Review of Economic Studies (1994) **61**, 265-289 0034-6527/94/00140265$02.00

A Simulation Estimator for Dynamic Models of Discrete Choice

V. JOSEPH HOTZ
University of Chicago

ROBERT A. MILLER
Carnegie Mellon University

SETH SANDERS
Carnegie Mellon University

and

JEFFREY SMITH
University of Chicago

First version received February 1992; *final version accepted September* 1993 (*Eds.*)

This paper analyses a new estimator for the structural parameters of dynamic models of discrete choice. Based on an inversion theorem due to Hotz and Miller (1993), which establishes the existence of a one-to-one mapping between the conditional valuation functions for the dynamic problem and their associated conditional choice probabilities, we exploit simulation techniques to estimate models which do not possess terminal states. In this way our Conditional Choice Simulation (CCS) estimator complements the Conditional Choice Probability (CCP) estimator of Hotz and Miller (1993). Drawing on work in empirical process theory by Pakes and Pollard (1989), we establish its large sample properties, and then conduct a Monte Carlo study of Rust's (1987) model of bus engine replacement to compare its small sample properties with those of Maximum Likelihood (ML).

1. INTRODUCTION

Following Miller (1982, 1984) and Wolpin (1984), there have been many applications of maximum likelihood (ML) estimation techniques to dynamic models of discrete choice. (See the survey by Eckstein and Wolpin (1989).) There are several reasons for estimating econometric models that are explicitly derived from dynamic choice-theoretic frameworks over those which have less explicit connections to economic theory. The fact that estimated parameters can be interpreted within an economic framework automatically provides a common language for economists to discuss the results (within and between empirical studies). Similarly, an economic interpretation can be readily attached to hypothesis tests. Finally, predictions about economic phenomena can be made by conducting exercises in comparative dynamics on the economic models that support the estimation.

Along with this growing literature, there has developed an awareness of the very high computational costs of undertaking ML in such models, which, in turn, has sparked interest in alternative methods of estimation. One of these is to replace the assumption that agents optimize with specifications of behaviour that yield computationally less burdensome decisions rules for estimation purposes. For example, Hotz and Miller (1986, 1988) restrict the space of feasible decision rules to index functions that are linear in the state variables. Stock and Wise (1990) also attempt to simplify the calculation of decision rules in estimation by incorrectly passing an expectations operator through the (nonlinear) maximization operator. Recently, Hotz and Miller (1993) developed a new strategy for estimating dynamic models of discrete choices which avoids the high cost of recursively computing the valuation function many times, a cost associated with ML estimation, without compromising the rationality assumption. Their method, which we refer to as the conditional choice probability (CCP) estimator, is based on an alternative representation of the valuation function. This representation expresses the valuation function as a weighted sum of the possible utility streams that might occur. Therein, the weights are conditional probabilities of the choices prescribed by sequential optimization of future realizations of stochastic variables. In addition, the unobserved components of these utility streams are corrected for the dynamic selection which arises from optimizing behaviour, expressing these corrections in terms of conditional choice probabilities. Given consistent estimates of the conditional choice probabilities and the probabilities determining the other stochastic variables, relatively straightforward estimators of the structural parameters can be formed.

The CCP estimation procedure can be applied to a wide range of stochastic problems in discrete choice. However, to illustrate it, Hotz and Miller (1993) focus on a model from a much more restrictive class. The defining characteristic of this class, called the terminal state property, is the existence of at least one action at each decision node (called a terminating action) which, if taken, would eliminate the differential impact of subsequent choices on the state variables over the remainder of the agent's horizon. Within this class, which includes optimal stopping problems,[1] the number of states for which conditional choice probabilities must be estimated is considerably reduced.

In dynamic choice models which do not have the terminal state property, calculation of valuation functions remains problematic, even when using the alternative representation. To characterize the utility of a current action, the econometrician must assess the expected utility of subsequent choices, where the latter are assumed to be made optimally. The number of these future feasible actions can become large, both in terms of the number feasible at a point in time and the number of periods remaining in an agent's horizon. In essence, the decision tree associated with each current action tends to have many branches. This increases the complexity of implementing the CCP estimator, as the conditional choice probabilities must be estimated for all these nodes (or future feasible choices).[2]

In this paper, we consider how to estimate models lacking this terminal state property. Rather than evaluating the expected utilities associated with all feasible future paths, we show that one need only consider those associated with a path of simulated future choices. These simulated paths are generated in a manner consistent with optimal decision-making by exploiting (estimates of) the future conditional choice probabilities and the transition probabilities governing outcomes. Estimating equations for the structural parameters of a model can be formed using the utilities associated with the simulated paths to form valuation functions. We call the resulting estimator the conditional choice simulation

1. For example, structural models of optimal retirement or sterilization choice have this property.

2. This lack of terminating actions also greatly increases the complexity of the maximum likelihood strategies as it entails integrating over all future paths.

(CCS) estimator, and show that this estimator is $N^{1/2}$ consistent and asymptotically normal for a sample of size N.

Like the CCP estimator, the CCS estimator requires using only unrestricted estimates of the conditional choice and transition probabilities. But, in contrast to the CCP estimator, the new estimator proposed here does not necessarily require estimation of probabilities for all feasible future choice and transitions. Instead, it requires the estimation only of those choice and transition probabilities associated with the nodes of some agent's simulated future path. While the usefulness of simulation in estimating structural models of sequential decision-making has been pointed out and exploited by others,[3] our approach differs from earlier work in that it avoids both the backwards recursion computation of valuation functions (by exploiting the representation developed in Hotz and Miller (1993)) and integration over all future paths (via simulation of a single future path).[4] Finally, our use of simulation methods in forming estimators is different from the applications presented in Pakes and Pollard (1989) and McFadden (1989) in that the simulated paths do not depend on the structural parameter estimates, a feature which facilitates estimation in two ways. First, new simulated paths are not generated for each different set of structural parameter values being evaluated in the estimation algorithm. Second, for a given sample the criterion function for evaluating the structural parameters is typically a smooth function, so derivative-based optimization algorithms can be applied.

Finally, in contrast to ML, both the CCP and CCS estimators separate the problem of estimating parameters which generated the data from the problem of solving the dynamic programming model for any given set of parameters. This means that, while the reduction in computer machine time is quite dramatic when either alternative to ML is used in estimation, substantial amounts of computer programming time may be required to solve for the optimal decision rules when undertaking comparative dynamic exercises. However, as Hotz and Miller (1993) demonstrate, this is not necessarily the case; it depends on the specific nature of the exercise under consideration.

The remainder of the paper is organized as follows. The next section describes the class of dynamic discrete-choice models to be considered and reviews the formulation of valuation functions developed in Hotz and Miller (1993). In Section 3, we develop the CCS estimator and establish its asymptotic properties for finite-horizon models. Section 4 extends these results to infinite-horizon Markov models. Then in Section 5, we present a Monte Carlo study of the small sample performance of this estimator in the renewal model estimated in Rust (1987). Several variations on the CCS estimator are implemented and compared to the maximum likelihood estimator.

2. THE MODEL AND REPRESENTING CONDITIONAL VALUATION FUNCTIONS

To maintain comparability with Hotz and Miller (1993), we restrict the analysis in this and the following section to finite-horizon, discrete-choice models. In each period

3. See Pakes and Pollard (1989), Berkovec and Stern (1991) and Altug and Miller (1991).

4. Our approach to estimation of dynamic models is quite similar to that used in Altug and Miller (1991). The main features distinguishing our work from Altug and Miller (1991) are that they use simulation methods to deal with common shocks hitting the population and assume a form of finite history dependence to reduce the computational burden associated with CCP estimation. This paper assumes that there are no aggregate shocks but does not impose the assumption of finite history dependence. These differences affect the proof strategies used to establish the respective results, because the criterion function which defines Altug and Miller's (1991) estimator is smooth in the parameters, while the criterion function for the CCS estimator is not continuous. Note that the parameter space includes both the structural parameters, θ, and the incidental choice and transition probabilities, ψ. As mentioned in the text, changing the estimates of the conditional choice probabilities (which occurs as the sample size increases), creates jumps in the simulated paths.

$t \in T \equiv \{1, \ldots, T\}$, a typical agent chooses an action for which there are J alternatives. Let

$$d_{tj} \in \{0, 1\}, \quad \text{for all } (t, j) \in T \times J, \tag{2.1}$$

denote the agent's decision about action j in period t, where $d_{tj} = 1$ indicates that action j is chosen and $d_{tj} = 0$ otherwise. We assume that the actions are mutually exclusive, meaning:

$$\sum_{j=1}^{J} d_{tj} = 1, \tag{2.2}$$

for all $t \in T$. Thus the agent's choice in period t can be summarized by the $J-1$ dimensional vector, $d_t = (d_{t1}, \ldots, d_{t,J-1})'$.

The agent conditions his choice in period t on his history, which includes his initial endowment of characteristics $b_0 \in \mathscr{B}$, and the history of realizations on outcome variables $(b_1, \ldots, b_{t-1})' \in \mathscr{B}$.[5] Consequently, each history has a Markov representation, $H_t \in \mathscr{H} \equiv \mathscr{B} \times \mathscr{B}^T$, where the elements are $(b, \ldots, b_{t-1})$ and the last $(T-t)$ elements are dummies to indicate the remaining (unspent) periods of the agent's life. Moreover, in many applications including the Monte Carlo study we undertake, the history of an agent can be characterized by a vector of much lower dimension than $(T-t)$. We assume the transition from H_t to H_{t+1} is either fully determined by action d_t, or generated from a known conditional probability distribution which depends upon the agent's history and his current choice. Accordingly, let $F_j(H_{t+1} \mid H_t)$ denote the probability that H_{t+1} occurs given $d_{tj} = 1$ and history H_t and denote by $F(H_{t+1} \mid H_t)$ the $J-1$ dimensional vector, $F_1(H_{t+1} \mid H_t), \ldots, F_{J-1}(H_{t+1} \mid H_t))'$.

The agent's objective is to maximize the expected value of a sum of a period-specific payoffs or utilities. Let u_{tj} denote the utility associated with choice j in period t. Without loss of generality, u_{tj} can be written as the sum of a deterministic component, $u_j^*(H_t)$, which depends upon the agent's history up to period t, and a stochastic component, ε_{tj}, which is mean independent of $u_j^*(H_t)$. Let $u^*(H_t) = (u_1^*(H_t), \ldots, u_J^*(H_t))'$ and $\varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tJ})'$ denote $J \times 1$ vectors of the deterministic and stochastic utility components, respectively. We assume the probability distribution function for ε_t, denoted by $G(\varepsilon_t \mid H_t)$, has a joint probability density function, $dG(\varepsilon_t \mid H_t)$. In particular applications, $u^*(H_t)$, as well as the $F(H_{t+1} \mid H_t)$ and $G(\varepsilon_t \mid H_t)$, may depend upon a vector of parameters which are the object of structural estimation.[6] For now, we focus on the structure of an agent's decision problem, introducing these parameters in the next section.

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

$$E\left(\sum_{s=t}^{T} \sum_{j=1}^{J} d_{sj}[u_j^*(H_s) + \varepsilon_{sj}] \mid H_t\right). \tag{2.3}$$

Let d_s^0 denote his optimal choice in period s (or, more accurately, the realization of an optimal decision rule of s). We define the condition valuation function associated with choice j made in period t as:

$$V_j(H_t) = E\left[\sum_{s=t+1}^{T} \sum_{j=1}^{J} d_{sj}^0[u_j^*(H_s) + \varepsilon_{sj}] \mid H_t, d_{tj} = 1\right]. \tag{2.4}$$

Ignoring ties, optimal decision-making implies that $d_{tk}^0 = 1$ if and only if:

$$k = \underset{j \in J}{\operatorname{argmax}} \, [u_j^*(H_t) + \varepsilon_{tj} + V_j(H_t)]. \tag{2.5}$$

5. The notational convention adopted here is that the realization b_t occurs at the end of period t.

6. As explained in Section 3, we shall assume that the regression function $u_j^*(H_t)$ is known up to this parameter vector for each $j \in J$ and a similar assumption will be made with respect to $G(\varepsilon_t \mid H_t)$.

From (2.5), it is obvious that the optimal decision rule depends on the differences in expected lifetime utility associated with the various choices, not their absolute levels. Define $v(H_t)$ as the $J-1$ dimensional vector of differences in the conditional valuation functions of agent n at time t. That is:

$$\begin{aligned} v(H_t) &\equiv (v_1(H_t), \ldots, v_{J-1}(H_t))' \\ &= (V_1(H_t) - V_J(H_t), \ldots, V_{J-1}(H_t) - V_J(H_t))'. \end{aligned} \tag{2.6}$$

The representation of $v(H_t)$ is based on a 1 to 1 mapping between $v(H_t)$ and the $J-1$ dimensional vector of conditional choice probabilities $p(H_t) \equiv (p_1(H_t), \ldots, p_{J-1}(H_t))'$ defined by their components:

$$p_k(H_t) = \Pr \{k = \underset{j \in J}{\operatorname{argmax}} [u_j^*(H_t) + \varepsilon_{tj} + V_j(H_t)] \mid H_t\} \tag{2.7}$$

for each $k \in \{1, \ldots, J-1\}$. Intuitively, $p_k(H_t)$ is the probability of taking action k conditional on the initial conditions and past outcomes.

The key result we exploit from Hotz and Miller (1993) is that $v(H_t)$ can be represented as a function of future conditional choice probabilities. To see this, note that $p(H_t)$ can always be expressed as a mapping from $v(H_t)$ and H_t, where the latter dependence on H_t arises because $dG(\varepsilon_t \mid H_t)$ varies with H_t. Proposition 1 in the Hotz and Miller paper proves that there exists an inverse to this mapping, here denoted by $q(p(H_t), H_t)$, where $p(\cdot)$ belongs to the $J-1$ dimensional simplex. That is:

$$\begin{aligned} v(H_t) &= q(p(H_t), G(\cdot \mid H_t), u^*(H_t)) \\ &= q(p(H_t), H_t). \end{aligned} \tag{2.8}$$

A feature evident from the first equality in (2.8) is that, given the value of the conditional choice probability vector $p(H_t)$, the dependence of $v(H_t)$ on H_t only arises through $G(\varepsilon_t \mid H_t)$ and $u^*(H_t)$, the components of the model characterizing the structure of the agent's decision problem. This feature turns out to play a key role in the estimation strategy of Hotz and Miller as well as in the one developed below.

The alternative representation of agent's conditional valuation functions also follows from Proposition 1 in Hotz and Miller (1993). Consider the expected utility an agent obtains in period t conditional on H_t and on behaving optimally, which is given by:

$$E(\textstyle\sum_{j=1}^{J} d_{tj}^0(H_t) + \varepsilon_{tj}] \mid H_t) = \textstyle\sum_{j=1}^{J} p_j(H_t)[u_j^*(H_t) + E(\varepsilon_{tj} \mid H_t, d_{tj} = 1)]. \tag{2.9}$$

The proposition in Hotz and Miller implies that the conditional expectation of the transitory component to current utility can be expressed as the following function of $q_t(p_t(H_t), H_t)$:

$$\begin{aligned} R_k(p(H_t), H_t) &\equiv E(\varepsilon_{tk} \mid H_t, d_{tk} = 1) \\ &= \int \varepsilon G_k(\xi_{1k}(H_t), \ldots, \xi_{Jk}(H_t) \mid H_t) d\varepsilon \\ &= \int \varepsilon G_k(\xi_{1k}(H_t), \ldots, \xi_{Jk}(H_t) \mid H_t) d\varepsilon, \end{aligned} \tag{2.10}$$

where $G_k(\varepsilon \mid H_t) \equiv \partial G((\varepsilon \mid H_t) \partial \varepsilon_k$ and

$$\xi_{jk}(H_t) \equiv \varepsilon + u_k^*(H_t) - u_j^*(H_t) + q_k(p(H_t), H_t) - q_j(p(H_t), H_t). \tag{2.11}$$

(Note that the function in (2.10) accounts for the selectivity of expected transitory utility components that arises in choices among actions.) Consequently, the agent's expected utility in period t can be expressed as:

$$U(p(H_t), H_t)=\sum_{j=1}^{J} p_j(H_t)[u_j^*(H_t)+R_j(p(H_t), H_t)]. \tag{2.12}$$

It follows that the conditional valuation function, $V_j(H_t)$, can be expressed as the following function of future choice probabilities and H_t:

$$V_j(H_t)=\sum_{s=t+1}^{T} E[U(H_s, p(H_s))\mid d_{tj}=1, H_t], \tag{2.13}$$

where the expectation on the right-hand side of (2.13) is taken over future histories, H_s for $s\in\{t+1, \ldots, T\}$.[7] In general, all of the time-varying expressions in summation in (2.13) must be determined in order to characterize the conditional valuation of action j in period t. Hotz and Miller (1993) show that for models in which the terminal state property (described in Section 1) holds, the formulation of (2.13) can be simplified. In particular, the conditional valuation asociated with a terminating action, action J, say, takes the following form:

$$V_J(H_t)=\sum_{s=t+1}^{T} E[E(u_{sJ}\mid H_s)\mid H_t], \tag{2.14}$$

and the valuations associated with all other actions j, $j\in\{1, \ldots, J-1\}$, can be expressed as:

$$\begin{aligned} V_j(H_t)=E[&U(p(H_{t+1}), H_{t+1})+V_J(H_{t+1}) \\ &+\sum_{k=1}^{J-1} p_k(H_{t+1})q_k(p(H_{t+1}), H_{t+1})\mid d_{tj}=1, H_t]. \end{aligned} \tag{2.15}$$

The essential feature induced by the terminal state property is that (2.14) and (2.15) do not depend upon future choices beyond period $t+1$; consequently, one needs only determine the choice probabilities associated with period $t+1$, a fact which greatly reduces the computational burden of calculating the conditional valuation functions.

Many models, however, do not possess terminal states. Consider, for example, the job-matching model in Miller (1984) in the case where there are just two jobs. Suppose the value of a match is revealed through experience on the job and a person maximizes his expected sum of discounted utility, or its monetary equivalent, by sequentially choosing between jobs. In Miller's setup, u_{tj} is assumed to be a normally distributed random variable with mean $\beta^t\gamma_{tj}$ and standard deviation $\beta^t\delta_{tj}$, where $\beta\in(0, 1)$ is some discount factor. In this case, u_{tj} represents the (discounted) utility that an agent receives in period t from working in job $j\in\{1, 2\}$. His beliefs about the job are characterized by $(\gamma_{tj}, \delta_{tj})$, which are updated over time according to Bayes' formula:

$$\begin{aligned} \gamma_{tj}&=\frac{\gamma_{t-1,j}\delta_{t-1,j}^{-2}+d_{t-1,j}u_{t-1,j}\sigma_j^{-2}}{\delta_{t-1,j}^{-2}+d_{t-1,j}\sigma_j^{-2}} \\ \delta_{tj}&=(\delta_{t-1,j}^{-2}+d_{t-1,j}\sigma_j^{-2})^{-2}, \end{aligned} \tag{2.16}$$

7. In independent work, Manski (1993) develops a class of discrete-choice models, nested in the framework laid out by Hotz and Miller (1993), and applicable to situations where the utility an agent would have received from actions not chosen does not depend on unobservables. In particular, Manski avoids the censoring problem that optimizing behaviour typically induces when unobservables are present, by not including the $R_j(p(H_t), H_t)$ terms in (2.12), and, thus, in (2.13).

where σ_j^2 is the variance of the noise about the unknown job-specific match quality. In this setting, the vector $(\gamma_{t1}, \gamma_{t2}, \delta_{t1}, \delta_{t2})$ is a sufficient statistic for H_t, u_{tj}^* reduces to $\beta^t\gamma_{tj}$ and $\varepsilon_{tj}=u_{tj}-\beta^t\gamma_{tj}$ is a normally distributed random variable with mean 0 and variance $\beta^{2t}(\delta_{tj}^2+\sigma_j^2)$. It follows that $v(H_t)=q(p(H_t), H_t)$ is the real-valued function:

$$q(p(H_t), H_t)=\beta^t\{(\gamma_{t2}-\gamma_{t1}+\delta_t^{1/2}\Phi^{-1}(p(H_t))\}, \tag{2.17}$$

and the dynamic selection correction term is:

$$R(p(H_t), H_t)=\beta^t[\delta_t^{-1/2}(\delta_{t1}^2+\sigma_1^2)]\frac{\phi(\Phi^{-1}(p(H_t)))}{p(H_t)} \tag{2.18}$$

where $\delta_t^2\equiv(\sigma_1^2+\sigma_2^2+\delta_{t1}^2\delta_{t2}^2)$, $\phi(\cdot)$ is the standard normal density function, and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function, evaluated at any $p\in(0, 1)$. In this model, the potential for changing jobs declines over time but it never disappears entirely. Therefore, neither of the jobs represents a terminal state.

A second example of a dynamic, discrete-choice model which does not possess the terminal state property is the model of welfare and labour force participation in Sanders (1993). In this model, a woman chooses whether to work in the labour force and whether to accept benefits from a public welfare programme in periods $t\in\{1, 2, \ldots, T\}$. Let $D_{t1}=1$ if the woman works in the labour force in period t and $D_{t1}=0$ otherwise; also, let $D_{t2}=1$ if she accepts welfare benefits (valued at W) and $D_{t2}=0$, otherwise. That is, in terms of the notation used in our general framework, a woman's period t choice vector, $d_t\equiv(d_{t1}, \ldots, d_{t4})'$, where:

$$\begin{aligned} d_{t1}&=D_{t1}D_{t2} & d_{t2}&=D_{t1}(1-D_{t2}) \\ d_{t3}&=(1-D_{t1})D_{t2} & d_{t4}&=(1-D_{t1})(1-D_{t2}) \end{aligned} \tag{2.19}$$

If the woman participates in the work force, she increases her current and future income prospects (the latter through a "learning-by-doing" human capital production process), but her current income is taxed at a (proportional) rate τ if she accepts welfare. We denote the woman's net wage earnings in period t, by $D_{t1}(1-\tau D_{t2})[\sum_{s=0}^{t-1}\gamma_s D_{t-s,1}]$, where γ_s measures the return to current earnings from working s periods ago. It follows that the vector $(d_{t-1}, d_{t-2}, \ldots, d_o)$ is a sufficient statistic for H_t. In our simplified version of Sanders' model, the woman's per period utility for each of alternative choices is:

$$\begin{aligned} u_{tj}=\beta^t(&\{(1-\tau d_{t1})[\textstyle\sum_{s=0}^{t-1}\gamma_s(d_{t-s,1}+d_{t-s,2})]-\alpha_1\}(d_{t1}+d_{t2}) \\ &+(W-\alpha_2)(d_{t3}+d_{t4})+\varepsilon_{tj}) \end{aligned} \tag{2.20}$$

where α_1 denotes the amount by which working lowers current utility (due to the loss of leisure time), α_2 is the amount by it is lowered if she accepts welfare benefits (due to the effects of stigma) and ε_{tj} represents a choice-specific unobservable utility component which is assumed to be independently distributed over choices and time periods according to a Type I Extreme Value distribution with location parameter of 0. The woman's optimization problem is to maximize the expected value of the sum of future period payoffs of the form in (2.20) by sequentially choosing d_t over her T-period lifetime. Given the distribution of the ε_{tj}'s, it follows that:

$$q_j(p(H_t), H_t)=\beta^t\ln(p_j(H_t)/p_J(H_t)]) \tag{2.21}$$

$$R_j(p(H_t), H_t)=\beta^t[\gamma-\ln(p_j(H_t))] \tag{2.22}$$

for $j \in \{1, 2, 3\}$, where $J = 4$ and γ is Euler's constant (≈ 0.577). Because of the recurring option of working in the labour force and the human capital accumulation process in this model, there are no terminal states. Sanders estimates this model using the CCS estimator developed below.

These two examples, plus Rust's (1987) renewal model considered in Section 5, represent models in the literature which lack the terminal state property Hotz and Miller (1993) exploited in their empirical study of contraceptive choices over the life cycle. This paper shows how their inversion theorem can be exploited in models which lack terminal states.

3. THE CCS ESTIMATOR

We now define the conditional choice simulation (CCS) estimator of the structural parameters associated with models of the class described in the previous section and characterize its asymptotic properties. Consider a cross-section of N agents (of different ages) drawn from a population in (calendar) period t whose behaviour is characterized by such a model. Adding an additional subscript to denote observations in the sample, let H_{nt}, d_{nt}, and b_{nt}, respectively, denote the history, choice, and realized outcome for the n-th agent in the sample in period t. Let A_{nt} denote the age of the n-th agent of period t and assume that all agents have a finite life of length T. We also introduce a $Q \times 1$ vector of parameters, denoted by $\theta \in \Theta$, which characterizes agents' preferences and which are the object of estimation.

The set of assumptions used to establish the large sample properties of the CCS estimator is as follows:

Assumption 1. $\mathscr{B}_0$ and $\mathscr{B}$ are finite sets with K and L elements, respectively. Since there are KL^{s-1} feasible histories leading up to period s, summing over $s \in \{1, \ldots, T\}$, it follows that $M = K(L^T - 1)/(L - 1)$. Accordingly, let $\psi_0 \equiv (P_0', F_0')'$, a $M(JK - 1) \times 1$ vector, denote the true values of the conditional choice and transition probabilities associated with the feasible histories.

Assumption 2. The probability distribution function for ε_{nt} may depend on θ_0, where $G(\varepsilon_{nt} \mid H_{nt}, \theta_0) \equiv G(\varepsilon_{nt} \mid H_{nt})$. Similarly, $u^*(H_{nt}, \theta_0) \equiv u^*(H_{nt})$ and $U(p(H_{nt}), H_{nt}, \psi_0, \theta_0) \equiv U(p(H_{nt}), H_{nt})$. Both $dG(\varepsilon_{nt} \mid H_t, \theta)$ and $u^*(H_t, \theta)$ are differentiable in θ.

Assumption 3. The $Q \times 1$ vector, θ_0, belongs to the interior of a closed compact set Θ.

Assumption 4. The population lives in a stationary environment. Consequently, the distribution functions generating H_{nt} and ε_{nt} are invariant over calendar times $t \in \{1, 2, \ldots\}$.

Assumption 5. The data consists of the finite sequence $\{H_{nt}, d_{nt}\}_{n=1}^{N}$ sampled independently over the population (random sample).[8]

8. Although the estimator is defined for a single cross-section (that is, one decision per agent coupled with his outcome history), it also applies to a panel (subject to the assumptions listed in the text). Since there are no aggregate shocks or unobserved state variables than carry over more than one period, all unobserved heterogeneity is specific to each person-calendar time pair, (n, t). Consequently, a panel of N' people over T' time periods is no more than a T' sequence of cross-sections, and, thus, has the same finite distributional properties as would a single cross-section sample of size $N'T'$. As such, the observational unit in a panel data set is the coordinate pair (n, t) and not the n-th agent.

The finiteness restriction on the state variables in Assumption 1 is used to apply results from Pakes and Pollard (1989) on the asymptotic properties of estimators using simulation methods to our context. Assumption 2 allows for the dependence of the probability distributions and the per period payoffs on θ_0 and, along with Assumption 3, provides regularity conditions on the functions and parameter space needed to establish consistency and asymptotic properties. Taken together, Assumptions 4 and 5 rule out the existence of unobserved state variables and decisions, the possibility of common or aggregate variation over (calendar) time, the existence of cohort differences across agents, and the possibility of serially correlated unobservables. These assumptions enable us to (synthetically) form cohorts from cross-sectional data on agents of different ages which we then use to form estimates of future choice and transition probabilities.

To estimate θ_0, we proceed in two stages. The first stage recursively simulates the future paths associated with taking each available action $k \in J$, using consistent estimates of the conditional choice probabilities in P_0 and transition probabilities in F_0 based on the relative frequencies of choices and outcome transitions observed in the data. We then form the expected discounted utilities associated with these simulated paths, as functions of θ, in order to estimate the conditional valuation functions of the actions taken in period t. In the second stage, we estimate θ_0 by minimizing a function of the orthogonality conditions associated with condition (2.9), substituting the simulated values for v_{nt}.

Consider the first stage. For each agent n, we simulate future paths associated with having chosen each of the actions $j \in J$ in period t. Suppose $d_{ntk}=1$. Given this choice, we first generate b_{nt} by partitioning the unit interval into L segments of length $F_k^{(N)}((H_{nt}, b^{(l)}) \mid {}_{nt})$, $l=1, \ldots, L$, where $F_k^{(N)}((H_{nt}, b^{(l)}) \mid H_{nt})$ is an estimate of $F_k((H_{nt}, b^{(l)}) \mid H_{nt})$, the probability of realizing $b^{(l)}$, given choice k and history, H_{nt}. To estimate these transition probabilities, one can use cell estimators of the following form:

$$F_k^{(N)}((H_{nt}, b^{(l)}) \mid H_{nt}) = \frac{\sum_{m=1}^{N} 1\{H_{mt}=H_{nt},\, b_{m,t+1}=b^{(l)}\} d_{mtk}}{\sum_{m=1}^{N} 1\{H_{mt}=H_{nt}\} d_{mtk}}, \; l=1, \ldots, L, \tag{3.1}$$

and where $1\{\cdot\}$ denotes the indicator function which equals 1 if the statement inside the parenthesis is true and 0 otherwise. Then the hypothetical outcome associated with choosing action k in period t, $b_{nt}^{(k,N)} \in \mathscr{B}$, is found by drawing a random variate, $\eta_{nt}^{(k,1)}$, from the (0, 1) uniform distribution and defining $b_{nt}^{(k,N)}$ according to:

$$b_{nt}^{(k,N)} = \sum_{i=1}^{L} b^{(i)} 1\left\{\sum_{l=0}^{i-1} F_k^{(N)}((H_{nt}, b^{(l)}) \mid H_{nt}) \leqq \eta_{nt}^{(k,1)} < \sum_{l=1}^{i} F_k^{(N)}((H_{nt}, b^{(l)}) \mid H_{nt})\right\}. \tag{3.2}$$

Conditional on history $H_{n,t+1}^{(k,N)} \equiv (H_{nt}, b_{nt}^{(k,N)})$, we next simulate the choice in period $t+1$ (when the n-th agent would be age $A_{nt}+1$). To do so, we make use of estimates of the conditional choice probabilities, $p(H_{m,t+1}^{(k,N)})$. Again, consider using a cell estimator defined as follows:

$$p^{(N)}(H_{n,t+1}^{(k,N)}) = \frac{\sum_{m=1}^{N} 1\{H_{mt}=H_{n,t+1}^{(k,N)}\} d_{mt}}{\sum_{m=1}^{N} 1\{H_{mt}=H_{n,t+1}^{(k,N)}\}} \tag{3.3}$$

Partitioning the unit interval into J segments of length $p_j^{(N)}(H_{n,t+1}^{(k,N)})$, $j=\{1,\ldots,J-1\}$, the hypothetical choice, $d_{n,t+1}^{(k,N)}$, is simulated by drawing a second random variable, $\eta_{n,t+1}^{(k,2)}$, from the (0, 1) uniform distribution and setting

$$d_{n,t+1}^{(k,N)}=\begin{pmatrix} 1\{0<\eta_{n,t+1}^{(k,2)}\leqq p_1^{(N)}(H_{n,t+1}^{(k,N)})\} \\ 1\{p_1^{(N)}(H_{n,t+1}^{(k,N)})<\eta_{n,t+1}^{(k,2)}\leqq\sum_{j=1}^{2}p_j^{(N)}(H_{n,t+1}^{(k,N)})\} \\ \vdots \\ 1\{\sum_{j=1}^{J-1}p_j^{(N)}(H_{n,t+1}^{(k,N)})<\eta_{n,t+1}^{(k,2)}\leqq 1\} \end{pmatrix}. \tag{3.4}$$

Given $d_{n,t+1}^{(k,N)}$, a $t+1$ outcome, $b_{n,t+1}^{(k,N)}$, is generated by taking another random draw, $\eta_{n,t+1}^{(k,1)}$, from the $(0,1)$ uniform distribution, using cell estimates for $F_k^{(N)}((H_{nt},b^{(l)})\mid H_{nt})$, $l=\{1,\ldots,L\}$ associated with that choice, and calculating (3.2). Then the hypothetical choice vector $d_{n,t+2}^{(k,N)}$ is simulated for period $t+2$. Continuing in this manner, we successively simulate outcomes and choices for each period through $t+T-A_{nt}$, using the two sequences of random variates, $\{\eta_{nt}^{(k,1)},\ldots,\eta_{n,t+T-A_{nt}-1}^{(k,1)}\}$ and $\{\eta_{n,t+1}^{(k,2)},\ldots,\eta_{n,t+T-A_{nt}}^{(k,2)}\}$. This process generates the sequence of histories, $\{H_{n,t+1}^{(k,N)},\ldots,H_{n,t+T-A_{nt}-1}^{(k,N)}\}$, associated with choosing k in period t. The above strategy for simulating such histories is repeated for each of the remaining possible actions $j\in\{1,\ldots,J-1\}$ which might be taken by the n-th agent in period t.

From these sequences we assign values to the lifetime utility differential between each action $k\in\{1,\ldots,J-1\}$ and the base action J, for any value of $\theta\in\Theta$ from the simulated $H_{n,s}^{(k,N)}$ histories. Writing $\psi^{(N)}$ for the cell estimates of the conditional choice and transition probabilities obtained from (3.1) and (3.3), we now define $v_k(\chi_n,\psi^{(N)},\theta)$, the simulated lifetime differential associated with taking action k versus J in period t, as:

$$v_k(x_n,\psi^{(N)},\theta)=\sum_{s=t+1}^{t+T-A_{nt}}[U(p^{(N)}(H_{ns}^{(k,N)}),H_{ns}^{(k,N)},\theta)-U(p^{(N)}(H_{ns}^{(J,N)}),H_{ns}^{(J,N)},\theta)], \tag{3.5}$$

where $x_n\in X$ denotes a vector associated with person n at date t, whose components are his age A_{nt}, his history at t, H_{nt} (as recorded in the data), and the individual specific realizations of the random variables $\{\eta_{nt}^{(k,1)},\ldots,\eta_{n,t+T-A_{nt}-1}^{(k,1)}\}$ and $\{\eta_{n,t+1}^{(k,2)},\ldots,\eta_{n,t+T-A_{nt}}^{(k,2)}\}$ that determine future hypothetical choices and outcomes for each action $k\in 1,\ldots,J$ he might have taken in period t.[9] In this equation, $U(p^{(N)}(H_{ns}^{(i,N)}),H_{ns}^{(i,N)},\theta)$ denotes the (simulated) expected utility person n would receive in period s by choosing action $i\in\{1,\ldots,J\}$ if he had accumulated history, $H_{ns}^{(i,N)}$, and the choice probabilities associated with that history were $p^{(N)}(H_{ns}^{(i,N)})$.

Define for each $n\in N$, the $(J-1)\times 1$ vector of these differentials as:

$$v(x_n,\psi^{(N)},\theta)\equiv(v_1(x_n,\psi^{(N)},\theta),\ldots,v_{J-1}(x_n,\psi^{(N)},\theta))'. \tag{3.6}$$

While the $J-1$ utility differentials $v_k(x_n,\psi^{(N)},\theta)$ vary with θ, the future paths, as determined by the simulated outcomes and choices, only vary as the sample changes. As we mentioned in the Introduction, this implies the simulated paths are computed only once for a given sample and are not simultaneously determined with the CCS estimator for θ_0.

9. It is convenient to define x_n as a vector of the same length for all observations. To do so, let x_n be represented as a $[1+M+J2T]$-dimensional vector equal to $(A_{nt},h_n',\eta_{n1}',\ldots,\eta_{nJ}')\prime$, where: (i) h_n is a vector of length M, the number of possible realizations of H, in which that element indexing the realization H_{nt} being equal to 1 and all the other elements set equal to 0 and (ii) η_{nk} is a vector of length $2T$, whose first elements $(T-A_{nt})$ elements are $\{\eta_{nt}^{(k,1)},\ldots,\eta_{n,t+T-A_{nt}-1}^{(k,1)}\}$, the next A_{nt} elements are equal to 0, the next $(T-A_{nt})$ elements are $\{\eta_{n,t+1}^{(k,2)},\ldots,\eta_{n,t+T-A_{nt}-1}^{(k,2)}\}$, and the remaining A_{nt} elements are set equal to 0.

Having simulated the differences in conditional valuation functions for the sample, we turn to the estimation of θ_0, the second stage of our estimation strategy. Let z_{nt} denote an $R \times 1$ vector of instruments and define the following (vector) function:

$$f(x_n, \psi, \theta) = z_{nt} \otimes [\textstyle\sum_{i=1}^{M} q(p^{(i)}, H_{nt}, \theta) 1\{H_{nt} = H^{(i)}\} - v(x_n, \psi, \theta)], \tag{3.7}$$

where $H^{(i)}$ denotes the history associated with the choice $p^{(i)}$, $i = 1, \ldots, M$. The componenents of z_{nt} must be (or converge to) random variables which are orthogonal to the difference between $q(p^{(i)}, H_{nt}, \theta)$ and $v(x_n, \psi, \theta)$; for example, the elements of H_{nt} meet such a requirement. For purposes of identification, we make one further assumption, namely:

Assumption 6. θ_0 is the unique solution to $E[f(x_n, \psi_0, \theta)] = 0$.

The CCS estimator, denoted $\theta^{(N)}$, minimizes a quadratic function in the sample analogues of (3.7) evaluated at $\psi^{(N)}$. On average, this estimator sets the simulated utility paths close to the corresponding functions, $q(p_n^{(N)}(H_{nt}), H_{nt}, \theta)$, given in (2.9), where the latter are evaluated at estimates of $p_n(H_{nt})$.[10] More formally, let W_N denote a $(J-1)R$-dimensional square weighting matrix which converges to a constant matrix, W. Then $\theta^{(N)} \in \Theta$ minimizes:

$$[N^{-1} \textstyle\sum_{n=1}^{N} f(x_n, \psi^{(N)}, \theta)]' W_N [N^{-1} \textstyle\sum_{n=1}^{N} f(x_n, \psi^{(N)}, \theta)]. \tag{3.8}$$

Establishing the consistency and asymptotic distribution of $\theta^{(N)}$ is complicated by the use of simulators for estimating the conditional valuation functions and the fact that these simulators are based on estimated values of ψ_0. The complication arises because, as N increases, changes in $\psi^{(N)}$ cause the simulated differentials, $v(x_n, \psi^{(N)}, \theta)$, to change in discontinuous ways. To deal with these jumps, we exploit the results on the asymptotic properties of simulation estimators in Pakes and Pollard (1989). The proofs are found in the Appendix, where we use the estimators of F_0 and P_0 given in (3.1) and (3.3), respectively, to estimate these incidental parameters. More precisely, the Appendix proves the following proposition for the CCS estimator:

Proposition 1. *$\theta^{(N)}$ converges in probability to θ_0 and $N^{1/2}(\theta^{(N)} - \theta_0)$ converges in distribution to a normal random variable with mean 0 and covariance matrix*

$$(\Gamma_{11}' W \Gamma_{11})^{-1} \Gamma_{11}' W' E(f_n f_n') W \Gamma_{11} (\Gamma_{11}' W \Gamma_{11})^{-1} + (\Gamma_{11}' W \Gamma_{11})^{-1} \Gamma_{12}' \Omega \Gamma_{12} (\Gamma_{11}' W \Gamma_{11})^{-1}, \tag{3.9}$$

where $f_n \equiv f(x_n, \psi_0, \theta)$, Γ_{11} is the $R(J-1) \times Q$ matrix,

$$\Gamma_{11} = E(z_{nt} \otimes \partial [q(p_{nt}^{(N)}, H_{nt}, \theta) - v(x_n, \psi_0, \theta_0)] / \partial \theta),$$

Γ_{12} is the $R(J-1) \times M(JK-1)$ matrix,

$$\Gamma_{12} = E(z_{nt} \otimes \partial [q(p_{nt}^{(N)}, H_{nt}, \theta) - v(x_n, \psi_0, \theta_0)] / \partial \psi),$$

and Ω is the $M(JK-1)$ dimensional covariance matrix of $\psi^{(N)}$ which is defined in the Appendix.

10. The form of this estimator represents a generalization of the Berkson–Theil estimator for estimating a logistic regression model with discrete regressors.

As can be seen in (3.9), there are two components to the covariance matrix of $\theta^{(N)}$. The first component, $(\Gamma'_{11}W\Gamma_{11})^{-1}\Gamma'_{11}W'E(f_n f'_n)W\Gamma_{11}(\Gamma'_{11}W\Gamma_{11})^{-1}$, arises from simulating the conditional valuation function rather evaluating all of the possible future choice probabilities which are feasible given H_{nt}. The second component, $(\Gamma'_{11}W\Gamma_{11})^{-1}\Gamma'_{12}\Omega\Gamma_{12}(\Gamma'_{11}W\Gamma_{11})^{-1}$, is due to sampling error, which is transmitted to the covariance of $\theta^{(N)}$ via the preliminary estimation of ψ_0, the conditional choice and transition probabilities. Consistent estimates of the terms in this covariance matrix, namely Γ_{11}^{-1}, $E(f_n f'_n)$, and Ω, can be formed using their corresponding sample analogues, evaluated at $(\psi^{(N)}, \theta^{(N)})$. In the case of Γ_{12}, we perturb ψ around $\psi^{(N)}$, simulate the valuation functions at the perturbed values and then calculate the changes in the estimated parameters, for each observation $n\in N$, in order to form a consistent estimate.

The precision of this estimator can be tightened by conducting more than one (set of) simulation(s). In the first stage, suppose we now independently simulate S future paths for each choice $j\in J$ (starting at period t), rather than constructing just one simulated path. Then, following McFadden (1989), it is straightforward to show that the only change in the asymptotic covariance matrix would be to replace $E(f_n f'_n)$ in (3,9) with $S^{-1}E(f_n f'_n)$. (For example, the first component of the covariance matrix falls to one half of its value when S is doubled.) Indeed, as S goes to ∞ (which is equivalent to calculating the expected utility of the remaining lifetime), the covariance matrix of $\theta^{(N)}$ converges to $(\Gamma'_{11}W\Gamma_{11})^{-1}\Gamma'_{12}\Omega\Gamma_{12}(\Gamma'_{11}W\Gamma_{11})^{-1}$, the second component in (3.9), which is the covariance matrix for the CCP estimator in Hotz and Miller (1993).

Finally, we note that several variations on the proposed estimator will also yield consistent (and asymptotically normal) estimates of θ_0. For example, in forming the $q(\cdot)$ functions, one could use any number of consistent estimators of F_0 and P_0 when forming (3.7), including the (smooth) simulators proposed by McFadden (1989) for discrete-choice models. Below, we report on results for several alternatives in our Monte Carlo study. In addition, alternative ways of estimating $v_k(H_{nt}, \psi_0, \theta_0)$ can be employed. For example, one could use either:

$$\sum_{j=1}^{J} d_{nsj}^{(i,N)}[u_j^*(H_{ns}^{(i,N)}, \theta)+R_j(p_s^{(N)}(H_{ns}^{(i,N)}), H_{ns}^{(i,N)}, \theta)]$$

or

$$\sum_{j=1}^{J} d_{nsj}^{(i,N)}[u_j^*(H_{ns}^{(i,N)}, \theta)+\varepsilon_{nsj}^{(i,N)}(\theta)]$$

in place of $U(p^{(N)}(H_{ns}^{(i,N)}), H_{ns}^{(i,N)}, \theta)$, in (3.5), where $d_{nsj}^{(i,N)}$, $j=1,\ldots,J$ are the elements of $d_{ns}^{(i,N)}$ and $\varepsilon_{nsj}^{(i,N)}(\theta)$ is a random draw from the probability distribution, $G(\varepsilon|H_{t+1}, \theta)$. Since only one of the elements in $d_{ns}^{(i,N)}$ is equal to 1 and the remaining are zeros, using either of the above expressions will generally require less computation than is involved in forming the expression for $U(p^{(N)}(H_{ns}^{(i,N)}), H_{ns}^{(i,N)}, \theta)$ in (3.5). This will be especially true when the set of actions (J) is large. The proof strategy for consistency and asymptotic normality of $\theta^{(N)}$ for these variants on the above estimator follows along similar lines to the one in the Appendix.

4. THE CCS ESTIMATOR IN INFINITE-HORIZON MODELS

With minimal work, the CCS estimator and the asymptotic properties just established can be extended to an important class of Markov models that have an infinite horizon. To demonstrate this, we make some notational changes, establish that the inversion theorem of Hotz and Miller (1993) still holds, define the CCS estimator in this new context, and extend Proposition 1 of this paper to cover this case. The latter is contained in Proposition 2 which concludes this section.

Much of the notation in the previous two sections remains intact. Rather than defining H_t as the actual history of outcomes as of the beginning of period t (along with the agent's initial conditions), we now interpret H_t as a sufficient statistic for that history and, in this way, maintain the assumption that the cardinality of $\mathscr{H}$ is finite. Therefore, as before, (H_t, ε_t) is the set of state variables upon which the agent's decisions are based. Similarly, the probability distribution functions, $G(\varepsilon \mid H_{nt})$ and $F(H_{n,t+1} \mid H_{nt})$ retain their previous meaning. However, rather than objective (2.3), we now assume that a typical agent maximizes:

$$E(\textstyle\sum_{s=t}^{\infty} \sum_{j=1}^{J} d_{sj}\beta^s[u_j^*(H_t) + \varepsilon_{sj}] \| H_t) \tag{4.1}$$

where $\beta \in (0, 1)$ is a constant discount factor. Assumptions 1 through 6, given in Section 3, remain unaltered.

The main difference between the CCS estimator in the infinite- versus finite-horizon case is the end point for the simulations. In the finite-horizon case, the sequences of actions are simulated up to the last period of each agent's life (T) and these sequences are used to impute the lifetime utility realizations that, in expectation, equal the conditional valuation functions when evaluated at the true parameter values. In practice, it is impossible to simulate the actions of an infinitely-lived agent. We provide two ways of circumventing this issue. As we demonstrate below, which alternative is chosen will typically depend on the computational aspects of the application at hand.

Analogous to the standard approach for solving infinite-horizon models of dynamic programming, we could approximate the utility achieved over an infinite lifetime with a truncated finite-horizon counterpart. Actions and outcomes are simulated for a finite number of time periods, denoted by T^*. To implement this version of the CCS estimator, we merely replace (3.5), which is an expression for the simulated difference of the conditional valuation functions, with:

$$\begin{aligned} v_k(x_n, \psi^{(N)}, \theta) = {} & \textstyle\sum_{s=t+1}^{t+T^*} \beta^{s-t} \\ & \times [U(p^{(N)}(H_{ns}^{(k,N)}), H_{ns}^{(k,N)}, \theta) - U(p^{(N)}(H_{ns}^{(J,N)}), H_{ns}^{(J,N)}, \theta)], \end{aligned} \tag{4.2}$$

The main drawback of this particular CCS estimator is that when β is close to 1, many periods must be included to ensure the properties of the estimator are not unduly affected by the finite-horizon approximation. Whereas ML also suffers from this deficiency, the alternative we now propose does not.

Instead of starting with any action $k \in \{1, \ldots, J\}$ and simulating outcomes and future choices until the expected value of remaining lifetime utility (discounted back to the present) is negligible, suppose we simulate only until the current history is revisited. Let $t_{kn}(N) > t$ denote the period in which the current state is first revisited and let $T_{kn}(N)$ be the first passage time back to it. Thus, the current state is revisited again in period $[t + T_{kn}(N)]$. The simulated sequence of realizations leading from period t to period $[t + T_{kn}(N)]$ is then repeated, ad infinitum, as a deterministic cycle with a periodicity of $T_{kn}(N)$ periods. Hence, the realized value of lifetime utility is just the realized value of utility received in the next $T_{kn}(N)$ periods, scaled up by the factor $(1 - \beta^{T_{kn}(N)})^{-1}$. By construction, the expectation of this stream of realized utilities equals the corresponding conditional valuation function, as in the finite-horizon model we analyzed in the previous sections. In this case, (3.5) becomes:

$$v_k(x_n, \psi^{(N)}, \theta)=\sum_{s=t+1}^{t_{kn}(N)} (1-\beta^{T_{kn}(N)})^{-1}\beta^{s-t} \times [U(p^{(N)}(H_{ns}^{(k,N)}), H_{ns}^{(k,N)}, \theta)-U(p^{(N)}(H_{ns}^{(J,N)}), H_{ns}^{(J,N)}, \theta)]. \tag{4.3}$$

Which version of the CCS estimator one wishes to apply depends on the specific nature of the application at hand. If the period length is quite short (which implies little discounting between adjacent periods), and there are only a few relevant features differentiating histories (meaning that M is quite small), the second version of the CCS estimator might be more practical. In addition, one could combine the two strategies for simulating lifetime utilities, by adopting the second approach for histories that occur relatively frequently and the first approach for other histories.

We summarize the above results for the estimation of infinite-horizon models in the following proposition:

Proposition 2. *Suppose that agents' preferences are characterized by* (4.1) *instead of* (2.6). *Then the inversion result of Proposition* 1 *in Hotz and Miller* (1993) *still applies. In addition, by replacing* (3.5) *with* (4.3), *the properties of the CCS estimator summarized in Proposition* 1 *of this paper apply.*

5. A MONTE CARLO STUDY

This final section investigates the small-sample properties of our estimator by undertaking a Monte Carlo study of simulated data based on the model of engine replacement in Rust (1987).[11] Because a full description of the model is given in Rust (1987), we confine ours to the essentials, concentrating on the econometric aspects.

In each month t, the owner (manager) decides whether or not to replace a bus's engine so as to minimize the discounted costs of maintaining it; thus, $J=2$. Let d_{nt1} index the action of replacing the n-th bus engine in month t, where $d_{nt1}=1$ if it is replaced and 0 otherwise and $d_{nt2}=(1-d_{nt1})$. The discounted monthly cost associated with maintaining a given bus engine is assumed to depend on H_{nt}, its accumulated mileage, in the following way:

$$u_{nt1}=-\beta^t(\theta_{01}+\theta_{02}\, H_{nt}+\varepsilon_{nt1}) \tag{5.1}$$

$$u_{nt2}=-\beta^t(\theta_{02}\, H_{nt}+\varepsilon_{nt2}), \tag{5.2}$$

where $\beta\in(0,1)$ is the discount factor, θ_{01} is a parameter indexing the (fixed) cost of replacing an engine and θ_{02} is a parameter indexing the variable cost per accumulated miles of operating a bus and ε_{ntj}, is a stochastic cost component assumed to be identically and independently distributed across (n, j, t) as Type I Extreme Value with location parameter 0. The law of motion governing mileage accumulation is:

$$H_{n,t+1}=b_{nt}+(1-d_{nt1})\, H_{nt}, \tag{5.3}$$

where b_{nt} is the (stochastic) mileage realized in month t. As in Rust (1987), we assume that b_{nt} is an independent and identically distributed random variable with (fixed)

11. A copy of our computer program, complete with Rust's bus replacement problem and a version of Sanders' welfare participation model, is available upon request from Professor Seth Sanders, The Heinz School, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

discrete support. In particular, $b_{nt} \in \{0, 1, 2\}$ and, although technically inconsistent, we follow Rust in assuming that H_{nt} takes on one of a discrete set of values representing fixed-length mileage intervals, so that $H_{nt} \in \{1, 2, \ldots, 90\}$. The transition probability function governing b_{nt} is of the form, $F(b^{(i)})$, i.e., the probability of realizations does not depend upon accumulated mileage. The bus manager sequentially chooses $\{d_{ns1}\}_{s=0}^{\infty}$ to maximize:

$$E_0\{\textstyle\sum_{s=0}^{\infty} -\beta^s[(\theta_{01}+\theta_{02}H_{ns}+\varepsilon_{ns1})d_{ns}+(\theta_{02}H_{ns}+\varepsilon_{ns2})(1-d_{ns1})]\}, \tag{5.4}$$

subject to (5.3) and the transition probability for H_{ns}. Denote by d_{nt1}^0 the optimal decision in month t which one can easily show depends only on $(H_{nt}, \varepsilon_{nt1}, \varepsilon_{nt2})$.

Adopting the CCS estimation strategy developed in Section 3, we simulate paths for the two choices for each observation in a random sample buses representing realizations of H_{nt} and d_{nt1}^0. Given the distribution of the ε_{tj}'s, it follows that the expressions for $q_j(p^{(N)}(H_{nt}), H_{nt})$ and $R_j(p_1^{(N)}(H_{nt}), H_{nt}, \theta), j=1, 2,$ are given in (2.21) and (2.22), and that the representation in (2.13) for montly costs, evaluated at some $\theta\in\Theta$, reduces to:

$$\begin{aligned} U(p_1^{(N)}(H_{nt}), H_{nt}, \theta) = -(0.9)^t\{&p_1^{(N)}(H_{nt}^{(1,N)})[\theta_1+\theta_2 H_{nt}^{(1,N)}+\gamma-\ln(p_1^{(N)}(H_{nt}^{(1,N)}))] \\ &+(1-p_1^{(N)}(H_{nt}^{(2,N)}))[\theta_2 H_{nt}^{(2,N)}+\gamma-\ln(1-p_1^{(N)}(H_{nt}^{(2,N)}))]\}, \end{aligned} \tag{5.5}$$

and the simulated difference in valuation functions in (3.5) specializes to:

$$v_1(x_n, F_0, P^{(N)}, \theta) = x_{n0}+x_{n1}\theta_1+x_{n2}\theta_2, \tag{5.6}$$

where

$$\begin{pmatrix} x_{n0} \\ \\ x_{n1} \\ x_{n2} \end{pmatrix} = \begin{pmatrix} -\sum_{s=t+1}^{50}(0.9)^s[(p_1^{(N)}(H_{ns}^{(1,N)})[\gamma-\ln(p_1^{(N)}(H_{ns}^{(1,N)}))] \\ \quad +(1-p_1^{(N)}(H_{ns}^{(2,N)}))[\gamma-\ln(1-p_1^{(N)}(H_{ns}^{(2,N)}))]] \\ -\sum_{\times s=t+1}^{50}(0.9)^s[(1+p_1^{(N)}(H_{ns}^{(1,N)})-p_1^{(N)}(H_{ns}^{(2,N)}))] \\ -\sum_{s=t+1}^{50}(0.9)^s[p_1^{(N)}(H_{ns}^{(1,N)})H_{ns}^{(1,N)}+(1-p_1^{(N)}(H_{ns}^{(2,N)}))H_{ns}^{(2,N)}] \end{pmatrix}$$

$T^*=50$, and $\beta=0{\cdot}9$. Given this restriction, (5.6) is linear in θ, the second-stage estimation problem can be conducted using (weighted) least squares on the 90 cells characterizing the observed histories, H_{nt}. More formally, let:

$$\begin{aligned} y_i &= N^{-1}\textstyle\sum_{n=1}^{N} 1\{H_{nt}=i\}[\ln(p_1^{(N)}(H_{nt}=i)/[1-p_1^{(N)}(H_{nt}=i)])-x_{n0}], \\ x_i &= N^{-1}\textstyle\sum_{n=1}^{N} 1\{H_{nt}=i\}\begin{pmatrix} x_{n1} \\ x_{n2}\end{pmatrix}, \end{aligned} \tag{5.7}$$

for each history $i\in\{1, \ldots, 90\}$. Then, a CCS estimator for θ_0 is:

$$\theta^{(N)} = (\textstyle\sum_{i=1}^{90} \kappa_i^{(N)} x_i' x_i)^{-1}(\textstyle\sum_{i=1}^{90} \kappa_i^{(N)} x_i' y_i). \tag{5.8}$$

where $\kappa_i^{(N)}$ is a weighting factor which converges in probability to some positive constant $k^{(i)}$ for each $i \in \{1, \ldots, 90\}$.[12]

As noted by a referee, there are several severe limitations to this Monte Carlo study. First, in order to compare the small-sample properties of the CCS estimator with the ML estimator, we have picked a specification which could be estimated easily for a large number of different samples using either estimation method. Such a criteria ruled out more complicated models which could only be handled with the CCS estimator. Second, as in Rust (1987), we chose to set β to a number instead of estimating it. This clearly limits the generality of our investigation, since the discount factor plays such a crucial role in evaluating the future payoffs of current actions.[13]

Creating the simulated data used in our Monte Carlo study involved four steps.[14] First, the true incremental mileage probabilities and the true probabilities of bus engine replacement conditional on bus mileage were obtained for each of the sets of underlying structural parameters we investigated. Next, the steady-state distribution of bus mileages was generated for each set of parameters in order to create a set of probabilities for bus mileages at the start of a bus-month. These sets of probabilities were then combined with a random number generator to create samples of bus-months of a selected size. Finally, the simulated data in each sample was aggregated by mileage cell to create a data set containing the cell count and the non-parametric engine replacement and mileage increment probabilities for each cell in each sample for each set of underlying parameters.

The sets of parameters used to generate these samples were as follows. The true transition probabilities for mileage increments were fixed at 0·349, 0·639 and 0·012 for $b_{nt} = 0, 1, 2$, respectively. The bus engine replacement probabilities are obtained by applying Rust's fixed point algorithm to each of the two sets of structural parameters. The first set of structural parameters had $\theta_{01} = 2{\cdot}0$ and $\theta_{02} = 0{\cdot}09$, while the second has $\theta_{01} = 8{\cdot}0$ and $\theta_{02} = 0{\cdot}09$. We refer to these as the "low" and "high" replacement cost regimes, respectively, indexing the differences in the fixed costs of replacing an engine across the two sets.

12. In terms of the notation and results developed in Section 3, the instruments, $z_{nt}^{(N)}$, and sample moments, $f_n(x_n, \psi, \theta)$, for the above problem can be expressed as:

$$z_{nt}^{(N)} = \sum_{i=1}^{90} 1\{H_{nt} = i\}\left[(N^{-1} \sum_{m=1}^{N} 1\{H_{mt} = i\})\kappa_i^{(N)}\begin{pmatrix} x_{n1} \\ x_{n2} \end{pmatrix}\right],$$

$$f_N(x_n, \psi, \theta) = \sum_{i=1}^{90} 1\{H_{nt} = i\}\left[(N^{-1} \sum_{m=1}^{N} 1\{H_{mt} = i\})\kappa_i^{(N)}\begin{pmatrix} x_{n1} \\ x_{n2} \end{pmatrix}\right.$$
$$\left.\otimes\left(\ln\left(\frac{p_1^{(N)}(H_{nt} = i)}{1 - p_1^{(N)}(H_{nt} = i)}\right) - (x_{n0} + x_{n1}\theta_1 + x_{n2}\theta_2)\right)\right],$$

respectively. It is straightforward to show that the Proposition in Section 3 applies to (5.8) where the terms in the covariance matrix given in (3.9) are:

$$f_n \equiv E(x_{nt} \mid H_{nt})\left[\sum_{i=1}^{90} 1\{H_{nt} = i\}\kappa_i\right] \otimes E\left(\ln\left(\frac{p_1^{(N)}(H_{nt} = i)}{1 - p_1^{(N)}(H_{nt} = i)}\right) - x_{n0} - x_{n1}\theta_{01} - x_{n2}\theta_{02}\right),$$

$$\Gamma_{11} \equiv E[E(x_{nt} \mid H_{nt})x_{nt}'],$$

$$\Gamma_{12} = \partial E(f_n)/\partial\psi.$$

13. There is nothing inherent in our method which precludes estimation of β. In fact, Hotz and Miller (1993) do estimate this parameter for a model of contraceptive choice, using the related CCP estimator. In the present context, our primary reason for not estimating β was the intractability it presented for implementing the ML estimator. Because we wanted to compare estimates produced by the latter method with those obtained with our CCS estimator, we resorted to fixing β.

14. We used the program provided to us by John Rust to generate these data sets, as well as to produce the ML estimates of θ_0 presented below.

The fixed-point algorithm provides values of the conditional valuation function corresponding to each set of parameters; these are then employed to obtain the true conditional probabilities.

Before presenting our estimation results, it is useful to briefly describe the characteristics of our simulated data sets. There are two noticeable differences across the low and high replacement cost regimes. The first concerns the replacement probabilities. In the low replacement cost regime, the first mileage category has a replacement probability of almost 12%, with the value rising to 30% around the fortieth cell and to 50% in the last cell. In contrast, in the high replacement cost regime, replacement probabilities begin at around 0·0003, and do not reach 1% until the fortieth mileage category. They peak at around 14% in the last cell. The second difference across the two regimes is in the steady-state distribution of buses over the (accumulated) mileage categories. In the high replacement cost regime, which corresponds roughly to the estimated values in Rust's paper, there are buses in every accumulated mileage category. While there are relatively more buses in the lower mileage categories, the distribution of bus-months is fairly even across the categories, with the number of bus-months per category dropping off fairly gently as one moves to higher mileage categories. However, in the low replacement cost regime, the steady-state distribution of bus-months by mileage is very uneven, with few (or no) buses observed for mileage categories beyond the twentieth and a steep decline in numbers per cell when going from the second on. While in finite samples, an increase in sample size does increase the counts of buses in the sparsely populated mileage categories for the low replacement cost regime, the rate of increase is trivial; in contrast, as sample size increases, the counts per category increase proportionately in the high replacement cost regime. As will be seen below, these two differences (in the replacement rates in certain mileage categories and the distribution of observed bus-months across the categories) across the two regimes play an important role in the success of particular methods used to implement the CCS estimator.

The results of our Monte Carlo investigation for estimating θ_0 are reported in Tables 1 through 4. Tables 1 and 2 present results for the low replacement cost specification using samples of 10,000 and 50,000 bus-months, respectively, while Tables 3 and 4 present the corresponding results for the high replacement cost case. For each estimator, we present the mean parameter estimates (averaged over 100 samples), the estimated asymptotic standard errors (evaluated with data from a single sample), and the corresponding empirical standard errors (using the sample standard deviation of the estimated parameters). All of the CCS estimates used a GLS-based weighting procedure in which:

$$\begin{aligned}\kappa_i^{(N)} &= \frac{G_1(x_{i0}+x_{i1}\theta_1^{(N)}+x_{i2}\theta_2^{(N)})}{[p_1^{(N)}(H=i)(1-p_1^{(N)}(H=i))]^{1/2}}\\ &= N[p_1^{(N)}(H=i)(1-p_1^{(N)}(H=i))]^{1/2}, \quad \text{for } i=1,\ldots,90, \end{aligned} \tag{5.9}$$

where $G_1(\cdot)$ is the density function for ε and $x_{i0}=N^{-1}\sum_{n=1}^{N} 1\{H_{nt}=i\}x_{n0}$.

For the first stage of the CCS estimator, we initially used cell estimators of the form given in (3.3) and (3.1), respectively, to estimate the replacement and mileage increment probabilities. We encountered mileage categories with no bus-months and categories in which there were no replacements when forming the estimates of some of the replacement probabilities. Such categories were not used in forming the second-stage estimator of θ_0 in (5.8), since the corresponding log-odds ratio is not defined when $p_{1i}^{(N)}=0$ or is, itself, undefined. This variant of the CCS estimator is labelled "CCS (Using Replacement Freq. for p_i's)" in the tables.

TABLE 1

Monte Carlo results for low replacement cost specification: $\theta_{01}=2\cdot0$ *and* $\theta_{02}=0\cdot09$ *samples of size* 10,000 *bus-months**

Estimator	Parameter	Mean estimate	Standard error	Emp. stand. dev.
Maximum likelihood	θ_1	2·015	0·048	0·048
	θ_2	0·091	0·019	0·020
CCS (Using replacement freq. for p_i's)	θ_1	1·986	0·052	0·049
	θ_2	0·071	0·020	0·021
CCS (Using true p_i's to form y_i's)	θ_1	1·958	0·017	0·013
	θ_2	0·057	0·008	0·009
CCS (Using Cox correction)	θ_1	1·784	1·685	0·778
	θ_2	0·019	0·145	0·099
CCS (Using kernel for p_i's; $\xi=0\cdot025$)	θ_1	1·795	0·037	0·040
	θ_2	0·022	0·013	0·021
CCS (Using kernel for p_i's; $\xi=0\cdot01$)	θ_1	1·925	0·043	0·046
	θ_2	0·046	0·015	0·020
CCS (Using kernel for p_i's; $\xi=0\cdot005$)	θ_1	1·960	0·046	0·049
	θ_2	0·055	0·016	0·020
CCS (Using kernel for p_i's; $\xi=0\cdot0025$)	θ_1	1·972	0·048	0·049
	θ_2	0·061	0·018	0·021
CCS (Using kernel for p_i's; $\xi=0\cdot001$)	θ_1	1·986	0·052	0·049
	θ_2	0·071	0·020	0·021
CCS (Drop sparse mileage categories)	θ_1	2·000	0·053	0·050
	θ_2	0·080	0·022	0·022

* Results are based on 100 replications of each specification. The estimated (asymptotic) standard errors were computed from the first sample only. All Conditional Choice Simulation (CCS) estimators used the GLS weighting factor described in text.

Consider the results for the low replacement cost case given in Tables 1 and 2. For either sample size, the means of the maximum likelihood (ML) estimates are very close to the true parameter values and are estimated precisely. With respect to the CCS estimator using the cell frequencies to estimate the replacement probabilities, the average estimates for the replacement cost parameter, θ_1, are reasonably close to the true value of 2·0. But the average estimates of the monthly maintenance cost parameter, θ_2, underestimate the true value of 0·09 by 21 and 9%, respectively, for the 10,000 and 50,000 bus-month samples. Finally, note that the average of the CCS estimates is within one standard deviation of the true value for either parameter and that the estimated standard errors are approximately the same as for the ML estimator, indicating little loss in (relative) efficiency from using this variant of the CCS estimator.

Turning to ML and initial CCS estimates for the high replacement cost regime presented in Tables 3 and 4, while the ML estimates are again close to their true values, both of the parameters are underestimated with the CCS estimator which uses the sample frequencies to estimate the p_{1i}'s. In particular, the θ_1 parameter is underestimated by 19 and 5%, respectively, for the 10,000 and 50,000 samples sizes while θ_2 is underestimated by 27 and 8%, respectively, for the corresponding sample sizes. Moreover, using either the estimated asymptotic standard errors or the empirical standard deviations, the average parameter estimates do not fall within two standard deviations of the corresponding true values; this is even true in the larger, 50,000 sample size case.

The observed bias associated with the sample frequency variant of the CCS estimator, especially in the high replacement cost regime, may be due to the omission from (5.8) of mileage categories in which there were no observed replacements or empty cells. This problem is not unique to our context. As has been noted in the literature on the Berkson–Theil Minimum χ^2 estimator applied to logit or probit models, zero cell probabilities,

TABLE 2

Monte Carlo results for low replacement cost specification: $\theta_{01}=2.0$ *and* $\theta_{02}=0.09$
samples of size 50,000 *bus-months**

Estimator	Parameter	Mean estimate	Standard error	Emp. stand. dev.
Maximum likelihood	θ_1	2·008	0·021	0·022
	θ_2	0·091	0·008	0·010
CCS (Using replacement freq. for p_i's)	θ_1	1·996	0·023	0·023
	θ_2	0·082	0·009	0·010
CCS (Using true p_i's to form y_i's)	θ_1	0·984	0·008	0·007
	θ_2	0·078	0·004	0·004
CCS (Using Cox correction)	θ_1	1·880	4·381	1·345
	θ_2	0·025	0·158	0·125
CCS (Using kernel for p_i's; $\xi=0{\cdot}025$)	θ_1	1·819	0·018	0·019
	θ_2	0·045	0·008	0·011
CCS (Using kernel for p_i's; $\xi=0{\cdot}01$)	θ_1	1·948	0·020	0·021
	θ_2	0·067	0·008	0·010
CCS (Using kernel for p_i's; $\xi=0{\cdot}005$)	θ_1	1·980	0·021	0·023
	θ_2	0·075	0·008	0·010
CCS (Using kernel for p_i's; $\xi=0{\cdot}0025$)	θ_1	1·990	0·022	0·023
	θ_2	0·079	0·009	0·010
CCS (Using kernel for p_i's; $\xi=0{\cdot}001$)	θ_1	1·996	0·023	0·023
	θ_2	0·082	0·009	0·010
CCS (Drop sparse mileage categories)	θ_1	2·001	0·024	0·023
	θ_2	0·089	0·009	0·010

* Results are based on 100 replications of each specification. The estimated (asymptotic) standard errors were computed from the first sample only. All Conditional Choice Simulation (CCS) estimators used the GLS weighting factor described in text.

TABLE 3

Monte Carlo results for high replacement cost specification: $\theta_{01}=8{\cdot}0$ *and* $\theta_{02}=0{\cdot}09$
samples of size 10,000 *bus-months**

Estimator	Parameter	Mean estimate	Standard error	Emp. stand. dev.
Maximum likelihood	θ_1	8·041	0·362	0·429
	θ_2	0·090	0·007	0·009
CCS (Using replacement freq. for p_i's)	θ_1	6·513	0·561	0·268
	θ_2	0·066	0·011	0·006
CCS (Using true p_i's to form y_i's)	θ_1	7·966	0·025	0·024
	θ_2	0·087	0·001	0·001
CCS (Using Cox correction)	θ_1	6·300	0·254	0·328
	θ_2	0·064	0·007	0·008
CCS (Using kernel for p_i's; $\xi=0{\cdot}025$)	θ_1	7·447	0·280	0·335
	θ_2	0·077	0·006	0·007
CCS (Using kernel for p_i's; $\xi=0{\cdot}01$)	θ_1	7·498	0·344	0·376
	θ_2	0·077	0·006	0·008
CCS (Using kernel for p_i's; $\xi=0{\cdot}005$)	θ_1	7·336	0·406	0·370
	θ_2	0·075	0·008	0·008
CCS (Using kernel for p_i's; $\xi=0{\cdot}0025$)	θ_1	7·041	0·475	0·330
	θ_2	0·071	0·009	0·008
CCS (Using kernel for p_i's; $\xi=0{\cdot}001$)	θ_1	6·512	0·560	0·286
	θ_2	0·066	0·011	0·006

* Results are based on 100 replications of each specification. The estimated (asymptotic) standard errors were computed from the first sample only. All Conditional Choice Simulation (CCS) estimators used the GLS weighting factor described in text.

empty cells, or, more generally, poorly estimated replacement probabilities can bias the estimates of the log-odds ratios in finite samples.[15] To investigate the potential impact of these sources of bias, we calculated the CCS estimator using the true replacement probabilities to form the y_i's in (5.7), continuing to use the estimates of the replacement probabilities

15. See Cox (1970), for example, on this point.

TABLE 4

Monte Carlo results for high replacement cost specification: $\theta_{01}=8\cdot 0$ *and* $\theta_{02}=0\cdot 09$ *samples of size* 50,000 *bus-months**

Estimator	Parameter	Mean estimate	Standard error	Emp. stand. dev.
Maximum likelihood	θ_1	8·025	0·181	0·202
	θ_2	0·090	0·004	0·004
CCS (Using replacement freq. for p_i's)	θ_1	7·582	0·311	0·183
	θ_2	0·083	0·006	0·004
CCS (Using true p_i's to form y_i's)	θ_1	7·990	0·011	0·007
	θ_2	0·089	0·0004	0·0002
CCS (Using Cox correction)	θ_1	7·263	0·051	0·175
	θ_2	0·073	0·002	0·005
CCS (Using kernel for p_i's; $\xi=0\cdot 025$)	θ_1	7·710	0·153	0·179
	θ_2	0·085	0·003	0·004
CCS (Using kernel for p_i's; $\xi=0\cdot 01$)	θ_1	7·879	0·190	0·194
	θ_2	0·087	0·004	0·004
CCS (Using kernel for p_i's; $\xi=0\cdot 005$)	θ_1	7·866	0·223	0·195
	θ_2	0·087	0·004	0·004
CCS (Using kernel for p_i's; $\xi=0\cdot 0025$)	θ_1	7·783	0·261	0·185
	θ_2	0·085	0·005	0·004
CCS (Using kernel for p_i's; $\xi=0\cdot 001$)	θ_1	7·582	0·311	0·183
	θ_2	0·083	0·006	0·004

* Results are based on 100 replications of each specification. The estimated (asymptotic) standard errors were computed from the first sample only. All Conditional Choice Simulation (CCS) estimators used the GLS weighting factor described in text.

to form the x_i's. The results of this exercise are labelled the "CCS (Using True p_{1i}'s to Form y_i's)" in each of the four tables. In each case, the resulting estimates are close to the true parameter values, even for the high replacement cost case (see Tables 3 and 4).[16] Thus, it appears that poorly estimated replacement probabilities—as opposed to poorly simulated functions—are accounting for much of the downward bias in the CCS estimates when simple cell frequencies are used to estimate the replacement probabilities.

In an attempt to reduce the bias in the CCS estimates of θ_0, we examined several procedures for estimating the replacement probabilities used to form log-odds ratios associated with the various mileage categories. Herein, we report on three. The first of these modifications is the standard correction for logistic models proposed by Cox (1970). This strategy involves adding an additional term to both the numerator and the denominator of the log-odds ratio. Mathematically, the definition is

$$\hat{y}_i^{\text{COX}}=\ln\left(\frac{\hat{p}_{1i}+\dfrac{1}{2N_i}}{1-\hat{p}_{1i}+\dfrac{1}{2N_i}}\right)-x_{i0}, \tag{5.10}$$

where $\hat{p}_{1i}$ is the original cell estimate of the replacement probability and N_i is the number of bus-months in mileage category i. It can be readily seen that as N_i increases, $\hat{y}_i^{\text{COX}}$ approaches the uncorrected dependent variable. This correction also allows for the inclusion of cells with an estimated replacement probability of zero, for which the log-odds

16. The corresponding estimates in Tables 1 and 2 are actually slightly more biased than the sample frequency variant of the CCS estimator, but this appears to be due, in the true p_i variant, to the inclusion of data on cells which had no observed replacements.

ratio would otherwise remain undefined. Examining the CCS estimates of θ_0 using this correction in Tables 1 through 4, one finds that they are even more biased than the CCS estimates using sample frequencies. While the additional bias is slight in the high replacement cost case (see Tables 3 and 4), it is substantial in the low replacement cost case (see Tables 1 and 2). For the latter regime, θ_1 is underestimated by 11 and 6% for the 10,000 and 50,000 sample sizes, respectively, while, for θ_2, the extent of under-estimation is 72 and 29% for the two respective sample sizes.

The second procedure we examined for estimating the replacement probabilities was use of a kernel estimator.[17] Put simply, each of the original cell estimates of the bus engine replacement probability was replaced by a weighted average of itself and the estimates from nearby cells. More precisely, the kernel estimator of p_{1i} was defined as follows:

$$\hat{p}_{1i}^{(N)}=\sum_{n=1}^{N}\left(\frac{\mathcal{K}\left(\frac{H_{nt}-i}{\xi}\right)d_{nt1}}{\sum_{n=1}^{N}\mathcal{K}\left(\frac{H_{nt}-i}{\xi}\right)}\right),\quad \text{for } i=1,\ldots,90, \tag{5.11}$$

where ξ is the bandwidth and the kernel function, $\mathcal{K}(\cdot)$, we actually use is given by:

$$\mathcal{K}\left(\frac{H_{nt}-i}{\xi}\right)=\phi\left(\frac{H_{nt}-i}{\xi}\right),$$

where $\phi(\cdot)$ is the standard normal density function. The log-odds ratios for each cell i were then calculated using this estimator of p_{1i}. This procedure implicitly relies on the model's implication that the true underlying replacement probability is a stable, continuous function of the bus mileage. Using larger bandwidths effectively increases the number of observations used to compute replacement probabilities in each cell.

We present results for the CCS estimator using kernel estimators of the p_{1i}'s for several alternative bandwidths in each of the four tables. (They are labelled "CCS (Using Kernel for p_{1i}'s; $\xi=a$)" for the alternative values, a, of the bandwidth.) Examining the results for the high replacement cost regime, we find that the averages of the estimates for this variant of the CCS estimator are less biased (in absolute value) than when sample frequencies are used to estimate the p_{1i}'s. While all of the bandwidths but the smallest reduce the bias in estimating both θ_1 and θ_2 (in the latter case, the use of the kernel smoothed estimates in the first stage of estimation simply reproduce the CCS estimates which use sample frequencies), a bandwidth of $\xi=0{\cdot}01$ produces estimates which are closest, on average, to the true parameter values in the high replacement cost regime. In contrast, using kernel estimation to produce p_{1i}'s in the log-odds ratios produces estimates of θ_0 which are biased downward to an even greater extent than those using sample frequencies in the low replacement cost case. The only exception to this is when a small bandwidth ($\xi=0{\cdot}001$) is used, which again just reproduces the estimates using sample frequencies.

The fact that the CCS estimator of θ_0 which uses kernel methods to estimate the p_{1i}'s is more biased in the low replacement cost regime suggests that there may be advantages, in certain cases, to excluding those categories which have few bus-months when calculating

17. We conjecture that using a flexible parametric form to extrapolate into sparse cells would yield similar results to the kernel smoothing procedure actually adopted. (The large sample properties are the same in both cases.)

(5.8). Recall from the preceding discussion that, unlike the high replacement cost specification, the bus-months tended to be concentrated in a small subset of the possible mileage categories in the low replacement cost regime. (In particular, we observed few or no bus-months in the twentieth through ninetieth mileage categories.) The kernel estimator constructs a weighted average of the replacement probabilities for these sparsely populated or unpopulated categories. But their use is likely to result in systematically biased estimates of the log odds ratios associated with these categories. This is so because the log-odds ratio is a concave function of the p_{1i}'s with the bias being greater the further the true p_{1i}'s are from 0.5.

Because of the potential for this bias, we examined a third procedure for estimating the replacement probabilities used to form the log-odds ratios in which those mileage categories with zero or low cell counts were excluded when forming (5.8). (More specifically, we excluded observations with high values of H_t and attached zero probabilities to the occurrence of these events.) This procedure was only undertaken for the low replacement cost regime, since it is only in this case that sparse cells were encountered; sample replacement frequencies were used to estimate the p_{1i}'s for the included categories.[18] Examining the entries labelled "CCS (Drop Sparse Mileage Categories)" in Tables 3 and 4, we find no bias for the average of the estimates of θ_1 in either the 10,000 or 50,000 sample size cases and no bias in the estimates of θ_2 in the larger samples. We do not find that the average estimate of θ_2 is underestimated by 11% in the 10,000 sample size case, but this is substantially less than the bias found for this parameter using the other two variants of the CCS estimator.

While necessarily tentative, given the restricted nature of the model considered, our Monte Carlo investigation suggests the following conclusions concerning the use of CCS estimators to estimate the structural parameters of dynamic, discrete-choice models. First, it appears that the potential for the greatest bias in the parameter estimates arises from using poorly estimated conditional choice probabilities to construct the $q(\cdot)$ functions. Bias resulting from the simulation of the conditional valuation functions appears to be much less important.

Second, our results provide some practical guidance as to how to generate estimates of the conditional choice probabilities sufficiently reliable to avoid large biases. A poor estimate of the conditional choice probability for a given history results from the interaction of the number of observations in the cell and the size of the true underlying choice probabilities. As the number of observations decreases and the true probabilities move away from 0.5, the estimates become more variable and the potential for bias in $q(\cdot)$ increases. When all of the histories have sufficient observations to produce reliable estimates of choice probabilities, cell estimates may be used. When all of the histories have roughly equal numbers of observations, but the true choice probabilities are suspected to be small (as evidenced by their infrequency in the data), then kernel smoothing appears to significantly improve the estimates of $q(\cdot)$ and, thereby, improve the estimates of θ_0 as well.[19] This is precisely what occurred in the high replacement cost regime examined above.

When the observations are distributed asymmetrically over the histories but the sample choice probabilities associated with the populated cells are neither extremely low

18. In particular, we excluded all but the first 20 mileage categories. Thus, in the samples of size 10,000, we ignored 58 observations (all clustered between mileage categories 20 and 28) and, in the samples of size 50,000, we ignored 309 observations. In the bigger sample, none of the buses achieved a mileage category of more than 42.

19. When using kernel methods, empirically-based procedures, such as cross-validation techniques, could be used to aid in the selection of appropriate bandwidths. See Silverman (1986) for more on these techniques.

or equal to zero, the omission of those histories with low numbers of observations appears to improve the estimates of the $q(\cdot)$'s, and thus of θ_0.[20] That deleting such histories can improve the estimates of θ_0 was shown in the case of the low replacement cost regime considered above. Finally, if the histories are a discrete approximation to a continuous underlying variable, as they are in Rust's model, then the quality of the resulting estimates of the conditional choice probabilities should be one factor guiding the choice of how many discrete categories to use in the approximation.

APPENDIX

Proof of Proposition 1. To prove the consistency and asymptotic distribution of $\theta^{(N)}$, it is convenient to formulate both the estimators for $\psi_0=(P_0, F_0)$ and θ_0 in terms of a set of orthogonality conditions. Define the $M(JK-1)\times 1$ vector, $g(x_n, \psi)$, which is used to form the cell estimators of the conditional choice probabilities, as:

$$g(x_n,\psi)=\begin{pmatrix} [d_{nt}-\sum_{i=1}^{M} 1\{H_{nt}=H^{(i)}\}p^{(i)}]\otimes\begin{pmatrix}1\{H_{nt}=H^{(1)}\}\\ \vdots \\ 1\{H_{nt}=H^{(M)}\}\end{pmatrix} \\ \begin{pmatrix}[F((H_{nt},b^{(1)})\mid H_{nt})-\sum_{i=1}^{M} 1\{H_{nt}=H^{(i)},b_{n,t+1}=b^{(1)}\}d_{nt}] \\ \vdots \\ [F((H_{nt},b^{(l)})\mid H_{nt})-\sum_{i=1}^{M} 1\{H_{nt}=H^{(i)},b_{n,t+1}=b^{(l)}\}d_{nt}] \\ \vdots \\ [F((H_{nt},b^{(L)})\mid H_{nt})-\sum_{i=1}^{M} 1\{H_{nt}=H^{(i)},b_{n,t+1}=b^{(L)}\}d_{nt}]\end{pmatrix}\otimes\begin{pmatrix}1\{H_{nt}=H^{(1)}\}\\ \vdots \\ 1\{H_{nt}=H^{(M)}\}\end{pmatrix}\end{pmatrix}. \tag{A.1}$$

It follows that our simulation estimator of (ψ_0, θ_0) is formed using the following $[R(J-1)+M(JK-1]\times 1$ vector of sample moments,

$$G_N(\psi,\theta)=N^{-1}\textstyle\sum_{n=1}^{N} h(x_n,\psi,\theta), \tag{A.2}$$

where

$$h(x_n,\psi,\theta)\equiv(f(x_n,\psi,\theta)', g(x_n,\psi)')'. \tag{A.3}$$

Finally, define the $[Q+M(JK-1)]\times WR[(J-1)+M(JK-1)]$ matrix A_N as

$$A_N=\begin{pmatrix}B_N & 0\\ 0 & I\end{pmatrix}, \tag{A.4}$$

where I is the $M(JK-1)$ identity matrix and B_N is the convergent $Q\times R(J-1)$ matrix:

$$B_N=(\textstyle\sum_{n=1}^{N} z_{nt}\otimes\partial[q(p_{nt}^{(N)}, H_{nt},\theta)-v(x_n,\psi^{(N)},\theta)]/\partial\theta)\,W_N. \tag{A.5}$$

Then $(\psi^{(N)}, \theta^{(N)})$ are implicitly defined by the $Q+M(JK-1)$ equations:

$$0=A_N G_N(\psi^{(N)}, \theta^{(N)}). \tag{A.6}$$

(Notice that this formulation exploits Newey's (1984) observation that sequential estimators may be expressed as the result of a joint estimation strategy in which a non-optimal weighting matrix is used.)

To prove $(\psi^{(N)}, \theta^{(N)})$ is consistent, we verify the three conditions in Corollary 3.2 of Pakes and Pollard (1989, p. 1039) as augmented by their Lemma 3.5 (on p. 1045). Condition (i) of their corollary is that:

$$\|A_N G_N(\psi^{(N)}, \theta^{(N)})\|\leqq o_p(1)+\inf_{(\psi,\theta)}\|A_N G_N(\psi,\theta)\|, \tag{A.7}$$

20. Note that the exclusion of histories for which there are few observations in the data in estimating θ_0 is obviously subject to the requirement that at least Q histories are included so that θ_0 can be identified.

where $\|\cdot\|$ is the Euclidean norm; it is satified by the definition $(\psi^{(N)}, \theta^{(N)})$. Our Assumptions 1 through 5 ensure $\Theta \times \Psi$ is compact and $G(\cdot)$ is continuous. These properties, in conjunction with our Assumption 6, imply that $\| \| G(\psi, \theta) \| > 0$ for all (ψ, θ) such that $\| (\psi, \theta) - (\psi_0, \theta_0) \| \geqq \delta$, where $\delta > 0$, which is Condition (ii) of Corollary 3.2 in Pakes and Pollard. Consistency of our estimator is established by verifying that (the uniform convergence) Condition (iii) of their Corollary 3.2 holds in our case. Note this condition is less stringent than requiring:

$$\sup_{(\psi,\theta)} \|G_N(\psi, \theta) - G(\psi, \theta)\| = o_p(1). \tag{A.8}$$

We shall presently show the class of functions defined by:

$$\mathscr{F} \equiv \{h(x, \psi, \theta) : (\psi, \theta) \in \Psi \times \Theta\}, \tag{A.9}$$

is Euclidean (in the sense that all its real-valued function components are).[21] Then, by Lemma (2.8) of Pakes and Pollard (1989, p. 1033), (A.8) is satisfied. Therefore, our estimator, $(\psi^{(N)}, \theta^{(N)})$, is consistent.

The large-sample distributional properties of $(\psi^{(N)}, \theta^{(N)})$ can be derived by checking that the Conditions (i)–(v) of Pakes and Pollard's Theorem 3.3 (p. 1040) are satisfied and by appealing to their Lemma 3.5 (p. 1045). By the definition of $(\psi^{(N)}, \theta^{(N)})$, Condition (i) is satisfied since $\| G_N(\psi^{(N)}, \theta^{(N)}) \| = 0_p(N^{-1/2})$. By our Assumption 2, $U(H_{nt}, \psi, \theta)$ is differentiable in (ψ, θ); hence $v(x_n, \psi, \theta)$ is too. Therefore, $G(\psi, \theta)$ is differentiable in (ψ, θ), with a derivative matrix of full rank, which is their Condition (ii). By a Central Limit Theorem, such as 7.1.2 in Chung (1974, p. 200), $N^{1/2}G_N(\psi_0, \theta_0)$ converges in distribution to a normal random variable centred at 0 with covariance V defined as:

$$V = \begin{pmatrix} E[f(x, \psi_0, \theta_0) f(x, \psi_0, \theta_0)'] & 0 \\ 0 & \Omega \end{pmatrix}, \tag{A.10}$$

where $\Omega = E[g(x, \psi_0) g(x, \psi_0)']$. (Note the blocks off the diagonal are 0 because the simulation errors are independent of differences between choices and their conditional expectations.) This is their Condition (iv). Condition (v) in Theorem 3.3 of Pakes and Pollard is simply our Assumption 3.

This leaves only (the equi-continuity) Condition (iii) of their Theorem to verify. We will show that $\mathscr{F}$ defined in (A.9) is Euclidean and that the parameterization is $\mathscr{L}^2$ continuous at $(\psi_0\ \theta_0)$ in the probability space from which the sample is drawn. Then noting:

$$\begin{aligned} &\|G_N(\psi, \theta) - G(\psi, \theta) - G_N(\psi_0, \theta_0)\| / \{N^{-1/2} + \|G_N(\psi, \theta)\| + \|G(\psi, \theta)\|\} \\ &\leqq \|N^{1/2}[G_N(\psi, \theta) - G(\psi, \theta)] - N^{1/2}[G_N(\psi_0, \theta_0) - G(\psi_0, \theta_0)]\|, \end{aligned} \tag{A.11}$$

it follows from Lemma 2.17 in Pakes and Pollard (1989, p. 1037) that Condition (iii) is met. To verify that the components of $h \in \mathscr{F}$ are Euclidean, we analyze $g(x_n, \psi)$ and $f(x_n, \theta, \psi)$ separately. With regard to $g(x_n, \psi)$, it is formed from differences and products of $(M+1)$ mappings from x_n to $\mathbb{R}$, namely d_{nt}, the constants $p^{(i)}$, and the indicator functions, $1\{H_{nt} = H^{(i)}\}$ for $i \in \{1, \ldots, M\}$. The class of functions generated by each of these component mappings by varying ψ through Ψ is Euclidean; hence, by Lemma 2.14 in Pakes and Pollard (1989, p. 1035), $g(x_n, \psi)$ is too. Turning now to $f(x_n, \psi, \theta)$, we first observe that neither z_{nt} nor $\{H_{nt} = H^{(i)}\}$ depend on the parameters (although this could be relaxed), so both are Euclidean in (ψ, θ). Appealing to Example 2.9 of Pakes and Pollard (1989, p. 1033) and their Lemma 2.15 (p. 1035), it follows that $q(\sum_{i=1}^{M} 1\{H_{nt} = H^{(i)} p_i^{(N)}, H_{nt}, \theta\})$ is Euclidean also. To show that $v(x_n, \psi, \theta)$ is Euclidean, we decompose it into a weighted linear combination of indicator functions like $1\{H^{(i)} = H_{ns}^{(k,N)}, H_{(j)} = H_{ns}^{(J,N)}\}$, where the weights depend on the paremeters (ψ, θ) only. In particular, for all $i \in \{1, \ldots, M\}$, we define $U_{ik}(\psi, \theta)$ by the identitity $U_{ik}(\psi, \theta) \equiv U(p^{(N)}(H^{(i)}), H^{(i)}, \psi^{(N)}, \theta)$ for any choice $k \in J$. Recalling (3.5), it then follows that:

$$\begin{aligned} v_k(x_{nk}, \psi^{(N)}, \theta) &= \textstyle\sum_{s=t+1}^{T} [U(p_s^{(N)}(H_{ns}^{(k,N)}), H_{ns}^{(k,N)}, \psi^{(N)}, \theta) - U(p_s^{(N)}(H_{ns}^{(J,N)}, H_{ns}^{(J,N)}, \psi^{(N)}, \theta)] \\ &= \textstyle\sum_{s=t+1}^{T} \sum_{i=1}^{M} \sum_{j=1}^{M} [U_{ik}(\psi, \theta) - U_{jJ}(\psi, \theta)] 1\{H^{(i)} = H_{ns}^{(k,N)}, H^{(j)} = H_{ns}^{(J,N)}\} \end{aligned} \tag{A.12}$$

Each of the terms on the right-hand side of (A.12) is Euclidean. Therefore $f(x_n, \psi, \theta)$ is too. Finally, $\mathscr{L}^2$ continuity follows from the fact that jumps occur in f only on a set of measure zero.

21. For a formal definition of Euclidean classes of functions, see Definition 2.7 in Pakes and Pollard (1989, p. 1032).

Appealing to Theorem 3.3 and Lemma 3.5 of Pakes and Pollard (1989, pp. 1040 and 1045, respectively), it follows that $N^{1/2}((\psi^{(N)}-\psi_0)', (\theta^{(N)}-\theta_0)')'$ is jointly distributed as a normal random variable with mean 0 and covariance matrix

$$(\Gamma' W^* \Gamma)^{-1} \Gamma' W^* V W^* \Gamma(\Gamma' W^* \Gamma)^{-1}, \tag{A.13}$$

where

$$W^* = \begin{pmatrix} W & 0 \\ 0 & I \end{pmatrix},$$

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ 0 & I \end{pmatrix},$$

and Γ_{11} and Γ_{12} are defined in the paper. It follows that the covariance matrix for $N^{1/2}(\theta^{(N)}-\theta_0)$ simplifies to (3.9). ||

Proof of Proposition 2. First, note that in their Appendix A, Hotz and Miller (1993) prove their Proposition 1 by establishing the inversion property for each H_t and $t \leqq T$. Since the conditional valuation functions are defined in the infinite-horizon problem for all such H_t and $t \in \{0, 1, \ldots,\}$, their proof applies, without further modification, to the infinite-horizon case. This establishes the existence of a mapping, denoted $q(p(H_t), H_t)$, with the same features as those attributed to (2.8) above. Second, none of the statements in Proposition 1 given in text are affected by substituting (4.3) for (3.5) in (3.10) and proceeding with this alternative definition of $f(x_n, \psi, \theta)$. This establishes the second part of Proposition 2. ||

Acknowledgements. We have benefited from comments by Dan Black, Bob Chirinko, Kermit Daniel, John Rust, anonymous referees, and workshop participants at the Universities of Chicago and Wisconsin. This research is supported by NICHD Grant R23-HD18935 and NSF Grant SES-8801631.

REFERENCES

ALTUG, S. and MILLER, R. (1990), "Human Capital Accumulation, Aggregate Shocks and Panel Data Estimation" (Working paper).

BERKOVEC, J. and STERN, S. (1991), "Job Exit Behavior of Older Men", *Econometrica* **59**, 189–210.

CHUNG, K. (1974) *A Course in Probability Theory* (Orlando, Fla.: Academic Press).

COX, D. (1970) *Analysis of Binary Data* (London: Chapman and Hall).

ECKSTEIN, Z. and WOLPIN, K. (1989), "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey", *Journal of Human Resources*, **24**, 562–598.

HOTZ, V. J. and MILLER, R. (1986), "The Economics of Family Planning" (unpublished manuscript).

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. and MILLER, R. (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, **60**, 497–529.

MANSKI, C. (1993), "Dynamic Choice in a Social Setting", *Journal of Econometrics* (forthcoming).

McFADDEN, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, **57**, 995–1027.

MILLER, R. (1982), "Job Specific Capital and Labor Mobility" (Ph.D. dissertation, University of Chicago).

MILLER, R. (1984), "Job Matching and Occupational Choice", *Journal of Political Economy*, **92**, 1086–1120.

NEWEY, W. (1984), "A Method of Moments Interpretation of Sequential Estimators", *Economics Letters*, **15**, 201–206.

PAKES, A. and POLLARD, D. (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, **57**, 1027–1058.

RUST, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, **57**, 999–1035.

SANDERS, S. (1993), "A Dynamic Model of Welfare Participation" (Ph.D. Dissertation, University of Chicago).

SILVERMAN, B. (1986) *Density Estimation for Statistics and Data Analysis* (London: Chapman and Hall).

STOCK, J. and WISE, D. (1990), "Pensions, the Option Value of Work, and Retirement", *Econometrica*, **58**, 1151–1180.

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

LINKED CITATIONS

A Simulation Estimator for Dynamic Models of Discrete Choice
V. Joseph Hotz; Robert A. Miller; Seth Sanders; Jeffrey Smith
The Review of Economic Studies, Vol. 61, No. 2. (Apr., 1994), pp. 265-289.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199404%2961%3A2%3C265%3AASEFDM%3E2.0.CO%3B2-9

[Footnotes]

[3] **Simulation and the Asymptotics of Optimization Estimators**
Ariel Pakes; David Pollard
Econometrica, Vol. 57, No. 5. (Sep., 1989), pp. 1027-1057.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198909%2957%3A5%3C1027%3ASATAOO%3E2.0.CO%3B2-R

[3] **Job Exit Behavior of Older Men**
James Berkovec; Steven Stern
Econometrica, Vol. 59, No. 1. (Jan., 1991), pp. 189-210.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199101%2959%3A1%3C189%3AJEBOOM%3E2.0.CO%3B2-Y

[7] **Conditional Choice Probabilities and the Estimation of Dynamic Models**
V. Joseph Hotz; Robert A. Miller
The Review of Economic Studies, Vol. 60, No. 3. (Jul., 1993), pp. 497-529.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199307%2960%3A3%3C497%3ACCPATE%3E2.0.CO%3B2-B

[13] **Conditional Choice Probabilities and the Estimation of Dynamic Models**
V. Joseph Hotz; Robert A. Miller
The Review of Economic Studies, Vol. 60, No. 3. (Jul., 1993), pp. 497-529.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199307%2960%3A3%3C497%3ACCPATE%3E2.0.CO%3B2-B

LINKED CITATIONS

[21] **Simulation and the Asymptotics of Optimization Estimators**
Ariel Pakes; David Pollard
Econometrica, Vol. 57, No. 5. (Sep., 1989), pp. 1027-1057.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198909%2957%3A5%3C1027%3ASATAOO%3E2.0.CO%3B2-R

References

Job Exit Behavior of Older Men
James Berkovec; Steven Stern
Econometrica, Vol. 59, No. 1. (Jan., 1991), pp. 189-210.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199101%2959%3A1%3C189%3AJEBOOM%3E2.0.CO%3B2-Y

The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey
Zvi Eckstein; Kenneth I. Wolpin
The Journal of Human Resources, Vol. 24, No. 4. (Autumn, 1989), pp. 562-598.
Stable URL:
http://links.jstor.org/sici?sici=0022-166X%28198923%2924%3A4%3C562%3ATSAEOD%3E2.0.CO%3B2-R

An Empirical Analysis of Life Cycle Fertility and Female Labor Supply
V. Joseph Hotz; Robert A. Miller
Econometrica, Vol. 56, No. 1. (Jan., 1988), pp. 91-118.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198801%2956%3A1%3C91%3AAEAOLC%3E2.0.CO%3B2-O

Conditional Choice Probabilities and the Estimation of Dynamic Models
V. Joseph Hotz; Robert A. Miller
The Review of Economic Studies, Vol. 60, No. 3. (Jul., 1993), pp. 497-529.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199307%2960%3A3%3C497%3ACCPATE%3E2.0.CO%3B2-B

LINKED CITATIONS

A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration
Daniel McFadden
Econometrica, Vol. 57, No. 5. (Sep., 1989), pp. 995-1026.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198909%2957%3A5%3C995%3AAMOSMF%3E2.0.CO%3B2-Z

Job Matching and Occupational Choice
Robert A. Miller
The Journal of Political Economy, Vol. 92, No. 6. (Dec., 1984), pp. 1086-1120.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198412%2992%3A6%3C1086%3AJMAOC%3E2.0.CO%3B2-4

Simulation and the Asymptotics of Optimization Estimators
Ariel Pakes; David Pollard
Econometrica, Vol. 57, No. 5. (Sep., 1989), pp. 1027-1057.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198909%2957%3A5%3C1027%3ASATAOO%3E2.0.CO%3B2-R

Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher
John Rust
Econometrica, Vol. 55, No. 5. (Sep., 1987), pp. 999-1033.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198709%2955%3A5%3C999%3AOROGBE%3E2.0.CO%3B2-N

Pensions, the Option Value of Work, and Retirement
James H. Stock; David A. Wise
Econometrica, Vol. 58, No. 5. (Sep., 1990), pp. 1151-1180.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199009%2958%3A5%3C1151%3APTOVOW%3E2.0.CO%3B2-1

An Estimable Dynamic Stochastic Model of Fertility and Child Mortality
Kenneth I. Wolpin
The Journal of Political Economy, Vol. 92, No. 5. (Oct., 1984), pp. 852-874.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198410%2992%3A5%3C852%3AAEDSMO%3E2.0.CO%3B2-F

NOTE: *The reference numbering from the original has been maintained in this citation list.*