
Conditional Choice Probabilities and the Estimation of Dynamic Models

Author(s): V. Joseph Hotz and Robert A. Miller

Source: *The Review of Economic Studies*, Jul., 1993, Vol. 60, No. 3 (Jul., 1993), pp. 497-529

Published by: Oxford University Press

Stable URL: <https://www.jstor.org/stable/2298122>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Oxford University Press is collaborating with JSTOR to digitize, preserve and extend access to *The Review of Economic Studies*

JSTOR

Conditional Choice Probabilities and the Estimation of Dynamic Models

V. JOSEPH HOTZ
University of Chicago

and

ROBERT A. MILLER
Carnegie Mellon University

First version received April 1989; final version accepted January 1993 (Eds.)

This paper develops a new method for estimating the structural parameters of (discrete choice) dynamic programming problems. The method reduces the computational burden of estimating such models. We show the valuation functions characterizing the expected future utility associated with the choices often can be represented as an easily computed function of the state variables, structural parameters, and the probabilities of choosing alternative actions for states which are feasible in the future. Under certain conditions, nonparametric estimators of these probabilities can be formed from sample information on the relative frequencies of observed choices using observations with the same (or similar) state variables. Substituting the estimators for the true conditional choice probabilities in formulating optimal decision rules, we establish the consistency and asymptotic normality of the resulting structural parameter estimators. To illustrate our new method, we estimate a dynamic model of parental contraceptive choice and fertility using data from the National Fertility Survey.

1. INTRODUCTION

Over the last several years there has been increasing interest in estimating structural models of dynamic discrete choice. Empirical applications have been undertaken in the areas of fertility (Wolpin (1984)), job search (Kiefer and Newmann (1979, 1981), Flinn and Heckman (1982), Lancaster and Chesher (1983), Wolpin (1987)), job matching (Miller (1982, 1984)), labour force participation (Eckstein and Wolpin (1989a), Gönül (1989)), Berkovec and Stern (1991)), patent renewal (Pakes (1986)) and the replacement of bus engines (Rust (1987))¹. These studies derive the stochastic process generating an agent's choice sequence from the solution to a dynamic optimization problem, which depends upon structural parameters characterizing the agent's preferences and her constraints. The estimation problem is to identify and consistently estimate the structural parameters from data on choices and other observed variables. Such estimates enable one to examine and forecast how exogenous changes in economic constraints affect choices.

In contrast to models with continuous choices which can be estimated from the first-order conditions, the optimal decision rules for dynamic discrete choice models are characterized by inequality conditions. This has prompted researchers to (numerically) solve the valuation function characterizing the optimal sequence of choices in order to

1. See Eckstein and Wolpin (1989b) for a recent survey of this fast growing field.

estimate the structural parameters of such models. Indeed, most of the studies cited above compute the valuation function using backwards recursion, not just once, but every time the parameters are evaluated in the estimation routine. Although several recent advances have been made that reduce these computational burdens (see Miller (1982, 1984), Wolpin (1984), Pakes (1986) and Rust (1987)), backwards recursion solutions remain extremely costly to implement. Such computational burdens have deterred researchers from estimating all but the most parsimonious specifications of structural models and from experimenting with alternative specifications. This limitation is potentially serious in light of findings, such as those of Flinn and Heckman (1982), which indicate that estimates for job search models appear very sensitive to alternative specifications of the model's underlying structure.

This paper presents a new estimator for such models, called the Conditional Choice Probability (CCP) estimator. Our approach does not require econometricians to explicitly solve the valuation functions used to characterize optimal decision rules via backwards recursion methods. It is based on a new representation of the valuation function which is expressed in terms of the utility payoffs, choice probabilities, and probability transitions of choices and outcomes that remain feasible in future periods. Under conditions presented below, this representation can be exploited to estimate the model's structural parameters by employing non-parametric estimates of these future choice probabilities and probability transitions in place of their true values. Coupling our representation of valuation functions with the semiparametric estimation procedures we develop in this paper, makes tractable empirical investigations of a wide class of dynamic discrete choice models previously considered too (computationally) expensive to analyse.

The paper is organized as follows. The class of dynamic discrete choice models investigated here is outlined in Section 2. Section 3 provides the representation theorem for valuation functions that ultimately allows us to avoid backwards recursion in estimation. The implication of this theorem is illustrated for a simple optimal stopping model in Section 4. Then, in Section 5, based on this new representation, we propose the CCP estimator for the underlying structural parameters and establish that it is $N^{1/2}$ -consistent and asymptotically normal, where N denotes sample size. The last three sections of the paper present an application of our approach by estimating a life cycle model of married couple's optimal contraceptive choice behaviour. Section 6 presents a model of the links between contraceptive choice and fertility, including the possibility of choosing (irreversible) sterilization. Although quite simple, the model nevertheless is encumbered with a large state space which would impose prohibitive computational costs if the maximum likelihood (ML) strategy previously employed in this literature was attempted. The data, described in Section 7, is for a sample of white married couples gathered in the National Fertility Survey of 1975. Finally, Section 8 reports the parameter estimates for several alternative specifications of the model, and examines some of their quantitative implications.

2. THE FRAMEWORK

The framework we investigate includes a wide class of dynamic, discrete choice models. Consider a typical agent making choices over time in an uncertain environment. She is assumed to choose one action from a set, \mathcal{C} , which contains up to J alternatives at each period, t , over a finite life of length T . (Without loss of generality, in this and the following two sections, we assume that calendar time and age are synonymous.) Her objective is to maximize the expected value of a sum of period-specific payoffs or utilities. Let $d_{ij} = 1$

indicate the agent chooses action j in t , and setting $d_{ij} = 0$ means she does something else. Then $\mathbf{d}_t = (d_{t1}, \dots, d_{tJ-1})'$ describes her action in period t . Alternatively expressed:

$$\begin{aligned} d_{ij} \in \{0, 1\} &\quad \text{for all } (t, j) \in T \times \mathcal{C}, \\ \sum_{j=1}^J d_{ij} &= 1 \quad \text{for all } t \in T. \end{aligned} \tag{2.1}$$

The action taken at period t typically affects the outcome, $b_t \in \mathcal{B}$, which arrives at the end of the period. Let $H_t = (b'_0, b_1, \dots, b_{t-1})'$ represent the agent's history as of the beginning of period t ; it includes a $L \times 1$ vector of the agent's initial endowment of characteristics, $b_0 \in \mathcal{B}_0$, and the agent's entire history of outcomes from period 1 through $t-1$.² We assume that \mathcal{B} is a finite set, but only that \mathcal{B}_0 is compact. The outcome, b_t , is either fully determined by action d_t , or generated according to the transition probabilities:

$$F_j(H_{t+1} | H_t), \tag{2.2}$$

where $H_{t+1} = (H_t, b_t)'$.³ In the context of econometric analyses, such distribution functions, as well as the agent's payoff functions described below, may be specified in terms of a vector of structural parameters. While this parameter vector is typically the focus of estimation, we shall first investigate the general structure of the decision problem, delaying the introduction of parametric representations until Section 5.

In each period t , there is a current utility or payoff, u_{ij} , associated with each choice j . Let $u_j^*(H_t) \equiv E(u_{ij} | H_t)$ denote the conditional expectation of u_{ij} , given H_t . It follows that:

$$u_{ij} = u_j^*(H_t) + \varepsilon_{ij}, \tag{2.3}$$

where the stochastic utility component, ε_{ij} , is, by construction, conditionally independent of H_t . Let $u^*(H_t) = (u_1^*(H_t), \dots, u_J^*(H_t))'$ and $\xi_t = (\varepsilon_{t1}, \dots, \varepsilon_{tJ})'$, respectively, denote $J \times 1$ vectors of deterministic and stochastic utility components. We write the distribution function of ξ_t as:

$$G(\xi_t | H_t) \tag{2.4}$$

and assume it has a well-defined, joint probability density function, $dG(\xi_t | H_t)$.

The agent sequentially chooses $\{d_t\}_{t \in T}$ to maximize the objective function:

$$E_0(\sum_{t=0}^T \sum_{j=1}^J d_{ij} [u_j^*(H_t) + \varepsilon_{ij}]). \tag{2.5}$$

Let $d_s^0 = (d_{s1}^0, \dots, d_{s,J-1}^0)'$ denote the agent's optimal choice in period s . We define the conditional valuation function associated with choosing j in period t as:

$$v_j(H_t) = E[\sum_{s=t+1}^T \sum_{j=1}^J d_{sj}^0 [u_{sj}^* + \varepsilon_{sj}] | H_t, d_{ij} = 1], \tag{2.6}$$

where $u_{ij}^* \equiv u_j^*(H_t)$ is adopted for notational simplicity. Optimal decision making implies that $d_{ik}^0 = 1$, if and only if:

$$k = \operatorname{argmax}_{j \in \mathcal{C}} [u_{ij}^* + \varepsilon_{ij} + v_{ij}], \tag{2.7}$$

2. Many problems have a finite state space representation, obviating the need to write down the whole history at each decision node. However the application we investigate in the latter parts of the paper does not enjoy this property, which justifies why a more general formulation is provided here.

3. Since H_t is a vector of length t , there is an argument for subscripting F (as well as many other mappings we define in the text) by t . An alternative notational convention is to express H_t as $(t, H_T^*)'$, where $H_T^* = (H_t, 0, \dots, 0)'$, and define $F(H_{t+1} | H_t)$, for example, on $T^2 \times \mathcal{B}_0^2 \times \mathcal{B}^{2T}$. In this way we avoid the notationally cumbersome subscripting.

where $v_{ij} \equiv v_j(H_t)$. Conditional on history H_t , the probability the agent chooses action k is therefore:

$$p_k(H_t) = \Pr \{k = \operatorname{argmax}_{j \in \mathcal{C}} [u_{ij}^* + \varepsilon_{ij} + v_{ij}] | H_t\}. \quad (2.8)$$

Let $\tilde{p}(H_t) \equiv (p_1(H_t), \dots, p_{J-1}(H_t))'$ denote the $(J-1)$ dimensional vector of conditional choice probabilities associated with the first $J-1$ actions in period t .

3. AN ALTERNATIVE REPRESENTATION OF CONDITIONAL VALUATION FUNCTIONS

In general, the conditional valuation function, $v_j(H_t)$, does not have a closed form solution. The standard practice is to exploit Bellman's (1957) equation and use backward recursion methods to obtain one. This section provides an alternative representation of $v_j(H_t)$ which will prove convenient when estimating a parametric representation of such models.

To derive this representation, note that (2.3) and (2.8) imply that the conditional probability of making choice 1, say, can be written as:

$$\begin{aligned} p_1(H_t) &= E(d_{t1}^0 = 1 | H_t) \\ &= \int_{\varepsilon_1=-\infty}^{\infty} \int_{\varepsilon_2=-\infty}^{\varepsilon_1+u_{t1}^*+v_{t1}-u_{t2}^*-v_{t2}} \cdots \int_{\varepsilon_J=-\infty}^{\varepsilon_1+u_{t1}^*+v_{t1}-u_{tJ}^*-v_{tJ}} dG(\varepsilon_1, \dots, \varepsilon_J | H_t) \\ &= \int_{-\infty}^{\infty} G_1(\varepsilon_1, [\varepsilon_1 + u_{t1}^* - u_{t2}^* + v_{t1} - v_{t2}], \dots, [\varepsilon_1 + u_{t1}^* - u_{tJ}^* + v_{t1} - v_{tJ}] | H_t) d\varepsilon_1, \end{aligned} \quad (3.1)$$

where $G_j(y_t | H_t) \equiv \partial G(y_t | H_t) / \partial u_{tj}$. Up to the normalization, $p_1(H_t)$, the integrand in the last line of (3.1) is the probability density function for ε_{t1} , given history H_t and $[u_{t1}^* + v_{t1} + \varepsilon_{t1}] \geq [u_{tk}^* + v_{tk} + \varepsilon_{tk}]$ for all $k \in \mathcal{C}$ (or, in words, when the first choice is optimal). (See McFadden (1981, p. 204) for example.) For each $k \in \mathcal{C}$, the expression corresponding to (3.1) is a positive, real-valued, mapping from the differences in conditional valuation functions associated with the optimal choice and the alternative actions. We now show that these differences can be expressed as functions of conditional choice probabilities.

Let $y = (v_1, \dots, v_{J-1})'$ be a $(J-1)$ -dimensional vector. For each $t \in T$ and $j \in \{1, \dots, J-1\}$, define the real-valued function, $Q_j(y, H_t)$, as:

$$\begin{aligned} Q_j(y, H_t) &= \int G_j([\varepsilon + u_{tj}^* - u_{t1}^* + v_j - v_1], \dots, [\varepsilon + u_{tj}^* - u_{t,j-1}^* + v_j - v_{j-1}], \varepsilon, \\ &\quad \times [\varepsilon + u_{tj}^* - u_{t,j+1}^* + v_j - v_{j+1}], \dots, [\varepsilon + u_{tj}^* - u_{tJ}^* + v_j] | H_t) d\varepsilon \end{aligned} \quad (3.2)$$

and $Q(y, H_t)$, a $(J-1)$ dimensional vector function, as:

$$Q(y, H_t) = (Q_1(y, H_t), \dots, Q_{J-1}(y, H_t))'. \quad (3.3)$$

If y comprises the differences in conditional valuation functions, namely,

$$y = (v_{t1} - v_{tJ}, \dots, v_{t,J-1} - v_{tJ})' \equiv y(H_t), \quad (3.4)$$

then $\tilde{p}(H_t) = Q(y(H_t), H_t)$. The cornerstone of our estimation strategy is to express $y(H_t)$ as a function of $\tilde{p}(H_t)$. This requires $Q(y, H_t)$ to be invertible in y . By the following proposition, proved in Appendix A, its inverse exists.

Proposition 1. *For each H_t the mapping $Q(\underline{y}, H_t)$ is invertible in \underline{y} .*

Proposition 1 enables one to express $v_j(H_t)$ in terms of the choice probabilities, transition probabilities and expected (per period) payoffs associated with future histories. To demonstrate this, we proceed in several steps. First, the expected optimal payoff in period t , conditional on history H_t , is:

$$E(\sum_{j=1}^J d_{ij}^0 [u_j^*(H_t) + \varepsilon_{ij}] | H_t) = \sum_{j=1}^J p_{ij} [u_j^*(H_t) + E(\varepsilon_{ij} | H_t, d_{ij}^0 = 1)], \quad (3.5)$$

where $p_{ij} \equiv p_j(H_t)$. To express $E(\varepsilon_{ij} | H_t, d_{ij}^0 = 1)$ as a function of conditional choice probabilities, Proposition 1 implies that the (unnormalized) conditional density function, appearing in (3.1), can be written as:

$$\begin{aligned} G_j([\varepsilon + u_j^* - u_{i1}^* + \nu_j - \nu_1], \dots, [\varepsilon + u_j^* - u_{ij-1}^* + \nu_j - \nu_{j-1}], \varepsilon, \\ \times [\varepsilon + u_j^* - u_{i,j+1}^* + \nu_j - \nu_{j+1}], \dots, [\varepsilon + u_j^* - u_{iJ}^* + \nu_j] | H_t) \\ = G_j([\varepsilon + u_j^* - u_{i1}^* + Q_{ij}^{-1} - Q_{i1}^{-1}], \dots, [\varepsilon + u_j^* - u_{i,j-1}^* + Q_{ij}^{-1} - Q_{i,j-1}^{-1}], \varepsilon, \\ \times [\varepsilon + u_j^* - u_{i,j+1}^* + Q_{ij}^{-1} - Q_{i,j+1}^{-1}], \dots, [\varepsilon + u_j^* - u_{iJ}^* + Q_{ij}^{-1}] | H_t) \end{aligned} \quad (3.6)$$

where $\nu_j \equiv \nu_{ij}(H_t)$, $Q_{ij}^{-1} \equiv Q_j^{-1}(\underline{p}_t, H_t)$, and $\underline{p}_t \equiv (p_{t1}, \dots, p_{t,J-1})'$. Therefore, the expectation of ε_{ij} , when j is the optimal action for history H_t , is:

$$\begin{aligned} W_j(\underline{p}_t, H_t) \\ \equiv E(\varepsilon_{ij} | H_t, d_{ij}^0 = 1) \\ = \int \varepsilon G_j([\varepsilon + u_j^* - u_{i1}^* + Q_{ij}^{-1} - Q_{i1}^{-1}], \dots, [\varepsilon + u_j^* - u_{i,j-1}^* + Q_{ij}^{-1} - Q_{i,j-1}^{-1}], \varepsilon, \\ \times [\varepsilon + u_j^* - u_{i,j+1}^* + Q_{ij}^{-1} - Q_{i,j+1}^{-1}], \dots, [\varepsilon + u_j^* - u_{iJ}^* + Q_{ij}^{-1}] | H_t) / p_j(H_t) d\varepsilon. \end{aligned} \quad (3.7)$$

Using (3.5) and (3.7), it follows that the agent's expected utility (or payoff) in period t , conditional on H_t , is:

$$U(\underline{p}_t, H_t) \equiv \sum_{k=1}^J p_{ik} [u_k^* + W_k(\underline{p}_t, H_t)]. \quad (3.8)$$

To complete the representation of $v_j(H_t)$, we characterize the sequences of choices and state variables which would be feasible for the agent in future periods if she was to choose action j in period t . Let $\mathcal{A}_s(H_t)$ denote the set of histories which remain feasible for some age s following t , given history H_t :

$$\mathcal{A}_s(H_t) = \{H_s : H_s = (H_t, b_t, \dots, b_{s-1})'\}. \quad (3.9)$$

Denote the conditional choice probabilities associated with this finite set of possible histories by the vector set, $\underline{p}(H_t)$:

$$\underline{p}(H_t) = \{\underline{p}(H_s) : H_s \in \mathcal{A}_s(H_t) \text{ for } s = t+1, \dots, T\}. \quad (3.10)$$

Then the agent's conditional valuation function for $d_{ij} = 1$ is given by the sum (over the periods $s \in \{t+1, \dots, T\}$ and over histories which might eventuate, $H_s \in \mathcal{A}_s(H_t)$) of the associated expected payoffs, $U_s \equiv U(\underline{p}_s, H_s)$, times the probability of each H_s occurring. The probability of H_s occurring, conditional on H_t and $d_{ij} = 1$, is, in turn, given by the product of the relevant conditional choice and state transition probabilities:

$$F_j(H_{t+1} | H_t) \prod_{r=t+1}^{s-1} [\sum_{k=1}^J p_{rk} F_k(H_{r+1} | H_r)], \quad (3.11)$$

where $H_s \in \mathcal{A}_s(H_r)$ for each $r \in \{t+1, \dots, s-1\}$. Adopting the abbreviated notation $F_{rk} = F_k(H_{r+1} | H_r)$, we have thus established that one can represent any conditional valuation function as a real valued mapping of the current history and future conditional choice probabilities; denoting this new representation by $V_j(H_t, p(H_t))$, it is defined as:

$$V_j(H_t, p(H_t)) = \sum_{s=t+1}^T \sum_{H_s \in \mathcal{A}_s(H_t)} \{ U_s [F_{tj} \prod_{r=t+1}^{s-1} (\sum_{k=1}^J p_{rk} F_{rk})] \}. \quad (3.12)$$

It immediately follows that $V_j(H_t, p(H_t)) = v_j(H_t)$.

Representation (3.12) can be further simplified in some circumstances. Consider histories where there exists at least one action, which, if taken next period, eliminates the differential impact of any subsequent choices on outcomes. Such histories are said to be terminal histories. More precisely, a history H_t is said to be terminal if and only if there exists at least one action, say $J \in \mathcal{C}$, (called a terminating action) which, if chosen in $t+1$, must be picked for all periods $s \in \{t+2, \dots, T\}$.⁴ A search model where agents cannot change jobs exemplifies dynamic discrete choice models with this terminal history property. Suppose the j^{th} action is a terminating action associated with some history H_t . Then the indirect utility associated with this action at all $H_{t+1} \in \mathcal{A}_{t+1}(H_t)$ simplifies to:

$$V_j(H_{t+1}) = \sum_{s=t+2}^T \sum_{H_s \in \mathcal{A}_s(H_{t+1})} E(u_{sj} | H_s) [\prod_{r=t+1}^{s-1} F_j(H_{r+1} | H_r) | H_s \in \mathcal{A}_s(H_r)] \quad (3.13)$$

and it follows from Proposition 1 that the conditional valuation functions at time t associated with the remaining choices may be expressed as:

$$\begin{aligned} V_j(H_t, p(H_t)) &= \sum_{H_{t+1} \in \mathcal{A}_{t+1}(H_t)} \{ U_{t+1} + V_j(H_{t+1}) + \sum_{k=1}^{J-1} p_{t+1,k} v_{t+1,k} \} F_j(H_{t+1} | H_t) \\ &= \sum_{H_{t+1} \in \mathcal{A}_{t+1}(H_t)} \{ U_{t+1} + V_j(H_{t+1}) + (\sum_{k=1}^{J-1} p_{t+1,k} Q_k^{-1}(p_{t+1}, H_{t+1})) \} F_j(H_{t+1} | H_t). \end{aligned} \quad (3.14)$$

Equation (3.14) shows that if H_t is a terminal history, then $V_j(H_t, p(H_t))$ is a function of the values taken on by $p(H_{t+1})$ and $V_j(H_{t+1})$ as H_{t+1} ranges over the elements in the set $\mathcal{A}_{t+1}(H_t)$. Note that (3.13) implies that the conditional probabilities, associated with future choices beyond $t+1$, do not enter the expressions for valuations of period t choices. Consequently, the existence of terminal states greatly reduces the number of future choice probabilities required to calculate conditional valuation functions.⁵

The new representation of conditional valuation functions has two uses: one in forming orthogonality conditions for estimation purposes and the other for interpreting the comparative dynamics associated with changes in the state variables. We conclude this section with a brief discussion of the latter. From the definition of $Q(\nu, H_t)$ and (3.12), it follows that:

$$p(H_t) = Q(\nu(H_t, p(H_t)), H_t) \quad (3.15)$$

Differentiating with respect a continuous component in b_0 , say b_{01} , yields a taxonomy of the various contributing pieces:

$$\frac{\partial p(H_t)}{\partial b_{01}} = \frac{\partial Q}{\partial \nu_t} \left(\frac{\partial \nu_t}{\partial b_{01}} \Big|_{\partial p_t=0} + \frac{\partial \nu_t}{\partial p(H_t)} \frac{\partial p(H_t)}{\partial b_{01}} \right) + \frac{\partial Q_t}{\partial b_{01}} \Big|_{\partial \nu_t=0}. \quad (3.16)$$

The last term in (3.16) captures changes in $p(H_t)$ due to the impact of b_{01} on current utility, holding constant the conditional valuation function. This effect is through two

4. All states associated with terminating actions are absorbing states.

5. The existence of a terminating action is a sufficient, but hardly necessary, condition for realizing the simplifications in the representations of conditional valuation function described in the text. Such gains accrue whenever one can readily obtain the valuation function associated with one or more of the choices.

channels: one through the effects of b_{01} on $y^*(H_t)$, the deterministic components of current utility and the other through those on $G(\xi_t | H_t)$, the probability distribution function of ξ_t . The other way in which b_{01} affects $p(H_t)$ is through its impact on expected future utility, namely y_t . Again, this overall effect (the first term in (3.16)) consists of the influence of b_{01} through several channels: one through its effect on y_t , holding $p(H_t)$ constant and, the other, through its effect on future choice probabilities $p(H_t)$, characterizing $U(p_t, H_t)$ and, ultimately, $V_j(H_t, p(H_t))$. The latter effects reflect the changes in the probabilities of reaching future nodes and adjustments to the dynamic selection correction terms in future periods.

As our application below demonstrates, all these terms can be estimated without computing the valuation function. Similar approximations can be derived to gauge the effects of discrete valued state variables. However, comparative dynamics which change the model's structure (the mappings for u_{ij}^* , F_{ij} and G_i) cannot always be handled this way. In predicting how an agent would react to a regime change, the critical question is whether a second agent now exists whose observed behaviour, in a distributional sense, mimics what the first would do under the new regime. If not, it seems necessary to compute the optimal decision rule under the new regime.

4. AN EXAMPLE: OPTIMAL STOPPING

Before discussing the CCP estimator in detail, we consider the form of $V_j(H_t, p(H_t))$ for a simple optimal stopping model to illustrate the content of Proposition 1. We preface the empirical investigations contained in the second half of the paper, by supposing a couple must decide when, over the course of their lifetime, to permanently sterilize and no longer be at risk to bear children. Prior to sterilizing, births occur according to (an exogenous) stochastic process. Each period, the couple receives a level of utility which depends on the number of their offspring; this payoff reflects the balance between the satisfaction derived from and the costs associated with rearing these children. The example ignores other forms of heterogeneity across couples; hence, \mathcal{B}_0 can be ignored in this illustration.

In terms of the notation developed above, let $d_{t1} = 1$ if the couple does not sterilize in period t and $d_{t2} = 1$ if it sterilizes. Because $J = 2$, $d_{t2} = 1 - d_{t1}$. Since sterilization is assumed irreversible and available at any t , every history H_t is terminal; thus $d_{t1} = 0$ implies $d_{s1} = 0$ for all $s \in \{t, \dots, T\}$. The period of the couple's lifetime in which sterilization takes place, τ , is called the stopping time (for childbearing).

The outcomes in this model are births. We let $b_s = 1$ if a child is born in period s , and let $b_s = 0$ otherwise. For simplicity, we assume a birth occurs at the end of period t to unsterilized women with probability $\alpha \in (0, 1]$. That is:

$$F_1((H_t, 1) | H_t) = \alpha. \quad (4.1)$$

Sterilized women cannot bear children, so:

$$F_2((H_t, 1) | H_t) = 0. \quad (4.2)$$

We assume that the nonstochastic component of the woman's utility in any period t depends only on the number of existing children. Consequently, this problem has a finite state space; the couple only has to keep track of periods since their marriage and current family size in making contraception decisions. Let \tilde{H}_t denote family size at t . In terms of the couple's birth history $\tilde{H}_t = H'_t \iota_t$, where ι_t denotes a $t \times 1$ vector of ones, or

more simply:

$$\tilde{H}_t \equiv \sum_{s=1}^t b_{t-s}. \quad (4.3)$$

For the sake of illustration, let the couple's current utility be quadratic in \tilde{H}_t :

$$u_1^*(H_t) = u_2^*(H_t) = \beta'(\delta_1 \tilde{H}_t + \delta_2 \tilde{H}_t^2) \quad (4.4)$$

where $\beta \in (0, 1)$ is the discount factor. The choice specific idiosyncratic component associated with each action, ε_{ij} , is assumed to be identically and independently distributed across (t, j) as a Type I Extreme Value random variable with location parameter 0. The couple's decision problem, then, is to sequentially choose $\{d_s\}_{s=0}^T$ (or, equivalently the optimal stopping time τ) which maximizes:

$$E_0[\sum_{t=0}^T \beta^s (\delta_1 \tilde{H}_s + \delta_2 \tilde{H}_s^2 + d_s \varepsilon_{s1} + (1 - d_s) \varepsilon_{s2})]. \quad (4.5)$$

To characterize the optimal decision rule for this model, we need the conditional valuation functions of each action $j \in \{1, 2\}$. Because sterilization is a terminating action, the value of setting $d_t = 0$ is just the expected discounted utility derived from the stock of existing children, \tilde{H}_t . The above assumptions imply this value is:

$$v_2(H_t) = \beta'(\gamma + \delta_1 \tilde{H}_t + \delta_2 \tilde{H}_t^2)/(1 - \beta^{T-t}), \quad (4.6)$$

where γ is Euler's constant (≈ 0.577). The discounted value of not sterilizing in period t , and remaining fertile at least one more peiod is:

$$v_1(H_t) = \max_{\{d_s\}_{s=t+1}^T} E[\sum_{s=t+1}^T \beta^s (\delta_1 \tilde{H}_s + \delta_2 \tilde{H}_s^2 + d_s \varepsilon_{s1} + (1 - d_s) \varepsilon_{s2}) | H_t]. \quad (4.7)$$

Aside from ε_{t1} and ε_{t2} , the only difference in expected future utility from the two actions the parents can take is due to the value of births. Therefore, the optimal decision rule is:

$$d_t = \begin{cases} 0, & \text{if } \varepsilon_{t1} - \varepsilon_{t2} \geq \beta^{-t} v_1(H_t) - \beta^{-t} v_2(H_t) \\ 1, & \text{if } \varepsilon_{t1} - \varepsilon_{t2} < \beta^{-t} v_1(H_t) - \beta^{-t} v_2(H_t) \end{cases} \quad (4.8)$$

where $\beta^{-t} v_j(H_t)$ is the current (undiscounted) conditional valuation function for history H_t and action j . This implies the conditional probability of choosing not to sterilize in period t is:

$$p_1(H_t) = \{1 + \exp [\beta^{-t} v_1(H_t) - \beta^{-t} v_2(H_t)]\}^{-1}. \quad (4.9)$$

Because sterilization can be undertaken at any time, all histories are terminal; hence, the representations of $v_1(H_t)$ and $v_2(H_t)$ that we seek take the form of (3.14) and (3.13), respectively. Since the right-hand side of (4.6) already corresponds to the form of V_{t2} in (3.13), we only need to derive the expression for $V_1(H_t, p(H_t))$. To proceed, note that $\mathcal{A}_{t+1}(H_t) = \{(H_t, 1), (H_t, 0)\}$ and the associated set of conditional choice probabilities is:

$$p(H_t) = \{p_1(H_t, 0), p_1(H_t, 1)\}. \quad (4.10)$$

Using (4.9), it follows from Proposition 1 that $Q^{-1}(p_1(H_{t+1}), H_{t+1})$ is:

$$Q^{-1}(p_1(H_{t+1}), H_{t+1}) = \beta^{t+1} \ln [p_1(H_{t+1})/p_2(H_{t+1})]. \quad (4.11)$$

To complete the expression for $V_1(H_t, p(H_t))$, we need to characterize the form of the $W_{t+1,j}$ functions associated with U_{t+1} in (3.8). Given the assumed distributions for ε_{t1} and ε_{t2} , these functions take the form:

$$\begin{aligned} W_j(p_{t+1}, H_{t+1}) &= E(\varepsilon_{t+1,j} | H_{t+1}, d_{t+1,j}^0 = 1) \\ &= E(\varepsilon_{t+1,j} | H_{t+1}, \beta^{t+1} \varepsilon_{t+1,1} - \beta^{t+1} \varepsilon_{t+1,2} > v_j(H_{t+1}) - v_k(H_{t+1})) \\ &= E(\varepsilon_{t+1,j} | H_{t+1}, \varepsilon_{t+1,j} > \varepsilon_{t+1,k} + \ln \{p_j(H_{t+1})/[1 - p_j(H_{t+1})]\}) \\ &= \gamma - \ln [p_j(H_{t+1})], \end{aligned} \quad (4.12)$$

for $j = 1, 2$, where γ is Euler's constant. Substituting (4.4), (4.6), (4.11) and (4.12) into (3.14) yields:

$$\begin{aligned} V_1(p(H_t), H_t) = & \beta^t \alpha \{ \delta_1(\tilde{H}_t + 1) + \delta_2(\tilde{H}_t + 1)^2 + \sum_{j=1}^2 p_j(H_t, 1)(\gamma - \ln [p_j(H_t, 1)]) \\ & + \beta[\gamma + \delta_1(\tilde{H}_t + 1) + \delta_2(\tilde{H}_t + 1)^2]/(1 - \beta^{T-t-1}) \\ & + p_1(H_t, 1)(\ln [p_1(H_t, 1)/p_2(H_t, 1)]) \} \\ & + \beta^t(1 - \alpha)\{\delta_1\tilde{H}_t + \delta_2\tilde{H}_t^2 + \sum_{j=1}^2 p_j(H_t, 0)(\gamma - \ln [p_j(H_t, 0)]) \\ & + \beta[\gamma + \delta_1\tilde{H}_t + \delta_2\tilde{H}_t^2]/(1 - \beta^{T-t-1}) \\ & + p_2(H_t, 0)(\ln [p_1(H_t, 0)/p_2(H_t, 0)])\}. \end{aligned} \quad (4.13)$$

This function consists of two expressions in braces: the one in first three lines gives the expected lifetime utility from period $t+1$ on if a child is born in period t and $H_{t+1} = (H_t, 1)$, weighted by the probability that such a birth will occur; the second, in the final three lines, gives the probability-weighted utility associated with the birth not occurring and $H_{t+1} = (H_t, 0)$. Each of these expressions, in turn, consists of the sum of: (a) the expected payoff in period $t+1$, U_{t+1} ; (b) the value of sterilizing at $t+1$, V_2 ; and (c) a term which adjusts for the fact that sterilization may not be optimal in $t+1$.⁶

Using $V_1(p(H_t), H_t)$ in place of $V_1(H_t)$ and the expression in (4.6) for $V_2(H_t)$, one can represent the conditional probability of choosing either action as a function of the couple's history, H , and the one period ahead choice probabilities $p(H_t)$. Consequently $p(H_t)$ and $v_2(H_t)$ are sufficient to summarize the expected future value of an action in period t . Provided we can obtain consistent estimates of the future choice probabilities cheaply, the representation developed here can be used to formulate estimators for α , β , δ_1 and δ_2 , the structural parameters of interest. We turn to the issue of estimation in the next section.

5. LARGE SAMPLE PROPERTIES OF THE CCP ESTIMATOR

This section addresses the issue of estimating structural parameters for the class of models described in Section 2 which exploit the representation of valuation functions developed in Section 3. We suppose that $y^*(H_t)$, $G(\varepsilon_t | H_t)$ and $F_j(H_{t+1} | H_t)$ are parameterized by a vector of structural parameters, $\theta_0 \in \Theta$, and propose a strategy for its estimation. The CCP estimator is obtained in two stages: we first formulate nonparametric estimators of future choice and transition probabilities and then use these incidental parameter estimates in a set of estimating equations which are solved for θ . We establish that the estimator is consistent, converges at a rate of $N^{1/2}$, and has a normal asymptotic distribution.

Suppose the model developed in Section 2 characterizes the behaviour of a population of agents from which we draw a random sample of size N at some point in calendar time, t say.⁷ We utilize an n subscript throughout to denote the variables and functions for the n^{th} agent in the sample and define A_{nt} to be the age of agent n at calendar time t (which implies her planning horizon is $T - A_{nt}$ periods). In developing the properties of estimators, we make the following additional assumptions:

Assumption 1. θ_0 is a $Q \times 1$ vector belonging to the interior of a compact set Θ .

Assumption 2. $y^*(H_{nt}, \theta)$, $F_j(H_{n,t+1} | H_{nt}, \theta)$ and $dG(\varepsilon_{nt} | H_{nt}, \theta)$ are differentiable in θ .

6. In (4.13), the (a) components for $(H_t, 1)$ and $(H_t, 0)$ are given in the first and third lines of (4.13), respectively, and the sum of corresponding (b) and (c) components are given in the other lines.

7. Samples of longitudinal data could also be treated in a similar manner; no new conceptual issues arise in that case.

Assumption 3. b_{ns} takes on one of K possible values when $t - A_{nt} < s \leq t + T - A_{nt}$.

Assumption 4. The elements in H_{nt} and d_{nt}^0 , the choices made at age A_{nt} , are observed without error for each agent, but ξ_{nt} is unobserved.

Assumption 5. The population lives in a stationary environment; consequently $F_j(H_{n,t+1} | H_{nt}, \theta_0)$, the transition probability generating b_{nt} and $G(\xi_{nt} | H_{nt}, \theta_0)$, the probability distribution governing the unobservables, are invariant across calendar time periods.

Assumption 6. Conditional on H_{nt} , the distribution ξ_{nt} is independently distributed across agents.

Assumptions 1 and 2 are regularity conditions on the functions and parameter space which enable the use of standard results when establishing consistency and asymptotic distributions. Assumption 3 restricts the feasible outcomes for b_{nt} to a finite set and Assumptions 4 through 6 allow us to (synthetically) form cohorts from cross-sectional data on agents (of different ages) which we use to estimate the future choice and transition probabilities.

To motivate the definition of the CCP estimator and help fix ideas, we begin by considering the (somewhat unrealistic) situation where the conditional choice and transitional probabilities entering the estimation equations are known. In this case θ_0 could be estimated by a Generalized Methods of Moments (GMM) strategy, using the representation of the conditional valuation functions developed in Section 3. Accordingly, we augment $p(H_{nt})$ to include the future (feasible) transition probabilities. Let this set be denoted by $\rho(H_{nt})$ and defined as:

$$\rho(H_{nt}) = \rho(H_{nt}) \cup \{F_1(H_{n,s+1} | H_{ns}), \dots, F_J(H_{n,s+1} | H_{ns}): H_{ns} \in \mathcal{A}_s(H_{nt}) \text{ for } s = t, \dots, T\}. \quad (5.1)$$

Also, let $V_j(H_{nt}, \rho(H_{nt}), \theta_0) \equiv V_j(H_{nt}, \rho(H_{nt}))$. Defining the corresponding representation of conditional choice probabilities as:

$$P_k(H_{nt}, \rho(H_{nt}), \theta) = \Pr \{k = \operatorname{argmax}_{j \in \mathcal{C}} [u_j^*(H_{nt}, \theta) + \varepsilon_{nj} + V_j(H_{nt}, \rho(H_{nt}), \theta)] | H_{nt}\}, \quad (5.2)$$

it follows immediately that:

$$P_j(H_{nt}, \rho(H_{nt}), \theta_0) = p_j(H_{nt}). \quad (5.3)$$

Let z_{nt} denote an $R \times 1$ vector of instruments, with $(J-1)R \geq Q$, and let α_N be a $Q \times (J-1)R$ random matrix which converges in N to the constant matrix α_0 . Define the $Q \times 1$ vector of orthogonality conditions as:

$$\pi_{1n}(\theta, \rho(H_{nt})) = \alpha_N(z_{nt} \otimes [d_{nt} - P(H_{nt}, \rho(H_{nt}), \theta)]). \quad (5.4)$$

where $P(H_{nt}, \rho(H_{nt}), \theta) \equiv (P_1(H_{nt}, \rho(H_{nt}), \theta), \dots, P_{J-1}(H_{nt}, \rho(H_{nt}), \theta))'$. The instruments z_{nt} are chosen so that θ_0 uniquely satisfies

$$E(z_{nt} \otimes [d_{nt} - P(H_{nt}, \rho(H_{nt}), \theta)]) = 0. \quad (5.5)$$

Under the assumption that $\rho(H_{nt})$ is known, a GMM estimator for θ_0 , denoted $\hat{\theta}_0^{(N)}$, can be obtained by constructing sample analogues of (5.4) and averaging them over $n \in \{1, \dots, N\}$.

Typically $\rho(H_{nt})$ is unknown and must be estimated. We consider two cases, depending on the nature of the initial conditions, b_{n0} . First, consider the case where \mathcal{B}_0 is a finite set with L elements. Recall that the set of feasible outcomes, \mathcal{B} , is also assumed to be finite (Assumption 3); thus $\mathcal{B}_0 \times \mathcal{B}^T$, the set of possible histories for H_{nt} , is itself finite. Consequently, there are only a finite number of transition probabilities to consider too, each of which has discrete support. In particular, $M \equiv L(K^{T+1}-1)(K-1)^{-1}$ such histories exist, generically denoted $H_m \in \mathcal{B}_0 \times \mathcal{B}^T$ for $m = \{1, \dots, M\}$. Accordingly, let $\underline{\rho}_m \equiv (\underline{\rho}'_m, F'_m, \dots, F'_{mJ})'$, where $\underline{\rho}_m \equiv p_m(H_m)$ denotes the $(J-1)$ -dimensional vector of associated choice probabilities and $\tilde{F}_{mj} \equiv F_j((H_m, b) | H_m)$ is the $(K-1)$ -dimensional vector of probabilities for the respective outcomes, conditional on history H_m and choice $j \in \{1, \dots, J\}$. Then one can use sample frequencies of the choices and subsequent outcomes from observations for each of the M possible histories to form estimates of $\underline{\rho} = (\underline{\rho}'_1, \dots, \underline{\rho}'_M)'$. More formally, define the indicator function, $I_1(H_m, H_{nt})$, for any two histories H_m and H_{nt} , as:

$$I_1(H_m, H_{nt}) = \begin{cases} 1 & \text{if } H_{nt} = H_m \\ 0 & \text{otherwise.} \end{cases} \quad (5.6)$$

Then unrestricted estimators of $\underline{\rho}_m$ and F_{mj} are defined, respectively, as:

$$\underline{\rho}_m^{(N)} = \frac{\sum_{n=1}^N I_1(H_m, H_{nt}) d_{nt}}{\sum_{n=1}^N I_1(H_m, H_{nt})} \quad \text{and} \quad \underline{F}_{mj}^{(N)} = \frac{\sum_{n=1}^N I_1(H_m, H_{nt}) d_{nj} b_{nt}}{\sum_{n=1}^N I_1(H_m, H_{nt}) d_{nj}}. \quad (5.7)$$

Let $\underline{\rho}^{(N)}$ denote the $M(JK-1)$ dimensional vector of estimators $(\underline{\rho}_1^{(N)}, \dots, \underline{\rho}_M^{(N)})'$. Noting $\rho(H_{nt}) \subseteq \{\underline{\rho}_m\}_{m=1}^M$ for all $n \in \{1, 2, \dots, N\}$, an estimator of θ_0 may be constructed using elements of $\underline{\rho}^{(N)}$ instead of $\rho(H_{nt})$ in forming the sample average of (5.4) and setting it to 0.

The large sample properties of this estimator, which we call the CCP estimator and denote by $\theta_1^{(N)}$, follow from the fact that $\underline{\rho}^{(N)}$ is itself a GMM estimator. Define the $M(JK-1)$ -dimensional vector:

$$\pi_{2n}(\underline{\rho}) = (\pi_{2n}(\underline{\rho}_1)', \pi_{2n}(\underline{\rho}_2)', \dots, \pi_{2n}(\underline{\rho}_M)')'.$$

Where the $(JK-1)$ dimensional vector, $\pi_{2n}(\underline{\rho}_m)$, is defined as:

$$\pi_{2n}(\underline{\rho}_m) = \begin{bmatrix} (d_{nt} - \underline{\rho}_m) \\ (b_{nt} - \tilde{F}_{m1}) d_{nt1} \\ \vdots \\ (B_{nt} - \tilde{F}_{mJ}) d_{ntJ} \end{bmatrix} I_1(H_m, H_{nt}), \quad (5.8)$$

for $m = \{1, \dots, M\}$. Then $(\underline{\rho}^{(N)}', \theta_1^{(N)})'$ is a GMM estimator solving:

$$N^{-1} \sum_{n=1}^N \begin{bmatrix} \pi_{1n}(\theta, \rho(H_{nt})) \\ \pi_{2n}(\underline{\rho}) \end{bmatrix} = 0. \quad (5.9)$$

From Theorems 2.1 and 3.1 in Hansen (1982, p. 1035 and p. 1042), the consistency and asymptotic normality of $\theta_1^{(N)}$ follows immediately. Defining:

$$\Omega_{ij} \equiv E(\pi_{in} \pi'_{jn})$$

$$\Gamma_{i1} \equiv E(\partial \pi_{in} / \partial \theta)$$

$$\Gamma_{22} \equiv E(\partial \pi_{2n} / \partial \underline{\rho}),$$

for $i \in \{1, 2\}$ and $j \in \{1, 2\}$, it follows (from Newey (1984), for example) that Σ_1 , the asymptotic covariance matrix for $\hat{\theta}_1^{(N)}$, is

$$\Sigma_1 = \Gamma_{11}^{-1}(\Omega_{11} + \Gamma_{21}(\Omega_{22} + \Omega_{21} + \Omega_{12})\Gamma_{21}')\Gamma_{11}^{-1}. \quad (5.11)$$

There are two computational drawbacks associated with the estimator defined in (5.9). First, the cell estimators in (5.7) are infeasible if \mathcal{B}_0 is a closed interval which supports a well defined probability density function for b_{n0} , because the probability of sampling two identical histories is 0. Second, even when \mathcal{B}_0 is finite, a strictly positive probability exists that not all cells are visited for any fixed sample size N . As a practical matter, this is manifested by empty or sparse cells which may render the estimator defined above non-operational.

To overcome these two limitations, nonparametric procedures can be used to estimate the incidental parameters, $\rho(H_{nt})$, and, ultimately, θ_0 . Accordingly, define the kernel function, $J(H_{nt})$, by a real-valued, bounded, symmetric differentiable function which integrates to 1 on the set of all possible histories. Also let $h_N \in \mathbb{R}^+$ denote the band-width associated with the kernel estimator for samples of size N . For each $H_{nt}^{(r)} \in \mathcal{A}(H_{nt})$, the kernel estimators for $p_{nr} \equiv p(H_{nt}^{(r)})$ and $F_{njr} \equiv F_j((H_{nt}^{(r)}, b) | H_{nt}^{(r)})$ are given by:

$$\begin{aligned} \hat{p}_{nr}^{(N)} &= \left[\sum_{m=1, m \neq n}^N d_{mt} J\left(\frac{H_{mt} - H_{nt}^{(r)}}{h_N}\right) \right] \left[\sum_{m=1, m \neq n}^N J\left(\frac{H_{mt} - H_{nt}^{(r)}}{h_N}\right) \right]^{-1} \\ \hat{F}_{njr}^{(N)} &= \left[\sum_{m=1, m \neq n}^N d_{mt} b_{nr} J\left(\frac{H_{mt} - H_{nt}^{(r)}}{h_N}\right) \right] \left[\sum_{m=1, m \neq n}^N d_{mt} J\left(\frac{H_{mt} - H_{nt}^{(r)}}{h_N}\right) \right]^{-1}. \end{aligned} \quad (5.12)$$

Theorems 1 and 3 of Bierens (1983, p. 701–702) establish the conditions, including that $h_N \rightarrow 0$ and $N^{1/2}h_N^k \rightarrow \infty$ as $N \rightarrow \infty$, under which $\hat{\rho}_{nr}^{(N)} \equiv (\hat{p}_{nr}^{(N)}, \hat{F}_{n1r}^{(N)}, \dots, \hat{F}_{nJr}^{(N)})$ is uniformly consistent for each (n, r) .⁸

Substituting $\hat{\rho}_{nr}^{(N)}$ for the corresponding components of $\rho(H_{nt})$ in (5.4), another variant on the CCP estimator of θ_0 , denoted $\hat{\theta}^{(N)}$, is obtained by, once again, forming the sample analogue of (5.4). Appendix B establishes that $\hat{\theta}^{(N)}$ is consistent, and that $N^{1/2}(\hat{\theta}^{(N)} - E[\hat{\theta}^{(N)}])$ is asymptotically normal but is not centred on 0. This asymptotic bias is due to the local averaging in the kernel estimator for the incidental parameters ($\hat{\rho}_{nr}^{(N)}$) used to construct $\hat{\theta}^{(N)}$. However, Appendix B also shows how to form a linear combination of estimators of this form for θ_0 to obtain a new estimator, $\hat{\theta}_2^{(N)}$, which is $N^{1/2}$ consistent, asymptotically normal and unbiased. Proposition 2 summarizes the results of this section and Appendix B.

Proposition 2. *If \mathcal{B}_0 is finite, $\hat{\theta}_1^{(N)}$, given by (5.9), is consistent and $N^{1/2}(\hat{\theta}_1^{(N)} - \theta_0)$ converges in distribution to a normal random variable with mean 0 and covariance matrix, Σ_1 , defined in (5.11). If \mathcal{B}_0 is compact (but not finite), then there exists a consistent estimator, $\hat{\theta}_2^{(N)}$, defined in (B.27), such that $N^{1/2}(\hat{\theta}_2^{(N)} - \theta_0)$ converges in distribution to a normal random variable with mean 0 and covariance matrix, Σ_2 , given in (B.21) of Appendix B.*

The assumptions invoked to establish the asymptotic properties of the CCP estimator limit its applicability to certain dynamic discrete choice contexts. First, consider the situation where some of the state variables are not observed. Although Assumption 4 rules out the existence of such unobserved state variables, it is quite natural for such

8. As noted above, one may wish to use the kernel estimators in (5.12), even in the case where all of the elements of H_{nt} finite-valued random variables. While the large sample properties of these choice and transition probability estimators still holds. We have not investigated their small sample properties. (See Hotz, Miller, Sanders and Smith (1991) for an investigation of the small sample properties of a closely related estimator.)

variables to play an essential role in models of interest. For example, in the job matching model estimated by Miller (1982, 1984), the beliefs an agent holds about the quality of his match are unobserved. Determined by nature and past choices, they are used to evaluate future prospects. In this model, the estimated choice probabilities cannot be conditioned on all the variables that help determine the agent's future decisions and their conditional valuation functions. One could, following Miller, solve the dynamic programming problem explicitly, derive the optimal decision rule to generate the stochastic process that characterizes the unobservables, and ultimately undertake ML estimation.

Assumptions 5 and 6 imply there are no common shocks which would produce, over time, correlated responses within the population of agents. Consequently many interesting questions about secular change, technological progress, and business cycles fall outside the scope of the CCP estimator as proposed herein. However, Altug and Miller (1991) recently have adapted our approach to a competitive economy where aggregate fluctuations are transmitted through prices which affect labour supply decisions and human capital accumulation. In principle, one could allow for such forms of aggregate variation by explicitly modeling the processes generating such aggregate processes and then estimating the resulting model with ML techniques. This approach has not been taken due to the substantial computational costs its implementation would entail.

Finally, the CCP estimator requires samples to be drawn from the population of all possible histories, H_m , in order to (nonparametrically) estimate the choice and transition probabilities in $\rho(H_m)$. (More formally, random sampling of agents at a point in calendar time coupled with Assumptions 5 and 6, ensure that all feasible histories have a positive probability of being included in any particular sample.) This requirement is yet another way in which the CCP estimator is, in principle, less versatile than ML. Because the latter method computes the optimal decision rule, it may be theoretically possible to parametrically identify dynamic models with data sets that only track the first few periods of the decision maker's problem. As with unobserved heterogeneity, the usefulness of pursuing ML in these estimation environments depends on the confidence one can place in specifying the structure of unobserved phenomena, in this case choices and outcomes occurring near the end of the decision tree never seen in the data set.

6. AN EMPIRICAL MODEL OF CONTRACEPTIVE CHOICE AND VOLUNTARY STERILIZATION

The remainder of the paper applies the preceding results to a dynamic model of contraceptive choice for a sample of white married couples surveyed in the National Fertility Survey of 1975 (NFS). The model explicitly deals with the option of voluntary sterilization, generalizing the one in Section 4. In recent years voluntary sterilization has become the most common method of family planning within the U.S.A. For example, among married women between the ages of 25 to 34 in the U.S., Potts (1988) reports that 40 percent rely on sterilization as their contraceptive method, which is twice as many as use the next most common contraceptive method, the pill.

While the model outlined in Section 4 captures the optimal stopping aspects of family formation, it lacks several features of the contraceptive choice decisions of married couples. First, it does not allow for any contraceptive control apart from sterilization. Yet temporary methods of contraceptive control are important as our data clearly show. Second, it does not explicitly characterize the structure of the payoffs parents derive from their children. For example, previous economic models of life choice fertility distinguish between the utility parents derive from the presence of children as well as the (economic)

costs parents face in rearing them. Moreover, while the model in Section 4 specified the parental utility depended only on the existing number of offspring, empirical evidence suggests that the utility parents receive from their children (see Hotz and Miller (1988)) and the costs associated with rearing them (see Espenshade (1980) and Lazear and Michael (1988)) depend on their ages as well as their total number.

The empirical application presented here allows for these features within a dynamic structural model of contraceptive choice. We estimate a model in which parents value the direct utility received from their children and from their own consumption. Their consumption of these two goods is constrained by their (limited and uncertain) income, the costs of rearing their own children, and their inability to control perfectly the extent and timing of childbearing. These features of the choice problem facing parents, as well as the possibility that utility from offspring and their costs vary with the number and age distribution of their children, would suggest that parents adapt their choice of contraceptive methods to the outcomes they have realized in previous periods and in anticipation of the future consequences of current actions.

In the model we estimate, the n^{th} couple can choose one of the following three contraceptive actions: voluntary sterilization, temporary contraception (such as the pill), and no contraception. That is:

$$\begin{aligned} d_{nt1} &= \begin{cases} 1 & \text{if the } n^{\text{th}} \text{ couple does not contracept at age } t \\ 0 & \text{otherwise,} \end{cases} \\ d_{nt2} &= \begin{cases} 1 & \text{if the } n^{\text{th}} \text{ couple contracepts at age } t \\ 0 & \text{otherwise,} \end{cases} \\ d_{nt3} &= \begin{cases} 1 & \text{if the } n^{\text{th}} \text{ couple sterilizes at age } t \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (6.1)$$

Contraceptive methods 1 and 2 only imperfectly control births. For these two choices, the (transition) probabilities characterizing the occurrence of a birth are assumed to depend on the mother's education, denoted E_{1n} , her age, A_{1nt} , and the contraceptive method used. That is, for $j \in \{1, 2\}$:

$$F_j(H_{n,t+1} | H_{nt}) = F_j(H_{n,t+1} | E_{1n}, A_{1nt}), \quad (6.2)$$

where, as in Section 4, $H_{n,t+1} = (H_{nt}, b_{nt})$ and $b_{nt} \in \{0, 1\}$ is an indicator variable for births to the n^{th} when the mother is A_{1nt} years old. We continue to maintain the assumption that sterilization is fully effective and terminating.⁹

Each period the couple receives a level of utility depending on their stock of children, their own consumption, and which contraceptive method they used. Let a_{nt} denote the service flow couple n receives from their children when the mother is A_{1nt} years old, and let c_{nt} denote parental consumption. We assume this payoff at age A_{1nt} is a linear function of each of these goods, plus an additive component measuring the utility specifically associated using contraceptive method j . These choice specific costs are assumed to depend upon a quadratic function of the mother's age, A_{1nt} , and her education, E_{1n} , as well as an independent (across (n, t, j)) random disturbance, ε_{ntj} , drawn from a Type I Extreme Value distribution with a zero mean. That is:

$$u_{ntj} = \beta' [a_{nt} + c_{nt} - \psi_1 a_{nt}^2 - \psi_2 c_{nt}^2 + \mu_{j0} + A_{1nt} \mu_{j1} + \mu_{j2} + E_{1n} \mu_{j3} + \varepsilon_{ntj}], \quad (6.3)$$

9. Thus $F_3(H_{n,t+1} | E_{1n}, A_{1nt}) = F_3(H_{n,t+2} | E_{1n}, A_{1n,t+1}) \dots = 0$ if $d_{nt3} = 1$. See Altug and Miller (1991) and Hotz, Miller, Sanders and Smith (1991) for structural applications of estimation strategies that avoid backwards induction by exploiting Proposition 1 but do not rely on terminal states.

where β is the subjective discount factor, (ψ_1, ψ_2) are curvature parameters characterizing parental preferences for consumption and service flows from children, and $\mu_j = (\mu_{j0}, \dots, \mu_{j3})$, are parameters characterizing the couple's preferences over the specific contraceptive methods for $j \in \{1, 2\}$; without loss of generality, we normalize μ_3 to 0. We further assume that the service flow from children, a_{nt} , is a linear function of past births:

$$a_{nt} = \sum_{s=0}^t \eta_s b_{n,t-s}, \quad (6.4)$$

where the parameter, η_s , is the service flow a child of age s yields. (We assume that the service flows from children aged 21 years and older are the same.)

We assume the couple faces a per period budget constraint, in which they must allocate their income in period t , y_{nt} , between their own consumption and the costs of rearing their children. The budget constraint is:

$$c_{nt} + e_{nt} = y_{nt}, \quad (6.5)$$

where e_{nt} denotes total expenditures on children in period t . Expenditures on children are assumed to depend on family structure in the following way:

$$e_{nt} = \sum_{s=0}^{20} \omega_s b_{n,t-s}, \quad (6.6)$$

where ω_s is the (exogenously-determined) level of expenditures required for a child of age s . (We assume that parents only make these expenditures for children who are less than 21 years of age.)

Parental income is assumed to be exogenously determined but stochastic. We suppose that y_{nt} is a function of the husband's age and education, denoted by A_{2nt} and E_{2n} , respectively, and a conditionally independent stochastic component, ξ_{nt} . More precisely, we use the following specification:

$$y_{nt} = \gamma_0 + A_{2nt}\gamma_1 + A_{2t}^2\gamma_2 + E_{2n}\gamma_3 + E_{2n}A_{2nt}\gamma_4 + E_{2n}A_{2nt}^2\gamma_5 + \xi_{nt}, \quad (6.7)$$

where $\gamma = (\gamma_0, \dots, \gamma_5)'$ is a parameter vector to be estimated, and ξ_{nt} is a zero mean and is uncorrelated over (n, t) .¹⁰ In our empirical analysis, we use husband's income to measure y_{nt} .¹¹

Substituting results from (6.4) through (6.7) into (5.3), we obtain:

$$\begin{aligned} u_{nj} = & \beta \{ \sum_{i=0}^t \eta_i b_{n,t-i} + \underline{x}'_{2nt} \gamma + \xi_{nt} - \sum_{i=0}^{20} \omega_i b_{n,t-i} + \psi_1 (\sum_{i=0}^{21} \eta_i b_{n,t-i})^2 \\ & - \psi_2 (\underline{x}'_{2nt} \gamma + \xi_{nt} - \sum_{i=0}^{20} \omega_i b_{n,t-i})^2 + \underline{x}'_{1nt} \mu_j + \varepsilon_{nj} \}, \end{aligned} \quad (6.8)$$

where the vectors, \underline{x}_{1nt} and \underline{x}_{2nt} , are:

$$\begin{aligned} \underline{x}_{1nt} &= (1, A_{nt}, A_{nt}^2, E_{1n})' \\ \underline{x}_{2nt} &= (1, A_{2nt}, A_{2nt}^2, E_{2n}, E_{2n}A_{2nt}, E_{2n}A_{2nt}^2)'. \end{aligned} \quad (6.9)$$

10. The failure to allow for serial correlation in ξ_{nt} conditional on H_{nt} , is clearly a limitation of the model. Nevertheless, the estimation strategy we have developed in this paper can deal with certain types of serial correlation in the specification of (6.7). If, for example, one were to adopt an AR(1) process to allow for serial correlation in the earnings process disturbances, two modifications would need to be made to the model presented in the text. First, because H_{nt} would contain lagged values of y_{nt} , the conditional expectations of the payoffs in future periods, $E(u_{nsj}|H_{ns})$ for $s > t$, would be now a function of $y_{n,t-1}$. Second, because y_{nt} would depend on $y_{n,t-1}$ (as well as a set of exogenous forcing variables), one would need to (non-parametrically) estimate a transition probability function for earnings, $F(y_{nt}|y_{n,t-1}, \underline{x}_{2nt})$, in addition to the transition functions for birth occurrences given in (6.2). Lacking adequate data on lagged earnings in the National Fertility Survey, we could not pursue the estimation of such a specification.

11. Restricting family income to the husband's earnings was done for several reasons. First, because of data limitations in the NFS, we only had measures of the earnings of the husband and wife; respondents were not asked about other sources of income. Second, we did not include the earnings of the wife's income in our measures of the couple's income, because of our concern that it is endogenously determined and intimately related to the childbearing decisions of the couple. In an earlier version of this paper, we presented results for a more elaborate structural model in which the wife's labour supply and labour earnings, were explicitly modeled along with the couple's contraceptive choices. These are available from the authors upon request.

From the description of this model, it is straightforward to see that $H_{nt} = (b'_{n0}, b_{n1}, \dots, b_{n,t-1})'$ where $b_{n0} = (t - A_{1nt}, t - A_{2nt}, E_{1n}, E_{2n})'$ and the vector of structural parameters we seek to estimate is given by:

$$\boldsymbol{\theta}_0 = (\beta, \psi_1, \psi_2, \eta_1, \dots, \eta_{21}, \mu'_1, \mu'_2, \omega_1, \dots, \omega_{20}, \gamma')'. \quad (6.10)$$

The couple choose a sequence of contraceptive actions to maximize the expected value of the sum of the per period payoffs, calculated using (6.8). To formulate estimating equations, note that the period t conditional choice probabilities for this model, expressed as functions of period $t+1$ choice probabilities and the couple's history, take the following form:

$$\begin{aligned} P_k(H_{nt}, \rho(H_{nt}), \boldsymbol{\theta}_0) \\ = \exp [x'_{1nt} \mu_k + \beta^{-t} V_k(H_{nt}, \rho(H_{nt}), \boldsymbol{\theta}_0)] \\ \times \{\sum_{j=1}^3 \exp [x'_{1nt} \mu_j + \beta^{-t} V_j(H_{nt}, \rho(H_{nt}), \boldsymbol{\theta}_0)]\}^{-1}, \end{aligned} \quad (6.11)$$

for $k \in \{1, 2\}$. Moreover, utilizing the expressions in (4.11) and (4.12), the difference in the conditional valuation function for $k \in \{1, 2\}$ and its counterpart for the terminating action (sterilization) takes the form:

$$\begin{aligned} \beta^{-t} V_k(\rho(H_{nt}), H_{nt}, \boldsymbol{\theta}_0) - \beta^{-t} V_3(\rho(H_{nt}), H_{nt}, \boldsymbol{\theta}_0) \\ = \lambda_0 F_{ntk} + \sum_{i=1}^{21} b_{n,t-i-1} F_{ntk} + \sum_{i=2}^{26} \lambda_i x_{2n,t-i} F_{ntk} + x'_{1nt} \eta_k \\ - \sum_{j=1}^3 \{p_j(H_{nt}, 1) \ln [p_j(H_{nt}, 1)] - p_j(H_{nt}, 0) \ln [p_j(H_{nt}, 0)]\} F_{ntk} \\ + \sum_{j=1}^2 \left\{ p_j(H_{nt}, 1) \ln \left(\frac{p_j(H_{nt}, 1)}{p_3(H_{nt}, 1)} \right) - p_j(H_{nt}, 0) \ln \left(\frac{p_j(H_{nt}, 0)}{p_3(H_{nt}, 0)} \right) \right\} F_{ntk} \\ - \sum_{j=1}^3 \{p_j(H_{nt}, 1) \ln [p_j(H_{nt}, 1)]\} + \sum_{j=1}^2 \left\{ p_j(H_{nt}, 0) \ln \left(\frac{p_j(H_{nt}, 0)}{p_3(H_{nt}, 0)} \right) \right\} \end{aligned} \quad (6.12)$$

where the parameter vector, $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_{26})'$ is a non-linear function of $\boldsymbol{\theta}_0$, the structural parameters, defined:

$$\begin{aligned} \lambda_0 &= \sum_{s=0}^{20} \beta^s [(a_s - \psi_1 a_s^2) - (\psi_2 e_s + \psi_3 e_s^2) + (\gamma_0 + \gamma_2^s + \gamma_3^s)(2\psi_3 e_s)] + \frac{\beta^{21}}{1-\beta} (a_{21} - \psi_1 a_{21}^2), \\ \lambda_i &= - \sum_{s=1}^{20-i} \beta^s [2\psi_1 a_s a_{s+1} + 2\psi_3 e_s e_{s+1}] - \sum_{s=21-i}^{20} \beta^s (2\psi_1 a_s a_{21}) - \frac{\beta^{21}}{1-\beta} (2\psi_1 a_{21}^2), \\ &\quad \text{for } i \in \{1, \dots, 21\}, \end{aligned}$$

$$\begin{aligned} \lambda_{22} &= \sum_{s=1}^{21} \beta^{s-1} (\gamma_1 + \gamma_4^s + \gamma_5^s)(2\psi_3 e_{s-1}), \\ \lambda_{23} &= \sum_{s=1}^{21} \beta^{s-1} (\gamma_2 + 2\gamma_3^s)(2\psi_3 e_{s-1}), \\ \lambda_{24} &= \sum_{s=1}^{21} \beta^{s-1} \gamma_3 (2\psi_3 e_{s-1}), \\ \lambda_{25} &= \sum_{s=1}^{21} \beta^{s-1} (\gamma_4 + 2\gamma_5^s)(2\psi_3 e_{s-1}), \\ \lambda_{26} &= \sum_{s=1}^{21} \beta^{s-1} \gamma_5 (2\psi_3 e_{s-1}). \end{aligned} \quad (6.13)$$

Given values for the transition probabilities, the right-hand side of (6.12) is linear in the lagged birth indicators and the household characteristics. This feature lends itself

to the following four-stage strategy for estimating of θ_0 : (1) obtain (nonparametric) estimates of the probability transitions and the conditional choice probabilities; (2) using these estimates to form (6.11) and, in turn, the choice probabilities in (6.12), estimate (λ, μ_1, μ_2) , and their standard errors, with the methods developed in Section 5; (3) use (6.7) to estimate γ , the parameter's characterizing the parent's income process; and (4) use the estimates from stages (2) and (3) to estimate the remaining structural parameters in θ_0 by minimum distance methods. We adopted this strategy because it is relatively cheap to compute.

While the procedures for undertaking Steps 2 through 4 are either well known or previously discussed, a few details about Step 1 are in order. For each observation $n \in \{1, \dots, N\}$ estimates of conditional choice probabilities were obtained for two hypothetical histories, namely $(H_{nt}, 0)$ and $(H_{nt}, 1)$, which characterize the couple's position in period $(t+1)$ in the event of a birth or not, respectively. For convenience and without loss of generality, we recast the state space in terms of the educational attainment of both parents, their ages, and the successive ages of their progeny by birth parity in place of the original H_{nt} space. That is, we expressed the kernel function in terms of $H_{nt}^* = (H_{nt}^{(1)}, \dots, H_{nt}^{(I)})$, where $H_{nt}^{(1)} = E_{1n}, H_{nt}^{(2)} = E_{2n}, H_{nt}^{(3)} = A_{1nt}, H_{nt}^{(4)} = A_{2nt}, H_{nt}^{(5)}$ is the age of the oldest child, ..., and $H_{nt}^{(I)}$ is the age of the $(I-4)^{\text{th}}$ oldest, with $(I-4)$ being the largest number of children belonging to any one family at t in the sample. The kernel function we used in our empirical analysis was of the form:

$$J\left(\frac{H_{nt}^*}{h_N}\right) = \prod_{i=1}^I \phi\left(\frac{H_{nt}^{(i)}}{\sigma_i h_N}\right), \quad (6.14)$$

where $\phi(\cdot)$ is the standard normal density function, σ_i is the sample standard deviation of $H_{nt}^{(i)}$ and h_N is the bandwidth. The bandwidth actually used in our analysis was $h_N = 1$. Preliminary analyses conducted with alternative bandwidths did not reveal a great deal of sensitivity in the estimates of θ_0 to this choice.

7. THE DATA

The data, taken from the National Fertility Survey (NFS), is a sample of white couples who were married over the period 1970 through 1975 inclusive. They were interviewed twice, in 1970 and at the beginning of 1976, and information was gathered on the births of their offspring, other demographic characteristics, as well as information on the husband's annual labour market earnings for the years 1970 and 1975. They were also asked to provide monthly records of their contraceptive utilization over the six year period. We aggregated this data to form annualized measures of the contraceptive choices, classifying each couples according to one of the three contraceptive actions described in the previous section.¹² Of a total of 2374 couples interviewed in both 1970 and 1976, we used data on 2088. The sample loss is due to missing data on demographic characteristics, husband's income, fertility, or contraceptive histories. We then formed couple-year observations from those at risk to bear children (that is, couples who had not yet sterilized).

12. The couple's contraceptive choice in each calendar year was constructed as follows. If either partner reported sterilizing in any month during a calendar year t , we recorded their contraceptive choice as sterilization. If sterilization did not occur in year t and the couple recorded a birth in year $t+1$, we used their contraceptive strategy in the month prior to the wife becoming pregnant (that is whether they were contracepting or not) as their year t contraceptive choice. Couples who did not sterilize and did not have a birth in year $t+1$, were further categorized; if no birth occurred in year t , we assigned the contraceptive strategy they followed in the majority of the months in year t ; alternatively, if a birth occurred in year t , we assigned their year t contraceptive choice as contracepting, on the grounds that women are infertile during the post-partum period.

A total of 796 couples voluntarily sterilized at some time over the period 1970 through 1975, resulting in a total of 10,257 couple-years for use in our analysis. Descriptive statistics on this sample are provided in Table I. With respect to their contraceptive practices, in any given year, approximately 18% of those couples at risk used no form of contraception, 74% used temporary contraceptive methods and 8% chose to sterilize.

The relationship between contraceptive choices and the characteristics of fecund couples is suggested by cross-tabulations provided in Table II. The left-hand panel shows approximately 25% of wives under the age of 21 use no form of contraception, this percentage falling to around 17 for wives in their thirties, and then rising. The percentage who contracept is highest among those women in the 21–25 age bracket, while the proportion of couples who sterilize rises from 1·5% for the youngest age group, peaking at 12% for couples with wives aged 36–40.

The middle panel displays the distribution of contraceptive choice by educational attainment of the wife. Couples with wives who did not attain a high school degree have the highest proportions of either using no contraceptive methods or choosing to sterilize, relative to the other educational groups. The high incidence of sterilization among women with less education partly reflects the fact that women with lower levels of educational attainment started their childbearing earlier. Also note that a relatively high proportion of college graduates do not contracept; perhaps these women delayed childbearing until after completing their education.

Because the model suggests that family size affects contraceptive choice, the relationship between these two variables is illustrated in the third panel of Table II. As might be expected, the incidence of the use of both temporary contraceptive methods and

TABLE I
Descriptive statistics
(Standard deviations in parentheses)

Variable	Year					
	1970	1971	1972	1973	1974	1975
Wife's age	28.257 (5.880)	28.777 (5.799)	29.461 (6.517)	30.416 (5.814)	31.368 (5.898)	32.234 (5.904)
Wife's education	12.559 (2.018)	12.583 (1.997)	12.623 (1.983)	12.627 (1.995)	12.649 (2.005)	12.673 (2.010)
Husband's age	30.307 (6.614)	31.302 (6.547)	31.936 (6.517)	32.914 (6.573)	33.872 (6.668)	33.740 (6.712)
Husband's education	13.057 (2.613)	13.091 (2.562)	13.090 (2.543)	13.099 (2.554)	13.109 (2.564)	13.147 (2.557)
No contraception	0.213 (0.409)	0.177 (0.381)	0.187 (0.390)	0.180 (0.385)	0.175 (0.389)	0.159 (0.366)
Contraception	0.696 (0.460)	0.737 (0.440)	0.755 (0.430)	0.744 (0.436)	0.745 (0.436)	0.770 (0.421)
Sterilization	0.091 (0.288)	0.086 (0.280)	0.057 (0.232)	0.075 (0.264)	0.080 (0.271)	0.071 (0.257)
Current births	0.148 (0.359)	0.148 (0.361)	0.119 (0.328)	0.121 (0.328)	0.101 (0.303)	0.087 (0.287)
Number of children in family	2.162 (1.536)	2.203 (1.495)	2.065 (1.414)	2.280 (1.395)	2.321 (2.321)	2.367 (1.367)
Husband's income	8108.16 (3106.71)					10517.88 (5133.47)
Number of sterilizations	191	163	99	123	121	99
Number of observations	2088	1897	1734	1635	1512	1391

TABLE II
*Contraceptive choices and demographic characteristics of couples**

		Wife's Age					
Contraception choice		<21	21-25	26-30	31-35	36-40	>40
No contraception		80 24.8	494 20.4	558 17.7	384 16.5	259 17.1	110 20.7
Contraception		238 73.7	1860 76.9	2342 74.5	1673 72.1	1073 70.7	390 73.3
Sterilization		5 1.6	65 2.7	244 7.8	264 11.3	186 12.2	32 6.0
Total		323	2419	3144	2321	1518	532

		Wife's Education				
Contraception choice		Less than high school	High school graduate	Some college	College graduate	More than college
No Contraception		351 23.5	946 17.6	289 15.7	237 20.1	62 17.7
Contraception		988 66.2	4030 74.8	1411 76.4	882 74.6	265 75.7
Sterilization		153 10.2	411 7.6	146 7.9	63 5.3	23 6.6
Total		1492	5387	1846	1182	350

		Number of Children					
Contraception choice		0	1	2	3	4	≥ 5
No contraception		374 41.2	651 29.9	471 13.6	210 10.1	92 10.1	87 12.0
Contraception		526 58.0	1492 68.4	2741 79.1	1606 77.5	677 74.2	534 74.0
Sterilization		7 0.8	37 1.7	252 7.3	256 12.4	143 15.7	101 14.0
Total		907	2180	3464	2072	912	722

* The top entry in each cell gives the number of observations, while the bottom one gives the column percentage.

sterilization are positively correlated with the total number of children, while the decision to use no contraceptive method is negatively correlated. These data also suggest contraceptive methods vary with the number and age distribution of the couple's existing stock of children. In results not reported here, couples with young children and those with older children were found much more likely to either contracept or sterilize than couples with either no children or with children between the ages of 4 and 16.

8. EMPIRICAL RESULTS

This section reports results from the estimation of the contraceptive choice model of Section 6, using the data described in Section 7. Beginning with the birth transition probabilities, or contraceptive failure rates, Table III presents estimates of the probability a birth occurs when a couple does not contracept, F_{nt1} , and the probability associated with using a temporary method, F_{nt2} . These were estimated for each of 28 alternative

TABLE III
Failure by wife's age and education
(Standard errors in parentheses)

No contraceptive Method Used

Wife's age	Wife's education					Marginal
	<High school	High school	Some college	College	>College	
<20	0.619 (0.106)	0.733 (0.114)	0.692 (0.266)			0.692
20-24	0.390 (0.055)	0.651 (0.031)	0.821 (0.051)	0.714 (0.085)	0.750 (0.216)	0.630
25-29	0.259 (0.049)	0.492 (0.032)	0.511 (0.052)	0.607 (0.053)	0.583 (0.101)	0.482
30-34	0.184 (0.044)	0.193 (0.031)	0.222 (0.062)	0.345 (0.064)	0.500 (0.158)	0.227
35-39	0.106 (0.045)	0.082 (0.024)	0.122 (0.051)	0.036 (0.035)	0.222 (0.138)	0.093
≥40	0.009 (0.010)	0.009 (0.010)	0.009 (0.010)	0.083 (0.080)	0.009 (0.010)	0.009
Marginal	0.251	0.390	0.435	0.444	0.453	0.378

Temporary Contraceptive Methods Used

Wife's age	Wife's education					Marginal
	<High school	High school	Some college	College	>College	
<20	0.210 (0.054)	0.149 (0.052)	0.173 (0.154)			0.173
20-24	0.039 (0.013)	0.084 (0.010)	0.103 (0.018)	0.089 (0.025)	0.081 (0.073)	0.081
25-29	0.512 (0.014)	0.055 (0.007)	0.067 (0.013)	0.076 (0.016)	0.063 (0.031)	0.060
30-34	0.026 (0.011)	0.031 (0.006)	0.035 (0.011)	0.023 (0.011)	0.012 (0.012)	0.030
35-39	0.028 (0.016)	0.019 (0.006)	0.017 (0.011)	0.009 (0.011)	0.023 (0.011)	0.019
≥40	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)	0.017 (0.016)	0.002 (0.003)	0.002
Marginal	0.048	0.049	0.056	0.052	0.026	0.050

cross-classifications of the wife's education and age. As might be expected, the failure rates for each contraceptive method fall with age. Overall, the estimated failure rates for couples who use no contraception rise with the wife's educational attainment, while, for those using temporary methods, failure rates rise and then decline. The non-monotonicity in the relationship between the latter failure rates and the wife's education seems to contradict one conclusion of Rosenzweig and Schultz (1989), namely that higher levels of a woman's educational attainment are associated with more efficacious use of contraceptive methods. Only for the 35-39 age group does this association between education and contraceptive failures appear to be negative (as their hypothesis would predict).

Estimates of the contraceptive failure rates and the conditional choice probabilities were used to obtain the estimates of λ , μ_1 , and μ_2 for the reduced form representation characterized in (6.11) and (6.12); these estimates are displayed in Table IV. (The estimated standard errors appropriately account for the prior estimation of the incidental parameters, as described in Section 5 and Appendix B. The effect of this correction on these estimates was negligible.) Note that almost all of the parameters are statistically significant at conventional significance levels. Based on a χ^2 test of the joint significance of $(\lambda_1, \dots, \lambda_{26})$, $(\mu_{11}, \mu_{12}, \mu_{13})$ and $(\mu_{21}, \mu_{22}, \mu_{23})$, we find that the set of conditioning variables included in the contraceptive choice decision rules do help explain the observed choices; the test statistic is 220 which, with 33 degrees of freedom, implies the null hypothesis that these coefficients are all zero is strongly rejected. The significance of the individual coefficients on lagged births indicates the age distribution of children, in addition to their total numbers, has an important impact on parental decisions about

TABLE IV
Reduced form
(Asymptotic standard errors in parentheses)

Parameter	Variable	Estimate	Parameter	Variable	Estimate	Parameter	Variable	Estimate
λ_0		43.776 (10.892)	λ_{12}	$n_{n,t-11}$	-2.427 (0.598)	λ_{24}	A_{2nt}^2	0.041 (0.011)
λ_1	b_{nt}	-5.445 (0.324)	λ_{13}	$b_{n,t-12}$	-2.706 (0.678)	λ_{25}	$E_{2n}A_{2nt}$	0.229 (0.056)
λ_2	$b_{n,t-1}$	-1.690 (0.192)	λ_{14}	$b_{n,t-13}$	-3.412 (0.823)	λ_{26}	$E_{2n}A_{2nt}^2$	-0.003 (0.0009)
λ_3	$b_{n,t-2}$	-1.238 (0.200)	λ_{15}	$b_{n,t-14}$	-2.786 (0.914)	μ_{10}		8.893 (1.735)
λ_4	$b_{n,t-3}$	-1.314 (0.216)	λ_{16}	$b_{n,t-15}$	-2.810 (1.097)	μ_{11}	A_{1nt}	-0.568 (0.106)
λ_5	$b_{n,t-4}$	-1.708 (0.237)	λ_{17}	$b_{n,t-16}$	-3.173 (1.386)	μ_{12}	A_{1nt}^2	0.008 (0.002)
λ_6	$b_{n,t-5}$	-0.990 (0.259)	λ_{18}	$b_{n,t-17}$	-3.308 (1.828)	μ_{13}	E_{1n}	-0.034 (0.627)
λ_7	$b_{n,t-6}$	-2.288 (0.283)	λ_{19}	$b_{n,t-18}$	-0.033 (2.416)	μ_{20}		9.131 (1.302)
λ_8	$b_{n,t-7}$	-1.981 (0.302)	λ_{20}	$b_{n,t-19}$	-6.697 (3.674)	μ_{21}	A_{1nt}	-5.599 (0.081)
λ_9	$b_{n,t-8}$	-2.234 (0.363)	λ_{21}	$b_{n,t-20}^+$	-11.308 (4.789)	μ_{22}	A_{1nt}^2	0.009 (0.001)
λ_{10}	$b_{n,t-9}$	-1.605 (0.399)	λ_{22}	E_{2n}	-3.725 (0.852)	μ_{23}	E_{1n}	0.074 (0.019)
λ_{11}	$b_{n,t-10}$	-2.562 (0.479)	λ_{23}	A_{2n}	-2.582 (0.709)			

contraceptive use. This importance of the age distribution of children was found in our earlier study (Hotz and Miller (1988)), which estimated an index model under the assumption that the unobserved disturbances were normally distributed (instead of Type I Extreme Value), ignored the effects of future conditional choice probabilities on parental decision rules (explicitly treated in this paper), and exploited a different data set (the Panel Study of Income Dynamics), over a longer period (1970 through 1979), and on a smaller sample (350 married couples).

Many implications of the underlying structural model emerge from an examination of its reduced form. Comparative dynamic exercises were conducted to analyze changes in contracept behaviour in response to changes in the values of the variables in H_{nt} . Consider, for example, the effects of differences in the wife's level of education, E_{1n} . In particular, it follows from (3.16) and (6.11) that in this case:

$$\frac{\partial P_{ntk}}{\partial E_{1n}} = P_{ntk} [1 + \sum_{j=1}^2 \exp [x'_{1n} \mu_j + \beta^{-t} \nu_{nj}]]^{-2} \left[\mu_{k4} + \beta^{-t} \frac{\partial \nu_{ntk}}{\partial \rho_{nt}} \frac{\partial \rho_{nt}}{\partial E_{1n}} \right] \quad (8.1)$$

Expression (8.1) consists of two components. The first, involving μ_{k4} , measures the (direct) effect of changes in the wife's education on choice probabilities through its effect on the per period payoffs. The second, which involves the partial of ν_{ntk} with respect to ρ_{nt} , captures the (indirect) effect of her education on future choice probabilities and failure rates. It can be calculated from (6.12); nonparametric estimation procedures were used to obtain consistent estimates of the derivative of $\rho(H_{nt})$.

The effect of changes in the wife's education on parental contraceptive choice, holding all else in H_{nt} constant, is qualitatively similar to the gross association presented in Table II. Holding other characteristics of a couple constant at their sample means, our model predicts that couples with wives who have an additional year of education are 16.5% less likely to not contracept, 5% less likely to sterilize, but 15% more likely to use temporary methods. We also find that the direct and indirect effects of increases in E_{1n} are in the same direction; thus, they reinforce each other to produce this overall effect.

With respect to the impact of the wife's age on the couple's contraceptive choices, our model predicts that as the wife grows older by one year the couple is 9.8% more likely to not contracept, 1.6% less likely to use contraceptive methods and 3% more likely to sterilize. In contrast to the results for changes in her education, the direct effects of the wife's ageing are qualitatively different than the indirect effects; holding other characteristics and $\rho(H_{nt})$ constant, couples are 0.4% more likely to sterilize as the wife ages by one year.

The discrepancy between the (estimated) direct and indirect effects of the wife's age on contraceptive choice calls attention to an important feature of our model of parental decision making. The wife's age (as well as her education) affects parental choices through two distinct avenues. One is through the effect of the wife's age on the payoffs parents receive from specific choices (as measured by the parameters μ_{k1} and μ_{k2}). Our estimates of the direct effects of the wife's age suggest that, as women get older, couples choose contraceptive actions which reduce the risks of a pregnancy. Such findings are consistent with a view that the direct effect of wife's age characterizes the impact of changes in the opportunity cost of the wife's time on contraceptive choice.

But, the wife's age also affects the failure rates of alternative contraceptive methods. Recall from Table III that the failure rates for both the no contraception and temporary method choices decline with the wife's age. Moreover, based on the reduced form estimates, a temporary and unexpected increase in the failure rate of a given contraception method leads parents to substitute away from that method.¹³ For example, in response

13. This is found by substituting our estimates of the reduced form and incidental parameters into (6.12), and then recalculating (6.11).

to a 1% increase in the fecundity for those not contracepting, the probability that couples choose this method would decline by 0·91% while the probability of their taking either of the other choices would each increase by 0·11%. As the failure rates of temporary contraceptive methods increase, the substitution away is weaker; in response to a 1% increase in its failure rate, there is a 0·03% reduction in its use and a corresponding increase of 0·11% in either the use of no contraceptive method or sterilization. Given our estimates of μ_2 , as wives age, holding the payoffs constant, couples would respond to the relative decline in the fecundity by contracepting less. The latter (indirect) effect, because of its magnitude, dominates the effects of the wife's ageing on payoffs and leads to the conclusion that, as wives age, couples more frequently chose not to contracept.

Based on the estimates in Table IV, couples with more educated or older husbands are more likely to use temporary methods and sterilization than other households, although the magnitude of these effects is much smaller than was found with respect to wives'. Couples whose husband has one additional year of education are 5% less likely to not contracept, 0·4% more likely to use temporary methods and 0·1% more likely to sterilize, while those with husbands who are one year older are 2·7% more likely to not contracept and 4% less likely to use either temporary methods or sterilization.

Turning to family composition and contraceptive choice behaviour, we considered how these choices vary with existing numbers of offspring and the spacing of previous births, by examining several offspring configurations. For couples with only 1 child, the predicted probability of not contracepting immediately after the birth is 3%, with 85% using a temporary contraceptive method and 12% choosing sterilization. As the child ages, the estimates imply that parents will increasingly choose to not contracept. By the time the child is 5 years old, the incidence of not contracepting more than doubles to 8% while the use of temporary methods and sterilization declines to 82% and 10%, respectively; at age 10, the corresponding probabilities are, 0·07, 0·82 and 0·11, respectively. To examine how contraceptive choice behaviour changes with family size, consider the case of a household with 2 children, where the first child is 5 years older than the second. The probability of not contracepting immediately after the birth of the second child, relative to the single child family, is lower (by 33%), while the probability of using either temporary methods is slightly lower (by 1%) and that for sterilization is slightly higher (by 12%). When the youngest child reaches 5 years of age, the probability of not contracepting rises to 5%, (that is less than in the single child household), and the probabilities of using temporary methods and sterilization would be predicted to increase to 82% and 13%, respectively. By the time the youngest child is 10, the choice probabilities are 0·04, 0·83 and 0·44%, respectively. Overall, the contraceptive choice decision rules derived from the Table IV estimates imply, holding all else constant, the more children a couple has had, especially past two, the more likely they are to use more effective methods of contraceptive control and that parents do alter their contraceptive strategies to space births and to diminish their chances of pregnancy at later stages of their life cycles as their children grow older.

To identify the structural parameters, estimates of equation (6.8), the husband's earnings equation, were first obtained; these estimates are found in Table V.¹⁴ The predicted life-cycle earnings profile displays the typical concave shape. Evaluated at the mean educational level, the husband's earnings rise early in the life cycle (at age 30, husband's earnings increase at an annual rate of 3·2%), peak at age 46, and decline thereafter. (For example, at age 60, earnings are declining at an annual rate of 2·9 percent.) Husbands with higher levels of education earn more; the rate of return to

14. The specification of the husband's earnings equation we estimated also included a dummy variable for the year 1970, denoted D_n . We included this dummy because the questions in the 1970 and 1975 NSF interviews were worded slightly differently.

education, evaluated at their respective sample means, is 6.2%. Our results also imply higher levels of education are associated with faster growth rates in earnings. Evaluated at the mean age, husbands with a ninth-grade education experience an annual growth rate of 0.2%, compared to 2.0% for those with 12 years of schooling, and 4.4% for those with 16 years of education.

TABLE V
Husband's earnings equation
(Asymptotic standard errors in parentheses)

Parameter	Variable	Estimate	Parameter	Variable	Estimate
γ_0		31,971.20 (2,578.97)	γ_4	$E_{2n}A_{2nt}$	169.87 (10.66)
γ_1	A_{2nt}	-1,485.60 (142.20)	γ_5	$E_{2n}A_{2nt}^2$	-1.88 (0.14)
γ_2	A_{2nt}^2	16.58 (1.92)	γ_6	D_n	1713.90 (50.44)
γ_3	E_{2n}	-2,944.96 (194.17)		σ_ξ	3715.81 (14.94)

Nested within the model developed in Section 6 are many structural specifications, which differ in the restrictions they impose on the parental utility function and the costs of raising children. Here we report the results from estimating three such specifications. Of the three, Model A is the most restrictive; it assumes service flows and child care expenditures are independent of the child's age:

$$\eta_s = \eta_0 \quad \text{for all } s \in \{1, \dots, 21\} \quad (8.2)$$

$$\omega_s = \omega_0 \quad \text{for all } s \in \{1, \dots, 20\}. \quad (8.3)$$

Model B relaxes the restrictions on expenditures by replacing (8.2) with:

$$\begin{aligned} \omega_2 &= \omega_3 & \omega_{10} &= \dots = \omega_{14} \\ \omega_4 &= \dots = \omega_6 & \omega_{15} &= \dots = \omega_{17} \\ \omega_7 &= \dots = \omega_9 & \omega_{18} &= \dots = \omega_{20}. \end{aligned} \quad (8.4)$$

The least restrictive specification we report on, Model C, assumes expenditures on children follow the pattern indicated by (8.4) but relaxes (8.2) as follows:

$$\begin{aligned} \eta_2 &= \eta_3 & \eta_{10} &= \dots = \eta_{14} \\ \eta_4 &= \dots = \eta_6 & \eta_{15} &= \dots = \eta_{17} \\ \eta_7 &= \dots = \eta_9. \end{aligned} \quad (8.5)$$

The estimates for these models are found in Table VI.

The overidentifying restrictions, associated with the reduced form mappings which (6.13) and (8.2) through (8.5) define for the respective specifications, form the basis for a χ^2 test. All three are strongly rejected; the significance level of the test statistic for

TABLE VI
Structural parameter estimates
 (Asymptotic standard errors in parentheses)

Parameter	Model A	Model B	Model C	Parameter	Model A	Model B	Model C	Parameter	Model A	Model B	Model C
$\eta_0, \dots, \eta_{21}^+$	0.284 (11,387.1)	5.220 (1.513)		$\omega_0, \dots, \omega_{20}$	-4.126 (11,337.0)			β	-0.722 (79.021)	0.049 (0.364)	0.646 (0.678)
η_0		1.495 (3.972)		ω_0		0.00002 (0.001)		μ_{10}	4.594 (1.537)	5.734 (1.585)	5.552 (1.609)
η_1		1.051 (1.380)		ω_1		-0.011 (0.266)	0.197 (2.187)	μ_{11}	-0.323 (0.929)	-0.420 (0.096)	-0.408 (0.099)
η_2, \dots, η_3		0.487 (0.495)		ω_2, ω_3		3.171 (48.047)	0.096 (1.538)	μ_{12}	0.005 (0.001)	0.006 (0.001)	0.006 (0.001)
η_4, \dots, η_6		0.249 (0.570)		$\omega_4, \dots, \omega_6$		-1139.060 (537.288)	-0.162 (0.994)	μ_{13}	-0.098 (0.024)	-0.062 (0.025)	-0.067 (0.026)
η_7, \dots, η_9		0.403 (0.729)		$\omega_7, \dots, \omega_9$		-805.936 (352.241)	-1.636 (30.008)	μ_{20}	5.112 (1.287)	5.473 (1.299)	5.309 (1.302)
$\eta_{10}, \dots, \eta_{14}$		0.241 (0.693)		$\omega_{10}, \dots, \omega_{14}$		-727.984 (328.695)	1.358 (15.100)	μ_{21}	-0.412 (0.080)	-0.432 (0.081)	-0.421 (0.081)
$\eta_{15}, \dots, \eta_{17}$		0.317 (1.062)		$\omega_{15}, \dots, \omega_{17}$		-685.862 (305.064)	34.852 (872.934)	μ_{22}	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)
η_{18}		0.947 (1.118)		$\omega_{18}, \dots, \omega_{20}$		-748.417 (338.929)	-54.344 (708.563)	μ_{23}	0.057 (0.019)	0.058 (0.019)	0.060 (0.020)
η_{19}		-0.589 (1.088)		ψ_1	-0.00003 (0.091)	0.039 (0.012)	1.439 (5.277)				
η_{20}		0.402 (3.311)		ψ_2	19.661 (5.725)	0.188 (1.251)	0.047 (1.251)				
η_{21}		1.595 (3.386)									

Model A is 0.182×10^{-16} , for Model B 0.703×10^{-14} and 0.0034 for Model C. Confronted with these test statistics, we did explore several other configurations of structural coefficients that are not nested by Model C but failed to identify a specification that was not rejected. This failure can be attributed to non-linearities in the reduced-form mapping which hindered our search to recover an exactly identified structural specification.

Despite their poor fits, as compared with the reduced form in Table IV, it is worthwhile to briefly compare our findings of the three structural specifications with each other, and with previously published work. Both sets of restrictions (8.1) and (8.2) are, individually and jointly, rejected against the relaxations (8.3) and (8.4), at the 0.01 level. These rejections underscore one advantage of our estimator, namely it enables one to investigate a richer set of structural specifications. The only other existing dynamic structural analyses of fertility, by Wolpin (1984) and Montgomery (1987), maintain the assumption of age-invariant service flows and expenditures on children in order to make their respective maximum likelihood estimators computationally feasible. On the other hand, many of the coefficients reported in Table VI are insignificant. Moreover the estimated annual expenditure per child (measured in 1975 \$US) are implausibly low. (For example, considerably higher estimates are obtained in the cost-of-children literature by Espenshade (1980) and Lazear and Michael (1988).)

Viewed as an application of techniques developed in the first five sections of this paper, several noteworthy points emerge. The existence of a reduced form which is no more difficult to estimate than a multinomial logit greatly simplifies the computation of dynamic, discrete-choice models. We argued the resulting estimates are helpful in explaining various behaviours across different household types, and conducted some limited comparative dynamic exercises with it. Moreover, adjustments to the estimated standard errors, arising from the prior estimation of the conditional choice and transition probabilities, were small. As we noted above, the estimated reduced form captures some of the systematic variation in the data. Indeed, the reduced form is so precisely estimated that it strongly rejects specifications which are sufficiently parsimonious for ML estimation. On the other hand, the non-linear mapping between the reduced-form and the structural parameters proved intractable; consequently, we failed to recover a specification which was not rejected. So, while the new representation of conditional valuation functions and the associated estimation techniques are not directly responsible for this shortcoming, the ease with which the reduced form of the model could be estimated and used to reject parsimoniously parameterized structural specifications, may be a harbinger.

APPENDIX A

Proof of Proposition 1. From (3.3) in the text, $\underline{Q}(\underline{y}, H_t)$ is the $(J - 1)$ dimensional vector function:

$$\underline{Q}(\underline{y}, H_t) = (Q_1(\underline{y}, H_t), \dots, Q_{J-1}(\underline{y}, H_t))', \quad (\text{A.1})$$

with components $j \in \{1, \dots, J - 1\}$ defined as:

$$Q_j(\underline{y}, H_t) = \int G_j(\varepsilon + \nu_j - \nu_1, \dots, \varepsilon + \nu_j | H_t) d\varepsilon. \quad (\text{A.2})$$

For each (t, H_t) and $W \subset \Delta^{J-1}$, we now define the correspondence $\underline{Q}^{-1}(W, H_t)$ as:

$$\underline{Q}^{-1}(W, H_t) = \{\underline{y} \in \mathbb{R}^{J-1}: \underline{Q}(\underline{y}, H_t) \in W\}. \quad (\text{A.3})$$

The proposition asserts that, for all $\underline{p} \in \Delta^{J-1}$ the set $\{\underline{y} \in \mathbb{R}^{J-1}: Q_i(\underline{y}, H_t) = p_i\}$ is either empty or consists of a singleton.

First we show the Inverse Function Theorem applies here. For any $t < T$, fix H_t^* at any value in $B_0 \times B^{T-1}$. By assumption $G(\xi_t | H_t^*)$ is equipped with a well-defined probability density function. Therefore the cross partial of $G(\xi_t | H_t^*)$ with respect to ξ_i exists for all $(i, j) \in \{1, \dots, J - 1\}^2$ with $i \neq j$, and is non-negative; denote

it by $G_{ij}(\varepsilon_t)$. From (A.2), $Q(\varepsilon, H_t^*)$ is differential with respect to ε , and has a square $(J-1)$ -dimensional derivative matrix function which is denoted by $D(\varepsilon)$. We now show that the matrix $D(\varepsilon^*)$ is invertible at any $\varepsilon^* \in \mathbb{R}^{J-1}$. For all $(i, j) \in \{1, \dots, J-1\}^2$ with $i \neq j$, define h_{ij} as:

$$h_{ij} = \int G_{ij}(\varepsilon + \nu_i^* - \nu_j^*, \dots, \varepsilon + \nu_j^*) d\varepsilon. \quad (\text{A.4})$$

Observe $h_{ij} > 0$ since $g_{ij} > 0$. By differentiating $Q_t(\varepsilon, H_t)$ with respect to ε at (ε^*, H_t^*) , one can verify:

$$D_{ij}(\varepsilon^*) = \begin{cases} \sum_{k=1, k \neq i}^J h_{ik} & \text{if } i = j, \\ -h_{ij} & \text{if } i \neq j. \end{cases} \quad (\text{A.5})$$

Define the matrices D_1 and D_2 as

$$D_1 = \begin{bmatrix} D_{11}(\varepsilon^*) & 0 & \cdots & 0 \\ 0 & D_{22}(\varepsilon^*) & & \\ \vdots & & & \\ 0 & 0 & \cdots & D_{J-1,J-1}(\varepsilon^*) \end{bmatrix}, \quad (\text{A.6})$$

$$D_2 = - \begin{bmatrix} 0 & D_{12}(\varepsilon^*) & \cdots & D_{1,J-1}(\varepsilon^*) \\ D_{12}(\varepsilon^*) & 0 & & D_{2,J-1}(\varepsilon^*) \\ \vdots & & & \\ D_{J-1,1}(\varepsilon^*) & D_{J-1,2}(\varepsilon^*) & \cdots & 0 \end{bmatrix}.$$

Because D_1 is diagonal it can be inverted. Also by construction

$$D(\varepsilon^*) = D_1 - D_2 = [I - D_2 D_1^{-1}] D_1. \quad (\text{A.7})$$

Let \tilde{h}_{ij} denote element (i, j) of $D_2 D_1^{-1}$. From (A.4) and (A.5)

$$\tilde{h}_{ij} = \begin{cases} 0 & \text{if } i = j, \\ h_{ij} / \sum_{k \neq j} h_{kj} & \text{if } i \neq j. \end{cases} \quad (\text{A.8})$$

Since $h_{ij} > 0$:

$$0 \leq \tilde{h}_{ij} \leq 1,$$

$$\sum_{i=1}^{J-1} \tilde{h}_{ij} < 1. \quad (\text{A.9})$$

It follows from Hadley (1973, p. 118) that the inverse of $[I - D_2 D_1^{-1}]$ exists. Therefore $D_1^{-1}[I - D_2 D_1^{-1}]^{-1}$ is the inverse of $D(\varepsilon^*)$. Therefore $D(\varepsilon^*)$ has a non-zero determinant.

Appealing to the Inverse Function Theorem, there exists an open neighbourhood $V^* \subset \mathbb{R}^{J-1}$ containing ε^* and an open set $W^* \subset \Delta^{J-1}$ containing $Q(\varepsilon^*, H_t^*)$ such that $Q: V \rightarrow W$ has a differentiable inverse $Q^{-1}(\cdot, H_t): W^* \rightarrow V^*$; see Theorem 2-11 in Spivak (1965, p. 35) for example.

The proof is completed by extending this local result to the whole space, using the Mean Value Theorem. Suppose that, contrary to the proposition there exist two (or more) points ε^* , and ε^{**} satisfying the equation $Q(\varepsilon^*, H_t^*) = Q(\varepsilon^{**}, H_t^*)$. Therefore $Q_1(\varepsilon^*, H_t^*) = Q_1(\varepsilon^{**}, H_t^*)$. Since $Q_1(\varepsilon, H_t^*)$ is differentiable in ε , by the Mean Value Theorem, found in Williamson, Cromwell and Trotter (1972, p. 275) for example, there exists some point ε^{***} such that $D_{1j}(\varepsilon^{***}) = 0$ for all $j \in \{1, \dots, J-1\}$. Hence the determinant of $D(\varepsilon^{***})$ is 0, which contradicts a statement we proved above. Hence no such ε^* and ε^{**} exist. ||

APPENDIX B

Proof of Proposition 2. The first half of Proposition 2 is proved in the text. Accordingly, consider the case where \mathcal{B}_0 is compact but not finite. Kernel estimators are used to estimate $p(H_{nt})$, the incidental parameter conditional choice probabilities associated with $\mathcal{A}(H_{nt})$, the set of histories which remain feasible after the occurrence of H_{nt} at subsequent ages $\{t+1, \dots, t+T-A_{nt}\}$. We now write $\mathcal{A}(H_{nt})$ as $\{H_n^{(1)}, \dots, H_n^{(\bar{n})}\}$, and denote the vector set of associated conditional choice probabilities by the, $p(H_{nt}) = p_n = \{p_{n1}, \dots, p_{n\bar{n}}\}$, where

$p_{nr} = p(H_n^{(r)})$ for each $r \in \{1, \dots, \tilde{n}\}$. To construct a kernel estimator for \hat{p}_{nr} , the indicator function $I_2(H_{mt}, H_{nt})$ is defined as

$$I_2(H_{mt}, H_{nt}) = \begin{cases} 1 & \text{if } A_{nt} = A_{mt} \text{ and } (b_{m1}, \dots, b_{m,t-1}) = (b_{n1}, \dots, b_{n,t-1}) \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B.1})$$

Notice $I_2(H_{mt}, H_{nt})$ partitions observations into sets of histories, where all the elements belonging to each set only differ by their initial conditions. Finally, let h_N , a positive real number, denote the bandwidth associated with the kernel estimator of \hat{p}_{nr} , and let $J(\cdot)$ be a bounded symmetric differentiable function such that $\int J(x)dx = 1$, for $x \in \mathbb{R}^L$ (where L is the dimension of b_{n0}). A kernel estimator of \hat{p}_{nr} is given by:

$$\hat{p}_{nr}^{(N)} = \left[\sum_{n=1, m \neq n}^N d_{mt} I_2(H_n^{(r)}, H_{mt}) J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \right] \left[\sum_{m=1, m \neq n}^N I_2(H_n^{(r)}, H_{mt}) J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \right]^{-1}. \quad (\text{B.2})$$

The transitional probabilities are estimated in a similar manner. Given H_{nt} , for each $H_n^{(r)} \in \mathcal{A}(H_{nt})$ and $j \in \{1, \dots, J\}$, denote the ($K - 1$ dimensional vector) probability transition $F_j(b | H_n^{(r)})$ by F_{nj} . It can be estimated by:

$$F_{nj}^{(N)} = \left[\sum_{n=1, m \neq n}^N b_{mt} d_{mj} I_2(H_n^{(r)}, H_{mt}) J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \right] \left[\sum_{m=1, m \neq n}^N d_{mj} I_2(H_n^{(r)}, H_{mt}) \right]^{-1}. \quad (\text{B.3})$$

Writing $F_n = \{F_{n1}, \dots, F_{nL}, \dots, F'_{n\tilde{n}1}, \dots, F'_{n\tilde{n}J}\}$, and $F_n^{(N)}$ for its corresponding estimator, we thus construct estimators $\rho_n^{(N)} = \{p_n^{(N)}, F_n^{(N)}\}$ of the incidental parameter sets ρ_n .

For each $n \in \{1, \dots, N\}$. Substituting $\rho_n^{(N)}$ for $\rho(H_{nt})$ in $\pi_{1n}(\theta, \rho(H_{nt}))$ in (5.4), an estimator of θ_0 , denoted $\hat{\theta}^{(N)}$, is obtained by forming sample analogues of (5.5).

To prove $\hat{\theta}^{(N)}$ is consistent, we start by noting that for all $n \in \{1, 2, \dots\}$ and $N \in \{2, 3, \dots\}$, both ρ_n and $\rho_n^{(N)}$ belong to a compact space \mathcal{S} . Since $\pi_{1n}(\theta, \rho_n)$ is continuous, $\pi_{1n}(\theta, \rho_n)$ is uniformly continuous in ρ_n for each $\theta \in \Theta$. By Theorem 1 of Bierens (1982, p. 701), $\rho_n^{(N)}$ converges uniformly to ρ_n in probability. For each $\theta \in \Theta$, therefore:

$$N^{-1} \sum_{n=1}^N \pi_{1n}[\theta, \rho_n^{(N)}] - N^{-1} \sum_{n=1}^N \pi_{1n}[\theta, \rho_n] = o_p(1). \quad (\text{B.4})$$

Hence $\hat{\theta}^{(N)}$ converges to $\theta_0^{(N)}$ in probability. By Theorem 2.1 of Hansen (1982, p. 1035), $\theta_0^{(N)}$ converges almost surely to θ_0 , so $\hat{\theta}^{(N)}$ converges to θ_0 in probability as claimed.

We now show $N^{1/2}(\hat{\theta}^{(N)} - \theta_0)$ is asymptotically distributed as a normal random variable. (Also see Newey (1991) and Pakes and Olley (1991) for some recent work in this area.) To establish this claim some extra notation is required. We first define the $Q \times Q$ matrices Γ_n and Γ_N , and also the $Q \times (JK - 1)$ matrices $\Gamma_{r(n)}$, as

$$\Gamma_n = (\partial \pi_{1n}^{(1)}(\theta_0, \rho_n)' / \partial \theta, \dots, \partial \pi_{1n}^{(Q)}(\theta_0, \rho_n)' / \partial \theta) \quad (\text{B.5})$$

$$\Gamma_N = N^{-1} \sum_{n=1}^N \Gamma_n \quad (\text{B.6})$$

$$\Gamma_{r(n)} = (\partial \pi_{1n}^{(1)}(\theta_0, \rho_n)' / \partial \rho_{r(n)}, \dots, \partial \pi_{1n}^{(Q)}(\theta_0, \rho_n)' / \partial \rho_{r(n)}). \quad (\text{B.7})$$

Second, let q_{1ntj} denote the joint mixed probability density function for (H_{nt}, d_{nj}) , and q_{2nt} the joint mixed probability density function for H_{nt} . That is:

$$q_{1ntj} = \partial \Pr[H_{nt}, d_{nj} = 1] / \partial b_{n0} \quad (\text{B.8})$$

$$q_{2nt} = \partial \Pr[H_{nt}] / \partial b_{n0}. \quad (\text{B.9})$$

The Kernel estimator for q_{1ntj} and q_{2nt} are respectively defined as:

$$q_{1ntj}^{(N)} = (N-1)^{-1} \sum_{m=1, m \neq n}^N h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) I_2(H_{mt}, H_{nt}) d_{mj} \quad (\text{B.10})$$

$$q_{2nt}^{(N)} = (N-1)^{-1} \sum_{m=1, m \neq n}^N h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) I_2(H_{mt}, H_{nt}). \quad (\text{B.11})$$

Notice that $q_{1ntj} = q_{2nt} p_{nj}$ and $q_{1ntj}^{(N)} = q_{2nt}^{(N)} p_{nj}^{(N)}$. Similarly we define:

$$q_{3ntjk}^{(N)} = (N-1)^{-1} \sum_{m=1, m \neq n}^N h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) I_2(H_{mt}, H_{nt}) d_{mj} b_{mtk}. \quad (\text{B.12})$$

Third, for each $r(n) \in \{1, \dots, \tilde{n}\}$, define the $(JK - 1)$ dimensional vector $d_m^{r(n)}$ by:

$$d_m^{r(n)} = \begin{bmatrix} p_{r(n)1}(d_{mt1}q_{1r(n)1}^{-1} - q_{2r(n)1}^{-1})I_2(H_{mt}, H_{nt}) \\ \vdots \\ p_{r(n),J-1}(d_{mt,J-1}q_{1r(n),J-1}^{-1} - q_{2r(n),J-1}^{-1})I_2(H_{mt}, H_{nt}) \\ F_{r(n)11}(b_{mt1}q_{3r(n)11}^{-1} - q_{1r(n)1}^{-1})I_2(H_{mt}, H_{nt})d_{mt1} \\ \vdots \\ F_{r(n)1,K-1}(b_{mt,K-1}q_{3r(n)1,K-1}^{-1} - q_{1r(n)1}^{-1})I_2(H_{mt}, H_{nt})d_{mt1} \\ F_{r(n)21}(b_{mt1}q_{3r(n)21}^{-1} - q_{1r(n)2}^{-1})I_2(H_{mt}, H_{nt})d_{mt2} \\ \vdots \\ F_{r(n)J,K-1}(b_{mt,K-1}q_{3r(n)J,K-1}^{-1} - q_{1r(n)J}^{-1})I_2(H_{mt}, H_{nt})d_{mtJ} \end{bmatrix}.$$

We shall presently use the following result:

$$\begin{aligned} & \frac{\partial p_{r(n)j}}{\partial q_{1r(n)j}}(q_{1r(n)j}^{(N)} - q_{1r(n)j}) + \frac{\partial p_{r(n)j}}{\partial q_{2r(n)j}}(q_{2r(n)j}^{(N)} - q_{2r(n)j}) \\ &= p_{r(n)j}(q_{1r(n)j}^{(N)}/q_{1r(n)j} - q_{2r(n)j}^{(N)}/q_{2r(n)j}) \\ &= (N-1)^{-1} \sum_{m=1, m \neq n}^N h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) d_m^{r(n)}, \end{aligned} \quad (\text{B.13})$$

where $d_m^{r(n)}$ is the j -th component of $d_m^{r(n)}$. An analogous result for the transition probabilities also holds.

Fourth, define $\rho_{nm}^{(N)}$, the $(JK - 1) \times 1$ symmetric kernel (vector), as

$$\rho_{nm}^{(N)} = h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) [(\sum_{r(n)=1}^{\tilde{n}} \Gamma_{r(n)} d_m^{r(n)}) + (\sum_{r(m)=1}^{\tilde{m}} \Gamma_{r(m)} d_n^{r(m)})]. \quad (\text{B.14})$$

We observe:

$$\begin{aligned} E[\rho_{nm}^{(N)'} \rho_{nm}^{(N)}] &= O(1) h_N^{-2k} E\left\{ J\left(\frac{b_{n0} - b_{m0}}{h_N}\right)^2\right\} \\ &= O(1) h_N^{-k} \\ &= o(N). \end{aligned} \quad (\text{B.15})$$

The bottom line in (B.15) follows from our requirement that $N^{1/2}h_N^k \rightarrow \infty$ (which implies $N^{-1/2}h_N^{-k} \rightarrow 0$ and therefore $N^{-1}h_N^{-k} \rightarrow 0$); the second line uses a change of variables, while the $O(1)$ term, a symmetric $J \times J$ matrix; is just:

$$2E[(\sum_{r(n)=1}^{\tilde{n}} \Gamma_{r(n)} d_m^{r(n)}) (\sum_{r(n)=1}^{\tilde{n}} \Gamma_{r(n)} d_m^{r(n)} + \sum_{r(m)=1}^{\tilde{m}} \Gamma_{r(m)} d_n^{r(m)})']. \quad (\text{B.16})$$

Lemma 3.1 in Powell, Stock and Stoker (1989), along with (B.15), gives us the following projection result:

$$N^{-1}(N-1)^{-1} \sum_{n=1}^{N-1} \sum_{m=n+1}^N \rho_{nm}^{(N)} = N^{-1} \sum_{n=1}^N E(\rho_{nm}^{(N)} | g_{nt}, H_{nt}) - E(\rho_{nm}^{(N)}/2) + o_p(N^{-1/2}). \quad (\text{B.17})$$

Fifth, define the parameter vector $\varrho_3^{(N)}$ as the solution to the equations:

$$\Gamma_N(\varrho_3^{(N)} - \varrho_0) = N^{-1} \sum_{n=1}^N \left[\pi_{1n}(\varrho_0, \rho_n) + \sum_{r(n)=1}^{\tilde{n}} \Gamma_{r(n)} (N-1)^{-1} h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \sum_{m=1, m \neq n}^N d_m^{r(n)} \right]. \quad (\text{B.18})$$

The asymptotic normality of $N^{1/2}(\varrho_3^{(N)} - \varrho_0)$ is established by first, showing $N^{1/2}(\varrho_3^{(N)} - \varrho_0)$ is asymptotically normal and second, proving $\varrho_3^{(N)} - \varrho_0 = o_p(N^{-1/2})$. The claim then follows from (x)(d) in Rao (1973),

p. 122). Expanding the definition of $\hat{\theta}_3^{(N)}$ given in (B.18) yields:

$$\begin{aligned}
 & \Gamma_N(\hat{\theta}_3^{(N)} - \theta_0) \\
 &= N^{-1} \sum_{n=1}^N \left[\pi_{1n}(\theta_0, \rho_n) + \sum_{r(n)=1}^{\tilde{n}} \Gamma_{r(n)}(N-1)^{-1} h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \sum_{m=1, m \neq n}^N d_m^{r(n)} \right] \\
 &= N^{-1} \sum_{n=1}^N \pi_{1n}(\theta_0, \rho_n) + N^{-1}(N-1)^{-1} \sum_{n=1}^N \sum_{m=1, m \neq n}^N \left[h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) (\sum_{r(n)=1}^{\tilde{n}} \Gamma_{r(n)} d_m^{r(n)}) \right] \\
 &= N^{-1} \sum_{n=1}^N \pi_{1n}(\theta_0, \rho_n) + N^{-1}(N-1)^{-1} \sum_{n=1}^{N-1} \sum_{m=n+1, m \neq n}^N \rho_{nm}^{(N)} \\
 &= N^{-1} \sum_{n=1}^N [\pi_{1n}(\theta_0, \rho_n) + E(\rho_{nm}^{(N)} | d_{nt}, H_{nt}) - E(\rho_{nm}^{(N)} / 2)] + o_p(N^{-1/2}). \tag{B.19}
 \end{aligned}$$

The second equality in (B.19) rearranges the terms; then we substitute $\rho_{nm}^{(N)}$ defined in (B.14) and appeal to the projection result (B.17). By assumption, the square matrix Γ is invertible. Consequently the inverse of Γ_N exists for sufficiently large N with arbitrarily high probability. Multiplying (B.19) on the left by $N^{1/2}\Gamma_N^{-1}$ thus establishes $N^{1/2}(\hat{\theta}_3^{(N)} - \theta_0)$ converges to the same distribution as:

$$N^{1/2}\Gamma_N^{-1} \sum_{n=1}^N [\pi_{1n}(\theta_0, \rho_n) + E(\rho_{nm}^{(N)} | d_{nt}, H_{nt}) - E(\rho_{nm}^{(N)} / 2)]. \tag{B.20}$$

(By (x)(d) on p. 123 of Rao (1973) the $o_p(1)$ term can be ignored.) Applying the Lindeberg-Feller theorem, (B.20) converges to a normal random variable with mean $-\Gamma^{-1}E(\rho_{nm}^{(N)} / 2)$ and covariance, $\Gamma^{-1}\Sigma_2\Gamma^{-1}$, where

$$\Sigma_2 = E\{[\pi_{1n}(\theta_0, \rho_n) + E(\rho_{nm}^{(N)} | d_{nt}, H_{nt})][\pi_{1n}(\theta_0, \rho_n) + E(\rho_{nm}^{(N)} | d_{nt}, H_{nt})]\}. \tag{B.21}$$

(See, for example, p. 123 and 138 of Rao (1973).)

We now show the difference between $\hat{\theta}^{(N)}$ and $\hat{\theta}_3^{(N)}$ is $o_p(N^{-1/2})$, thus establishing the asymptotic normality of $\hat{\theta}^{(N)}$ as well. Expanding the sample moments to the orthogonality conditions we obtain:

$$\hat{\Gamma}_N(\hat{\theta}_0 - \hat{\theta}^{(N)}) = N^{-1} \sum_{n=1}^N \left[\pi_n(\theta_0, \rho_n) + \sum_{r(n)=1}^{\tilde{n}} \hat{\Gamma}_{r(n)}(N-1) h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \sum_{m=1, m \neq n}^N d_m^{r(n)} \right] \tag{B.22}$$

where $\hat{\Gamma}_N$ and $\hat{\Gamma}_{r(n)}$ are in (uniformly) close neighbourhoods of Γ_N and $\Gamma_{r(n)}$, respectively. Subtracting (B.18) from (B.22), it now follows that:

$$\begin{aligned}
 & \hat{\Gamma}_N(\hat{\theta}_0 - \hat{\theta}^{(N)}) - \Gamma_N(\hat{\theta}_0 - \hat{\theta}_3^{(N)}) \\
 &= (\hat{\Gamma}_N - \Gamma_N)(\hat{\theta}_0 - \hat{\theta}_3^{(N)}) + \hat{\Gamma}_N(\hat{\theta}_3^{(N)} - \hat{\theta}^{(N)}) \\
 &= N^{-1}(N-1)^{-1} \sum_{n=1}^N \sum_{r(n)=1}^{\tilde{n}} \left[(\hat{\Gamma}_{r(n)} - \Gamma_{r(n)}) h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \sum_{m=1, m \neq n}^N d_m^{r(n)} \right]. \tag{B.23}
 \end{aligned}$$

Since $\hat{\Gamma}_N$ is a consistent estimator of Γ_N , it follows that:

$$(\hat{\Gamma}_N - \Gamma_N)(\hat{\theta}_0 - \hat{\theta}_3^{(N)}) = o_p(1)O_p(N^{-1/2}). \tag{B.24}$$

Also, since $\hat{\Gamma}_{r(n)}$ converges to $\Gamma_{r(n)}$ uniformly in probability,

$$\begin{aligned}
 & N^{-1}(N-1) \sum_{n=1}^N \sum_{r(n)=1}^{\tilde{n}} \left[(\hat{\Gamma}_{r(n)} - \Gamma_{r(n)}) h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \sum_{m=1, m \neq n}^N d_m^{r(m)} \right] \\
 &= o_p(1) \left[N^{-1}(N-1)^{-1} \sum_{n=1}^N \sum_{r(n)=1}^{\tilde{n}} h_N^{-k} J\left(\frac{b_{m0} - b_{n0}}{h_N}\right) \sum_{m=1, m \neq n}^N d_m^{r(m)} \right] \\
 &= o_p(1)O_p(N^{-1/2}). \tag{B.25}
 \end{aligned}$$

(The bottom line of (B.25) can be deduced from a similar argument to the one applied to the double summation over $\rho_{nm}^{(N)}$. See (B.14) through (B.17) and the surrounding discussion.) Noting $o_p(1)O_p(N^{-1/2}) = o_p(N^{-1/2})$ it follows from (B.22) through (B.25) that:

$$o_p(N^{-1/2}) = \hat{\Gamma}_N(\hat{\theta}_3^{(N)} - \hat{\theta}^{(N)}) = (\Gamma_N + o_p(1))(\hat{\theta}_3^{(N)} - \hat{\theta}^{(N)}).$$

Since Γ_N is invertible

$$(\hat{\theta}_3^{(N)} - \hat{\theta}^{(N)}) = o_p(N^{-1/2}) + o_p(1)(\hat{\theta}_3^{(N)} - \hat{\theta}^{(N)}) = o_p(N^{-1/2}). \tag{B.26}$$

(The second equality of (B.26) is established by showing the contrary hypothesis is false.)

There is no reason to expect $E[\rho_{nm}^{(N)}] = 0$. To eliminate any asymptotic bias, we form a linear combination of estimators which take the form of $\theta_3^{(N)}$, defining $\theta_2^{(N)}$ as:

$$\theta_2^{(N)} = (\theta^{(N)} - \sum_{g=1}^{G-1} c_g \theta_g^{(N)}) / (1 - \sum_{g=1}^{G-1} c_g), \quad (\text{B.27})$$

where:

- (i) $G = 2k + 1$;
- (ii) $\theta_g^{(N)}$ for $g = 1, \dots, G - 1$, is an estimator formed in the manner described above, where the bandwidth used in the kernel function is $h_{Ng} = \psi_g h_N$, where $\psi_1, \dots, \psi_{G-1}$ are distinct but otherwise arbitrary positive constants and $\rho_{mn}^{(N,g)}$ is defined analogously for $\rho_{mn}^{(N)}$; and
- (iii) c_1, \dots, c_{G-1} are a set of weights given by:

$$\begin{bmatrix} c_1 \\ \vdots \\ c_{G-1} \end{bmatrix} = \begin{bmatrix} \psi_1, \dots, \psi_{G-1} \\ \vdots \\ \psi_1^{G-1}, \dots, \psi_{G-1}^{G-1} \end{bmatrix}^{-1} \begin{bmatrix} h_N^{-1} \\ \vdots \\ h_N^{-G+1} \end{bmatrix}. \quad (\text{B.28})$$

By inspection $\theta_2^{(N)}$ is a consistent estimator of θ_0 and, adapting the proof strategy of Powell, Stock and Stoker (1989) it can be shown that $N^{1/2}(\theta_2^{(N)} - \theta_0)$ has a limiting multivariate distribution with mean $\bar{\theta}$, and covariance matrix $\Sigma_2^{(N)}$, defined:

$$\Sigma_2^{(N)} = \Gamma^{-1}(4E\{E[\rho_{nm}^{(N)} | d_{nt}, H_{nt}] E[\rho_{nm}^{(N)} | d_{nt}, H_{nt}]'\})\Gamma^{-1}, \quad (\text{B.29})$$

where $\rho_{nm}^{(N)}$ is of the form given in (B.14), using the higher-order kernel function,

$$\bar{J}(\mathbf{b}_{n0}, \mathbf{b}_{m0}, h_N) = \left[J\left(\frac{\mathbf{b}_{n0} - \mathbf{b}_{m0}}{h_N}\right) - \sum_{g=1}^{G-1} c_g (\psi_g)^{-k} J\left(\frac{\mathbf{b}_{n0} - \mathbf{b}_{m0}}{\psi_g h_N}\right) \right] [1 - \sum_{g=1}^{G-1} c_g]^{-1}, \quad (\text{B.30})$$

in place of $J([\mathbf{b}_{n0} - \mathbf{b}_{m0}] / h_N)$. More specifically, observe from (B.27) that

$$N^{1/2}(\theta_2^{(N)} - \theta_0) = (1 - \sum_{g=1}^{G-1} c_g)^{-1} [N^{1/2}(\theta^{(N)} - \theta_0) - \sum_{g=1}^{G-1} c_g N^{1/2}(\theta_g^{(N)} - \theta_0)]. \quad (\text{B.31})$$

Since $N^{1/2}(\theta^{(N)} - \theta_0)$ and $N^{1/2}(\theta_g^{(N)} - \theta_0)$ are asymptotically normal, $N^{1/2}(\theta_2^{(N)} - \theta_0)$ converges in distribution to a normal random variable with mean:

$$-\Gamma^{-1} E[\rho_{nm}^{(N)} - \sum_{g=1}^{G-1} c_g \rho_{nm}^{(N,g)}] / (1 - \sum_{g=1}^{G-1} c_g). \quad (\text{B.32})$$

For notational convenience, define the $Q \times 1$ vector function, $\mathcal{L}^{(N)}(\mathbf{b}_{n0}, \mathbf{b}_{m0})$, as:

$$\mathcal{L}^{(N)}(\mathbf{b}_{n0}, \mathbf{b}_{m0}) = E[\sum_{r=1}^R \Gamma_{r(n)} d_n^{r(n)} | \mathbf{b}_{n0}, \mathbf{b}_{m0}]. \quad (\text{B.33})$$

Also define the real valued vector function, $\xi(\mathbf{h}, N)$, as

$$\xi(\mathbf{h}, N) = \int J(\mathbf{x}) \mathcal{L}^{(N)}(\mathbf{b}_{n0}, \mathbf{b}_{n0} + h\mathbf{x}) F(\mathbf{b}_{n0} + h\mathbf{x}) d\mathbf{b}_{n0} d\mathbf{x}, \quad (\text{B.34})$$

where $F(\mathbf{b}_{n0})$ is the density function for \mathbf{b}_{n0} . We assume that $\xi(\mathbf{h}, N)$ is differentiable $(G+1)$ times in \mathbf{h} at the point $\mathbf{h} = 0$, and denote by $\xi^{(i)}(0, N)$ the value of the i^{th} derivative. Hence

$$\begin{aligned} E[\rho_{nm}^{(N,g)}] &= (\psi_g h_N)^{-k} \int J\left(\frac{\mathbf{b}_{n0} - \mathbf{b}_{m0}}{\psi_g h_N}\right) \mathcal{L}^{(N)}(\mathbf{b}_{n0}, \mathbf{b}_{m0}) F(\mathbf{b}_{n0}) F(\mathbf{b}_{m0}) d\mathbf{b}_{n0} d\mathbf{b}_{m0} \\ &= \int J(\mathbf{x}) \mathcal{L}^{(N)}(\mathbf{b}_{n0}, \mathbf{b}_{n0} + \psi_g h_N \mathbf{x}) F(\mathbf{b}_{n0} + \psi_g h_N \mathbf{x}) d\mathbf{b}_{n0} d\mathbf{x} \\ &= \xi(\psi_g h_N, N) \\ &= \sum_{i=1}^{G-1} \psi_g^i h_N^i \xi^{(i)}(0, N) + O(h_N^G). \end{aligned} \quad (\text{B.35})$$

Substituting (B.35) into (B.32), the asymptotic mean of $N^{1/2}(\theta_2^{(N)} - \theta_0)$ is therefore

$$\{\sum_{i=1}^{G-1} (1 - \sum_{g=1}^{G-1} \psi_g^i h_N^i c_g) \xi^{(i)}(0, N) + O(h_N^G)\} / [2(1 - \sum_{g=1}^{G-1} c_g)], \quad (\text{B.36})$$

which is $O(h_N^G)$, because $\sum_{g=1}^{G-1} \psi_g^i h_N^i c_g = 1$ for each $i \in \{1, \dots, G-1\}$, by (B.28). Recall that h_N is chosen such that $N^{1/2} h_N^k \rightarrow \infty$, and $N h_N^{2k} \rightarrow \infty$ as $N \rightarrow \infty$. However, we are free to set h_N such that $N h_N^{2k+1} \rightarrow 0$ as $N \rightarrow \infty$. Since $G = 2k + 1$, it follows immediately that $N^{1/2} O(h_N^G)$ is $o(1)$. Therefore, $\theta_2^{(N)}$ is asymptotically unbiased as claimed.

To derive Σ_2 , the asymptotic covariance matrix for $\theta_2^{(N)}$, observe:

$$\begin{aligned} \Gamma_N(\theta_0 - \theta_2^{(N)}) - N^{-1} \sum_{n=1}^N \pi_{1n}(\theta_0, \rho_n) + o_p(N^{-1/2}) \\ &= N^{-1} \sum_{n=1}^N E[\rho_{nm}^{(N,g)} - \sum_{g=1}^{G-1} c_g \rho_{mn}^{(N,g)} | d_{nt}, H_{nt}] / (1 - \sum_{g=1}^{G-1} c_g) \\ &\equiv N^{-1} \sum_{n=1}^N E[\bar{\rho}_{nm}^{(N)} | d_{nt}, H_{nt}]. \end{aligned} \quad (\text{B.37})$$

The first equality follows from (B.19), applied also to each $g \in \{1, \dots, G-1\}$, and uses the fact that $N^{1/2}(\theta_g^{(N)} - \theta_g^{(N)})$ is $O_p(N^{-1/2})$ and $N^{1/2}(\theta_2^{(N)} - \theta_0)$ is asymptotically unbiased. Thus:

$$(\theta_0 - \theta_2^{(N)}) = \Gamma_N^{-1} N^{-1} \sum_{n=1}^N \{\pi_n(\theta_0, \rho_n) + E[\bar{\rho}_{mn}^{(N)} | d_{nt}, H_{nt}] + o_p(N^{-1/2})\} \quad (\text{B.38})$$

from which formula (B.21) comes since $\bar{\rho}_{mn}^{(N)}$ converges pointwise to the same limit as $\rho_{mn}^{(N)}$. ||

Acknowledgements. We have benefited from the comments of Ricardo Barros, Kim Balls, Daniel Black, Adrian Pagan, Stephen Spear, two referees, as well as participants in workshops at the Universities of Chicago, Michigan, Minnesota, Rochester, Toronto, Virginia, Western Ontario and Wisconsin as well as Brown, Carnegie Mellon and Duke Universities. We also wish to thank Kermit Daniel and especially Jeff Smith for their comments and research assistance. This research was supported by NICHD grant R23-HD18935.

REFERENCES

- ALTUG, S. and MILLER, R. (1991), "Human Capital, Aggregate Shocks and Panel Data Estimation" (Discussion Paper, Economic Research Center NORC, 91-1).
- BELLMAN, R. (1957) *Dynamic Programming* (Princeton: Princeton University Press).
- BERKOVEC, J. and STERN, S. (1991), "Job Exit Behavior of Older Men", *Econometrica*, **59**, 189–210.
- BIERENS, H. (1983), "Uniform consistency of Kernel Estimators of a Regression Function under Generalized Conditions", *Journal of the American Statistical Association*, **78**, 669–707.
- CHUNG, K. (1974) *A Course in Probability Theory* (New York: Academic Press).
- ECKSTEIN, Z. and WOLPIN, K. (1982a), "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience", *Review of Economic Studies*, **56**, 375–390.
- ECKSTEIN, Z and WOLPIN, K. (1989b), "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey", *Journal of Human Resources*, **24**, 562–598.
- ESPENSHADE, T. (1984) *Investing in Children* (Washington: Urban Institute).
- FLINN, C. and HECKMAN, J. (1982), "New Methods for Analyzing Structural Models of Labor Force Dynamics", *Journal of Econometrics*, **18**, 115–168.
- GÖNÜL, F. (1989), "Dynamic Labor Force Participation Decisions of Males in the Presence of Layoffs and Uncertain Job Offers", *Journal of Human Resources*, **24**, 195–220.
- HADLEY, G. (1973) *Linear Algebra* (Reading, MA: Addison-Wesley Publishing Co).
- HANSEN, L. (1982), "Large Sample Properties of Methods of Moments Estimators", *Econometrica*, **50**, 1029–1054.
- HOTZ, V. J. and MILLER, R. (1988), "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply", *Econometrica*, **56**, 91–118.
- HOTZ, V. J., MILLER, R., SANDERS, S., and SMITH, J. (1991), "A Simulation Estimator for Dynamic Discrete Choice Models" (Unpublished manuscript, University of Chicago and Carnegie Mellon University).
- KIEFER, N. and NEUMANN, G. (1979), "An Empirical Job-Search Model with a Test of the Constant Reservation Wage Hypothesis", *Journal of Political Economy*, **87**, 89–108.
- KIEFER, N. and NEUMANN, G. (1981), "Individual Effects in a Nonlinear Model: Explicit Treatment of Heterogeneity in the Empirical Job Search Literature", *Econometrica*, **49**, 965–980.
- KOLMOGOROV, A. and FOMIN, S. (1970) *Introductory Real Analysis* (New York: Dover Publications).
- LANCASTER, T. and CHESHER, A. (1983), "An Econometric Analysis of Reservation Wages", *Econometrica*, **51**, 1661–1676.
- LAZEAR, E. and MICHAEL, R. (1988) *Allocation of Income within the Household* (Chicago: University of Chicago Press).
- McFADDEN, D. (1981), "Econometric Models of Probabilistic Choice" in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data and Econometric Applications*, (Cambridge: MIT Press), 198–272.
- MILLER, R. (1982), "Job Specific Capital and Labor Mobility" (Ph. D. dissertation, Chicago).
- MILLER, R. (1984), "Job Matching and Occupational Choice", *Journal of Political Economy*, **92**, 1086–1120.
- MONTGOMERY, M. (1987), "A Dynamic Model of Contraceptive Choice" (Unpublished Manuscript, SUNY-Stony Brook).
- NEWHEY, W. (1984), "A Method of Moments Interpretation of Sequential Estimators", *Economic Letters*, **14**, 201–206.
- NEWHEY, W. (1991), "The Asymptotic Variance of Semiparametric Estimators" (MIT Working Paper #583).
- PAKES, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", *Econometrica*, **54**, 755–784.
- PAKES, A. and OLLEY, S. (1991), "A Limit Theorem for a Smooth Class of Semiparametric Estimators" (mimeo, Yale University).
- POTTS, M. (1988), "Birth Control Methods in the United States", *Family Planning Perspectives*, **20**, 288–297.

- POWELL, J., STOCK, J. and STOCKER, T. (1989), "Semiparametric Estimation of Index Coefficients", *Econometrica*, **57**, 1403-1480.
- PRAKASA RAO, B. (1983) *Nonparametric Functional Estimation* (Orlando: Academic Press).
- RAO, C. (1973) *Linear Statistical Inference and Its Applications*, 2nd Ed. (New York: John Wiley & Sons).
- ROSENZWEIG, M. and SCHULTZ, T. P. (1985), "The Demand and Supply of Births", *American Economic Review*, **75**, 992-1015.
- RUST, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, **55**, 999-1034.
- SPIVAK, M. (1965) *Calculus on Manifolds* (Redwood City: Addison-Wesley).
- WILLIAMSON, R., CROMWELL, R. and TROTTER, H. (1972) *Calculus of Vector Functions* (Englewood Cliffs: Prentice-Hall).
- WOLPIN, K. (1984), "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality", *Journal of Political Economy*, **92**, 852-874.
- WOLPIN, K. (1987), "Estimating a Structural Search Model: The Transition from School to Work", *Econometrica*, **55**, 801-819.