economic Research, Stanford, California, 1975.
- Heckman, J. and R.Willis, "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Married Women," JPE 5, No.1 (February 1977), 27-58.

Lucas, Robert, E.Jr., "Economic Policy Evolution: A Critique" in K.Brunner and A.H.Meltzer (eds.) The Phillips Curve and Labor Markets Carnegie-Rochester Conference.

Marschak, Jacob, "Measurement for Policy and Predictions," in W.C.Hood and T.C.Koopman (eds.) Studies in Econometric Method, Cowles Monograph No.14 (New York: Wiley, 1953).

McFadden, I.P., "Control Choice Analysis: A Survey," Annals of Economic and Social Measurement, 5, No.4, 1976.

Orazem, Pefer, F., "Two Models of School Achievement and Attendance with Application to Segregated Schools," Ph.D Dissertation, Yale University, 1983.

Wolpin, Kenneth, I., "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," forthcoming Journal of Political Economy.

Wallace, T.Dudley, L.A., Ibreu, "Full-Time Schooling in Life-Cycle Models of Human Capital Accumulation," Journal of Political Economy, February, 1976.