Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models

Author(s): Victor Aguirregabiria and Pedro Mira

Source: *Econometrica*, Jul., 2002, Vol. 70, No. 4 (Jul., 2002), pp. 1519-1543

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/3082006

Econometrica, Vol. 70, No. 4 (July, 2002), 1519–1543

SWAPPING THE NESTED FIXED POINT ALGORITHM: A CLASS OF ESTIMATORS FOR DISCRETE MARKOV DECISION MODELS

BY VICTOR AGUIRREGABIRIA AND PEDRO MIRA[1]

This paper proposes a new nested algorithm (NPL) for the estimation of a class of discrete Markov decision models and studies its statistical and computational properties. Our method is based on a representation of the solution of the dynamic programming problem in the space of conditional choice probabilities. When the NPL algorithm is initialized with consistent nonparametric estimates of conditional choice probabilities, successive iterations return a sequence of estimators of the structural parameters which we call K-stage policy iteration estimators. We show that the sequence includes as extreme cases a Hotz-Miller estimator (for $K = 1$) and Rust's nested fixed point estimator (in the limit when $K \to \infty$). Furthermore, the asymptotic distribution of all the estimators in the sequence is the same and equal to that of the maximum likelihood estimator. We illustrate the performance of our method with several examples based on Rust's bus replacement model. Monte Carlo experiments reveal a trade-off between finite sample precision and computational cost in the sequence of policy iteration estimators.

KEYWORDS: Discrete Markov decision models, nested algorithms, policy iteration, maximum pseudo-likelihood estimators.

1. INTRODUCTION

OVER THE LAST TWO DECADES, there has been considerable progress in the econometrics of discrete choice dynamic programming models. The estimation of these models has enhanced our understanding of individual and firm behavior and it has proved a useful tool in the quantitative assessment of public policy proposals in numerous areas.[2] However, the computational burden of estimation has been an important constraint with substantial implications for empirical work: the model specification has to be very parsimonious in terms of state variables and time-invariant explanatory variables and this may lead to important misspecifications. Thus a major contribution of Rust (1987, 1988) was to introduce the *conditional independence assumption* and the *nested fixed point algorithm* (NFXP). The former defined a class of models in which the cost of multiple integration over unobserved state variables and the dimensionality of the

[1] We are grateful to Miguel Delgado, Bob Miller, Ariel Pakes, John Rust, seminar participants at Universidad Carlos III, Carnegie Mellon University, CREST, University of Chicago, the referees and the editor for their useful comments. The Gauss code with the estimation procedures in this paper can be found at the web address: http://people.bu.edu/vaguirre/programs/npl.html.

[2] See Wolpin (1996). For excellent surveys on the structural estimation of dynamic discrete choice models and their empirical applications, see Eckstein and Wolpin (1989), Pakes (1994), Rust (1994b), and Miller (1997).

dynamic programming problem were greatly reduced; the latter described how to obtain maximum likelihood estimates. His work resulted in a spate of applications; examples include Rust (1987), Ahn (1995), Rust and Phelan (1997), and Gilleskie (1998).[3] NFXP is a full solution method that re-solves the fixed point problem in value function space at each parameter value, and derives the choice probabilities from the fixed point. Hotz and Miller (1993) pioneered an even simpler estimator than NFXP for this class of problems. They observed that it is not necessary to fully solve the fixed point problem in order to estimate the structural parameters. Since there is a one to one mapping between the normalized value functions and conditional choice probabilities, nonparametric estimates of choice probabilities can be inverted into nonparametric estimates of the value functions and these estimates can be combined with the current period utility functions to obtain new choice probabilities that do not require solving the fixed point problem.[4] Hotz and Miller showed that minimizing a sample criterion function based on these choice probabilities leads to a consistent, asymptotically normally distributed estimator, i.e., the *conditional choice probabilities* estimator (CCP). The CCP estimator has been used in empirical applications by Hotz and Miller (1993), Slade (1998), Altug and Miller (1998), and Aguirregabiria (1999), among others.

Previous conventional wisdom was that the CCP estimator achieved a significant computational gain at the expense of efficiency, both in finite samples and asymptotically.[5] Thus, researchers had the choice between two extremes: a full solution NFXP estimator with the attendant computational burden, or the much faster but less efficient CCP estimator. In this paper we propose a new estimation procedure that bridges the gap between these two estimation strategies. In particular, we show that an appropriately formed CCP estimator is asymptotically equivalent to the NFXP estimator and that for any finite sample a recursive extension of this particular CCP estimator delivers the NFXP estimator.

We build on Hotz and Miller's work to show that the original fixed point problem in "value function space" can be reformulated as a fixed point problem in "probability space." That is, for any value of the structural parameters θ we prove that the vector of conditional choice probabilities P_θ associated with the solution of the dynamic programming problem can be obtained as the unique fixed point of a mapping in probability space: $P_\theta = \Psi_\theta(P_\theta)$. Moreover, at the fixed point P_θ the Jacobian matrix of Ψ_θ is always zero: $\partial\Psi_\theta(P_\theta)/\partial P = 0$. The policy iteration mapping $\Psi_\theta(P)$ is the cornerstone of our estimation procedure and the "zero

[3] The Conditional Independence Assumption restricts the structure of serial correlation of unobserved state variables. See Miller (1984), Pakes (1986), and Berkovec and Stern (1991) for early examples of applications that modeled important sources of unobserved heterogeneity. Keane and Wolpin (1994, 1997) developed a simulation and interpolation method that has been used to obtain approximate maximum likelihood estimates for models with serially correlated unobservables.

[4] Magnac and Thesmar (2002) exploit this insight to analyze nonparametric identification of this class of models. A similar estimation method that avoids nested solution of the dynamic programming problem was proposed by Manski (1991, 1993).

[5] See previous discussions on efficiency of the CCP estimator by Eckstein and Wolpin (1989), and Rust (1994a, p. 154), and the Monte Carlo study by Hotz et al. (1994).

Jacobian property" is the key to its finite sample and asymptotic properties. In our nested pseudo-likelihood algorithm (NPL) the inner algorithm maximizes in θ a pseudo-likelihood function based on choice probabilities $\Psi_\theta(P)$ where P is an estimate of choice probabilities. The outer algorithm is a fixed point algorithm that computes $\Psi_\theta(P)$ at the current parameter estimates to update the estimate of P. When the NPL algorithm is initialized with consistent nonparametric estimates of conditional choice probabilities, successive iterations return a sequence of estimators of the structural parameters that we call *K-stage policy iteration (PI)* estimators. We show that the sequence includes as extreme cases a CCP estimator (for $K = 1$) and a root of the likelihood equations (in the limit when $K \to \infty$). Furthermore, the asymptotic distribution of all the estimators in the sequence is the same and equal to that of the NFXP estimator. The intuition behind this result is quite simple. The maximum likelihood estimator is the value of θ that maximizes a likelihood function with choice probabilities P_θ. The class of PI estimators recasts the estimation problem in terms of choice probabilities $\Psi_\theta(P^*)$, where P^* represents the *true* (population) conditional choice probabilities. These are treated as "nuisance" parameters and replaced by root-n consistent and asymptotically normal estimates at each stage. However the zero Jacobian property has the double implication that: (i) the pseudo-information matrix is block-diagonal in the parameters of interest and the nuisance parameters; and (ii) the score and pseudo-score variables are identical at the true parameter values. It follows from (i) that the asymptotic distribution of the K-stage PI estimator does not depend on the asymptotic distribution of the estimator of the nuisance parameters in P^*, and from (ii) that this asymptotic distribution is the same as that of the ML estimator.

Therefore contrary to conventional wisdom a particular CCP estimator is asymptotically equivalent to the maximum likelihood estimator. If we are only concerned about computation costs we might as well stop at $K = 1$. However, the finite sample performance of a PI estimator might depend on the number of stages K. We illustrate the potential trade-off between finite sample precision and computational cost with several Monte Carlo experiments based on Rust's bus replacement model. We find that the finite sample properties of the estimator do indeed improve monotonically with K. Also, the benefits of using the 2-stage PI estimator instead of the 1-stage estimator are very significant, but the additional gains in precision obtained by using the ML estimator instead of 2-stage PI are small.

Finally, we also compare the computational cost of the NFXP and NPL algorithms for maximum likelihood estimation. We may want to obtain ML estimates out of concern about finite sample properties of CCP and K-stage PI estimators, or it may be that initial nonparametric estimates of conditional choice probabilities are not available.[6] In this case our NPL algorithm can be initialized with an

[6] For instance, this is the case when there is unobserved time-invariant heterogeneity in individual's preferences, or when there are unobservable aggregate shocks.

arbitrary vector of choice probabilities. Although it is clear that it will not return a *sequence* of consistent estimators of structural parameters, on convergence it will still produce a root of the likelihood equations. Both NPL and NFXP are nested algorithms that combine a (pseudo) likelihood climbing routine with fixed point iterations that solve the dynamic programming problem. By *swapping the nesting of the two algorithms* NPL avoids repeated solution of the dynamic programming problem at the expense of a larger number of pseudo-likelihood climbing iterations. Therefore, it is intuitive that NPL may be faster than NFXP if fixed point iterations are more costly than pseudo-likelihood climbing iterations. This was indeed the case in our experiments with Rust's bus replacement model, where NPL produced ML estimates 5 to 15 times faster than NFXP.

The rest of the paper is organized as follows. In Section 2 we first review a class of discrete Markov decision processes and several useful results inherited from the literature; we then define the Policy Iteration operator and establish its properties. In Section 3 we describe the NPL algorithm and show that on convergence it is equivalent to NFXP. In Section 4 we introduce the class of K-stage PI estimators and we obtain their asymptotic properties. Section 5 compares the performance of the NFXP and NPL algorithms using the bus engine data set in Rust (1987) and presents a Monte Carlo study that illustrates the precision in finite samples of the different policy iteration estimators. We conclude in Section 6 with a summary of our results. Proofs of propositions are provided in the Appendix.

2. DISCRETE MARKOV DECISION PROCESSES

2.1. *Definitions and Assumptions*

We define a discrete decision process and present our basic assumptions. There are two types of variables: the vector of state variables s and a control variable a that belongs to a finite set of mutually exclusive choice alternatives $A = \{1, 2, \ldots, J\}$. Time is discrete and it is indexed by t. At each period t an agent observes s_t and chooses a_t in order to maximize the expected sum of current and future discounted utilities. Future values of some state variables are uncertain for the agent. His beliefs about uncertain future states can be represented by a Markov transition probability $p(s_{t+1}|s_t, a_t)$. The time horizon of the decision problem is infinite. Utility is time separable and $u(s_t, a_t)$ represents the one-period utility function. The parameter $\beta \in (0, 1)$ is the time discount factor.

Under certain regularity conditions, Blackwell's theorem establishes the following properties. First, there exists a stationary, Markovian optimal decision rule $\delta(s_t)$: the decision at period t is the same as the decision at period $t + j$ if $s_t = s_{t+j}$. Therefore, we omit the subindex t for the rest of this section and we use s' to denote the vector of next period's state variables. Second,

$\delta(s) = \arg\max_{a\in A}\{u(s,a) + \beta \int V(s')p(ds'|s,a)\}$ where the *value function* $V()$ is the unique solution of the Bellman equation:[7]

$$V(s) = \max_{a\in A}\left\{u(s,a) + \beta \int V(s')p(ds'|s,a)\right\}. \tag{1}$$

We assume that the researcher knows the utility and the transition probability functions up to a vector of parameters θ. From an econometric point of view we distinguish two types of state variables: $s = (x, \epsilon)$. The subvector x groups variables that are observed by both the agent and the researcher, whereas the subvector ϵ includes those state variables that are observed only by the agent. Given data on observable state variables and the actual choices made by agents, our goal is to obtain an estimate of θ. We introduce two assumptions concerning the role of observable and unobservable state variables that have been widely used in the literature (see Rust (1994b)). These assumptions contribute to simplify the estimation problem considerably. Although the primitives and operators introduced here are all functions of the parameter vector θ, in this section we are mostly concerned with their properties for a given θ; we therefore omit it as an argument in order to simplify the notation.

ASSUMPTION 1 (Additivity): *The one period utility function is additively separable in the observable and unobservable components*: $u(s,a) = u(x,a) + \epsilon(a)$, *where* $\epsilon(a)$ *is the a*th *component of the* $J \times 1$ *vector* ϵ. *The support of* $\epsilon(a)$ *is the real line for all a.*

ASSUMPTION 2 (Conditional Independence): *The transition probability of the state variables factors as* $p(x', \epsilon'|x, a, \epsilon) = g(\epsilon'|x')f(x'|x,a)$, *where* $g(.)$ *has finite first moments and is continuous and twice differentiable in* ϵ'.

ASSUMPTION 3 (Finite Domain of Observable State Variables): $x \in X = \{x^1, \ldots, x^M\}$.

We can exploit Assumptions 1–2 to obtain versions of the value functions and the Bellman operator in which the unobservables are integrated out. These versions will prove more useful than equation (1) in the analysis of the estimation problem. Let $V_\sigma(x)$ hereafter denote the expectation of the value function conditional on the state variables x: $V_\sigma(x) \equiv \int V(x,\epsilon)g(d\epsilon|x)$, where σ represents parameters that characterize the distribution of the ϵ's. Under Assumptions 1–3, V_σ solves the smooth Bellman equation:

$$V_\sigma(x) = \int \max_{a\in A}\left[u(x,a) + \epsilon(a) + \beta \sum_{x'} V_\sigma(x')f(x'|x,a)\right] g(d\epsilon|x). \tag{2}$$

The smoothed Bellman operator $\Gamma_\sigma(.)$ defined by the right-hand side of this functional equation is a contraction mapping (see the proof of Proposition l(a) in the Appendix). It follows that V_σ is its unique fixed point: $V_\sigma = \Gamma_\sigma(V_\sigma)$.

[7] In this paper the current period utility is not bounded. This requires a generalization of Blackwell's Theorem. See Rust (1988).

The *conditional choice probability* $P(a|x)$ is the probability that alternative a is the optimal choice given the vector of observable state variables x:

$$(3) \qquad P(a|x) = \int I\left\{a = \arg\max_{j \in A}[v(x,j) + \epsilon(j)]\right\} g(d\epsilon|x)$$

where $v(x,a)$ is the *choice-specific value function* $u(x,a) + \beta \sum_{x'} V_\sigma(x') f(x'|x,a)$.

2.2. *From Conditional Choice Probabilities to Value Functions*

We now review several results that relate conditional choice probabilities and value functions. In the next subsection we build on these results to show how the original fixed point problem resulting from the Bellman equation in "value space" can be solved using an operator in "probability space."

First, it can be shown that the choice probabilities conditional on any value of x are uniquely determined by the vector of normalized value functions or utility differences $\{\tilde{v}(x,a) : a > 1\}$, where $\tilde{v}(x,a)$ is defined as $v(x,a) - v(x,1)$. That is, *there exists a vector mapping* $Q_x(.)$ *such that* $\{P(a|x) : a > 1\} = Q_x(\{\tilde{v}(x,a) : a > 1\})$, where, without loss of generality, we exclude the probability of alternative one. For instance, if unobservables have independent-across-alternatives extreme value distributions, the jth component of this function takes the well known logistic form $Q_x^j(\{\tilde{v}(x,a)\}) = \exp\{\tilde{v}(x,j)/\sigma\}\,[1 + \sum_{a=2}^{J} \exp\{\tilde{v}(x,a)/\sigma\}]^{-1}$.

A general representation of the mapping $Q_x()$ can be obtained from McFadden's (1981) *social surplus function:*

$$(4) \qquad S(\{v(x,a), a \in A\}, x) = \int \max_{a \in A}[v(x,a) + \epsilon(a)]\, g(d\epsilon|x).$$

The social surplus function computes expected utility, in the multinomial choice-random utility framework, as a function of the set of choice-specific utilities. Clearly, $V_\sigma(x) = S(\{v(x,a), a \in A\}, x)$. The *Williams-Daly-Zachary (WDZ) theorem* establishes the following properties of the social surplus function:[8] (i) it is strictly convex; (ii) it is additive: for any scalar α, $S(\alpha + \{v(x,a)\}, x) = \alpha + S(\{v(x,a)\}, x)$; and (iii) its gradient is equal to the vector of conditional choice probabilities: $P(a|x) = \partial S(\{v(x,a)\}, x)/\partial v(x,a)$. Using properties (ii) and (iii) we obtain the representation $Q_x^j(\{\tilde{v}(x,a) : a > 1\}) = \partial S([0, \{\tilde{v}(x,a) : a > 1\}], x)/\partial v(x,j)$.

Second, Hotz-Miller's invertibility proposition states that continuous differentiability and the strict convexity of the social surplus function imply that, for every x, $Q_x()$ is invertible.

Third, Hotz and Miller (1993) also showed how invertibility can be exploited to obtain alternative *representations of the value function in terms of choice*

[8] See Rust (1994b, Theorem 3.1, and references therein). The *strict* convexity of the social surplus function is actually a strengthening of the WDZ theorem which follows from the unbounded support and continuity of the distribution of unobservables.

probabilities. We now restate one such representation in our framework.[9] First, notice that the Bellman equation (2) can be rewritten as

$$V_\sigma(x) = \sum_{a\in A} P(a|x)\Big\{u(x,a) + E[\epsilon(a) \mid x,a] + \beta \sum_{x'} f(x'|x,a)V_\sigma(x')\Big\} \tag{5}$$

where $E[\epsilon(a) \mid x,a]$ is the expectation of the unobservable $\epsilon(a)$ conditional on the optimal choice of alternative a:

$$E[\epsilon(a) \mid x,a] = [P(a|x)]^{-1} \int \epsilon(a) I\{\tilde{v}(x,a) + \epsilon(a) \geq \tilde{v}(x,j) + \epsilon(j),\ j \in A\} g(d\epsilon|x). \tag{6}$$

Clearly, the conditional expectations $E[\epsilon(a)|.]$ are functions of the utility differences $\tilde{v}(x,a)$. Since the mapping Q_x from utility differences into choice probabilities is invertible, it follows that these conditional expectations can be expressed as functions of the choice probabilities. We denote these functions by $e_x(a, \{P(j|x)\})$. In the case of extreme value unobservables they have the closed form $e_x(a, \{P(j|x)\}) = \gamma - \ln(P(a|x))$ where γ is Euler's constant. Let us substitute these functions into (5) and stack the M equations for each possible value of x. In compact matrix notation we get

$$V_\sigma = \sum_{a\in A} P(a) * [u(a) + e(a,P) + \beta F(a) V_\sigma] \tag{7}$$

where $*$ is the Hadamard (or element-by-element) product; V_σ is the $M \times 1$ vector describing the value function $V_\sigma(x)$; P is the $M(J-1) \times 1$ vector of conditional choice probabilities, alternative one excluded; $P(a), u(a)$, and $e(a,P)$ are $M \times 1$ vectors that stack the corresponding elements at all states for alternative a; and $F(a)$ is the $M \times M$ matrix of conditional transition probabilities $f(x'|x,a)$. This system of fixed point equations can be solved for the value function to obtain V_σ as a function of P:

$$V_\sigma = \varphi(P) = (I_M - \beta F^U(P))^{-1} \Big\{ \sum_{a\in A} P(a) * [u(a) + e(a,P)] \Big\} \tag{8}$$

where $F^U(P) = \sum_{a\in A} P(a) * F(a)$ is the $M \times M$ matrix of unconditional transition probabilities induced by P.

2.3. *The Fixed Point Problem in Probability Space*

We now introduce a policy iteration operator Ψ in the space of conditional choice probabilities. This operator is the cornerstone of our estimation algorithm. In the preceding subsections we defined conditional choice probabilities in terms

[9] The representation we restate here is one of two representations proposed by Hotz-Miller for finite horizon models; it was used by Aguirregabiria (1999) in an infinite horizon Markov model. See our comment on the alternative representation in Section 4.

of the value function in equation (3) and we showed that the value function can be written in terms of conditional choice probabilities in equation (8). Substituting equation (8) into equation (3) for all choices and states we get:

$$P = \Psi(P) \equiv \Lambda(\varphi(P)). \tag{9}$$

$\varphi()$ is a policy valuation operator that maps an $M(J-1)\times 1$ vector of conditional choice probabilities into an $M \times 1$ vector in value function space using Hotz and Miller's representation. $\Lambda()$ is a policy improvement operator that maps an $M \times 1$ vector in value function space into an $M(J-1)\times 1$ vector of conditional choice probabilities. It stacks all choice probabilities associated with the value function, as defined in equation (3).

The policy iteration operator Ψ can be evaluated at any vector of conditional choice probabilities, *optimal or not*. For an arbitrary vector of choice probabilities P^0, the valuation operator $\varphi(P^0)$ returns the value function corresponding to the arbitrary behavior represented by P^0. For an arbitrary vector of values, say V_σ^0, the policy improvement mapping $\Lambda(V_\sigma^0)$ returns the optimizing agent's choice probabilities under the assumption that expected utilities as of next period are given by the vector V_σ^0. Thus the composite mapping $\Psi(P^0)$ should be interpreted as giving the current optimal choice probabilities of an agent whose future behavior will be to randomize over alternatives according to P^0.[10]

Notice that by the definitions in equations (3) and (8) we have that $V_\sigma = \varphi(P)$ and $P = \Lambda(V_\sigma)$; therefore it is clear that the set of optimal choice probabilities P is a fixed point of Ψ. Thus the original fixed point problem in "value space" can be reformulated as a fixed point problem in "probability space." The following propositions establish the relationship between the two fixed point problems and an important property of the policy iteration operator.

PROPOSITION 1: *Under Assumptions* 1–3:

(a) *Ψ has a unique fixed point P.*

(b) *The sequence $P^K = \Psi(P^{K-1})$, $K = 1, \ldots, \infty$, converges to P for any P^0.*

(c) *Equivalence of Ψ and Newton iterations: For any P^0, consider the pair of linked sequences $\{V_\sigma^K, P^K\}$ defined by $V_\sigma^K = \varphi(P^K)$, $P^{K+1} = \Lambda(V_\sigma^K)$. Clearly, $P^K = \Psi(P^{K-1})$. Then, $\{V_\sigma^K\}$ is the sequence of Newton iterations converging to the unique solution of the Bellman equation* (2).

PROPOSITION 2: *Under Assumptions* 1–3, *the Jacobian matrices of $\varphi(.)$ and $\Psi(.)$ are zero at the fixed point P.*

Proposition 2 establishes that at the fixed point it is not possible to increase expected utility by changing choice probabilities; that is, the optimal choice probabilities maximize the valuation operator locally. As a consequence, the Jacobian

[10] The interpretation of Ψ as a policy iteration operator is made explicit in the proof of Proposition 1. We provide a reformulation of the original DDP in which unobserved state variables are 'integrated out,' conditional choice probabilities are the (continuous) control variables, V_σ is the value function, and Ψ is the policy iteration operator.

of the policy iteration operator is zero. This result is the key to the properties of the algorithm and the sequential estimators we propose in this paper.

We end this section with remarks on the role of Assumptions 1–3. Assumption 3 allows us to describe functions as vectors in a Euclidean space and to use compact matrix notation, but we believe it is not essential for the results in this section.[11] On the contrary, Assumptions 1 and 2 are central to our analysis. Assumption 2 (Conditional Independence) is necessary for future utility differences not to depend on current unobservable state variables. Together with Assumption 1 (Additivity), this plays a crucial role in Hotz and Miller's representation of the value functions that we use to define the policy iteration operator. Assumption 1 can be relaxed to allow for multiplicative separability between observable and unobservable components. The differentiability in Assumption 2 is needed in order to make the policy iteration operator a differentiable function of choice probabilities.

3. MAXIMUM LIKELIHOOD ESTIMATION AND NESTED ALGORITHMS

Let θ_u, θ_g, and θ_f be the vectors of unknown parameters in the utility function u, the density of unobservables g, and the conditional transition probability function f, respectively. That is, $\theta \equiv (\theta_u, \theta_g, \theta_f)$. Hereafter, we incorporate the subindex θ, or a component of θ, in those functions that depend on the structural parameters. In order to guarantee the existence, consistency, and asymptotic normality of the ML estimator, we impose smoothness of the primitives with respect to θ.

ASSUMPTION 4: *$u_{\theta_u}(x,a)$, $g_{\theta_g}(\epsilon|x)$, and $f_{\theta_f}(x'|x,a)$ are continuous and twice differentiable with respect to θ.*

Suppose our data set consists of a cross-section of observations from a random sample of individuals $\{x_i, a_i, x_i' : i = 1, \ldots, n\}$. Under Assumption 2, the log-likelihood function of the model can be decomposed into conditional choice probability and transition probability terms as follows:

$$l(\theta) = l_1(\theta) + l_2(\theta_f) = \sum_{i=1}^{n} \ln P_\theta(a_i|x_i) + \sum_{i=1}^{n} \ln f_{\theta_f}(x_i'|x_i, a_i). \tag{10}$$

Consistent estimates of the conditional transition probability parameters θ_f can be obtained from transition data without having to solve the Markov decision model. In the rest of the paper we focus on the estimation of $\alpha \equiv (\theta_u, \theta_g)$ given initial consistent estimates of θ_f obtained from likelihood $l_2(\theta_f)$.[12] This two-stage

[11] For instance, the representation of the value function $V_\sigma = \varphi(P)$ in (8) still holds with continuous state variables but, instead of inverting a matrix, the inversion of an infinite-dimensional linear operator is required. This can be done via approximation methods; see Rust (1996). However, the implications of the approximation error on the statistical properties of the estimators are unknown in this literature.

[12] The discount factor β is assumed to be known.

estimation strategy, which reduces the computational burden of estimation, was used in Rust (1987) and Hotz and Miller (1993).[13]

Let α^* denote the *true* value of α hereafter. The (partial) MLE of α^* can be computed using Rust's well known *nested fixed point algorithm* (NFXP). In this procedure, an 'inner' fixed point algorithm computes the conditional choice probabilities $P_\theta = \Psi_\theta(P_\theta)$ and their derivatives for given parameter values. The 'outer' algorithm feeds on this solution and maximizes the likelihood with respect to α using the BHHH method.[14] We propose an alternative nested procedure:

Nested Pseudo Likelihood Algorithm (NPL):

Let $\hat{\theta}_f$ be an estimate of θ_f. Start with an initial guess for the conditional choice probabilities, $P^0 \in [0,1]^{MJ}$. At iteration $K \geq 1$, apply the following steps:

Step 1: Obtain a new pseudo-likelihood estimate of α, α^K, as

$$\alpha^K = \arg\max_{\alpha \in \Theta} \sum_{i=1}^{n} \ln \Psi_{(\alpha,\hat{\theta}_f)}(P^{K-1})(a_i|x_i) \tag{11}$$

where $\Psi_\theta(P)(a|x)$ is the element (a,x) of $\Psi_\theta(P)$.

Step 2: Update P using the 'arg max' from step 2, i.e.

$$P^K = \Psi_{(\alpha^K,\hat{\theta}_f)}(P^{K-1}). \tag{12}$$

Iterate in K until convergence in P (and α) is reached.

In our nested procedure we swap the order of the two algorithms. That is, the outer algorithm (step 2) iterates on Ψ to solve the fixed point problem, and the inner algorithm (step 1) maximizes a pseudo-likelihood function. In the inner algorithm a fixed P^{K-1} is used to construct the pseudo value function $\varphi_{(\alpha,\hat{\theta}_f)}(P^{K-1})$ and choice probabilities $\Psi_{(\alpha,\hat{\theta}_f)}(P^{K-1})$ that approximate the exact choice probabilities $P_{(\alpha,\hat{\theta}_f)}$. The outer algorithm updates the value of P and checks for convergence to a fixed point. Note that each outer policy iteration is performed with updated parameter values.

The NFXP algorithm always converges to a root of the likelihood equations. We show that NPL satisfies a weaker version of the same property.[15]

[13] See Rust (1987, 1994b) for a 3-stage extension that is asymptotically equivalent to full maximum likelihood.

[14] See Rust (1988) for a description of NFXP. The inner algorithm is a *hybrid fixed point algorithm* that begins with successive approximation iterations in the Bellman operator Γ_σ and then switches to Newton-Kantorovich iterations in order to guarantee convergence. Results (b) and (c) in our Proposition 1 suggest that the initial successive approximation iterations are actually not necessary for convergence.

[15] However, neither NFXP nor NPL guarantees convergence to a *global* maximum of the likelihood. In both cases, the researcher should start the algorithms from different points and keep track of the maxima. Note also that we are implicitly assuming that convergence of NPL's inner algorithm is not an issue. NFXP converges if the BHHH method with optimal step is used in its outer algorithm. The same method can be used in NPL's inner algorithm.

PROPOSITION 3 (Equivalence of NFXP and NPL): *Suppose the pseudo-likelihood maximization problems in* (11) *have unique interior solutions for any sample and any value of* P. *Then, if NPL converges, it does so to a root of the likelihood equations.*

Proposition 3 is an application of Proposition 2. In particular, Proposition 2 implies that $\partial P_\theta/\partial\theta' = \partial\Psi_\theta(P_\theta)/\partial\theta'$, and therefore the pseudo-score is exactly equal to the score when the pseudo-score is evaluated at the fixed point of the policy iteration operator. If the problem is sufficiently smooth, then on convergence NPL delivers a fixed point pair; that is, $P_{NPL} = \Psi_{(\alpha_{NPL},\hat{\theta}_f)}(P_{NPL})$. It follows that the NPL algorithm has found a root of the likelihood equation.[16]

We now give an example that illustrates practical issues in the implementation of the NPL algorithm. Consider a class of models where: (i) the unobservable ϵ's have independent across alternatives, extreme value distributions; and (ii) there is *multiplicative separability* between x and θ_u in the utilities, i.e., $u_{\theta_u}(x,a) = h(x,a)'\eta(\theta_u)$ where $h(x,a)$ and $\eta(\theta_u)$ are known vector-valued functions with dimension p. Define $H(a)$ as the $M \times p$ matrix with rows $h(x,a)$ for each value of x. In this case, the policy/iteration operator is

$$\Psi_{(\alpha,\theta_f)}(P)(a|x) = \frac{\exp\{\tilde{h}_{(\theta_f,P)}(x,a)'\eta(\theta_u) + \tilde{e}_{(\theta_f,P)}(x,a)\}}{\sum_{j=1}^{J}\exp\{\tilde{h}_{(\theta_f,P)}(x,j)'\eta(\theta_u) + \tilde{e}_{(\theta_f,P)}(x,j)\}}$$

where $\tilde{h}_{(\theta_f,P)}(x,a)'$ is a row of the matrix

$$H(a) + \beta F(a)\left(I_M - \beta F^U_{\theta_f}(P)\right)^{-1}\sum_{j=1}^{J} P(j) * H(j),$$

and $\tilde{e}_{(\theta_f,P)}(x,a)$ is an element of the vector

$$\beta F(a)\left(I_M - \beta F^U_{\theta_f}(P)\right)^{-1}\sum_{j=1}^{J} P(j) * [\gamma - \ln P(j)].$$

This model has several features that make the use of NPL specially advantageous. First and most important, it is very convenient that the extreme value assumption gives closed-form conditional expectation functions $e(j,P) = \gamma - \ln P(j)$. In general computing $e()$ would involve first inverting the $Q_x()$ mappings to obtain utility differences and then solving the multiple integration problem in (6) given utility differences. For distributions other than the extreme value this

[16] Note that this result is conditional on convergence and that convergence itself has not been proved. However, a weak convergence result can be established. Under the regularity conditions of Proposition 4 we can show that the probability that NPL converges locally goes to 1 as the sample size increases. By local convergence we mean that if P^0 is sufficiently close to $\widehat{P}_n$, then P^K converges to $\widehat{P}_n$, in K and α^K converges to $\hat{\alpha}_n$, where $\hat{\alpha}_n$ is the maximum likelihood estimator and $\widehat{P}_n$ are the corresponding choice probabilities. More details are available from the authors.

may be a serious computational problem.[17] Second, the extreme value assumption also implies that integration over unobservables during pseudo-likelihood estimation has a simple (logistic) closed form. Third, since $\{\tilde{h}_{(\theta_f, P)}(x, a)\}$ and $\{\tilde{e}_{(\theta_f, P)}(x, a)\}$ do not depend on α, they are fixed over a pseudo-likelihood estimation. To obtain $\Psi_{(\alpha, \theta_f)}(P)(a|x)$ for different values of α we do not have to repeat the inversion and multiplication of large matrices that is required for policy valuation. And fourth, the pseudo-likelihood function is globally concave in $\eta(\theta_u)$, which guarantees convergence of the hill-climbing pseudo-likelihood iterations for any initial value of θ_u.[18] In contrast, to compute the probabilities $P_{(\alpha, \theta_f)}(a|x)$ that enter the likelihood function we have to invert and multiply $M \times M$ matrices repeatedly in policy iteration. Furthermore, the likelihood function is not globally concave in α, nor in a transformation of α. Therefore, convergence of NFXP's outer BHHH algorithm may require the use of optimal steps with a significant increase of the computational cost of estimation.[19] In Section 5 we show that NPL is indeed much faster than NFXP for Rust's bus replacement problem, which is an example of this class of models.

Finally, note that Proposition 2 is crucial to obtain equivalence of NFXP and NPL. Based on this, it is straightforward to see that the equivalence result also holds for full likelihood versions of NFXP and NPL.[20] Also note that finite horizon models are covered in our infinite horizon framework if the decision period t is included among the observable state variables.[21] Next section studies the properties of NPL when it is initialized with a consistent estimator of the true conditional choice probabilities.

4. SEQUENTIAL POLICY ITERATION ESTIMATORS

Let $\hat{\theta}_f$ denote a consistent estimator of conditional transition probability parameters, and let P^0 be a consistent, nonparametric estimator of the true conditional choice probabilities P^*. Consider using P^0 as an initial guess in our NPL

[17] In the context of static discrete choice models of product differentiation, Berry (1994) shows that the mapping from utility differences to market shares (i.e., choice probabilities) is invertible and that for any vector of choice probabilities the inverse is the unique fixed point of a mapping in utility space. Berry, Levinsohn, and Pakes (1995) solve this inverse problem in the estimation of a model of product differentiation in the automobile industry.

[18] Given the pseudo MLE of $\eta(\theta_u)$, we can obtain an estimator of θ_u using minimum distance. This is the pseudo MLE of θ_u if $\eta()$ is a one-to-one function.

[19] Without multiplicative separability in the utility function, the pseudo likelihood function will not in general be globally concave. Furthermore, to obtain the conditional choice value functions $v(x, a)$ for different values of α we have to multiply $M \times M$ matrices. However, it is still true that we do not have to invert the matrix $I - \beta F^U(P)$, which is the main computational cost of policy valuation.

[20] See the Appendix. However, our discussion above suggests that the potential gains from using NPL instead of NFXP are much larger in a partial likelihood context.

[21] In infinite horizon problems the inverse of the matrix $(I - \beta F^U)$, which is used in the valuation operator, has a very special structure. On the other hand, backward induction reduces the cost of NFXP's inner algorithm. Therefore, it is harder to conjecture what the potential gains of using NPL are in this case. However, all the results in Propositions 1–4 of this paper apply, beginning with the key 'zero-jacobian' property of Proposition 2.

algorithm. Performing one, two, and in general K iterations of the NPL algorithm yields a sequence $\{\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^K\}$ of statistics that can be used as estimators of α^*. We call them *sequential policy iteration (PI) estimators*. Thus, for $K \geq 1$, the K-stage PI estimator is defined as:

$$\hat{\alpha}^K = \arg\max_{\alpha \in \Theta} \sum_{i=1}^{n} \ln \Psi_{(\alpha, \hat{\theta}_f)}(P^{K-1})(a_i|x_i) \tag{13}$$

where $P^K = \Psi_{(\hat{\alpha}^K, \hat{\theta}_f)}(P^{K-1})$, and P^0 is a consistent, nonparametric estimator of the true conditional choice probabilities P^*.

In this section we study the asymptotic statistical properties of this sequence of estimators. The main result is in Proposition 4, which shows that for any value of K the PI estimators are consistent and asymptotically equivalent to the partial MLE of α^*. More formally we have the following proposition.

PROPOSITION 4: *Let P^* be the true set of conditional choice probabilities, and let f^* be the true conditional transition probability of x. Let Θ_α, Θ_f, and $\Theta_P \equiv [0,1]^{(J-1)M}$ be the set of possible values of α, θ_f, and P, respectively. Consider the following regularity conditions*:

(a) *Θ_α and Θ_f are compact sets.*

(b) *$\Psi_{(\alpha, \theta_f)}(P)$ is continuous and twice continuously differentiable in α, θ_f, and P.*

(c) *$\Psi_{(\alpha, \theta_f)}(P)(a|x) > 0$ for any $(a, x) \in A \times X$ and any $\{\alpha, \theta_f, P\} \in \Theta_\alpha \times \Theta_f \times \Theta_P$.*

(d) *$\{a_i, x_i', x_i\}$ for $i = 1, 2, \ldots, n$ are independently and identically distributed, and $\Pr(x_i = x) > 0$ for any $x \in X$.*

(e) *There is a unique $\theta_f^* \in \text{int}(\Theta_f)$ such that $f_{\theta_f^*}(x'|x, a) = f^*(x'|x, a)$ for all (x', x, a).*

(f) *There is a unique $\alpha^* \in \text{int}(\Theta_\alpha)$ such that, for any $(a, x) \in A \times X$, $P_{(\alpha^*, \theta_f^*)}(a|x) = P^*(a|x)$. Furthermore, for any $\alpha \neq \alpha^*$ the set $\{(a, x) : \Psi_{(\alpha, \theta_f^*)}(P^*)(a|x) \neq P^*(a|x)\}$ is nonempty.*

(g) *$\hat{\theta}_f$ and P^0 are strongly consistent estimators of θ_f^* and P^*, respectively, and*

$$\left[(1/\sqrt{n}) \sum_{i=1}^{n} \partial \ln \Psi_{(\alpha^*, \theta_f^*)}(P^*)(a_i|x_i)/\partial \alpha', \right.$$
$$\left. \sqrt{n}(\hat{\theta}_f - \theta_f^*)', \sqrt{n}(P^0 - P^*)' \right]' \to_d N(0, \Omega).$$

Under these conditions, a policy iteration estimator, $\hat{\alpha}^K$, is root-n-consistent, asymptotically normal, and asymptotically equivalent to the partial maximum likelihood estimator, i.e., $\sqrt{n}(\hat{\alpha}^K - \alpha^*) \to_d N(0, V^*)$ where

$$V^* = \Omega_{00}^{-1} + \Omega_{00}^{-1}\{H_f \Omega_{f0} + \Omega_{0f} H_f' + H_f V(\hat{\theta}_f) H_f'\}\Omega_{00}^{-1}$$

and

$$\Omega_{00} = E\left(\frac{\partial \ln P_i^*}{\partial \alpha}\frac{\partial \ln P_i^*}{\partial \alpha'}\right), \quad H_f = E\left(\frac{\partial \ln P_i^*}{\partial \alpha}\frac{\partial \ln P_i^*}{\partial \theta_f'}\right),$$

$$\Omega_{f0} = \Omega_{0f}' = \left[E\left(\frac{\partial \ln f_i^*}{\partial \theta_f}\frac{\partial \ln f_i^*}{\partial \theta_f'}\right)\right]^{-1} E\left(\frac{\partial \ln f_i^*}{\partial \theta_f}\frac{\partial \ln P_i^*}{\partial \alpha'}\right),$$

with $P_i^* \equiv P_{(\alpha^*, \theta_f^*)}(a_i|x_i)$ *and* $f_i^* \equiv f_{\theta_f^*}(x_i'|x_i, a_i)$.

The K-stage PI estimator is a particular case of a *quasi-generalized M-estimator* as defined in Gourieroux and Monfort (1995), with (θ_f^*, P^*) as the vector of nuisance parameters. The proof of Proposition 4 uses an induction argument. We show that if P^{K-1} is root-n-consistent and asymptotically normal, then $\sqrt{n}(\hat{\alpha}^K - \alpha^*) \to_d N(0, V^*)$, and $P^K \equiv \Psi_{(\hat{\alpha}^K, \hat{\theta}_f)}(P^{K-1})$ is also root-n-consistent and asymptotically normal. Since P^0 is root-n-consistent and asymptotically normal by assumption, the proof is complete.[22]

The regularity conditions (b) and (c) follow from our assumptions in Sections 2 and 3. The assumption of iid observations in condition (d) makes the proof of Proposition 4 easier, but it can be relaxed.[23] Conditions (e) and (f) are identification assumptions. In particular, (f) implies that α^* uniquely maximizes in α both $E[\ln \Psi_{(\alpha, \theta_f^*)}(P^*)(a_i|x_i)]$ and $E[\ln P_{(\alpha, \theta_f^*)}(a_i|x_i)]$. That is, this condition implies that α^* is identified from the partial likelihood as well as from the pseudo-likelihood that is used to obtain the K-stage estimators.

Notice that the asymptotic variance of a K-stage PI estimator does not depend on the variance of the initial nonparametric estimator of the choice probabilities, P^0. This asymptotic property of the PI estimators results from Proposition 2. In particular, in a Taylor expansion of the first order conditions that define a K-stage PI estimator, the term associated with P^{K-1} is

$$\left[(1/n)\sum_{i=1}^{n} \partial^2 \ln \Psi_{(\hat{\alpha}^K, \hat{\theta}_f)}(P^{K-1})(a_i|x_i)/\partial\theta\,\partial P'\right]\sqrt{n}(P^{K-1} - P^*).$$

The expression within brackets converges in probability to

$$E(\partial^2 \ln \Psi_{(\alpha^*, \theta_f^*)}(P^*)(a|x)/\partial\theta\,\partial P'),$$

[22] Given our assumptions that x is discrete and $\Pr(x_i = x^m) > 0$ *for any* $x^m \in X$, obtaining $\sqrt{n}$-consistent nonparametric estimators is straightforward. Delgado and Mora (1995a, b) show that in nonparametric regression with discrete regressors frequency or 'cell' estimators and nearest-neighbor estimators are $\sqrt{n}$-consistent under very weak conditions. Frequency estimators for empty x-cells in a given sample are defined to be zero. Bierens (1983) proves that in a nonparametric regression with discrete explanatory variables, the kernel estimator is root-n-consistent.

[23] We consider a cross-section for the sake of simplicity, but our results here and in Section 3 of the paper can be extended to the case of panel datasets as long as the conditional independence assumption holds along both dimensions of the panel.

which by Proposition 2 and the equivalence of the information matrix is zero; and $\sqrt{n}(P^{K-1} - P^*)$ converges to a vector of random variables with finite variances. Therefore, this term does not have any effect on the asymptotic distribution of $\sqrt{n}(\hat{\alpha}^K - \alpha^*)$. Furthermore, this asymptotic distribution is the same as that of the MLE because the score and pseudo-score variables are identical at the true parameter values.

Using consistent estimators of α^* and θ_f^*, it is straightforward to obtain a consistent estimator of V^*. Furthermore, though V^* depends on expectations that involve partial derivatives for P_{θ^*}, by Proposition 2 it is possible to estimate consistently V^* by using $\partial \Psi_{(\hat{\alpha}^K, \hat{\theta}_f)}(P^{K-1})/\partial\alpha$ instead of $\partial P_{(\hat{\alpha}^K, \hat{\theta}_f)}/\partial\alpha$. That is, the estimation of V^* does not require one to solve once the fixed point problem to obtain the vector of choice probabilities $P_{(\hat{\alpha}^K, \hat{\theta}_f)}$ and its partial derivatives.

It is simple to verify that Proposition 4 can be extended to a full maximum likelihood context, i.e., the joint estimation of α^* and θ_f^*. However, as mentioned in Section 3 we focus on partial maximum likelihood estimation because we believe it is in this context that the potential computational gains of the NPL algorithm and PI estimators are greatest.

It is clear from Proposition 3 that as the NPL algorithm converges in P^K the corresponding K-stage PI estimators converge to the NFXP estimator. The family of PI estimators also encompasses a *conditional choice probability* (CCP) estimator (Hotz and Miller (1993)). The CCP estimators were defined as the values of α that solve systems of equations of the form:

$$\sum_{i=1}^{n}\sum_{j=1}^{J} Z_i^j \left[I(a_i = j) - \widetilde{P}_{(\alpha, \hat{\theta}_f)}(P^0)(j|x_i) \right] = 0 \tag{14}$$

where $\{Z_i^j\}$ are vectors of instrumental variables (e.g., functions of x_i) and $\widetilde{P}$ denotes conditional choice probability functions that use Hotz and Miller's invertibility result to simplify the computation of the future component of value functions. It is straightforward to verify that the 1-stage PI estimator is a CCP estimator when $\widetilde{P}()$ is the policy iteration operator and $Z_i^j = \partial \ln \Psi_{(\alpha, \hat{\theta}_f)}(P^0)(j|x_i)/\partial\alpha$. For problems with terminating actions, Hotz and Miller proposed $\widetilde{P}$ functions based on a second representation of the value functions. This alternative representation is not equivalent to the valuation operator in equation (8). In particular, it does not satisfy Proposition 2 and as a result the corresponding CCP estimators are not asymptotically equivalent to the partial MLE.[24]

[24] If $a = 1$ is the terminating action, Hotz and Miller's second representation of the value function in terms of conditional choice probabilities is

$$\varphi_2(P)(x) = \sum_{j=1}^{J} P(j|x) e(j, \{P(d|x)\}) + v(1, x) + \sum_{j=2}^{J} P(j|x) Q_j^{-1}(\{P(d|x)\}).$$

For problems with the terminating action property, this is simpler to compute than the valuation operator $\varphi(P)$. Note that $\varphi(P) = \varphi_2(P)$ only at the fixed point $P = \Psi(P)$. Hotz and Miller (1993)

Given the asymptotic equivalence of the PI estimators and the partial MLE, the analysis of the relative performance of these estimators in finite samples is of interest. Unless there is a loss of precision, why not use the computationally inexpensive 1-stage PI estimator? An important limitation of Hotz-Miller's estimator is that initial nonparametric estimates of the conditional choice probabilities can be very imprecise, and this lack of accuracy might be transmitted, in finite samples, to the estimates of the structural parameters (see Eckstein and Wolpin (1989), Rust (1994a), and Hotz et al. (1994)). Our K-stage estimator may overcome this problem by iterating K times in the smooth policy iteration operator. The computational cost of these $(K-1)$ additional iterations is equal to the cost of $(K-1)$ policy iterations and $(K-1)$ pseudo-maximum likelihood estimations. Therefore, for intermediate values of K we get asymptotically equivalent estimators that are still cheaper to obtain than the MLE, yet potentially more precise in finite samples than the Hotz-Miller estimator. Is a 2-stage or 3-stage estimator enough to obtain significant gains in precision with respect to Hotz-Miller? We address this issue with a Monte Carlo experiment in Section 5.2.

5. THE PERFORMANCE OF THE NPL ALGORITHM AND THE PI ESTIMATORS: AN EXAMPLE

5.1. *Relative Speed of NPL and NFXP*

In order to illustrate the performance of our NPL algorithm in maximum likelihood estimation, we use Rust's well known bus replacement model and dataset.[25] We obtain *partial* ML estimates using the NPL and the NFXP algorithms for different specifications of the model according to the dimension of the state space (i.e., from 100 to 1100 cells) and the number of structural parameters (i.e., 2 and 4).[26]

Clearly, the CPU time required by these algorithms will depend on our choice of the values that initialize them, i.e., initial guesses of structural parameters for NFXP and initial conditional choice probabilities for NPL. In order to make the initial values for the two algorithms comparable, we consider a researcher who obtains these guesses from the data using a Hotz-Miller estimator. That is, the initial vector of structural parameters for NFXP is the Hotz-Miller estimator $\hat{\theta}_1$, and the initial conditional choice probabilities for NPL are the ones obtained from the Hotz-Miller estimator $\Psi_{\hat{\theta}_1}(P^0)$.

The results from this experiment can be summarized as follows. An important feature of this example is that fixed point iterations (i.e., policy iterations) are

described the CCP estimator in terms of $\varphi(P)$ but they used $\varphi_2(P)$ in estimation of their contraceptive choice model.

[25] See Rust (1987). Rust's model has been used in other studies to evaluate the performance of alternative algorithms and estimators, e.g. Hotz et al. (1994) and Rust (1997).

[26] The NFXP algorithm that we use has two features that contribute very significantly to improve its computational efficiency. First, we use a closed-form expression for the gradient of the likelihood function. Second, at each outer iteration, we use "smart guesses" for the vector of choice probabilities that initialize the policy iteration algorithm. Both features have been considered by Rust (1987, 1988).

much more expensive than likelihood and pseudo-likelihood climbing iterations. In fact, when the number of cells in the state space is larger than 200, policy iterations represent almost 100% of the estimation time for both algorithms. Therefore, it is very relevant to assess to what extent the use of NPL instead of NFXP reduces the number of policy iterations in the estimation procedure. The size of the state space does not affect the number of policy iterations in any of the two algorithms. For the model with two parameters the ratio of the number of policy iterations of NFXP and NPL is 5.5, and for the model with four parameters this ratio is equal to 9. Therefore, in these examples, NPL is 5.5 and 9 times faster, respectively, than the NFXP algorithm. Note that the use of the Hotz-Miller estimator to produce 'comparable' initial guesses actually undervalues the relative performance of the NPL algorithm. When we considered cases in which initial guesses were very poor, NPL always converged and the number of policy iterations decreased relative to NFXP with equally arbitrary initial guesses.[27] For instance, when we used a vector of zeros as an initial guess for the NFXP algorithm and random draws from a $U(0,1)$ to initialize NPL, the ratio of policy iterations NFXP/NPL was 10 for the model with two parameters and 15 with four parameters.

A final warning seems appropriate given the limited scope of this experiment. The main features of this example are extreme value distributed unobservables, partial likelihood estimation, infinite horizon, binary choice, and multiplicative separability between parameters and state variables in the utility functions. The relative speed of NPL and NFXP could be different in other applications.

5.2. *The Precision of PI Estimators: A Monte Carlo Exercise*

In Section 4 we established that all K-stage PI are asymptotically equivalent. We now illustrate their behavior in finite samples using a Monte Carlo experiment. We used Rust's bus engine model with parameters equal to ML estimates as the DGP. For sample sizes 1000, 5000, and 10000, we generated 1000 samples and for each of them we obtained the sequence of PI estimators and its limit, the ML estimator. The size of the state space grid is 200. Since the experiments with different sample sizes provided very similar results, we only report the experiment with 1000 observations.

Table I presents summary statistics for this experiment. For the 1, 2, and 3-stage PI estimators, we report the mean and median of the absolute estimation error and the empirical standard deviations, in all cases relative to the corresponding statistic for the MLE. We find that the finite sample properties of the estimators improve monotonically with K. The 1-PI estimator performs poorly and there are very significant benefits of doing more than one policy iteration.

[27] In fact, NFXP did not always converge even when initial values were within a 1% confidence region around the ML estimate. NFXP is guaranteed to converge if we compute an optimal step in the hill climbing iteration. However, the use of an optimal step requires one to solve the dynamic programming problem several times at each hill climbing iteration, and this increases enormously the computational cost of estimation.

TABLE I
MONTE CARLO EXPERIMENT

Experiment design	
Model:	Bus engine replacement model (Rust)
Parameters:	$\theta_0 = 10.47$; $\theta_1 = 0.58$; $\beta = 0.9999$
State space:	201 cells
Number observations:	1000
Number replications:	1000
Initial probabilities:	Kernel estimates

Monte Carlo distribution of MLE
(In parenthesis, percentages over the true value of the parameter)

	θ_0	θ_1
Mean Absolute Error:	2.08 (19.9%)	0.17 (29.0%)
Median Absolute Error:	1.56 (14.9%)	0.13 (22.7%)
Std. dev. estimator:	2.24 (21.4%)	0.16 (26.9%)
Policy iterations (avg.):	6.2	

Monte Carlo distribution of PI estimators (relative to MLE)
(All entries are 100* (K-PI statistic-MLE statistic)/MLE statistic)

		Estimators		
Parameter	Statistics	1-PI	2-PI	3-PI
θ_0	Mean AE	4.7%	1.6%	0.3%
	Median AE	14.2%	0.2%	−0.3%
	Std. dev.	6.8%	1.2%	0.3%
θ_1	Mean AE	18.7%	1.5%	0.2%
	Median AE	25.1%	0.7%	0.6%
	Std. dev.	11.0%	1.3%	0.2%

More interestingly, the relative performance of the 2-PI is excellent. Most of the benefits of additional policy iterations are obtained when one goes from 1 to 2 iterations.[28]

In Table II we study the discrepancy between the empirical standard deviations of the estimators in the Monte Carlo distributions and the average estimate of their asymptotic standard error. We can see that for the 2-stage and 3-stage this discrepancy is small and of the same order of magnitude as for the MLE. However, for the l-stage estimator the estimated asymptotic standard error is clearly downward biased.

Therefore, when we construct expected value functions using choice probabilities that are closer to the fixed point, our estimates become much more precise. In this application, two policy iterations are enough to get choice probabilities

[28] We also considered the performance of these estimators when the initial values for the choice probabilities are very imprecise (i.e., kernel estimator plus a noise term). The findings from this experiment, which we do not report here, actually reinforce the results in Table I: poor initial guesses have a very serious effect on the 1-PI estimator but a relatively small effect on the 2-PI, which is still close to the MLE.

TABLE II

RATIO BETWEEN ESTIMATED STANDARD ERRORS AND STANDARD DEVIATION OF MONTE CARLO DISTRIBUTION

Parameters	Statistics	Estimators			
		1-PI	2-PI	3-PI	MLE
θ_0	Ratio	0.801	1.008	1.027	1.022
θ_1	Ratio	0.666	1.043	1.076	1.065

that are close enough to the fixed point. A possible interpretation of this finding is based on the qualitative differences between going from 1 to 2 policy iterations and going from 2 to more iterations. The second policy iteration is the first one in which we incorporate the structure of the model to obtain the choice probabilities used to compute the pseudo-social surplus. These probabilities incorporate parametric assumptions on preferences and the distribution of unobservables that are not contained in the initial nonparametric estimates. All subsequent policy iterations impose (recursively) that the choice probabilities should be a fixed point of the policy iteration mapping, but no further assumptions about the functional form of the primitives are incorporated.

6. CONCLUDING REMARKS

We have proposed a new nested algorithm (NPL) for the estimation of the class of discrete Markov decision models with the conditional independence assumption. Our method is based on a representation of the solution of the dynamic programming problem in the space of conditional choice probabilities. When the NPL algorithm is initialized with consistent nonparametric estimates of conditional choice probabilities, successive iterations return a sequence of estimators of the structural parameters that we call *K-stage policy iteration estimators*. We show that the sequence includes as extreme cases a Hotz-Miller CCP estimator (for $K = 1$) and Rust's nested fixed point estimator (in the limit when $K \to \infty$). Furthermore, the asymptotic distribution of all the estimators in the sequence is the same and equal to that of the maximum likelihood estimator. If consistent nonparametric estimates are not available and NPL is initialized with arbitrary choice probabilities, on convergence it still produces a root of the likelihood equations. We illustrated the performance of our method with several examples based on Rust's bus replacement model. We found that NPL produces maximum likelihood estimates 5 to 15 times faster than NFXP. In Monte Carlo experiments we found that the finite sample properties of K-stage PI estimators improve monotonically with K. Also, the benefits of using the 2-stage PI estimator instead of the l-stage estimator were very significant, but the additional gains in precision obtained by using the ML estimator instead of 2-stage PI were small. To what extent these encouraging results will generalize beyond our simple example is an open question. In particular, it remains to be seen whether CCP and PI

estimators will be useful once we relax the assumption that unobservable state variables have iid extreme value distributions.

Dept. of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, U.S.A.; vaguirre@bu.edu; http://people.bu.edu/vaguirre

and

CEMFI, Casado del Alisal, 5, 28014 Madrid, Spain; mira@cemfi.es; http://www.cemfi.es/~mira

Manuscript received March, 1999; final revision received July, 2001.

APPENDIX: PROOFS OF PROPOSITIONS

For the sake of simplicity, we omit x as an argument in functions $Q_x(.), Q_x^{-1}(.), S(., x), e_x(j, .)$, $v(j, x)$, and $\tilde{v}(j, x)$. In order to prove Propositions 1 and 2 we need the following Lemmas.

LEMMA 1: *The mappings $Q(.), Q^{-1}(.)$, and $e(j, .)$ are differentiable.*

PROOF: Differentiability of the three mappings follows from Assumptions 1 and 2 on the distribution of unobservables, from application of the inverse function and chain rule theorems, and from the strict convexity of the social surplus function.

LEMMA 2: *Let p^0 denote an arbitrary vector of $J-1$ choice probabilities. Then*

$$\frac{\partial}{\partial p}\left[\sum_{j=1}^{J} p_j^0 e(j, p^0)\right] = -Q^{-1}(p^0).$$

PROOF: Let v and $\tilde{v}$ be arbitrary vectors of J choice-specific utilities and $J-1$ utility differences, respectively, and let $W_j(\tilde{v})$ represent the expectation $E(\epsilon(j)|x, j)$ as a function of the $J-1$ utility differences. Since the $Q()$ mapping is invertible we can write $e(j, p^0) = W_j(Q^{-1}(p^0))$. Therefore, differentiating $\sum_{j=1}^{J} p_j^0(j, p^0)$, we obtain

$$\text{(Ap.1)} \qquad \frac{\partial}{\partial p}\left[\sum_{j=1}^{J} p_j^0 e(j, p^0)\right] = [-i; I]W(\tilde{v}) + \left[\frac{\partial Q(\tilde{v})'}{\partial \tilde{v}}\right]^{-1} \frac{\partial W(\tilde{v})'}{\partial \tilde{v}} \begin{pmatrix} 1 - i'p^0 \\ p^0 \end{pmatrix}$$

where i is a vector of ones.

Now, consider the following representation of the surplus function:

$$S(v) = \sum_{j=1}^{J} Q_j(\tilde{v})\{v(j) + W_j(\tilde{v})\}.$$

In matrix notation, we can write $S(v) = [1 - i'Q(\tilde{v}); Q(\tilde{v})'](v + W(\tilde{v}))$. Differentiating with respect to $\tilde{v}'$ and using the WDZ theorem, it is possible to show that the right-hand side of (Ap. 1) is zero.

LEMMA 3: *Let P^0 be an arbitrary set of conditional choice probabilities, and define the mapping*

$$H(P^0, V_\sigma) = \sum_{a \in A} P^0(a) * [u(a) + e(a, P^0) + \beta F(a) V_\sigma].$$

Let p_m^0 be the column vector containing the $(J-1)$ choice probabilities in P^0 associated with state $x = x_m$. Let $\tilde{v}_m$ be the vector of utility differences associated with $x = x_m$. Then: (a) for any arbitrary P^0, $H(P^0, V_\sigma)$ is a contraction in V_σ; (b) $\partial H_n / \partial p_m^0 = 0$, for $n \neq m$; (c) $\partial H_m / \partial_m^0 = -Q^{-1}(p_m^0) + \tilde{v}_m$; and (d) H_m is a concave function of p_m^0.

PROOF: (a) is straightforward from the definition of $H()$ with modulus β. Given that H_m does not depend on the probabilities for states different to m, (b) follows trivially. (c) results directly from Lemma 2. Note that the convexity of the social surplus implies that the Jacobian matrix $\partial Q(\tilde{v})/\partial \tilde{v}'$ is positive semidefinite; this and (c) imply (d).

PROOF OF PROPOSITION 1(a): *The set of optimal choice probabilities P defined in equation* (3) *is the unique fixed point of the operator Ψ.* We first note that the smoothed Bellman operator Γ_σ has a unique fixed point. Rust, Traub, and Wozniakowski (2002) show that the additivity of the social surplus makes the smooth Bellman operator "quasilinear" and this is used to prove that Γ_σ is a contraction. Second, we show that if P is a fixed point of Ψ, then $\varphi(P)$ is a fixed point of Γ_σ. Let P be a fixed point of Ψ, i.e., $P = \Lambda(\varphi(P))$. Then, by the definition of mapping Γ_σ in equation (2) we have that

$$\begin{aligned}\Gamma_\sigma(\varphi(P)) &= \sum_{a\in A} \Lambda(a;\varphi(P)) * [u(a) + e(a;\Lambda(\varphi(P)) + \beta F(a)\varphi(P))] \\ &= \sum_{a\in A} P(a) * [u(a) + e(a;P) + \beta F(a)\varphi(P)] = \varphi(P).\end{aligned}$$

Now suppose that P^1 and P^2 are different fixed points of Ψ. Since $\varphi(P^1)$ and $\varphi(P^2)$ are fixed points of Γ_σ, uniqueness of the fixed point of Γ_σ implies that $\varphi(P^1) = \varphi(P^2)$. But then, $P^1 = \Psi(P^1) = \Lambda(\varphi(P^1)) = \Lambda(\varphi(P^2)) = \Psi(P^2) = P^2$, a contradiction. Therefore, uniqueness of the fixed point of Ψ follows from uniqueness of the fixed point of Γ_σ.

Finally, note that $P \equiv \Lambda(V_\sigma)$ and $V_\sigma = \varphi(P)$. Therefore, $P = \Lambda(\varphi(P)) = \Psi(P)$; i.e., $\Psi()$ does have a fixed point, which is P.

PROOF OF PROPOSITION 1(b): *For any arbitrary P^0, the sequence $P^K = \Psi(P^{K-1})$, $K = 1, \ldots, \infty$, converges to the fixed point P.* Consider the Markov decision process defined as follows. (a) The state variable is the observable state vector of the original problem, x. (b) The decision variable is a vector of probabilities, p, that belongs to the interior of the J-dimensional simplex. (c) The current period return is $U(x,p) \equiv \sum_j p_j[u(x,j) + e(j;p)]$. (d) The transition probabilities $f_p(x'|x,p) \equiv \sum_j p_j f(x'|x,j)$. We show that in this transformed problem φ, Λ, and Ψ are indeed policy valuation, policy improvement, and policy iteration operators. Checking that φ is the valuation operator for the transformed problem is trivial. Now notice that the policy improvement step involves solving $\max_{p'} H(P', \varphi(P))$. By Lemma 3, it is straightforward to see that $\Lambda(\varphi(P))$ satisfies the first order conditions for an interior solution to this problem. Since the objective function is globally concave by Lemma 3(d), a solution does exist and it is $\Psi(P)$. Finally, notice that the transformed problem is a Markov decision process with discrete state space and continuous action space and that a maximizing decision rule always exists in the Bellman equation. Therefore, Theorem 6.4.6 in Puterman (1994) applies and policy iteration is guaranteed to converge.

PROOF OF PROPOSITION 1(c): *Equivalence of Ψ and Newton iterations.* The vector V_σ solves the functional equation (2), $V_\sigma = \Gamma_\sigma(V_\sigma)$. Therefore, it is a zero of the operator $T(V_\sigma) \equiv V_\sigma - \Gamma_\sigma(V_\sigma)$; i.e., $T(V_\sigma) = 0$. Newton iterations are defined by

$$V_\sigma^{K+1} = V_\sigma^K - [\partial T(V_\sigma^K)/\partial V_\sigma']^{-1} T(V_\sigma^K) = V_\sigma^K - [I - \partial \Gamma_\sigma(V_\sigma^K)/\partial V_\sigma']^{-1}[V_\sigma^K - \Gamma_\sigma(V_\sigma^K)].$$

Notice we can write the Bellman operator as $\Gamma_\sigma(V_\sigma) = H(\Lambda(V_\sigma), V_\sigma)$, and by Lemma 3, $\partial H(\Lambda(V_\sigma), V_\sigma)/\partial V_\sigma' = \beta F(V_\sigma)$, and $\partial H(\Lambda(V_\sigma), V_\sigma)/\partial P' = 0$. Therefore, Newton iterations take the form

$$V_\sigma^{K+1} = [I - \beta F(V_\sigma^K)]^{-1} \sum_j \Lambda(j, V_\sigma^K) * \{u(j) + e(j, \Lambda(V_\sigma^K))\} = \varphi(\Lambda(V_\sigma^K)).$$

So Newton's algorithm for this problem consists of iterative application of the same policy improvement and policy valuation operators Λ and φ that define Ψ.

PROOF OF PROPOSITION 2: *Zero Jacobian matrices for φ and Ψ at the fixed point.* By definition, for any arbitrary P^0

$$\varphi(P^0) = \sum_{j=1}^{J} P^0(j) * \{u(j) + e(j, P^0) + \beta F(j)\varphi(P^0)\}.$$

Therefore, $\varphi(P^0) = H(P^0, \varphi(P^0))$. Differentiating on both sides with respect to p_m and collecting terms, one can show that

$$\partial\varphi(P^0)/\partial p_m' = [I - \beta F^U(P^0)]^{-1}\partial H(P^0, \varphi(P^0))/\partial p_m'.$$

By Lemma 3(c) $\partial H_m/\partial p_m = -Q^{-1}(p_m^0) + \tilde{v}_m$. Let P be the fixed point of Ψ. Then, $\tilde{v}_m = Q^{-1}(\Psi_m(P))$. So at the fixed point $\tilde{v}_m = Q^{-1}(p_m)$ and $\partial H_m/\partial p_m' = 0$. Since $\partial H_n/\partial p_m' = 0$ for $n \neq m$, we get $\partial\varphi(P)/\partial p_m' = 0$. Therefore, $\partial\Psi(P)/\partial p_m' = (\partial\Lambda(\varphi(P))\partial V_\sigma')\partial\varphi(P)/\partial p_m' = 0$ at the fixed point.

PROOF OF PROPOSITION 3: Let $\hat{\alpha}$ be a root of the (partial) likelihood equations

$$\sum_{i=1}^{N} \partial \ln P_{(\hat{\alpha}, \hat{\theta}_f)}(a_i|x_i)/\partial\alpha = 0.$$

Since P_θ is the unique fixed point of the mapping Ψ_θ, we can apply the implicit function theorem and Proposition 2 to obtain:

$$\partial P_\theta/\partial\theta' = [I - \partial\Psi_\theta(P_\theta)/\partial P']^{-1}\partial\Psi_\theta(P_\theta)/\partial\theta' = \partial\Psi_\theta(P_\theta)/\partial\theta'. \tag{Ap.2}$$

Therefore we can write the likelihood equations as

$$\sum_{i=1}^{N} \partial \ln \Psi_{(\hat{\alpha}, \hat{\theta}_f)}(P_{(\hat{\alpha}, \hat{\theta}_f)})(a_i|x_i)/\partial\alpha = 0. \tag{Ap.3}$$

Now suppose the sequence $\{P^K\}$ from the NPL algorithm converges to a vector P_{NPL}. By the Theorem of the maximum, the mapping (11) that defines the pseudo-maximum likelihood estimator is continuous in P. Therefore, the sequence $\{\hat{\alpha}^K\}$ converges. Let $\hat{\alpha}_{NPL}$ be the limit. Then $\hat{\alpha}_{NPL}$ maximizes the pseudo log-likelihood function $\sum_{i=1}^{N} \ln \Psi_{(\alpha, \hat{\theta}_f)}(P_{NPL})(a_i|x_i)$, and at $\hat{\alpha}_{NPL}$ the following marginal conditions of optimality hold:

$$\sum_{i=1}^{N} \partial \ln \Psi_{(\hat{\alpha}_{NPL}, \hat{\theta}_f)}(P_{NPL})(a_i|x_i)/\partial\alpha = 0. \tag{Ap.4}$$

But convergence of the sequences on both sides of equation (12) implies that P_{NPL} is the fixed point of the policy iteration mapping for $\theta = (\hat{\alpha}_{NPL}', \hat{\theta}_f')'$: i.e., $P_{NPL} = P_{\{\hat{\alpha}_{NPL}, \hat{\theta}_f\}}$. It is clear that $\hat{\alpha}_{NPL}$ is a root of the likelihood equations.

The same result also holds in the context of full likelihood estimation. The parameter vector is now $\theta' = (\alpha', \theta_f')$. The term $\partial l_2(\theta)/\partial\theta = (0'; \partial l_2(\theta_f)/\partial\theta_f')'$ is added to the first order conditions. Since exactly the same term is added to both likelihood and pseudo-likelihood equations, intuitively it should not have any effect on equivalence. More, explicitly, the likelihood equations are

$$\sum_{i=1}^{N} \partial \ln \Psi_{\hat{\theta}}(P_{\hat{\theta}})(a_i|x_i)/\partial\theta + \partial l_2(\hat{\theta})/\partial\theta = 0,$$

which are again the first order conditions satisfied by NPL's limit pair (θ_{NPL}, P_{NPL}) in a full likelihood context.

PROOF OF PROPOSITION 4: Define $\gamma \equiv (\theta, P)$, and let $\Gamma \equiv \Theta_\alpha \times \Theta_f \times \Theta_P$ be the set of possible values of γ. Let n denote sample size, and consider the functions: $\tilde{Q}_n(\gamma) \equiv (1/n)\sum_{i=1}^{n} \ln \Psi_\theta(P)(a_i|x_i)$, $Q_n^K(\alpha) \equiv \tilde{Q}_n(\alpha, \hat{\theta}_{fn}, P_n^{K-1})$, and $\tilde{Q}_\infty(\gamma) \equiv E[\ln \Psi_\theta(P)(a_i|x_i)]$.

(A) *Consistency of* $\hat{\alpha}^K$: The proof of consistency proceeds in stages:

(A.1) If P_n^{K-1} and $\hat{\theta}_{fn}$ are consistent, then $Q_n^K(\alpha)$ converges a.s. and uniformly in α to a deterministic function $Q_\infty(\alpha)$.

(A.2) If P_n^{K-1} and $\hat{\theta}_{fn}$ are consistent, then $\hat{\alpha}_n^K \equiv \arg\max_{\alpha\in\Theta} Q_n^K(\alpha)$ converges a.s. to α^*.

(A.3) For $K \geq 1$, if $P_n^{K-1} \to_{a.s.} P^*$ and $\hat{\theta}_{fn} \to_{a.s.} \theta_f^*$, then $P_n^K \to_{a.s.} P^*$.

By condition (g), $\hat{\theta}_{fn} \to_{a.s.} \theta_f^*$ and $P_n^0 \to_{a.s.} P^*$; a simple induction argument based on (A.2) and (A.3) completes the proof.

(A.1) By Lemma 24.1 in Gourieroux and Monfort (1995, Vol. II, p. 392), we have that if: (i) $\widetilde{Q}_n(\gamma)$ converges a.s. and uniformly in γ to $\widetilde{Q}_\infty(\gamma)$; (ii) $\widetilde{Q}_\infty(\gamma)$ is uniformly continuous in γ; and (iii) $\{\hat{\theta}_{fn}\}$ and $\{P_n^{K-1}\}$ converge a.s. to θ_f^* and P^*, respectively; then $\widetilde{Q}_n(\alpha, \hat{\theta}_{fn}, P_n^{K-1})$ converges a.s. and uniformly in α to $\widetilde{Q}_\infty(\alpha, \theta_f^*, P^*) \equiv Q_\infty(\alpha)$.

By regularity conditions (a) and (b) $\widetilde{Q}_\infty(\gamma)$ is continuous on a compact set, so it is uniformly continuous, i.e., (ii) holds. Condition (iii) holds by assumption. In order to prove that condition (i) holds, note that $|\ln \Psi_\theta(P)(a|x)|$ is bounded. Furthermore, if $H_*(.,.)$ is the true probability distribution of (a, x) and $H_n(.,.)$ is the empirical distribution of (a, x) in a sample of size n, then $H_n(j, x) \to_{a.s.} H_*(j, x)$ by condition (d).

(A.2) By Property 24.2 in Gourieroux and Monfort (1995, Vol. II, p. 387), if: (i) $Q_n^K(\alpha)$ converges a.s. and uniformly in α to $Q_\infty(\alpha)$; and (ii) $Q_\infty(\alpha)$ has a unique maximum in Θ_α at α^*; then $\hat{\alpha}_n^K \equiv \arg\max_{\alpha\in\Theta} Q_n^K(\alpha)$ converges a.s. to α^*. We have proved (i) in (A.1). Condition (f) implies (ii) by the information inequality.

(A.3) By definition $P_n^K = \Psi_{(\hat{\alpha}_n^K, \hat{\theta}_{fn})}(P_n^{K-1})$. By (A.2) $\hat{\alpha}_n^K \to_{a.s.} \alpha^*$. Since Ψ is continuous in γ, by the Slutsky theorem $P_n^K \to_{a.s.} \Psi_{(\alpha^*, \theta_f^*)}(P^*) = P^*$.

(B) *Asymptotic distribution of* $\hat{\alpha}^K$: Let P_n be a consistent estimator of P^* such that

$$\left[\sqrt{n}\partial\widetilde{Q}_n(\gamma^*)/\partial\alpha'; \sqrt{n}(\hat{\theta}_{fn} - \theta_f^*)'; \sqrt{n}(P_n - P^*)'\right]' \to_d N(0, \Omega),$$

and define $\hat{\alpha}_n \equiv \arg\max_{\alpha\in\Theta} \widetilde{Q}_n(\alpha, \hat{\theta}_{fn}, P_n)$. Then:

(B.1) $\sqrt{n}(\hat{\alpha}_n - \alpha^*) \to_d N(0, V^*)$, and V^* only depends on the upper left $r \times r$ submatrix of Ω where r is the dimension of the parameter vector $(\alpha'; \theta_f')$;

(B.2) $\left[\sqrt{n}\partial\widetilde{Q}_n(\gamma^*)/\partial\alpha'; \sqrt{n}(\hat{\theta}_{fn} - \theta_f^*)'; \sqrt{n}(\Psi_{(\hat{\alpha}_n, \hat{\theta}_{fn})}(P_n) - P^*)'\right]' \to_d N(0, \Omega^*)$, and the upper left $r \times r$ submatrices of Ω and Ω^* are identical.

The proof is completed by induction using (B.1) and (B.2) and condition (g).

(B.1) Given conditions (b) and (f) and the definition of $\hat{\alpha}_n$, the first order conditions of optimality imply that with probability approaching one $\partial\widetilde{Q}_n(\hat{\gamma}_n)/\partial\alpha = 0$, where $\hat{\gamma}_n \equiv (\hat{\alpha}_n, \hat{\theta}_{fn}, P_n)$. By condition (b), $\widetilde{Q}_n(.)$ is twice continuously differentiable and we can apply the stochastic mean value theorem to $\partial\widetilde{Q}_n(.)/\partial\alpha_j$ between $\hat{\gamma}_n$ and γ^*. If $\{\bar{\gamma}_n^j\}$ are the p mean values, then $\bar{\gamma}_n^j \to_{a.s.} \gamma^*$ and $\partial^2\widetilde{Q}_n(\bar{\gamma}_n^j)/\partial\alpha_j\partial\gamma' \to_p \partial^2\widetilde{Q}_\infty(\gamma^*)/\partial\alpha_j\partial\gamma'$ (see Amemiya (1985, Theorems 4.2.1 and 4.1.5)). Notice that condition (f) implies that $\partial^2\widetilde{Q}_\infty(\gamma^*)/\partial\alpha\partial\alpha'$ is a nonsingular (negative definite) matrix. Finally, by condition (g) and the Mann-Wald Theorem, it is straightforward to show that $\sqrt{n}(\hat{\alpha}_n - \alpha^*) \to_d N(0, V^*)$, where

$$V^* = \left(\frac{\partial^2\widetilde{Q}_\infty(\gamma^*)}{\partial\alpha\partial\alpha'}\right)^{-1}\left(I; \frac{\partial^2\widetilde{Q}_\infty(\gamma^*)}{\partial\alpha\partial\theta_f'}; \frac{\partial^2\widetilde{Q}_\infty(\gamma^*)}{\partial\alpha\partial P'}\right)\Omega$$
$$\times\left(I; \frac{\partial^2\widetilde{Q}_\infty(\gamma^*)}{\partial\alpha\partial\theta_f'}; \frac{\partial^2\widetilde{Q}_\infty(\gamma^*)}{\partial\alpha\partial P'}\right)'\left(\frac{\partial^2\widetilde{Q}_\infty(\gamma^*)}{\partial\alpha\partial\alpha'}\right)^{-1}.$$

Notice that

$$\partial^2\widetilde{Q}_\infty(\gamma^*)/\partial\alpha\partial\gamma' = E(\partial\ln\Psi_{\theta^*}(P^*)(a_i|x_i)/\partial\alpha \times \partial\ln\Psi_{\theta^*}(P^*)(a_i|x_i)/\partial\gamma')$$

by equivalence of the information matrix. By Propositions 1 and 2, for any pair (a, x): (i) $\Psi_{\theta^*}(P^*)(a|x) = P_{\theta^*}(a|x)$; (ii) $\partial\Psi_{\theta^*}(P^*)(a|x)/\partial P = 0$; and (iii) $\partial\Psi_{\theta^*}(P^*)(a|x)/\partial\theta = \partial P_{\theta^*}(a|x)/\partial\theta$. Then, $\partial^2\widetilde{Q}_\infty(\gamma^*)/\partial\alpha\partial P' = 0$ and we get

$$V^* = \Omega_{00}^{-1} + \Omega_{00}^{-1}\{H_f\Omega_{f0} + \Omega_{0f}H_f' + H_f V(\hat{\theta}_f)H_f'\}\Omega_{00}^{-1}$$

where

$$\Omega_{00} = E\left(\frac{\partial \ln P_{\theta^*}(a_i|x_i)}{\partial\alpha}\frac{\partial \ln P_{\theta^*}(a_i|x_i)}{\partial\alpha'}\right); \quad H_f = E\left(\frac{\partial \ln P_{\theta^*}(a_i|x_i)}{\partial\alpha}\frac{\partial \ln P_{\theta^*}(a_i|x_i)}{\partial\theta_f'}\right)$$

$$\Omega_{f0} = \Omega_{0f}' = \left[E\left(\frac{\partial \ln f_{\theta_f^*}(x_i'|x_i,a_i)}{\partial\theta_f}\frac{\partial \ln f_{\theta_f^*}(x_i'|x_i,a_i)}{\partial\theta_f'}\right)\right]^{-1} E\left(\frac{\partial \ln f_{\theta_f^*}(x_i'|x_i,a_i)}{\partial\theta_f}\frac{\partial \ln P_{\theta^*}(a_i|x_i)}{\partial\alpha'}\right).$$

It is simple to verify that V^* is the variance of the asymptotic distribution of the partial maximum likelihood estimator of α^*.

(B.2) Define

$$\omega_n^{K+1} \equiv \left[\sqrt{n}\partial\widetilde{Q}_n(\gamma^*)/\partial\alpha'; \sqrt{n}(\hat{\theta}_{fn} - \theta_f^*)'; \sqrt{n}(P_n^K - P^*)'\right]'.$$

We have $P_n^K \equiv \Psi_{(\hat{\alpha}_n^K, \hat{\theta}_{fn})}(P_n^{K-1})$. By a stochastic mean value theorem we can write $P_n^K - P^*$ as a function of $[\hat{\alpha}_n^K - \alpha^*; \hat{\theta}_{fn} - \theta_f^*; P_n^{K-1} - P^*]$, and by the same expansion used in (B.1) we express $(\hat{\alpha}_n^K - \alpha^*)$ as a function of $[\partial\widetilde{Q}_n(\gamma^*)/\partial\alpha; \hat{\theta}_{fn} - \theta_f^*; P_n^{K-1} - P^*]$. Therefore, we can obtain $\omega_n^{K+1} = A_n^K\omega_n^K$, where A_n^K depends on the mean values. It is possible to show that $A_n^K \to_p A^* < \infty$. It follows that if ω_n^K is asymptotically normal, then ω_n^{K+1} is also asymptotically normal. Furthermore, the upper-left $r \times r$ submatrix of A_n^K is the identity matrix. Therefore, the upper-left $r \times r$ submatrices in the variances of ω_n^{K+1} and ω_n^K are equal.

REFERENCES

AGUIRREGABIRIA, V. (1999): "The Dynamics of Markups and Inventories in Retailing Firms," *The Review of Economic Studies*, 66, 275–308.

AHN, N. (1995): "Measuring the Value of Children by Sex and Age Using a Dynamic Programming Model," *The Review of Economic Studies*, 62, 361–379.

ALTUG, S., AND R. MILLER (1998): "The Effect of Work Experience on Female Wages and Labour Supply," *Review of Economic Studies*, 65, 45–85.

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

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

BERRY, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242–262.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

BIERENS, H. (1983): "Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions," *Journal of the American Statistical Association*, 78, 669–707.

DELGADO, M., AND J. MORA (1995a): "Nonparametric and Semiparametric Estimation with Discrete Regressors," *Econometrica*, 63, 1477–1484.

——— (1995b): "On Asymptotic Inferences in Nonparametric and Semiparametric Models with Discrete and Mixed Regressors," *Investigaciones Economicas*, 19, 435–468.

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

GILLESKIE, D. (1998): "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66, 1–45.

GOURIEROUX, C., AND A. MONFORT (1995): *Statistics and Econometric Models*. Cambridge: Cambridge University Press.

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

HOTZ, J., R. A. MILLER, S. SANDERS, AND J. SMITH (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

KEANE, M. P., AND K. I. WOLPIN (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *The Review of Economics and Statistics*, 76, 648–672.

——— (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

MANSKI, C. (1991): "Nonparametric Estimation of Expectations in the Analysis of Discrete Choice under Uncertainty," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen. Cambridge: Cambridge University Press, 259–275.

——— (1993): "Dynamic Choice in Social Settings: Learning from the Experience of Others," *Journal of Econometrics*, 58, 121–136.

MCFADDEN, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge, MA: MIT Press, 198–272.

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

——— (1997): "Estimating Models of Dynamic Optimization with Microeconomic Data," in *Handbook of Applied Econometrics: Microeconomics (Vol 2)*, ed. by H. Pesaran and P. Smidth. Oxford, U.K.: Blackwell, 246–299.

PAKES, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.

——— (1994): "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions," in *Advances in Econometrics, Proceedings of the Sixth World Congress of the Econometric Society*, ed. by C. Sims. Cambridge: Cambridge University Press, 171–259.

PUTERMAN, M. (1994): *Markov Decision Processes*. New York: Wiley.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (1988): "Maximum Likelihood Estimation of Discrete Control Processes," *SIAM Journal of Control and Optimization*, 26, 1006–1024.

——— (1994a): "Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes," in *Advances in Econometrics, Proceedings of the Sixth World Congress of the Econometric Society*, ed. by C. Sims. Cambridge: Cambridge University Press, 119–170.

——— (1994b): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics, Volume 4*, ed. by R. E. Engle and D. McFadden. Amsterdam: North-Holland, 3082–3139.

——— (1996): "Numerical Dynamical Programming in Economics," in *Handbook of Computational Economics*, ed. by H. Amman, D. Kendrick, and J. Rust. Amsterdam: Elsevier-North Holland, Chapter 14.

——— (1997): "A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-Random, and Deterministic Discretizations," Manuscript, Department of Economics, Yale University.

RUST, J., AND C. PHELAN (1997): "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, 65, 781–831.

RUST, J., J. TRAUB, AND H. WOZNIAKOWSKI (2002): "Is There a Curse of Dimensionality for Contraction Fixed Points in the Worst Case?" *Econometrica*, 70, 285–329.

SLADE, M. (1998): "Optimal Pricing with Costly Adjustment: Evidence from Retail-Grocery Prices," *Review of Economic Studies*, 65, 87–107.

WOLPIN, K. (1996): "Public-Policy Uses of Discrete-Choice Dynamic Programming Models," *American Economic Review*, 86, 427–432.