Econometrica Supplementary Material

SUPPLEMENT TO "CONDITIONAL CHOICE PROBABILITY ESTIMATION OF DYNAMIC DISCRETE CHOICE MODELS WITH UNOBSERVED HETEROGENEITY"

(*Econometrica*, Vol. 79, No. 6, November 2011, 1823–1867)

BY PETER ARCIDIACONO AND ROBERT A. MILLER

IN THIS SUPPLEMENT, we provide proofs of the lemmas and theorems as well as a more detailed description of the Monte Carlo exercises.

A. PROOFS

PROOF OF LEMMA 1: Equation (3.3) implies that $V_t(z_t)$ can be written as

$$
\begin{aligned}
V_t(z_t) &= \sum_{j=1}^{J} \int d_{jt}^{o}(z_t, \varepsilon_t) \\
&\quad \times \left[u_{jt}(z_t) + \varepsilon_{jt} + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1}|z_t) \right] g(\varepsilon_t)\, d\varepsilon_t \\
&= \sum_{j=1}^{J} p_{jt}(z_t) \left[u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1}|z_t) \right] \\
&\quad + \sum_{j=1}^{J} \int d_{jt}^{o}(z_t, \varepsilon_t) \varepsilon_{jt} g(\varepsilon_t)\, d\varepsilon_t \\
&= \sum_{j=1}^{J} p_{jt}(z_t) v_{jt}(z_t) + \sum_{j=1}^{J} \int d_{jt}^{o}(z_t, \varepsilon_t) \varepsilon_{kt} g(\varepsilon_t)\, d\varepsilon_t.
\end{aligned}
\tag{A.1}
$$

Subtracting $v_{kt}(z_t)$ from both sides yields

$$
\begin{aligned}
V_t(z_t) - v_{kt}(z_t) &= \sum_{j=1}^{J} p_{jt}(z_t) v_{jt}(z_t) \\
&\quad + \sum_{j=1}^{J} \int d_{jt}^{o}(z_t, \varepsilon_t) \varepsilon_{jt} g(\varepsilon_t)\, d\varepsilon_t - v_{kt}(z_t) \\
&= \sum_{j=1}^{J} p_{jt}(z_t) [v_{jt}(z_t) - v_{kt}(z_t)] \\
&\quad + \sum_{j=1}^{J} \int d_{jt}^{o}(z_t, \varepsilon_t) \varepsilon_{jt} g(\varepsilon_t)\, d\varepsilon_t.
\end{aligned}
\tag{A.2}
$$

 DOI: 10.3982/ECTA7743

From Proposition 1 of Hotz and Miller (1993, p. 501), there exists a mapping $\psi_k^{(1)}(p)$ for each $j \in \{1, \dots, J\}$ such that

$$\psi_j^{(1)}[p_t(z_t)] = v_{jt}(z_t) - v_{1t}(z_t), \tag{A.3}$$

which implies

$$v_{jt}(z_t) - v_{kt}(z_t) = \psi_j^{(1)}[p_t(z_t)] - \psi_k^{(1)}[p_t(z_t)]. \tag{A.4}$$

Hotz and Miller (1993) also proved that (A.3) implies the existence of a mapping $\psi_j^{(2)}(p)$ for each $j \in \{1, \dots, J\}$ such that

$$\psi_j^{(2)}[p_t(z_t)] = \int d_{jt}^o(z_t, \varepsilon_t)\varepsilon_{jt} g(\varepsilon_t)\, d\varepsilon_t. \tag{A.5}$$

Substituting (A.4) and (A.5) into (A.2) completes the proof:

$$\begin{aligned} \psi_k[p_t(z_t)] &\equiv \sum_{j=1}^{J} p_{jt}(z_t)\big\{\psi_j^{(1)}[p_t(z_t)] - \psi_k^{(1)}[p_t(z_t)]\big\} + \sum_{j=1}^{J} \psi_j^{(2)}[p_t(z_t)] \\ &= V_t(z_t) - v_{kt}(z_t). \end{aligned} \tag{A.6}$$

Q.E.D.

PROOF OF THEOREM 1: The proof is by backward induction. We first establish that it holds when the time horizon is T and where the decision is made at T' and when $T' + 1 = T$. We then show that if it holds for a generic T', where $1 < T' < T$, then it also holds at $T' - 1$, completing the proof. Noting that $v_{kT}(z_T) \equiv u_k(z_T)$ for all $k \in \{1, \dots, J\}$ and $z_T \in \{1, \dots, Z\}$, including those in the decision rule $d_{kT'}^*(z_{T'}, j)$, and noting that when $T' + 1 = T$, equation (3.6) can be expressed as

$$\begin{aligned} v_{jT'}(z_{T'}) &= u_{jT'}(z_{T'}) + \beta \sum_{z_T=1}^{Z} \sum_{k=1}^{J} \big[u_{kT}(z_T) + \psi_k[p_T(z_T)]\big] \\ &\quad \times d_{kT'}^*(z_{T,j}) f_{jT'}(z_T | z_{T'}), \end{aligned} \tag{A.7}$$

which establishes that the theorem holds for $t = T - 1$.

Setting T' such that $1 < T' < T$ and assuming (3.8) holds implies

$$\begin{aligned} v_{jT'}(z_{T'}) &= u_{jT'}(z_{T'}) + \sum_{\tau=T'+1}^{T} \sum_{k=1}^{J} \sum_{z_\tau=1}^{Z} \beta^{\tau-T'} \big[u_{k\tau}(z_\tau) + \psi_k[p_\tau(z_\tau)]\big] \\ &\quad \times d_{k\tau}^*(z_\tau, j) \kappa_{\tau-1}^*(z_\tau | z_{T'}, j). \end{aligned} \tag{A.8}$$

Moving back to $T'-1$, equation (3.6) implies

$$v_{jT'-1}(z_{T'-1}) = u_{jT'-1}(z_{T'-1}) + \sum_{z_{T'}=1}^{Z}\sum_{k=1}^{J}\big[v_{kT'}(z_{T'}) + \psi_k[p_{T'}(z_{T'})]\big] \times \kappa^*_{T'-1}(z_{T'}|z_{T'-1}, j). \tag{A.9}$$

Substituting for $v_{kT'}(z_{T'})$ in (A.9) with (A.8) completes the proof:

$$\begin{aligned} v_{jT'-1}(z_{T'-1}) &= u_{jT'-1}(z_{T'-1}) \\ &\quad + \sum_{\tau=T'}^{T}\sum_{k=1}^{J}\sum_{z_\tau=1}^{Z}\beta^{\tau-T'-1}\big[u_{k\tau}(z_\tau) + \psi_k[p_\tau(z_\tau)]\big] \\ &\quad \times d^*_{k\tau}(z_\tau, j)\kappa^*_{\tau-1}(z_\tau|z_{T'-1}, j). \end{aligned} \tag{A.10}$$

Now consider the infinite horizon problem. For $t < T'$, we can express

$$\begin{aligned} v_{jt}(z_t) = u_{jt}(z_t) &+ \sum_{\tau=t+1}^{T'}\sum_{k=1}^{J}\sum_{z_\tau=1}^{Z}\beta^{\tau-t}\big[u_{k\tau}(z_\tau) + \psi_k[p_\tau(z_\tau)]\big] \\ &\times d^*_{k\tau}(z_\tau, j)\kappa^*_{\tau-1}(z_\tau|z_t, j) \\ &+ \sum_{z_{T'+1}=1}^{Z} V_{T'+1}(z_{T'+1})\kappa^*_{T'}(z_{T'+1}|z_t, j). \end{aligned} \tag{A.11}$$

We can bound $|V_{T'+1}(z_{T'+1})|$ by $\overline{V}$, which implies

$$\begin{aligned} &\left|\sum_{k=1}^{J}\sum_{z_{T'+1}=1}^{Z} V_{T'+1}(z_{T'+1})\kappa^*_{T'}(z_{T'+1}|z_t, j)\right| \\ &\quad \le \sum_{k=1}^{J}\sum_{z_{T'+1}=1}^{Z} |V_{T'+1}(z_{T'+1})|\kappa^*_{T'}(z_{T'+1}|z_t, j) \\ &\quad \le \overline{V} \end{aligned}$$

since

$$\sum_{z_{T'+1}=1}^{Z}\kappa^*_{T'}(z_{T'+1}|z_t, j) = 1 \quad \text{and} \quad \kappa^*_{T'}(z_{T'+1}|z_t, j) \ge 0 \quad \text{for all } \{z_{T'+1}, z_t\}.$$

It now follows from (A.11) that for all T',

$$\left| v_{jt}(z_t) - u_{jt}(z_t) - \sum_{\tau=t+1}^{T'} \sum_{k=1}^{J} \sum_{z_\tau=1}^{Z} \beta^{\tau-t}\big[u_{k\tau}(z_\tau) + \psi_k[p_\tau(z_\tau)]\big] \times d^*_{k\tau}(z_\tau, j)\kappa^*_{\tau-1}(z_\tau|z_t, j) \right| \le \beta^{T'-t+1}\overline{V}.$$

Since $\beta < 1$, the term $\beta^{T'-t+1}\overline{V} \to 0$ as $T' \to \infty$, proving the theorem. *Q.E.D.*

PROOF OF LEMMA 2: Define $v_j \equiv \ln Y_j$ and let $G_j(\varepsilon) \equiv \partial G(\varepsilon)/\varepsilon_j$. Let $\overline{\mathcal{H}} \equiv \mathcal{H}(e^{v_1}, e^{v_2}, \dots, e^{v_J})$. Since $\mathcal{H}(Y_1, Y_2, \dots, Y_J)$ is homogeneous of degree 1 and, therefore, the partial derivative $\mathcal{H}_j(Y_1, Y_2, \dots, Y_J)$ is homogeneous of degree 0, we have

$$\begin{aligned} &G_j(v_j + \varepsilon_j - v_1, \dots, v_j + \varepsilon_j - v_J) \\ &\quad = \mathcal{H}_j(e^{v_1}, \dots, e^{v_J}) \exp[-\overline{\mathcal{H}} e^{-v_j - \varepsilon_j}] e^{-v-\varepsilon_j}. \end{aligned}$$

From Theorem 1 of McFadden (1978, p. 80), integrating over $G_j(\varepsilon_t)$ yields the conditional choice probability

$$\begin{aligned} p_j &= \int G_j(v_j + \varepsilon_j - v_1, \dots, \varepsilon_j, \dots, v_j + \varepsilon_j - v_J)\, d\varepsilon_j \\ &= e^{v_j - v_1}\mathcal{H}_j[1, e^{v_2 - v_1}, \dots, e^{v_J - v_1}]/\mathcal{H}[1, e^{v_2 - v_1}, \dots, e^{v_J - v_1}] \\ &\equiv \phi_j(1, e^{v_2 - v_1}, \dots, e^{v_J - v_1}). \end{aligned} \tag{A.12}$$

By Proposition 1 of Hotz and Miller (1993), we can invert the vector function

$$\begin{pmatrix} p_2 \\ \vdots \\ p_J \end{pmatrix} = \begin{pmatrix} \phi_2(1, e^{v_2 - v_1}, \dots, e^{v_J - v_1}) \\ \vdots \\ \phi_J(1, e^{v_2 - v_1}, \dots, e^{v_J - v_1}) \end{pmatrix}$$

to make the vector of differences $(v_j - v_J)$ the subject of the equation. This is given by $\psi_j^{(1)}(p)$ from equation (A.3), implying that $\phi^{-1}(p)$ is given by

$$\phi_j^{-1}(p) = \exp\big[\psi_j^{(1)}(p)\big]. \tag{A.13}$$

Noting that $\psi_1^{(1)}(p) = 0$, we define $\phi_1^{-1}(p) = 1$.

The expected contribution of the disturbance from the jth choice is

$$\begin{aligned}
&\int d_j \varepsilon_k \, dG(\varepsilon) \\
&\quad = \int \varepsilon_j G_j(v_j + \varepsilon_j - v_1, \dots, v_j + \varepsilon_j - v_J)\, d\varepsilon_j \\
&\quad = \mathcal{H}_j(e^{v_1}, \dots, e^{v_J}) \int \varepsilon_j \exp[-\overline{\mathcal{H}} e^{-v_j - \varepsilon_j}] e^{-\varepsilon_j}\, d\varepsilon_j \\
&\quad = e^{v_j - v_1} \mathcal{H}_j(1, e^{v_2 - v_1}, \dots, e^{v_J - v_1})(\gamma - (v_j - v_1) + \ln \overline{\mathcal{H}})/\overline{\mathcal{H}}.
\end{aligned}$$

Substituting the formula for p_j and evaluating $\overline{\mathcal{H}}$ at $\phi^{-1}(p)$ implies that $\psi_j^{(2)}(p)$ from (A.5) can now be expressed as

$$\psi_j^{(2)}(p) = p_j\big[\gamma - \ln \phi_j^{-1}(p) + \ln \mathcal{H}[1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p)]\big]. \tag{A.14}$$

Substituting (A.13) and (A.14) into the definition of $\psi_j(p)$ given in equation (A.6) yields

$$\begin{aligned}
\psi_j(p) = {}& \sum_{k=1}^{J} p_k \{\ln \phi_k^{-1}(p) - \ln \phi_j^{-1}(p)\} \\
&+ \sum_{k=1}^{J} p_k \big(\gamma - \ln \phi_k^{-1}(p) + \ln \mathcal{H}[1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p)]\big).
\end{aligned}$$

Simplifying the expression completes the proof:

$$\psi_j(p) = \ln \mathcal{H}[1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p)] - \ln \phi_j^{-1}(p) + \gamma. \tag{A.15}$$

Q.E.D.

PROOF OF LEMMA 3: From (3.18),

$$\mathcal{H}(Y_1, \dots, Y_J) = \mathcal{H}_0(Y_1, \dots, Y_K) + \left(\sum_{j \in \mathcal{J}} Y_j^{1/\sigma}\right)^{\sigma}. \tag{A.16}$$

For all $j \in \mathcal{J}$, the formula for $\phi_j(Y)$ for the nested logit components is:

$$\phi_j(Y) = Y_j^{1/\sigma} \frac{\left(\sum_{j \in \mathcal{J}} Y_j^{1/\sigma}\right)^{\sigma - 1}}{\mathcal{H}(Y_1, \dots, Y_J)}.$$

Let $\phi^{-1}(p) \equiv (\phi_2^{-1}(p), \dots, \phi_J^{-1}(p))$ denote the inverse of $\phi(Y) \equiv (\phi_1(Y), \dots, \phi_{J-1}(Y))$. Then from (A.12),

$$p_j \equiv \phi_j[\phi^{-1}(p)] = [\phi_j^{-1}(p)]^{1/\sigma} \frac{\left(\sum_{k \in \mathcal{J}} [\phi_k^{-1}(p)]^{1/\sigma}\right)^{\sigma-1}}{\mathcal{H}(1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p))}. \tag{A.17}$$

Summing over $k \in \mathcal{J}$ and taking the quotient yields

$$\frac{p_j}{\sum_{k \in \mathcal{J}} p_k} = \frac{[\phi_j^{-1}(p)]^{1/\sigma}}{\sum_{k \in \mathcal{J}} [\phi_k^{-1}(p)]^{1/\sigma}},$$

which implies by direct verification that

$$\phi_j^{-1}(p) = A p_j^{\sigma}, \tag{A.18}$$

where A is unknown but greater than zero.

Substituting for $\phi_j^{-1}(p)$ in (A.17) with (A.18), we obtain, for each choice $j \in \mathcal{J}$,

$$\begin{aligned} p_j &= A^{1/\sigma} p_j \frac{\left(\sum_{k \in \mathcal{J}} A^{1/\sigma} p_k\right)^{\sigma-1}}{\mathcal{H}(1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p))} \\ &= A p_j \frac{\left(\sum_{j \in \mathcal{J}} p_j\right)^{\sigma-1}}{\mathcal{H}(1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p))}, \end{aligned}$$

which implies

$$\mathcal{H}(1, \phi_2^{-1}(p), \dots, \phi_J^{-1}(p)) = A\left(\sum_{k \in \mathcal{J}} p_k\right)^{(\sigma-1)}. \tag{A.19}$$

We can now substitute (A.19) and (A.18) into the expression for $\psi_j(p)$ given in (A.15), completing the proof:

$$\begin{aligned} \psi_j(p) &= \ln\left[A\left(\sum_{j \in \mathcal{J}} p_j\right)^{(\sigma-1)}\right] - \ln[A p_j^{\sigma}] + \gamma \\ &= \gamma - \sigma \ln(p_j) - (1-\sigma) \ln\left(\sum_{k \in \mathcal{J}} p_k\right). \end{aligned}$$

Q.E.D.

PROOF OF THEOREM 2: (i) For convenience, we consolidate the structural parameters into the vector $\lambda \equiv (\theta, \pi)$. Denote the true parameters and conditional choice probabilities by λ_0 and p_0, respectively. Let $l(\lambda, p)$ denote the corresponding vector of likelihoods associated with each choice probability, implying $p_0 = l(\lambda_0, p_0)$. For each $\mathcal{N}$, define $\Lambda_{\mathcal{N}}$ as the set of parameters solving (4.5) at $p = \widehat{p}$, where $(\widehat{\theta}, \widehat{\pi}, \widehat{p})$ simultaneously satisfies (4.6):

$$\Lambda_{\mathcal{N}} \equiv \left\{ \lambda_{\mathcal{N}} : \lambda_{\mathcal{N}} = \arg\max_{\lambda} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \ln[L(d_n, x_n | x_{n1}; \lambda, p_{\mathcal{N}})] \right.$$
$$\left. \text{where } p_{\mathcal{N}} = l(\lambda_{\mathcal{N}}, p_{\mathcal{N}}) \right\}.$$

Also define the set of parameters that maximize the corresponding expected log likelihood subject to the same constraint as

$$\Lambda_1 \equiv \left\{ \lambda_1 : \lambda_1 = \arg\max_{\lambda} E\{\ln[L(d_n, x_n | x_{n1}; \lambda, p_1)]\} \right.$$
$$\left. \text{where } p_1 = l(\lambda_1, p_1) \right\}.$$

By definition, $\lambda_0 \in \Lambda_1$ because (λ_0, p_0) solves

$$\lambda_0 = \arg\max_{\lambda} E\{\ln[L(d_n, x_n | x_{n1}; \lambda, p_0)]\} \quad \text{where } p_0 = l(\lambda_0, p_0).$$

Appealing to the continuity of $L(d_n, x_n | x_{n1}; \lambda, p_{\mathcal{N}})$ and $p(\lambda_{\mathcal{N}})$, the weak uniform law of large numbers implies the existence of a sequence $\widehat{\lambda}_{\mathcal{N}} \in \Lambda_{\mathcal{N}}$ converging to λ_0. Now consider sequences $\widetilde{\lambda}_{\mathcal{N}} \in \Lambda_{\mathcal{N}}$ that converge to other elements in Λ_1, say $\lambda_1 \neq \lambda_0$. The assumption of identification implies that for all $\lambda_1 \neq \lambda_0$,

$$E\{\ln[L(d_n, x_n | x_{n1}; \lambda_0, p_0)]\} > E\{\ln[L(d_n, x_n | x_{n1}; \lambda_1, p_1)]\}.$$

By continuity and the law of large numbers,

$$\frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \ln[L(d_n, x_n | x_{n1}; \widehat{\lambda}_{\mathcal{N}}, \widehat{p}_{\mathcal{N}})]$$
$$> \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \ln[L(d_n, x_n | x_{n1}; \widetilde{\lambda}_{\mathcal{N}}, \widetilde{p}_{\mathcal{N}})] + o_p(1).$$

This proves that choosing the element which maximizes the criterion function, $\widehat{\lambda}_{\mathcal{N}}$, from the set of fixed points, $\Lambda_{\mathcal{N}}$, selects a weakly consistent estimator for λ_0.

(ii) For each t, define the joint distribution of (x, s), induced by the parameter vector (λ, p) and the data, as

$$P_{\mathcal{N}t}(x, s, \widehat{\lambda}, \widehat{p}) \equiv \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left[I(x_{nt} = x) \frac{\widehat{L}_{nt}(s_{nt} = s)}{\widehat{L}_n} \right].$$

By the law of large numbers, for each x, the $X \times S - 1$ dimensional random variable $P_{\mathcal{N}t}(x, s, \widehat{\lambda}, \widehat{p})$ converges in probability to

$$P_t(x, s, \widehat{\lambda}, \widehat{p}) \equiv E\left[I(x_{nt} = x) \frac{\widehat{L}_{nt}(s_{nt} = s)}{\widehat{L}_n} \right].$$

Similarly, the joint distribution of (j, x, s) is defined at t as

$$P_{\mathcal{N}t}(j, x, s, \widehat{\lambda}, \widehat{p}) \equiv \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left[d_{njt} I(x_{nt} = x) \frac{\widehat{L}_{nt}(s_{nt} = s)}{\widehat{L}_n} \right],$$

which converges in probability to

$$P_t(j, x, s, \widehat{\lambda}, \widehat{p}) \equiv E\left[I(d_{njt} = 1) I(x_{nt} = x) \frac{\widehat{L}_{nt}(s_{nt} = s)}{\widehat{L}_n} \right].$$

Let $P_{\mathcal{N}}(\lambda, p)$ denote the $T \times (J-1) \times X \times S$ dimensional vector formed from components $P_{\mathcal{N}t}(j, x, s, \lambda, p)/P_{\mathcal{N}t}(x, s, \lambda, p)$ and let $P(\lambda, p)$ denote the vector of corresponding limit points $P_t(j, x, s, \lambda, p)/P_t(x, s, \lambda, p)$. Then the parameters that solve the fixed point characterized by (4.5) and (4.8) are elements of the set defined by

$$\Lambda'_{\mathcal{N}} \equiv \left\{ \lambda_{\mathcal{N}} : \lambda_{\mathcal{N}} = \arg\max_{\lambda} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \ln[L(d_n, x_n | x_{n1}; \lambda, p_{\mathcal{N}})] \right.$$
$$\left. \text{where } p_{\mathcal{N}} = P_{\mathcal{N}}(\lambda_{\mathcal{N}}, p_{\mathcal{N}}) \right\}$$

and, similar to part (i), elements in $\Lambda'_{\mathcal{N}}$ converge weakly to elements in the set

$$\Lambda'_1 \equiv \left\{ \lambda_1 : \lambda_1 = \arg\max_{\lambda} E\{\ln[L(d_n, x_n | x_{n1}; \lambda, p_1)]\} \right.$$
$$\left. \text{where } p_1 = P(\lambda_1, p_1) \right\}.$$

Noting that $(\lambda_0, p_0) \in \Lambda'_1$, the arguments in part (i) can be repeated to complete the proof that the fixed point solution in $\Lambda'_{\mathcal{N}}$ that achieves the highest value of (4.4) is consistent. *Q.E.D.*

A.1. *Asymptotic Covariance Matrix*

The asymptotic covariance matrix of our estimators is derived from Taylor expansions of two sets of equations: the first order conditions of (4.5) for λ and a set of equations that solve the conditional choice probability nuisance parameter vector p. The first order conditions of (4.5) can be written as

$$\text{(A.20)} \qquad \frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}} L_{\lambda n}(\widehat{\lambda}, \widehat{p}) = 0,$$

where

$$L_{\lambda n}(\lambda, p) \equiv \frac{\partial[\ln L(d_n, x_n|x_{n1}; \lambda, p)]}{\partial\lambda}.$$

Since $(\widehat{\lambda}, \widehat{p})$ is consistent and $L_{\lambda n}(\lambda, p)$ is continuously differentiable, we can expand (A.20) around (λ_0, p_0) to obtain

$$\text{(A.21)} \qquad N^{-1/2}\sum_{n=1}^{\mathcal{N}} L_{\lambda n}(\lambda_0, p_0) - A_\lambda\sqrt{\mathcal{N}}(\widehat{\lambda} - \lambda_0) - A_p\sqrt{\mathcal{N}}(\widehat{p} - p_0) = o_p(1),$$

where

$$A_\lambda \equiv \lim_{\mathcal{N}\to\infty}\left[\frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}}\frac{\partial L_{\lambda n}(\lambda_0, p_0)}{\partial\lambda}\right],$$

$$A_p \equiv \lim_{\mathcal{N}\to\infty}\left[\frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}}\frac{\partial L_{\lambda n}(\lambda_0, p_0)}{\partial p}\right].$$

The first estimator sets $\widehat{p}$ to solve (4.6) for each (j, t, x, s). Stacking $l_{jt}(x, s; \lambda, p)$ for each choice (j, t) (time indexed in the nonstationary case) and each value (x, s) of state variables to form $l(\lambda, p)$, the $(J-1)\times T\times X\times S$ vector function of the CCP parameters (λ, p), our estimator satisfies the $(J-1)T\times S$ additional parametric restrictions $l(\widehat{\lambda}, \widehat{p}) = \widehat{p}$. From the identity

$$0 = l(\widehat{\lambda}, \widehat{p}) - \widehat{p} = l(\lambda_0, p_0) - p_0,$$

we expand the second equation to the first order and rearrange to obtain

$$\text{(A.22)} \qquad (I - l_p)\sqrt{\mathcal{N}}(\widehat{p} - p_0) - l_\lambda\sqrt{\mathcal{N}}(\widehat{\lambda} - \lambda_0) = o_p(1),$$

where

$$l_\lambda \equiv \frac{\partial l(\lambda_0, p_0)}{\partial\lambda}, \qquad l_p \equiv \frac{\partial l(\lambda_0, p_0)}{\partial p}.$$

Using (A.22), we substitute out $\sqrt{\mathcal{N}}(\widehat{p}-p_0)$ in (A.21), which yields

$$\sqrt{\mathcal{N}}(\widehat{\lambda}-\lambda_0)=(B_1'B_1)^{-1}B_1'N^{-1/2}\sum_{n=1}^{\mathcal{N}}L_{\lambda n}(\lambda_0,p_0)+o_p(1),$$

where

$$B_1\equiv A_\lambda+A_p(I-l_p)^{-1}l_\lambda.$$

Appealing to the central limit theorem and using the fact that

$$A_\lambda\equiv E[L_{\lambda n}(\lambda_0,p_0)L_{\lambda n}(\lambda_0,p_0)'],$$

the asymptotic covariance matrix for $\sqrt{\mathcal{N}}(\widehat{\lambda}-\lambda_0)$ is thus

$$(B_1'B_1)^{-1}B_1'A_\lambda B_1(B_1B_1')^{-1}.$$

In the second estimator, the condition that $l(\widehat{\lambda},\widehat{p})=\widehat{p}$ is replaced by the $(J-1)T\times S$ equalities in (4.8). Define

$$Q_{njtxs}(\lambda,p)\equiv[p_{jt}(x,s)-d_{njt}]I(x=x_{nt})\frac{L_n(s_{nt}=s)}{L_n},$$

where $L_n\equiv L(d_n,x_n|x_{n1};\lambda,p)$, and $L_n(s_{nt}=s)$ is given by (4.7) evaluated at (λ,p). For each sample observation n, stack $Q_{njtxs}(\lambda,p)$ to form the $(J-1)T\times S$ dimensional vector $Q_n(\lambda,p)$. In vector form, (4.8) can then be expressed as

$$\frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}}Q_n(\widehat{\lambda},\widehat{p})=0.$$

We form the vector $h_n(\lambda,p)$, the expected outer product of $h_n(\lambda,p)$, and its square derivative matrix:

$$h_n(\lambda,p)\equiv\begin{bmatrix}L_{\lambda n}(\lambda,p)\\ Q_n(\lambda,p)\end{bmatrix},\qquad \Omega=E[h_n(\lambda_0,p_0)h_n(\lambda_0,p_0)'],$$

$$\Gamma=E\left[\frac{\partial h_n(\lambda_0,p_0)}{\partial\lambda}\frac{\partial h_n(\lambda_0,p_0)}{\partial p}\right].$$

From Hansen (1982, Theorem 3.1) or Newey and McFadden (1994, Theorem 6.1), it now follows that $\sqrt{\mathcal{N}}(\widehat{\lambda}-\lambda_0)$ is asymptotically normally distributed with mean zero and covariance matrix given by the top left square block of $\Gamma^{-1}\Omega\Gamma^{-1'}$ with dimension λ.

B. ADDITIONAL INFORMATION ON THE MONTE CARLO EXERCISES

All simulations were conducted in Matlab version 7.5 on the Duke Economics Department 64-bit batch cluster. The code was not parallelized. The cluster and the operating system of Matlab ensure one processor is dedicated to each Matlab job. All nonlinear optimization was done using Matlab's canned optimization routine fminunc, with the default values used to determine convergence. No derivatives were used in the maximization routines for the structural parameters. Convergence for the EM algorithm was determined by comparing log likelihood values 25 iterations apart. The algorithm was stopped when this difference was less than 10^{-7} for two successive iterations.

B.1. *Optimal Stopping*

This subsection provides further computational details about the optimal stopping problem. We discuss the data generating process as well as updating the conditional choice probabilities and the parameters governing the initial conditions.

B.1.1. *Data Creation*

For the true parameter values and the transition matrix for mileages implied by equation (7.2) and reported in the first column of Table I, we obtain the value functions by backward recursion for every possible mileage, observed permanent characteristic, unobserved state, and time. We draw permanent observed and unobserved characteristics from discrete uniform distributions with support 101 and 2, respectively, and start each bus at zero mileage. Given the parameters of the utility function, the value function, and the permanent observed and unobserved states, we calculate the probability of a replacement occurring in the first period. We then draw from a standard uniform distribution. If the draw is less than the probability of replacement, the decision in the first period is to replace. Otherwise we keep the engine. Conditional on the replacement decision, we draw a mileage transition using equation (7.2). Continuing this way, decisions and mileage transitions are simulated for 30 periods.

B.1.2. *The Likelihood*

Conditional on the permanent observed state, the mileage, and the unobserved state s, the likelihood of a particular decision at time t takes a logit form. The likelihoods for the FIML and CCP cases are, respectively, given by

$$\begin{aligned}\mathcal{L}_t(d_t|x_t, s; \theta) = \Bigg(d_{1t} + d_{2t}\exp\bigg[u_2(x_t, s, \theta)\\ + \beta\sum_{x_{t+1}} V(x_{t+1}, s, \theta)(f_2(x_{t+1}|x_t) - f_1(x_{t+1}|x_t))\bigg]\Bigg)\end{aligned}$$

$$\Big/\left(\exp\left[u_2(x_t, s, \theta) + \beta \sum_{x_{t+1}} V(x_{t+1}, s, \theta) \times (f_2(x_{t+1}|x_t) - f_1(x_{t+1}|x_t))\right] + 1\right),$$

$$\mathcal{L}_t(d_t|x_t, s, p; \theta) = \left(d_{1t} + d_{2t} \exp\left[u_2(x_t, s, \theta) - \beta \sum_{x_{t+1}} \ln[p_{1t+1}(x_{t+1}, s)] \times (f_2(x_{t+1}|x_t) - f_1(x_{t+1}|x_t))\right]\right) \Big/\left(\exp\left[u_2(x_t, s, \theta) - \beta \sum_{x_{t+1}} \ln[p_{1t+1}(x_{t+1}, s)] \times (f_2(x_{t+1}|x_t) - f_1(x_{t+1}|x_t))\right] + 1\right).$$

When s is unobserved, the log likelihood for a particular bus history is found by first taking the product of the likelihoods over time, conditional on type, and then summing across the types inside the logarithm. Thus in the FIML case, the likelihood is[1]

$$L(d_n|x_n; \theta, \pi) = \sum_{s=1}^{2} \prod_{t=11}^{30} \pi(s|x_1)\mathcal{L}_t(d_t|x_t, s; \theta).$$

In the CCP case, $\mathcal{L}_t(d_t|x_t, s; \theta)$ is replaced by $\mathcal{L}_t(d_t|x_t, s, p; \theta)$.

B.1.3. *Conditional Choice Probability Estimates*

We approximate equation (5.9), the second estimator for the conditional choice probabilities, with a flexible logit, where the dependent variable is d_{1t}. There are five cases:

Case 1. To obtain the estimates reported in column 4 of Table I (when s is ignored), we estimate the CCP's using $W_{1t} \equiv (1, x_{1t}, x_{1t}^2, x_2, x_2^2, x_{1t}x_2)$ as regressors in a logit.[2]

[1]Since we are taking products of potentially small probabilities, numerical issues can arise. These can be solved by scaling up the $\mathcal{L}_t(d_t|x_t, s; \theta, p)$ terms by a constant factor. However, in neither of our Monte Carlos was this an issue.

[2]For a sufficiently large but finite sample, we can saturate the finite set of regressors with a flexible logit that yields numerically identical estimates as the weighted bin estimator presented in the text.

Case 2. For the parameters reported in column 3, W_{1t} is fully interacted with $W_{2t} \equiv (1, s, t, st, t^2, st^2)$, which is 36 parameters to estimate in the logit generating the CCP's. Since s is observed, this flexible logit is estimated once.

Case 3. When s is unobserved, the flexible logit described in the previous case is estimated at each iteration of the EM algorithm; at the mth iteration, the conditional probabilities of being in each observed state, $q_s^{(m)}$, are used to weight the flexible logit.

Cases 4 and 5. For the last two columns, where there are aggregate effects, we fully interact the first set of variables with the s, but not t and t^2. Instead, we include time dummies, but given the moderate sample size, we did not interact them with the other variables. Hence the logits were estimated with 21 regressors: 12 combinations of x_{1t}, x_2, and s, as well as 9 time dummies.

B.1.4. *Initial Conditions*

Initial probabilities are specified as a flexible function of the first-period observables, denoted by W_0. Included in W_0 are the mileage at the first observed time period for the bus, x_{11}, and the permanent observed characteristic, x_2. The prior probability of being in unobserved state 2 during the first observed period in the data, $t = 1$, given the data for n is

$$\pi(2|x_1) = \frac{\exp(W_0\delta)}{1 + \exp(W_0\delta)}.$$

At iteration m, we calculate the likelihood for each data point conditional on the unobserved state. Under FIML,

$$L\big(d_n, s_n = s|x_n; \theta^{(m)}\big) = \prod_{t=11}^{30} \pi(s|x_1)\mathcal{L}_t\big(d_t|x_t, s; \theta^{(m)}\big).$$

The iterate $\delta^{(m+1)}$ solves

$$\delta^{(m+1)} = \arg\max_{\delta} \sum_{n=1}^{\mathcal{N}} \ln\left(\sum_{s=1}^{2} \pi(s|x_{n1})L\big(d_n, s_n = s|x_n; \theta^{(m)}\big)\right).$$

In the CCP case, we replace $\mathcal{L}_t(d_t|x_t, s; \theta^{(m)})$ with $\mathcal{L}_t(d_t|x_t, s, p^{(m)}; \theta^{(m)})$ and replace $L(d_n, s_n = s|x_n; \theta^{(m)})$ with $L(d_n, s_n = s|x_n, p^{(m)}; \theta^{(m)})$.[3]

B.1.5. *Creation of Time-Varying Intercepts*

In the case where the replacement costs varied over time (column 8 of Table I), we created the data by drawing values for the intercept from a normal

[3]The saturation argument we mentioned in the previous footnote applies here too.

distribution with standard deviation of 0.5. The value of θ_{0t+1}, is set to $0.7\theta_{0t}$ plus the value drawn at $t+1$ from the normal distribution.

B.2. *Entry/Exit*

We now turn to the details of the Monte Carlo experiment for the dynamic game. Again we describe the data creation as well as the variables used in both the conditional choice probabilities and the reduced form controls for the initial conditions problem.

B.2.1. *Data Creation*

The first step in creating the data is to obtain the probability of entering for every state. Equation (7.8) gives the flow payoff for being in the market conditional on the choices of the other firms. Note that the expected flow payoff of entering depends on the probabilities of other firms entering. Given initial guesses for the probability of exiting in each state, we form all the possible combinations of the entry decisions of the other firms using equation (7.4). We then substitute equations (7.4) and (7.8) into equation (7.5) to form the expected flow payoff of staying in or entering the market in every state. Since the transitions on the state variables conditional on the entry/exit decisions are known, we have all the pieces to form equation (7.9). Given equation (7.9), the Type 1 extreme value assumption implies that the probability of exiting is $1/(1+\exp(v_2^{(i)}(x_t, s_t))$. We can then update the entry/exit probabilities used to form equation (7.4). We then iterate on equations (7.4), (7.5), (7.9), and the logit probability of exiting until a fixed point is reached.[4]

The observed permanent market characteristics and the initial unobserved states were drawn from a discrete uniform distribution. We then began each market with no incumbents and simulated the model forward. We then removed the first 10 periods of data from the sample.

B.2.2. *The Likelihood*

We now derive the likelihood at time t for market n of the observed decisions and price process given the data and the parameters. Note that x_{nt+1}, which includes the permanent market characteristic as well as the incumbency status of each of the firms, is a deterministic function of x_{nt} and y_{nt}. The likelihood contribution for the ith firm at time t conditional on unobserved state s_t is

$$l^{(i)}\big(d_t^{(i)}, x_t, s_t; \theta, \pi\big) = \frac{d_{1t} + d_{2t}\exp[v_2^{(i)}(x_t, s_t, p, \theta, \pi)]}{1+\exp[v_2^{(i)}(x_t, s_t, p, \theta, \pi)]}.$$

[4]Multiple equilibria may be a possibility. This issue did not cause any problems for this set of Monte Carlo data.

Denote $E(y_t) = \alpha_0 + \alpha_1 x_1 + \alpha_2 s_t + \alpha_3 \sum_{i=1}^{I} d_{2t}^{(i)}$. Denoting n as the market, the likelihood of the data in market n at time t conditional on s_t is

$$\text{(B.1)} \quad \mathcal{L}_t(d_{nt}, y_{nt}|x_{nt}, s_t; \theta, \alpha, \pi, p) = \phi\left(\frac{y_{nt} - E(y_{nt})}{\sigma}\right)\prod_{i=1}^{I} l_n^{(i)}\big(d_{nt}^{(i)}, x_{nt}, s_t; \theta, \pi\big),$$

where $\phi(\cdot)$ is the density function of the standard normal distribution and σ is the standard deviation of η_t.

We can then substitute equation (B.1) into equation (4.3) to obtain the likelihood of the data for a particular market:

$$\text{(B.2)} \quad L(d_n, y_n, x_n|x_{n1}; \theta, \alpha, \pi, p) = \sum_{s_1=1}^{S}\sum_{s_2=1}^{S}\cdots\sum_{s_{\mathcal{T}}=1}^{S}\Bigg[\pi(s_1|x_{n1})\mathcal{L}_1(d_{n1}, y_{n1}|x_{n1}, s_1; \theta, \alpha, \pi, p) \times \left(\prod_{t=2}^{\mathcal{T}} \pi(s_t|s_{t-1})\mathcal{L}_t(d_{nt}, y_{nt}|x_{nt}, s_t; \theta, \alpha, \pi, p)\right)\Bigg].$$

To clarify the number of calculations required to form the expression in equation (B.2) for a particular market, we specify equation (B.2) using matrix notation. Denote A_{nt} as a $1 \times S$ vector with components given by the likelihood at $t = 1$ conditional on a particular unobserved states times the initial probability of being in the unobserved state:

$$\text{(B.3)} \quad \mathcal{A}_{n1} = [\pi(1|x_{n1})\mathcal{L}_1(d_{n1}, y_{n1}|x_{n1}, 1; \theta, \alpha, \pi, p) \quad \cdots \quad \pi(S|x_{n1})\mathcal{L}_1(d_{n1}, y_{n1}|x_{n1}, S; \theta, \alpha, \pi, p)].$$

If $\mathcal{T} = 1$, summing over the elements of A_{n1} gives $L(d_n, y_n, x_n|x_{n1}; \theta, \alpha, \pi, p)$. For $t > 1$, we form an $S \times S$ matrix where the (i, j) element gives the probability of moving from $s_{t-1} = i$ to $s_t = j$ times the likelihood contribution at t conditional on being in unobserved state j:

$$\text{(B.4)} \quad \mathcal{A}_{nt} = \begin{bmatrix} \pi(1|1)\mathcal{L}_t(d_{nt}, y_{nt}|x_{nt}, 1; \theta, \alpha, \pi, p) & \cdots & \pi(1|S)\mathcal{L}_t(d_{nt}, y_{nt}|x_{nt}, S; \theta, \alpha, \pi, p) \\ \vdots & \ddots & \vdots \\ \pi(S|1)\mathcal{L}_t(d_{nt}, y_{nt}|x_{nt}, 1; \theta, \alpha, \pi, p) & \cdots & \pi(S|S)\mathcal{L}_t(d_{nt}, y_{nt}|x_{nt}, S; \theta, \alpha, \pi, p) \end{bmatrix}.$$

Taking A_{n1} times A_{n2} gives a $1 \times S$ vector of the joint likelihood of the data and being in each of the unobserved states. We define A_n as the product of A_{nt}

over $\mathcal{T}$:

$$A_n = \prod_{t=1}^{\mathcal{T}} A_{nt}.$$

Then A_n is a row vector with S elements, with each element giving the joint likelihood of the data and being in a particular unobserved state at $\mathcal{T}$. To form A_n, an $S \times S$ matrix is multiplied by a $1 \times S$ matrix $\mathcal{T}$ times. Let the sth element be denoted by $A_n(s)$. The likelihood for the nth market is then given by

$$L(d_n, y_n, x_n | x_{n1}; \theta, \alpha, \pi, p) = \sum_{s=1}^{S} A_n(s).$$

B.2.3. *Obtaining Conditional Choice Probabilities*

Four sets of CCP's are used in this Monte Carlo:

(i) When s_t is ignored (column 3 of Table II), we specify the conditional probability of exiting at $t+1$ as a flexible function of the observed variables, W_{1ti} (for the ith firm in a given market at time t). The variables included in W_{1ti} are combinations of the permanent market characteristic, x_1, whether the firm is active in period t, $d_{2t}^{(i)}$, and the number of firms in the market at t:

$$W_{1ti} \equiv \left(1, x_1, x_1^2, d_{2t}^{(i)}, \sum_{i'} d_{2t}^{(i)}, \left(\sum_{i'} d_{2t}^{(i)}\right)^2\right).$$

We then estimate a logit on the probability of exiting using the variables in W_{1ti} as controls.

(ii) When s_t is observed (column 2), we add the variables in W_{2ti} to the logit, where

$$W_{2ti} \equiv \left(s_t, s_t x_1, s_t d_{2t}^{(i)}, s_t \sum_{i'} d_{2t}^{(i)}\right),$$

implying that 10 parameters govern the CCP's.

(iii) When the conditional choice probabilities are updated with the data (column 5) and when we use the two-stage method (column 6), we use the variables in both W_{1t} and W_{2t}. In both these cases, the mth iteration uses the conditional probabilities of being in each unobserved state, $q_{nst}^{(m)}$, as weights in the logit estimation.

(iv) Finally, when the CCP's are updated with the model (columns 4 and 7), we update the probability of exiting using the logit formula for the likelihood:

$$p_{1t}^{(i,m+1)}(x, s) = \frac{1}{1 + \exp[v_{2t}^{(i)}(x, s, p^{(m)}, \theta^{(m)}, \pi^{(m+1)})]}.$$

B.2.4. *Initial Conditions*

There is an initial conditions problem in the stationary equilibrium, because the distribution of s_1 depends on the the distribution of the observed states. We estimate this distribution jointly with the other parameters of the model. Since the unobserved state applies at the market level of aggregation, the relevant endogenous variable is the lagged number of firms in the initial period. We regress the lagged number of firms in the initial period on a flexible function of the characteristics of the market—in this exercise, a constant, x_1—and x_1^2. Denote the residual from this regression as ζ. We then approximate the initial probability of being in unobserved state s for the nth market using a multinomial logit form:

$$\pi(s|x_{n1}) = \frac{\exp([1 \quad \zeta_n]\delta_s)}{\sum_s \exp([1 \quad \zeta_n]\delta_s)}.$$

With δ_1 set to zero, there are eight parameters to be estimated. We estimate $\pi(s|x_{n1})$ at each iteration using a similar procedure to Section B.1.4, now allowing for the fact that the unobserved states follow a Markov transition. Despite this additional complication, the algorithm is the same: calculate the likelihood given each initial unobserved state and take it as a given when maximizing to update δ.

Dept. of Economics, Duke University, 201A Social Science, Durham, NC 27708-0097, U.S.A.; psarcidi@econ.duke.edu

and

Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, U.S.A.

Manuscript received February, 2008; final revision received April, 2011.