Econometrica, Vol. 80, No. 5 (September, 2012), 2303–2319

SEQUENTIAL ESTIMATION OF STRUCTURAL MODELS WITH A FIXED POINT CONSTRAINT

BY HIROYUKI KASAHARA AND KATSUMI SHIMOTSU[1]

This paper considers the estimation problem of structural models for which empirical restrictions are characterized by a fixed point constraint, such as structural dynamic discrete choice models or models of dynamic games. We analyze a local condition under which the nested pseudo likelihood (NPL) algorithm converges to a consistent estimator, and derive its convergence rate. We find that the NPL algorithm may not necessarily converge to a consistent estimator when the fixed point mapping does not have a local contraction property. To address the issue of divergence, we propose alternative sequential estimation procedures that can converge to a consistent estimator even when the NPL algorithm does not.

KEYWORDS: Contraction, dynamic games, nested pseudo likelihood, recursive projection method.

1. INTRODUCTION

EMPIRICAL IMPLICATIONS OF ECONOMIC THEORY are often characterized by fixed point problems. Upon estimating such models, researchers typically consider a class of extremum estimators with a fixed point constraint $P = \Psi(\theta, P)$. For example, if $P = \{P(a|x)\}$ are the conditional choice probabilities, and the sample data are $\{a_m, x_m\}_{m=1}^M$, then maximizing $M^{-1}\sum_{m=1}^M \ln P(a_m|x_m)$ subject to $P = \Psi(\theta, P)$ gives the maximum likelihood estimator (MLE, hereafter).[2]

The fixed point constraint $P = \Psi(\theta, P)$ summarizes the set of structural restrictions of the model that is parametrized by a finite-dimensional vector $\theta \in \Theta$.[3] In principle, we may compute the MLE by the nested fixed point algorithm (Rust (1987)), which repeatedly solves all the fixed points of $P = \Psi(\theta, P)$ at each candidate parameter value. The major obstacle of applying such an estimation procedure lies in the computational burden of solving the fixed point problem for a given parameter.

[1]We are grateful to the co-editor and three anonymous referees whose comments greatly improved the paper. The authors thank Victor Aguirregabiria, David Byrne, Hide Ichimura, Kenneth Judd, Vadim Marmer, Lealand Morin, and seminar participants at the Bank of Japan, FEMES, New York Camp Econometrics, NASM, SITE, Vienna Macroeconomic Workshop, Boston University, Michigan, Montreal, Hitotsubashi, HKU, HKUST, Johns Hopkins, SETA, Tokyo, UBC, UWO, Yale, Yokohama National University, and Xiamen for helpful comments. The authors thank the SSHRC for financial support.

[2]For simplicity, we assume that the distribution function of x_m is known. In many structural models, the distribution function of x_m can be estimated using only the data of $\{x_m\}_{m=1}^M$.

[3]Examples of the operator $\Psi(\theta, P)$ include, among others, the policy iteration operator for a single-agent dynamic programming model (e.g., Rust (1987), Hotz and Miller (1993), Aguirregabiria and Mira (2002), Kasahara and Shimotsu (2008)), the best response mapping of a game (e.g., Aguirregabiria and Mira (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2008)), and the fixed point operator for a recursive competitive equilibrium (e.g., Aiyagari (1994); Krusell and Smith (1998)).

 DOI: 10.3982/ECTA8291

To reduce the computational cost, Hotz and Miller (1993) developed a representation of the value function in terms of choice probabilities, and proposed a two-step estimator that does not require solving the fixed point problem at each trial parameter value. Building on the idea of Hotz and Miller (1993), a number of recent papers in empirical industrial organization develop two-step estimators for models with multiple agents (e.g., Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2008), Bajari, Chernozhukov, Hong, and Nekipelov (2009)). These two-step estimators may suffer from substantial finite sample bias, however, when the choice probabilities are poorly estimated in the first step.[4]

To address the limitations of two-step estimators, Aguirregabiria and Mira (2002, 2007; henceforth AM07) developed a recursive extension of the two-step method of Hotz and Miller (1993), called the *nested pseudo likelihood (NPL) algorithm*. With $P = \{P(a|x)\}$ denoting the vector of conditional choice probabilities, the NPL algorithm starts from an initial estimate $\tilde{P}_0$ and iterates the following steps until $j = k$:

Step 1. Given $\tilde{P}_{j-1}$, update θ by $\tilde{\theta}_j = \arg\max_{\theta\in\Theta} M^{-1} \sum_{m=1}^{M} \ln[\Psi(\theta, \tilde{P}_{j-1})](a_m| x_m)$.

Step 2. Update $\tilde{P}_{j-1}$ using the obtained estimate $\tilde{\theta}_j$: $\tilde{P}_j = \Psi(\tilde{\theta}_j, \tilde{P}_{j-1})$.

The estimator $\tilde{\theta}_1$ is a version of Hotz and Miller's two-step estimator, called the *pseudo maximum likelihood (PML) estimator*. As AM07 showed, it is often the case that evaluating the mapping $\Psi(\theta, P)$ for a fixed value of P across different values of θ is computationally inexpensive and implementing Step 1 of the NPL algorithm is easy. This recursive method can be applied to models with unobserved heterogeneity, and the limit of the sequence of estimators is more efficient than the two-step estimators *if it converges to a consistent fixed point*.[5]

While the NPL algorithm provides an attractive apparatus for empirical researchers, its convergence is a concern, as recognized by AM07 (p. 19). Indeed, little is known about its convergence properties except that, in some examples, the NPL algorithm converges to a point distance away from the true value or fails to converge even after a large number of iterations, as shown in Pesendorfer and Schmidt-Dengler (2010; henceforth PS10) and Su and Judd (2012). In view of this mixed evidence and its practical importance, it is imperative that we understand the convergence properties of the NPL algorithm.

Su and Judd (2012) and Su (2012) proposed a constrained optimization approach called the Mathematical Program with Equilibrium Constraints

[4]See, for example, simulation results in Aguirregabiria and Mira (2007) and Pakes, Ostrovsky, and Berry (2007).

[5]Two-step estimators can be applied to models with unobserved heterogeneity when an initial consistent estimator of the type-specific conditional choice probabilities is available. Kasahara and Shimotsu (2009) derived sufficient conditions for nonparametric identification of a finite mixture model of dynamic discrete choices.

(MPEC) for structural estimation. The simulation study by Su (2012) illustrates that it is computationally feasible to use state-of-the-art constrained optimization solvers to estimate a discrete choice game of incomplete information with multiple equilibria. The MPEC approach is a promising approach that can be used even when the NPL algorithm is locally unstable, although we have yet to see how successfully it can be applied to estimate empirically relevant dynamic game models such as the models of AM07.

In the first of our two main contributions, this paper derives the conditions under which the NPL algorithm converges to a consistent estimator when it is started from a neighborhood of the true value. We show that a key determinant of the convergence of the NPL algorithm is the *contraction* property of the mapping Ψ. Intuitively, the faster the mapping achieves contraction, the closer the value obtained after one iteration is to the fixed point, and the NPL algorithm works well if the mapping satisfies a good contraction property. Using the model of dynamic games of AM07 and the model of PS10 as examples, we show how the features of a model are related to the convergence property of the NPL algorithm.

As our second contribution, we propose alternative algorithms that are implementable even when the original NPL algorithm does not converge to a consistent estimator. The first algorithm replaces $\Psi(\theta, P)$ in the second step of the NPL algorithm with $\Lambda(\theta, P) = [\Psi(\theta, P)]^\alpha P^{1-\alpha}$, which has a better contraction property than Ψ under some conditions. The second algorithm decomposes the space of P into the unstable subspace and its orthogonal complement based on the eigenvectors of $\partial\Psi(\theta, P)/\partial P'$, and takes a Newton step on the unstable subspace. The third algorithm uses multiple iterations of a fixed point mapping to gain efficiency.

The rest of the paper is organized as follows. Section 2 analyzes the convergence properties of the NPL algorithm and analyzes two examples. Section 3 develops an alternative algorithm. Simulation results are reported in Section 4, and the conclusion follows. The Supplemental Material (Kasahara and Shimotsu (2012)) contains the proofs and further results, including additional alternative algorithms, models with permanent unobserved heterogeneity, and additional Monte Carlo results.

2. THE MODEL AND THE NESTED PSEUDO LIKELIHOOD ALGORITHM

2.1. *Asymptotic Properties of the NPL Estimator*

We consider a class of parametric discrete choice models of which restrictions are characterized by fixed point problems. As in AM07, we assume that the data come from a cross-section of M geographically separated markets over T periods and are stationary over time. Hence, the data are given by $\{a_{mt}, x_{mt} : m = 1, \dots, M; t = 1, \dots, T\}$, where m is the market subindex, $a_{mt} \in A$ denotes the choice variable, and $x_{mt} \in X$ denotes the observable state variable. We assume that the support of (a_{mt}, x_{mt}) is finite, $A \times X =$

$\{a^1, a^2, \ldots, a^{|A|}\} \times \{x^1, x^2, \ldots, x^{|X|}\}$.[6] All limits are taken as $M \to \infty$ unless stated otherwise.

Let $P \equiv \{P(a_{mt} = a|x_{mt} = x) : (a, x) \in A \times X\}$ denote the distribution of a_{mt} conditional on x_{mt} in market m at period t. Accordingly, P is represented by an L-dimensional vector, where $L = |A||X|$. The model is parametrized with a K-dimensional vector $\theta \in \Theta$, and the fixed point constraint $P = \Psi(\theta, P)$ summarizes the restrictions of the model. For each θ, the operator $\Psi(\theta, P)$ maps the space of conditional choice probabilities B_P into itself. The true conditional choice probability P^0 is one of the fixed points of the operator $\Psi(\theta, P)$ evaluated at the true parameter value θ^0. Given θ, the Jacobian $\nabla_{P'}\Psi(\theta, P)$ is an $L \times L$ matrix, where $\nabla_{P'} \equiv (\partial/\partial P')$. To save space, we denote the Jacobian matrices *evaluated at the true value* (θ^0, P^0) as $\Psi_P^0 \equiv \nabla_{P'}\Psi(\theta^0, P^0)$ and $\Psi_\theta^0 \equiv \nabla_{\theta'}\Psi(\theta^0, P^0)$. Let $\|\cdot\|$ denote the Euclidean norm.

We collect the assumptions employed in AM07. Define $Q_M(\theta, P) \equiv M^{-1} \times \sum_{m=1}^{M}\sum_{t=1}^{T} \ln \Psi(\theta, P)(a_{mt}|x_{mt})$, $\tilde{\theta}_M(P) \equiv \arg\max_{\theta\in\Theta} Q_M(\theta, P)$, $Q_0(\theta, P) \equiv EQ_M(\theta, P)$, $\tilde{\theta}_0(P) \equiv \arg\max_{\theta\in\Theta} Q_0(\theta, P)$, and $\phi_0(P) \equiv \Psi(\tilde{\theta}_0(P), P)$. Define the set of population NPL fixed points as $Y_0 \equiv \{(\theta, P) \in \Theta \times B_P : \theta \in \tilde{\theta}_0(P)$ and $P \in \phi_0(P)\}$. See AM07 for details. Denote the sth-order derivative of a function f with respect to all of its parameters by $\nabla^s f$. Let $\mathcal{N}$ denote a closed neighborhood of (θ^0, P^0).

ASSUMPTION 1: (a) *The observations* $\{a_{mt}, x_{mt} : m = 1, \ldots, M; t = 1, \ldots, T\}$ *are independent across* m *and stationary over* t, *and* $\Pr(x_{mt} = x) > 0$ *for all* $x \in X$. (b) $\Psi(\theta, P)(a|x) > 0$ *for any* $(a, x) \in A \times X$ *and any* $(\theta, P) \in \Theta \times B_P$. (c) $\Psi(\theta, P)$ *is twice continuously differentiable*. (d) Θ *is compact and* B_P *is a compact and convex subset of* $[0, 1]^L$. (e) *There is a unique* $\theta^0 \in \mathrm{int}(\Theta)$ *such that* $P^0 = \Psi(\theta^0, P^0)$. (f) (θ^0, P^0) *is an isolated population NPL fixed point*. (g) $\tilde{\theta}_0(P)$ *is a single-valued and continuous function of P in a neighborhood of* P^0. (h) *The operator* $\phi_0(P) - P$ *has a nonsingular Jacobian matrix at* P^0.

Assumptions 1(b) and 1(c) imply that $E\sup_{(\theta,P)\in\Theta\times B_P} \|\nabla^2 \ln \Psi(\theta, P)(a_{mt}|x_{mt})\|^r < \infty$ for any positive integer r. Assumption 1(e) is a standard identification condition. Assumptions 1(f) and 1(g) correspond to assumptions (v) and (vi) in Proposition 2 of AM07, respectively.

The PML estimator is $\hat{\theta}_{\mathrm{PML}} = \arg\max_{\theta\in\Theta} Q_M(\theta, \hat{P}_0)$, where $\hat{P}_0$ is an initial consistent estimator of P^0. Proposition 1 of AM07 showed that the PML estimator is consistent under Assumption 1. Also, when $\hat{P}_0$ satisfies

[6] It would be interesting to extend our analysis to models with continuously distributed variables. The asymptotic analysis of the NPL estimator in such models may become substantially complicated, however, because it involves functional derivatives of mappings such as $\tilde{\theta}_M(P)$. We conjecture that, under suitable regularity conditions, the NPL estimator is asymptotically normal and Propositions 1 and 2 hold if matrices such as Ψ_P^0 and M_{Ψ_θ} are replaced with corresponding operators. A detailed analysis is left for future research.

$\sqrt{M}(\hat{P}_0 - P^0) \to_d N(0, \Sigma)$, the PML estimator is asymptotically normal with asymptotic variance $V_{\text{PML}} = (\Omega_{\theta\theta})^{-1} + (\Omega_{\theta\theta})^{-1}\Omega_{\theta P}\Sigma(\Omega_{\theta P})'(\Omega_{\theta\theta})^{-1}$, where $\Omega_{\theta\theta} \equiv E(\nabla_\theta s_m \nabla_{\theta'} s_m)$ and $\Omega_{\theta P} \equiv E(\nabla_\theta s_m \nabla_{P'} s_m)$ with $s_m \equiv \sum_{t=1}^{T} \ln \Psi(\theta^0, P^0)(a_{mt}|x_{mt})$.

As discussed in the Introduction, Aguirregabiria and Mira (2002, 2007) developed a recursive extension of the PML estimator called the NPL algorithm. Starting from an initial estimator of P^0, the NPL algorithm generates a sequence of estimators $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^{k}$, which we call the *NPL sequence*. If the NPL sequence converges, its limit satisfies the conditions

$$\text{(1)} \qquad \check{\theta} = \arg\max_{\theta \in \Theta} Q_M(\theta, \check{P}) \quad \text{and} \quad \check{P} = \Psi(\check{\theta}, \check{P}).$$

A pair $(\check{\theta}, \check{P})$ that satisfies these two conditions in (1) is called an *NPL fixed point*. There could be multiple NPL fixed points. The *NPL estimator*, denoted by $(\hat{\theta}_{\text{NPL}}, \hat{P}_{\text{NPL}})$, is defined as the NPL fixed point with the highest value of the pseudo likelihood among all the NPL fixed points.

Proposition 2 of AM07 establishes the consistency of the NPL estimator $\hat{\theta}_{\text{NPL}}$ under Assumption 1. Thus, the NPL estimator is a consistent NPL fixed point. The NPL estimator is asymptotically normal with asymptotic variance $V_{\text{NPL}} = [\Omega_{\theta\theta} + \Omega_{\theta P}(I - \Psi_P^0)^{-1}\Psi_\theta^0]^{-1}\Omega_{\theta\theta}\{[\Omega_{\theta\theta} + \Omega_{\theta P}(I - \Psi_P^0)^{-1}\Psi_\theta^0]^{-1}\}'$, while the asymptotically efficient "one-step" MLE can be obtained from the NPL estimator by a one-step update (see p. 29 of AM07 for details). The NPL estimator does not depend on the initial estimator of P^0 and is more efficient than the PML estimator, especially when the initial estimator of P^0 is imprecise.

While AM07 illustrated that the estimator obtained as a limit of the NPL sequence performs very well relative to the PML estimator in their simulation, they neither provided the conditions under which the NPL sequence converges to a consistent NPL fixed point nor analyzed how quickly the convergence occurs. On the other hand, PS10 presented an example in which the NPL sequence converges to an NPL fixed point that is a distance away from the true value. To date, little is known about the conditions under which the NPL sequence converges to a consistent NPL fixed point, that is, the NPL estimator.

2.2. *Convergence Properties of the NPL Algorithm*

We now analyze the conditions under which the NPL algorithm produces the NPL estimator when started from a neighborhood of the true value.

ASSUMPTION 2: (a) *Assumption* 1 *holds*. (b) $\Psi(\theta, P)$ *is three times continuously differentiable in* $\mathcal{N}$. (c) $\Omega_{\theta\theta}$ *is nonsingular*.

Let $P_{a,x}^0$ denote an $L \times 1$ vector whose elements are the probability mass function of (a_{mt}, x_{mt}) arranged conformably with $\Psi(a|x)$. Let $\Delta_P \equiv$

$\mathrm{diag}(P^0)^{-2}\,\mathrm{diag}(P^0_{a,x})$.[7] With this notation, we may write $\Omega_{\theta\theta} = T\Psi^{0\prime}_\theta \Delta_P \Psi^0_\theta$ and $\Omega_{\theta P} = T\Psi^{0\prime}_\theta \Delta_P \Psi^0_P$. Define $M_{\Psi_\theta} \equiv I - \Psi^0_\theta(\Psi^{0\prime}_\theta \Delta_P \Psi^0_\theta)^{-1}\Psi^{0\prime}_\theta \Delta_P$, and define the spectral radius of A as $\rho(A) \equiv \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$. Then $(M_{\Psi_\theta}\Psi^0_P)^k \to 0$ as $k \to \infty$ if and only if $\rho(M_{\Psi_\theta}\Psi^0_P) < 1$ (Horn and Johnson (1985), Theorem 5.6.12).[8] As the following propositions show, $\rho(M_{\Psi_\theta}\Psi^0_P)$ determines the local convergence and the local divergence of the NPL sequence.

PROPOSITION 1: *Suppose that Assumption 2 holds and that* $\rho(M_{\Psi_\theta}\Psi^0_P) < 1$. *Then, there exists a neighborhood* $\mathcal{N}_1$ *of* P^0 *such that, for any initial value* $\tilde{P}_0 \in \mathcal{N}_1$, *we have* $\lim_{k\to\infty} \tilde{P}_k = \hat{P}_{\mathrm{NPL}}$ *almost surely.*

Let H be an $L \times L$ matrix of the generalized eigenvectors of $M_{\Psi_\theta}\Psi^0_P$ such that its first r columns correspond to the eigenvalues of $M_{\Psi_\theta}\Psi^0_P$ that are greater than 1 in modulus and each column of H has a length of 1. Split H^{-1} as $H^{-1} = \binom{H_1}{H_2}$, where H_1 is $r \times L$. For a constant c, we define a set $V(c) \equiv \{P \in [0,1]^L : \|H_1(P - \hat{P}_{\mathrm{NPL}})\| \le c\|H_2(P - \hat{P}_{\mathrm{NPL}})\|\}$. When $c = 0$, the set $V(0)$ reduces to a hyperplane $H_1(P - \hat{P}_{\mathrm{NPL}}) = 0$ spanned by the eigenvectors of $M_{\Psi_\theta}\Psi^0_P$ associated with eigenvalues that are no greater than 1 in modulus, on which the NPL sequence is nondivergent.

PROPOSITION 2: *Suppose that Assumption 2 holds and that* $\rho(M_{\Psi_\theta}\Psi^0_P) > 1$. *Then, for any* $c > 0$, *there exists a neighborhood* $\mathcal{N}_c$ *of* P^0 *such that, for any* $\tilde{P}_{j-1} \in \mathcal{N}_c \setminus V(c)$, *we have* $\|H_1(\tilde{P}_j - \hat{P}_{\mathrm{NPL}})\| > \|H_1(\tilde{P}_{j-1} - \hat{P}_{\mathrm{NPL}})\|$ *and* $\tilde{P}_j \notin V(c)$ *almost surely. Consequently, for any initial value* $\tilde{P}_0 \in \mathcal{N}_c \setminus V(c)$, *the NPL sequence does not converge to* $\hat{P}_{\mathrm{NPL}}$ *almost surely if it stays in* $\mathcal{N}_c$.

REMARK 1: In single-agent dynamic models, the Jacobian matrix Ψ^0_P is zero (Aguirregabiria and Mira (2002), Proposition 2). Consequently, the NPL method is always stable at $(\hat{\theta}_{\mathrm{NPL}}, \hat{P}_{\mathrm{NPL}})$. Proposition 7 in the Supplemental Material shows that $\tilde{P}_j - \hat{P}_{\mathrm{NPL}} = O(M^{-1/2}\|\tilde{P}_{j-1} - \hat{P}_{\mathrm{NPL}}\| + \|\tilde{P}_{j-1} - \hat{P}_{\mathrm{NPL}}\|^2)$ almost surely, which implies that the convergence rate is faster than linear. See Kasahara and Shimotsu (2008) for further details.

The matrix M_{Ψ_θ} represents the effect of updating θ in the first step of the NPL algorithm, whereas Ψ^0_P is the Jacobian of updating P in the second step.

[7]In a multiplayer model of a dynamic game in which unobserved state variables are independent across players, such as the model of AM07, Δ_P is simplified as $\mathrm{diag}(P^0)^{-1}\,\mathrm{diag}(f_x)$, where f_x is an $L \times 1$ vector whose elements are the probability mass function of x_i arranged conformably with $P(a|x)$.

[8]$\rho(A) \le \|A\|$ holds for any matrix A and any matrix norm $\|\cdot\|$. Therefore, $\|A\| < 1$ is a sufficient but not necessary condition for the convergence of A^k to zero.

When $\rho(M_{\Psi_\theta}\Psi_P^0) > 1$, an NPL sequence starting from $\mathcal{N}_c \setminus V(c)$ converges to $\hat{P}_{\text{NPL}}$ only if the NPL sequence first moves outside $\mathcal{N}_c$ and then moves either to $\hat{P}_{\text{NPL}}$ from outside of $\mathcal{N}_c$ or to $\mathcal{N}_c \cap V(c)$. The constant c in Proposition 2 can be chosen to be as small as desired, and doing so makes $\mu(\mathcal{N}_c \cap V(c))/\mu(\mathcal{N}_c)$ arbitrarily small, where μ is a Lebesgue measure. The case with $\rho(M_{\Psi_\theta}\Psi_P^0) = 1$ corresponds to a boundary case. The linear difference equation in Proposition 1 cannot fully characterize the local property of the fixed point, which depends on the details of the model (see, e.g., pp. 348–351 of Strogatz (1994)).

In general, given the nonlinear nature of the mapping Ψ, its local behavior may not fully characterize its global convergence property. For instance, even when $\rho(M_{\Psi_\theta}\Psi_P^0) > 1$, the NPL sequence may move away from the NPL fixed point initially and then move back to the NPL fixed point from a distance away. When the NPL sequence diverges away from the NPL estimator, an analysis of nonlinear dynamics (see, e.g., Chapter 10 of Strogatz (1994)) suggests three representative possibilities. First, as PS10 illustrated, the NPL sequence may converge to a NPL fixed point that is different from the NPL estimator. Second, as our simulation suggests, it may converge to a stable cycle. Third, the NPL sequence might never settle down to a fixed point or a period orbit.

2.3. *The Relation Between $\rho(M_{\Psi_\theta}\Psi_P^0)$ and $\rho(\Psi_P^0)$*

The condition $\rho(M_{\Psi_\theta}\Psi_P^0) < 1$ plays an important role for the convergence of the NPL algorithm. Because Ψ_P^0 is often closely related to the characteristics of the economic model, we want to find a bound of $\rho(M_{\Psi_\theta}\Psi_P^0)$ in terms of $\rho(\Psi_P^0)$. Since M_{Ψ_θ} is idempotent, M_{Ψ_θ} is diagonalizable as $M_{\Psi_\theta} = SDS^{-1}$, where the first $L - K$ diagonal elements of D are 1 and the other elements of D are zero. From the properties of the eigenvalues, we have $\rho(M_{\Psi_\theta}\Psi_P^0) = \rho(SDS^{-1}\Psi_P^0) = \rho(DS^{-1}\Psi_P^0 S)$. In our context, typically $L \gg K$ because the dimension of the state variable is much larger than the number of parameters. Consequently, D is close to an identity matrix, and we expect that $DS^{-1}\Psi_P^0 S \simeq S^{-1}\Psi_P^0 S$, which implies that the dominant eigenvalues of $M_{\Psi_\theta}\Psi_P^0$ and Ψ_P^0 are close to each other.[9] In our dynamic game model with $L = 144$ and $K = 2$, we find that $\rho(M_{\Psi_\theta}\Psi_P^0)$ is very similar to $\rho(\Psi_P^0)$ (see Table I).

2.4. *Simplex Restriction on P*

Since P represents probabilities, the elements of P must satisfy a simplex-type restriction, and this restriction needs to be imposed in parameterizing $\Psi(\theta, P)$. Suppose a has $J + 1$ support points, and split P into P^+ and P^-, where P^+ corresponds to the first to Jth choices, whereas P^- corresponds to

[9]If $\lambda(A)$ is an algebraically simple eigenvalue of A, then $\lambda(A+\Delta)/\lambda(A) = (y^H\Delta x)/(y^H Ax) + (\|\Delta\|^2)$, where x and y are a right- and left- $\lambda(A)$ eigenvector of A. See, for example, Theorem 6.3.12 of Horn and Johnson (1985).

TABLE I

THE SPECTRAL RADIUS OF Ψ_P^0 AND Λ_P^0[a]

θ_{RN}	α	$\rho(\Psi_P^0)$	$\rho(\Lambda_P^0)$	$\rho(M_{\Psi_\theta}\Psi_P^0)$	$\rho(M_{\Lambda_\theta}\Lambda_P^0)$
1	0.9407	0.3365	0.2572	0.2916	0.2557
2	0.8830	0.6925	0.4945	0.5949	0.4936
4	0.8250	1.1839	0.8017	1.1799	0.8046
6	0.7730	1.4788	0.9161	1.4777	0.9153

[a]The second column reports the optimal choice of α under which Λ_P^0 has the smallest spectral radius.

the $(J+1)$th choice. Let $\mathbf{1}_k$ denote a k-vector of ones; then the simplex restriction implies $P^- = \mathbf{1}_{\dim(P^-)} - \mathcal{E}P^+$ for a matrix $\mathcal{E}$ of zeros and ones defined appropriately. $\Psi(\theta, P)$ satisfies an analogous simplex restriction by its construction. Split $\Psi(\theta, P)$ analogously, and write P and $\Psi(\theta, P)$ as

$$P = \begin{pmatrix} P^+ \\ P^- \end{pmatrix} = \begin{pmatrix} P^+ \\ \mathbf{1}_{\dim(P^-)} - \mathcal{E}P^+ \end{pmatrix} = P(P^+), \tag{2}$$

$$\begin{aligned} \Psi(\theta, P) = \Psi(\theta, P(P^+)) &= \begin{pmatrix} \Psi^+(\theta, P^+) \\ \Psi^-(\theta, P^+) \end{pmatrix} \\ &= \begin{pmatrix} \Psi^+(\theta, P^+) \\ \mathbf{1}_{\dim(P^-)} - \mathcal{E}\Psi^+(\theta, P^+) \end{pmatrix}. \end{aligned} \tag{3}$$

Note from (3) that the derivative of $\Psi(\theta, P)$ with respect to P^- is zero.

The following proposition shows that the restrictions (2) and (3) do not affect the validity of Propositions 1 and 2. Define $\Psi_\theta^+ \equiv \nabla_{\theta'}\Psi^+(\theta^0, P^{0+})$, $\Psi_{P^+}^+ \equiv \nabla_{P^{+\prime}}\Psi^+(\theta^0, P^{0+})$, and $M_{\Psi_\theta}^+ \equiv I_{\dim(P^+)} - \Psi_\theta^+(\Psi_\theta^{+\prime}\Delta_P^+\Psi_\theta^+)^{-1}\Psi_\theta^{+\prime}\Delta_P^+$, where $\Delta_P^+ \equiv U'\Delta_P U$ with $U \equiv [I_{\dim(P^+)} \vdots -\mathcal{E}']'$.

PROPOSITION 3: *Suppose that $\tilde{P}_0$ satisfies the simplex restriction* (2). *Then, Propositions* 1 *and* 2 *hold, and the nonzero eigenvalues of $M_{\Psi_\theta}\Psi_P^0$ and Ψ_P^0 are the same as the nonzero eigenvalues of $M_{\Psi_\theta}^+\Psi_{P^+}^+$ and $\Psi_{P^+}^+$, respectively.*

Therefore, in practice, it suffices to check the eigenvalues of $M_{\Psi_\theta}^+\Psi_{P^+}^+$ to examine the convergence property of the NPL algorithm.

2.5. *Examples*

The following two examples illustrate Propositions 1–3.

EXAMPLE 1—A Dynamic Discrete Game by PS10: PS10 presented a game in which the global behavior of the NPL mapping can be analytically derived.

There are two firms with a binary choice $a_i \in \{0, 1\}$ for $i = 1, 2$, where $a_i = 1$ indicates firm i is active. The model has no state variable. The conditional choice probability is summarized by $P^+ = (P_1^+, P_2^+)'$, where P_i^+ denotes firm i's probability of choosing $a_i = 1$. The model has one parameter, θ, and the true parameter value θ^0 is in the interior of the parameter space $\Theta = [-10, -1]$. The data are generated from a unique symmetric equilibrium, $P_1^+ = P_2^+ = 1/(1 - \theta^0)$. When P^+ is in a neighborhood of the equilibrium, the mapping Ψ^+ takes the form

$$\Psi^+(\theta, P^+) = \begin{pmatrix} \Psi_1^+(\theta, P^+) \\ \Psi_2^+(\theta, P^+) \end{pmatrix} = \begin{pmatrix} 1 + \theta P_2^+ \\ 1 + \theta P_1^+ \end{pmatrix}.$$

PS10 showed that the NPL sequence converges to one of the inconsistent NPL fixed points if the initial estimate does not satisfy $P_1^+ = P_2^+$; if the initial estimate does satisfy $P_1^+ = P_2^+$, then the NPL sequence converges to the NPL estimator in one iteration.

We apply our local analysis to their model. With the definition of $\Psi_{P^+}^+$ and $M_{\Psi_\theta}^+$, a direct calculation gives

$$\Psi_{P^+}^+ = \begin{pmatrix} 0 & \theta^0 \\ \theta^0 & 0 \end{pmatrix}, \quad M_{\Psi_\theta}^+ = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$
$$M_{\Psi_\theta}^+ \Psi_{P^+}^+ = \frac{\theta^0}{2}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The eigenvalues of $\Psi_{P^+}^+$ are θ^0 and $-\theta^0$, and the eigenvalues of $M_{\Psi_\theta}^+ \Psi_{P^+}^+$ are 0 and $-\theta^0$. Because all the eigenvalues of $\Psi_{P^+}^+$ are outside the unit circle, the fixed point mapping $P^+ = \Psi^+(\theta, P^+)$ has no convergent path. Multiplying $M_{\Psi_\theta}^+$ annihilates the eigenvector of $\Psi_{P^+}^+$ associated with θ^0 but does not change the spectral radius of $\Psi_{P^+}^+$. Consequently, the NPL operator inherits the instability of $\Psi(\theta, P)$. From Proposition 2, the NPL sequence diverges away from the NPL estimator in the neighborhood of (θ^0, P^0) if the initial estimator does not lie on the convergent trajectory $P_1^+ = P_2^+$. These local results are weaker than the global results in PS10, but are consistent with their findings.

In PS10, it was assumed that $\theta^0 < -1$. However, if $\theta^0 \in (-1, 0)$, then $\rho(M_{\Psi_\theta}^+ \Psi_{P^+}^+) < 1$, and the NPL sequence locally converges to the NPL estimator. The range of the parameter values for which the NPL operator is stable corresponds to a small interaction between agents, where $\theta^0 = 0$ implies no interaction. When $\theta^0 = -1$, then $\rho(M_{\Psi_\theta} \Psi_P) = 1$ and we cannot apply our local analysis.[10]

The stability property of $\Psi(\theta, P)$ may not completely characterize the stability property of the NPL operator because of the effect of $\tilde{\theta}_M(\tilde{P}_{j-1})$ in the

[10]In the model of PS10, there exists a unique globally stable population NPL fixed point when $\theta^0 = -1$.

NPL algorithm. Now, suppose that firm i's payoff is given by $\theta_i + \varepsilon_i^1$ if both firms are active, so that the model has two parameters θ_1 and θ_2. Suppose that the true parameter value is $\theta_1^0 = \theta_2^0 = \theta_0^0$, and the data are generated from $P_1^+ = P_2^+ = 1/(1-\theta_0^0)$ as before, although we do not impose $\theta_1 = \theta_2$ in the estimation. Then,

$$\Psi^+(\theta, P^+) = \begin{pmatrix} 1+\theta_1 P_2^+ \\ 1+\theta_2 P_1^+ \end{pmatrix}, \quad \Psi_{P^+}^+ = \begin{pmatrix} 0 & \theta_1^0 \\ \theta_2^0 & 0 \end{pmatrix},$$

$$M_{\Psi_\theta}^+ = M_{\Psi_\theta}^+ \Psi_{P^+}^+ = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The eigenvalues of $\Psi_{P^+}^+$ are $\pm\sqrt{\theta_1^0\theta_2^0} = \pm\theta_0^0$, while the eigenvalues of $M_{\Psi_\theta}^+ \Psi_{P^+}^+$ are equal to zero. If $\theta_0^0 \in (-10, -1)$, the data are generated from the unstable symmetric equilibrium studied in PS10. The NPL sequence converges locally, however, because multiplying $M_{\Psi_\theta}^+$ annihilates both of the eigenvectors of $\Psi_{P^+}^+$ associated with θ_0^0 and $-\theta_0^0$.

EXAMPLE 2—A Dynamic Discrete Game by AM07: Consider the model of dynamic discrete games in Section 2 of AM07, with two firms and a binary choice $a_{it} \in A = \{0, 1\}$ for $i = 1, 2$. The primitives of the model are the profit functions $\Pi_i(a_{it}, a_{-it}, x_t; \theta)$'s, the transition probability function $f(x_{t+1}|a_{1t}, a_{2t}, x_t)$, the probability density function $g(\varepsilon_{it}; \theta)$ of independent and identically distributed (i.i.d.) private information, and the discount factor β. $x_t \in X = \{x^1, \dots, x^{|X|}\}$ is common knowledge and consists of $(S_t, a_{1,t-1}, a_{2,t-1})$, where S_t follows an exogenous Markov process, so that the profit functions are $\Pi_i(a_{it}, a_{-it}, S_t, a_{i,t-1}, a_{-i,t-1}; \theta)$. $f(x_{t+1}|x_t, a_t)$ and β are assumed to be known.

The vector of the conditional choice probability is given by $P = (P_1', P_2')'$, where $P_i = (P_i(0|x^1), \dots, P_i(0|x^{|X|}), P_i(1|x^1), \dots, P_i(1|x^{|X|}))'$. The equilibrium of the model is a fixed point of the mapping $\Psi(\theta, P)$ defined in equation (15) in AM07. In the Supplemental Material, we show that the Jacobian of $\Psi(\theta, P)$ evaluated at (θ^0, P^0) takes the form

$$\Psi_P^0 = \begin{pmatrix} 0 & \nabla_{P_2'}\Psi_1(\theta^0, P^0) \\ \nabla_{P_1'}\Psi_2(\theta^0, P^0) & 0 \end{pmatrix}, \tag{4}$$

where the diagonal blocks of Ψ_P^0 are zero from Proposition 2 of Aguirregabiria and Mira (2002).

The form of Ψ_P^0 in (4) suggests that the best response mapping $\Psi(\theta, P)$ is locally stable at an equilibrium if $\nabla_{P_{-i}'}\Psi_i(\theta^0, P^0)$'s are sufficiently small.[11] The

[11] As pointed out by Su and Judd (2012), the stability of the best response mapping at an equilibrium is not among the common notions pertaining to the stability of equilibria in game theory literature. PS10 and Aguirregabiria and Nevo (2010) provided two contrasting views on the possibility of using the stability of the best response mapping as a refinement concept.

following proposition shows how the local convergence condition is related to the size of the interaction between agents.

ASSUMPTION 3: (a) *Assumption* 2 *holds*. (b) $\Pi_i(a_{it}, a_{-it}, x_t; \theta)$ *is twice continuously differentiable in* θ *for* $i = 1, 2$. (c) *There exists* $\theta^* \in \Theta$ *such that* $\Pi_i(a_{it}, a^{\dagger}_{-it}, S_t, a_{i,t-1}, a^{\dagger}_{-i,t-1}; \theta^*) = \Pi_i(a_{it}, a^{\ddagger}_{-it}, S_t, a_{i,t-1}, a^{\ddagger}_{-i,t-1}; \theta^*)$ *for any* $(a^{\dagger}_{-it}, a^{\ddagger}_{-it}, a^{\dagger}_{-i,t-1}, a^{\ddagger}_{-i,t-1}) \in A^4$ *for* $i = 1, 2$. (d) *There exists* $\theta^{\diamond} \in \Theta$ *such that* $\Pi_i(a_{it}, a^{\dagger}_{-it}, x_t; \theta^{\diamond}) = \Pi_i(a_{it}, a^{\ddagger}_{-it}, x_t; \theta^{\diamond})$ *for any* $(a^{\dagger}_{-it}, a^{\ddagger}_{-it}) \in A^2$, *for* $i = 1, 2$.

Assumption 3(c) implies that, under θ^*, neither a current nor a past action of the competitor affects a firm's profit function. Assumption 3(d) is weaker than Assumption 3(c) and implies that, under $\theta^{\diamond}$, a current action of the competitor does not affect a firm's profit function.

PROPOSITION 4: (a) *Suppose that Assumptions* 3(a)–(c) *hold*. *Then*, *there exists a neighborhood* $\mathcal{N}^*$ *of* θ^* *such that*, *for any* $\theta^0 \in \mathcal{N}^*$, *there is a Markov perfect equilibrium* $P^0 = \Psi(\theta^0, P^0)$ *that satisfies* $\rho(M_{\Psi_\theta} \Psi^0_P) < 1$. (b) *Suppose that Assumptions* 3(a), 3(b), *and* 3(d) *hold*. *Then*, *there exists a neighborhood* $\mathcal{N}^{\diamond}$ *of* $(\beta, \theta) = (0, \theta^{\diamond})$ *such that*, *for any* $(\beta, \theta^0) \in \mathcal{N}^{\diamond}$, *there is a Markov perfect equilibrium* $P^0 = \Psi(\theta^0, P^0)$ *that satisfies* $\rho(M_{\Psi_\theta} \Psi^0_P) < 1$.

Therefore, a Markov perfect equilibrium for which the local convergence holds exists if the contemporaneous interaction between firms is small and either (i) the dynamic interaction between firms is small, or (ii) the discount factor is small.

3. ALTERNATIVE SEQUENTIAL LIKELIHOOD-BASED ESTIMATORS

When $\Psi(\theta, P)$ is not a contraction in a neighborhood of (θ^0, P^0), the NPL algorithm may not produce a consistent estimator. This section discusses alternative estimation algorithms that are implementable even in such cases.

Consider a class of mappings obtained as a log-linear combination of $\Psi(\theta, P)$ and P:

$$[\Lambda(\theta, P)](a|x) \equiv \big\{[\Psi(\theta, P)](a|x)\big\}^{\alpha} P(a|x)^{1-\alpha},$$

for all $(a, x) \in A \times X$. In numerical analysis, this is known as the relaxation method.[12] We consider the *NPL-Λ algorithm* that updates θ as in the first step of the original NPL algorithm but updates P using $\Lambda(\theta, P)$ in place of $\Psi(\theta, P)$ in the second step. P is a fixed point of $\Psi(\theta, P)$ if and only if it is a fixed point of $\Lambda(\theta, P)$. Therefore, in view of equation (1), the original NPL algorithm and

[12]Başar (1987) applied the relaxation method to find a Nash equilibrium. Ljungqvist and Sargent (2004, p. 574) also suggested applying the relaxation method to the model of Aiyagari (1994).

the NPL-Λ algorithm share the same set of NPL fixed points. Define $\Lambda_P^0 \equiv \nabla_{P'}\Lambda(\theta^0, P^0)$.

PROPOSITION 5: (a) *Suppose that the real part of every eigenvalue of Ψ_P^0 is smaller than* 1. *Then there exists $\alpha > 0$ such that $\rho(\Lambda_P^0) < 1$.* (b) *Suppose that the real part of every eigenvalue of Ψ_P^0 is greater than* 1. *Then there exists $\alpha < 0$ such that $\rho(\Lambda_P^0) < 1$.*

Therefore, when the real part of every eigenvalue of Ψ_P^0 is smaller than 1 (or greater than 1), we may choose the value of α so that $\Lambda(\theta, P)$ becomes locally contractive even when $\Psi(\theta, P)$ is not locally contractive.[13] Once an appropriate value of α is determined, the NPL-Λ algorithm converges to the NPL estimator under weaker conditions than for the original NPL algorithm at a similar computational cost.[14]

In the model of PS10, setting $\alpha = 1/(1+\theta^0)$ gives $\rho(M^+_{\Lambda_\theta}\Lambda^+_{P^+}) = 0$ for $\theta^0 \in (-10, -1) \cup (-1, 0)$ and the local convergence condition holds, where $\Lambda^+(\theta, P^+) = \{\Psi^+(\theta, P^+)\}^\alpha (P^+)^{1-\alpha}$ and $M^+_{\Lambda_\theta}$ is defined analogously to $M^+_{\Psi_\theta}$.[15]

In the Supplemental Material, we discuss two additional algorithms, the Recursive Projection Method (RPM) and the q-NPL algorithm. The RPM converges locally for any eigenvalues of Ψ_P^0, but it is computationally more intensive than the relaxation method. The q-NPL algorithm improves the efficiency of the estimates from the relaxation method and RPM algorithm.

4. MONTE CARLO EXPERIMENTS

We consider a dynamic game model of market entry and exit studied in Section 4 of AM07. We set the number of firms $N = 3$. The profit of firm i operating in market m in period t is equal to $\tilde{\Pi}_{it}(1) = \theta_{RS} \ln S_{mt} - \theta_{RN} \ln(1 + \sum_{j\neq i} a_{jmt}) - \theta_{FC,i} - \theta_{EC}(1 - a_{im,t-1}) + \varepsilon_{imt}(1)$, whereas its profit is $\tilde{\Pi}_{it}(0) =$

[13]When $\alpha < 0$, the elements of $\Lambda(\theta, P)$ may take values greater than 1 if $\Psi(\theta, P)$ is very small and (θ, P) is away from (θ^0, P^0). In practice, when $\alpha < 0$, we may modify Step 2 as $\tilde{P}_j(a|x) = \min\{[\Lambda(\tilde{\theta}_j, \tilde{P}_{j-1})](a|x), 1 - \varepsilon\}$ for a small $\varepsilon > 0$ to avoid such a possibility. When all the eigenvalues of Ψ_P^0 are real and smaller than 1, the optimal α is given by Judd (1998, p. 80) as $\alpha^* = 2/(2 - \lambda_{\max} - \lambda_{\min})$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of Ψ_P^0. In general, to optimally choose the value of α, we need to evaluate the Jacobian matrix Ψ_P^0 and all of its eigenvalues, say, using the PML estimator. In practice, when the evaluation of Ψ_P^0 is too costly, choosing $\alpha \approx 0$ leads to a locally contracting Λ from the proof of Proposition 5.

[14]Step 1 of the NPL-Λ algorithm is identical to that of the NPL algorithm because both algorithms update θ by maximizing the same objective function, while the computational cost of Step 2 of the NPL-Λ is mostly determined by the cost of evaluating the mapping $\Psi(\tilde{\theta}_j, \tilde{P}_{j-1})$.

[15]A direct calculation gives

$$M^+_{\Lambda_\theta}\Lambda^+_{P^+} = \frac{1 - \alpha(1+\theta^0)}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{for} \quad \alpha = 1/(1+\theta^0).$$

$\varepsilon_{imt}(0)$ if the firm is not operating. We assume that $\{\varepsilon_{imt}(0), \varepsilon_{imt}(1)\}$ follow i.i.d. Type I extreme value distribution, and S_{mt} follows an exogenous first-order Markov process $f_S(S_{m,t+1}|S_{mt})$.[16] The discount factor is set to $\beta = 0.96$, and the parameter values are given by $\theta_{RS} = 1.0$, $\theta_{EC} = 1.0$, $\theta_{FC,1} = 1.0$, $\theta_{FC,2} = 0.9$, and $\theta_{FC,3} = 0.8$. The parameter θ_{RN} determines the degree of strategic substitutabilities among firms and is the main determinant of the dominant eigenvalue of Ψ_P^0. All of the eigenvalues of Ψ_P^0 are inside the unit circle for $\theta_{RN} = 1$ and 2, while the smallest eigenvalues are less than -1 for $\theta_{RN} = 4$ and 6. We therefore let θ_{RN} take on a value of 2 or 4 across experiments and examine the performance of different estimators. We estimate θ_{RS} and θ_{RN}, leaving the other parameters fixed at the true values.

We apply nonlinear equation solvers to find all of the solutions for $P = \Psi(\theta^0, P)$. Across 1000 random initial values of P, the nonlinear equation solvers always find an identical solution upon successful convergence.[17] This suggests that $\Psi(\theta^0, P)$ has a unique fixed point. Similarly, the nonlinear equation solvers always find an identical solution for $P - \phi_0(P) = 0$ upon successful convergence, suggesting the unique NPL fixed point.

To generate an observation, we first randomly draw $x_m = \{S_{m1}, a_{1m0}, a_{2m0}, a_{3m0}\}$ from the steady-state distribution implied by the model. Then, given x_m, we draw $\{a_{1m1}, a_{2m1}, a_{3m1}\}$ from the equilibrium conditional choice probabilities. We replicate 1000 simulated samples for each of $M = 500$, 2000, and 8000 observations.

For the mapping Λ, we set $\alpha = 0.825$, which minimizes the spectral radius of Λ_P^0. As shown in Table I, the spectral radii of $M_{\Psi_\theta}\Psi_P^0$ and $M_{\Lambda_\theta}\Lambda_P^0$ are very similar to those of Ψ_P^0 and Λ_P^0, respectively. Thus, in view of Propositions 1 and 2, the convergence property of the NPL algorithm is primarily determined by the dominant eigenvalue of Ψ_P^0 and Λ_P^0.

Table II compares the performance of the two-step (PML) estimator and sequential estimators generated by the following four sequential algorithms evaluated at $k = 50$ iterations: (i) the NPL algorithm using Ψ, (ii) the NPL-Λ algorithm, (iii) the approximate RPM algorithm using $\Gamma(\theta, P, \eta)$ with $\delta = 0.5$, and (iv) the approximate q-NPL using $\Lambda^q(\theta, P, \eta)$ with $q = 4$. They are denoted by

[16]The state space for the market size S_{mt} is $\{2, 6, 10\}$. The transition probability matrix of S_{mt} is given by

$$\begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.2 & 0.6 & 0.2 \\ 0.0 & 0.2 & 0.8 \end{bmatrix}.$$

[17]We use two different nonlinear equation solvers in Matlab: "c05nb" from the NAG Toolbox for Matlab and "fsolve" from the Optimization Toolbox. For $\theta_{RN}^0 = 4$, the nonlinear equation solver "c05nb" successfully found a solution for the system of the nonlinear equation $P = \Psi(\theta^0, P)$ in 568 out of 1000 cases, and all of the 568 solutions were identical. The result was similar when we used "fsolve" in place of "c05nb."

TABLE II

BIAS AND RMSE

	Estimator	$M = 500$		$M = 2000$		$M = 8000$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
				$\theta_{RN} = 2$			
$\hat{\theta}_{RS}$	PML	−0.2215	0.2698	−0.0717	0.1112	−0.0229	0.0474
	NPL-Ψ	−0.0151	0.1347	−0.0002	0.0660	−0.0023	0.0323
	NPL-Λ	−0.0151	0.1347	−0.0002	0.0660	−0.0023	0.0323
	RPM	−0.0174	0.1331	−0.0028	0.0642	−0.0027	0.0320
	q-NPL-Λ^q	−0.0117	0.1240	0.0002	0.0606	−0.0018	0.0305
$\hat{\theta}_{RN}$	PML	−0.7895	0.9604	−0.2565	0.3949	−0.0828	0.1687
	NPL-Ψ	−0.0467	0.4705	−0.0009	0.2339	−0.0095	0.1130
	NPL-Λ	−0.0467	0.4705	−0.0009	0.2339	−0.0095	0.1130
	RPM	−0.0544	0.4642	−0.0102	0.2274	−0.0111	0.1116
	q-NPL-Λ^q	−0.0358	0.4280	0.0002	0.2131	−0.0079	0.1052
				$\theta_{RN} = 4$			
$\hat{\theta}_{RS}$	PML	−0.1280	0.1557	−0.0341	0.0514	−0.0082	0.0207
	NPL-Ψ	−0.0095	0.0676	−0.0062	0.0490	−0.0005	0.0408
	NPL-Λ	0.0028	0.0575	−0.0006	0.0294	−0.0003	0.0143
	RPM	0.0029	0.0576	−0.0012	0.0284	0.0000	0.0136
	q-NPL-Λ^q	0.0015	0.0542	−0.0009	0.0277	0.0000	0.0136
$\hat{\theta}_{RN}$	PML	−0.7713	0.9094	−0.1964	0.2599	−0.0462	0.0937
	NPL-Ψ	−0.1417	0.2572	−0.1414	0.2314	−0.0918	0.1612
	NPL-Λ	0.0241	0.1424	−0.0001	0.0739	0.0013	0.0352
	RPM	0.0249	0.1604	−0.0003	0.0841	0.0014	0.0342
	q-NPL-Λ^q	0.0228	0.1351	0.0000	0.0690	0.0014	0.0328

"PML," "NPL-Ψ," "NPL-Λ," "RPM," and "q-NPL-Λ^q," respectively. We report the bias and the root mean squared error (RMSE) of $\hat{\theta}_{RN}$ and $\hat{\theta}_{RS}$ across different estimators.

For $\theta_{RN} = 2$, the NPL-Ψ performs substantially better than the PML across different sample sizes, and the NPL-Λ and NPL-Ψ converge to the same estimate. On the other hand, when $\theta_{RN} = 4$, the NPL-Ψ performs substantially worse than the NPL-Λ, reflecting divergence. Further, as the sample size increases from 500 to 8000, the RMSE of the NPL-Λ decreases approximately at the rate of $M^{1/2}$, but the RMSE of the NPL-Ψ decreases at a much slower rate. For $\theta_{RN} = 4$ and $M = 8000$, the RMSE of the NPL-Ψ is even larger than that of the PML. Across different sample sizes and parameters, the RPM and the q-NPL-Λ^q outperform the NPL-Ψ.

The first four rows of Table III compare the RMSE across the estimators of θ_{RN} generated by different algorithms after $k = 2, 5, 10, \ldots, 25$ iterations when $M = 8000$. For $\theta_{RN} = 2$, the RMSE changes little after $j = 5$ iterations across all the algorithms, indicating their convergence. For $\theta_{RN} = 4$, the RMSE of the

TABLE III

RMSE OF $\hat{\theta}_{RN,k}$ FOR $k = 2, 5, 10, \dots, 25$ AT $M = 8000$[a]

		$k=2$	$k=5$	$k=10$	$k=15$	$k=20$	$k=25$
				$\theta_{RN} = 2$			
$\tilde{\theta}_{RN,k}$	NPL-Ψ	0.1196	0.1133	0.1130	0.1130	0.1130	0.1130
	NPL-Λ	0.1227	0.1131	0.1130	0.1130	0.1130	0.1130
	RPM	0.1401	0.1122	0.1120	0.1118	0.1117	0.1116
	q-NPL-Λ^q	0.1061	0.1051	0.1052	0.1052	0.1052	0.1052
RMSE of $(\tilde{\theta}_{RN,k+1} - \tilde{\theta}_{RN,k})$		0.0532	0.0041	0.0003	0.0000	0.0000	0.0000
RMSE of $(\tilde{\theta}_{RN,k+2} - \tilde{\theta}_{RN,k})$		0.0505	0.0017	0.0001	0.0000	0.0000	0.0000
				$\theta_{RN} = 4$			
$\tilde{\theta}_{RN,k}$	NPL-Ψ	0.0713	0.0748	0.0807	0.1235	0.1299	0.1593
	NPL-Λ	0.0651	0.0363	0.0353	0.0352	0.0352	0.0352
	RPM	0.0600	0.0357	0.0350	0.0341	0.0343	0.0342
	q-NPL-Λ^q	0.0366	0.0332	0.0328	0.0328	0.0328	0.0328
RMSE of $(\tilde{\theta}_{RN,k+1} - \tilde{\theta}_{RN,k})$		0.1272	0.1106	0.1551	0.2037	0.2410	0.2624
RMSE of $(\tilde{\theta}_{RN,k+2} - \tilde{\theta}_{RN,k})$		0.0310	0.0152	0.0157	0.0132	0.0101	0.0076

[a]The last two rows report the RMSE of $(\tilde{\theta}_{RN,k+1} - \tilde{\theta}_{RN,k})$ and $(\tilde{\theta}_{RN,k+2} - \tilde{\theta}_{RN,k})$ for NPL-Ψ.

NPL-Ψ sequence increases with the number of iterations, whereas our proposed estimators converge after 10 iterations. The last two rows of Table III report the RMSE of the first and the second differences of the NPL-Ψ sequence so as to examine its possible convergence to a two-period cycle. When $\theta_{RN} = 4$, the NPL-Ψ sequence does not converge to a NPL fixed point, but it gradually converges *every other iteration*, suggesting its convergence toward a two-period cycle.

5. CONCLUDING REMARKS AND EXTENSION

This paper analyzes the convergence properties of the NPL algorithm to estimate a class of structural models characterized by a fixed point constraint. We demonstrate how the local convergence property of the NPL algorithm is related to the feature of an economic model, and show that a key determinant is the contraction property of the fixed point mapping.

In practice, the convergence condition may be violated. In such a case, the NPL algorithm will not converge to a consistent estimator even if it is started from a neighborhood of the true parameter value. We develop alternative sequential estimators that can be used even when the original fixed point mapping is not locally contractive. As our simulations illustrate, these alternative estimators work well even when the original fixed point mapping is not a contraction, and their performance is substantially better than that of the two-step PML estimator.

Our convergence analysis is local. In a model with multiple NPL fixed points, whether the sequential algorithms analyzed in this paper can be used to obtain a consistent NPL fixed point depends on the initial value of P. Thus, when a reliable initial estimate is not available, it is recommended to repeatedly apply the NPL algorithm with different initial values. A closely related unresolved issue is the size of the domain of attraction for these sequential algorithms. For instance, if the q-NPL algorithm has a smaller domain of attraction than the NPL algorithm, then the finite sample properties of the q-NPL estimator may be worse than those of the NPL estimator. Examining such a possibility is an important future topic.

REFERENCES

AGUIRREGABIRIA, V., AND P. MIRA (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70 (4), 1519–1543. [2303,2304,2307,2308,2312]

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75 (1), 1–53. [2303,2304,2307]

AGUIRREGABIRIA, V., AND A. NEVO (2010): "Recent Development in Empirical IO: Dynamic Demand and Dynamic Games," Working Paper 419, University of Toronto. [2312]

AIYAGARI, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109 (3), 659–684. [2303,2313]

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (5), 1331–1370. [2304]

BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2009): "Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game," Report, Stanford University. [2304]

BAŞAR, T. (1987): "Relaxation Techniques and Asynchronous Algorithms for On-Line Computation of Noncooperative Equilibria," *Journal of Economic Dynamics and Controls*, 11, 531–549. [2313]

HORN, R. A., AND C. R. JOHNSON (1985): *Matrix Analysis*. Cambridge: Cambridge University Press. [2308,2309]

HOTZ, J., AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529. [2303,2304]

JUDD, L. J. (1998): *Numerical Methods in Economics*. Cambridge, MA: MIT Press. [2314]

KASAHARA, H., AND K. SHIMOTSU (2008): "Pseudo-Likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, 146, 92–106. [2303,2308]

——— (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77 (1), 135–175. [2304]

——— (2012): "Supplement to 'Sequential Estimation of Structural Models With a Fixed Point Constraint'," *Econometrica Supplemental Material*, 80, http://www.econometricsociety.org/ecta/Supmat/8291_proofs.pdf; http://www.econometricsociety.org/ecta/Supmat/8291_data_and_programs.zip. [2305]

KRUSELL, P., AND A. SMITH, JR. (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106 (5), 867–896. [2303]

LJUNGQVIST, L., AND T. J. SARGENT (2004): *Recursive Macroeconomic Theory* (Second Ed.). Cambridge, MA: MIT Press. [2313]

PAKES, A., M. OSTROVSKY, AND S. BERRY (2007): "Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples)," *RAND Journal of Economics*, 38 (2), 373–399. [2303,2304]

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–928. [2303,2304]

——— (2010): "Sequential Estimation of Dynamic Discrete Games: A Comment," *Econometrica*, 78, 833–842. [2304]

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033. [2303]

STROGATZ, S. H. (1994): *Nonlinear Dynamics And Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Reading, MA: Addison-Wesley. [2309]

SU, C.-L. (2012): "Estimating Discrete-Choice Games of Incomplete Information: A Simple Static Example," Working Paper, University of Chicago. [2304,2305]

SU, C.-L., AND K. L. JUDD (2012): "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica* (forthcoming). [2304,2312]

Dept. of Economics, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada; hkasahar@mail.ubc.ca

and

Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku Tokyo, 113-0033, Japan; shimotsu@e.u-tokyo.ac.jp.

Manuscript received December, 2008; final revision received July, 2011.