Sequential Estimation of Structural Models with a Fixed Point Constraint*

Hiroyuki Kasahara
Department of Economics
University of Western Ontario
hkasahar@uwo.ca

Katsumi Shimotsu
Department of Economics
Queen's University
shimotsu@econ.queensu.ca

December 2, 2008

Abstract

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 the conditions under which the nested pseudo-likelihood (NPL) algorithm achieves convergence and derive its convergence rate. We find that the NPL algorithm may not necessarily converge when the fixed point mapping does not have a local contraction property. To address the issue of non-convergence, we propose alternative sequential estimation procedures that can achieve convergence even when the NPL algorithm does not. Upon convergence, some of our proposed estimation algorithms produce more efficient estimators than the NPL estimator.

Keywords: contraction, dynamic games, nested pseudo likelihood, recursive projection method.
JEL Classification Numbers: C13, C14, C63.

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 $P = \Psi(\theta, P)$ in the space of probability distributions:

$$\max_{\theta \in \Theta} \quad Q_n(P) \quad \text{s.t.} \quad P = \Psi(\theta, P). \tag{1}$$

*We are grateful to Victor Aguirregabiria, David Byrne, Kenneth Judd, Vadim Marmer, Lealand Morin, Whitney Newey, and seminar participants at Far Eastern Summer Meeting of the Econometric Society, New York Camp Econometrics, North American Summer Meeting of the Econometric Society, Vienna Macroeconomic Workshop, University of British Columbia, University of Michigan, Hitotsubashi University, Johns Hopkins University, University of Tokyo, University of Western Ontario, Yale University, and Yokohama National University for helpful comments. The authors thank the SSHRC for financial support.

For example, if $P = \{P(a|x)\}$ is the conditional choice probabilities, and the sample data are $\{a_i, x_i\}_{i=1}^n$, then setting $Q_n(P) = n^{-1} \sum_{i=1}^n \ln P(a_i|x_i)$ gives the maximum likelihood estimator, whereas setting $Q_n(P) = -\left[n^{-1} \sum_{i=1}^n g(a_i, x_i; P)\right]' W \left[n^{-1} \sum_{i=1}^n g(a_i, x_i; P)\right]$ gives the generalized method of moments estimator under the moment condition $E[g(a_i, x_i; P^0)] = 0$, where W is a weighting matrix and P^0 is the true conditional choice probabilities.

The fixed point constraint $P = \Psi(\theta, P)$ in (1) summarizes the set of structural restrictions of the model that is parametrized by a finite vector $\theta \in \Theta$.[1] The sample data are generated from a fixed point of the operator $\Psi(\theta, \cdot)$ evaluated at the true parameter θ^0. Examples of the operator $\Psi(\theta, \cdot)$ 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, 2008a), the operator defined by the best response function of a game (e.g., Aguirregabiria and Mira, 2007; Pakes, Ostrovsky and Berry, 2007; Pesendorfer and Schmidt-Dengler, 2008), and the operator to define the fixed point problem for a recursive competitive equilibrium in dynamic macroeconomic models (e.g., Aiyagari, 1994; Krusell and Smith, 1998).

In principle, we may estimate the parameter θ in (1) by the nested fixed point algorithm (Rust, 1987), which repeatedly solves the fixed point P_θ of $P = \Psi(\theta, P)$ at each parameter value to maximize the objective function $Q_n(P_\theta)$ with respect to θ. The major practical obstacle of applying such an estimation procedure lies in the computational burden of solving the fixed point problem for a given parameter.

To reduce the computational burden, Hotz and Miller (1993) developed a simpler two-step estimator that does not require solving the fixed point problem for each trial value of the parameter. A number of recent papers in empirical industrial organization build on the idea of Hotz and Miller (1993) to 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 and Hong, 2006). These two-step estimators may suffer from substantial finite sample bias, however, when the choice probabilities are poorly estimated in the first step.

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*. Starting from an initial estimate $\tilde{P}_0$, the NPL algorithm iterates the following steps until $j = k$:

Step 1: Given $\tilde{P}_{j-1}$, update θ by $\tilde{\theta}_j = \arg\max_{\theta \in \Theta} n^{-1} \sum_{i=1}^n \ln[\Psi(\theta, \tilde{P}_{j-1})](a_i|x_i)$.

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

[1] In applications, many fixed point problems can be reformulated in terms of the space of probability distributions. For example, the restrictions of a dynamic programming model are often formulated as a fixed point problem in the value function space (i.e., Bellman equation), but we may reformulate it as a fixed point problem in the space of probability distributions using the policy iteration operator.

The estimator $\tilde{\theta}_1$ is a version of Hotz and Miller's two-step estimator, called the pseudo maximum likelihood (PML) estimator. AM07 showed that their recursive method can be applied to models with unobserved heterogeneity in the context of dynamic games, and the limit of a sequence of estimators generated by the NPL algorithm is more efficient than the two-step estimators *if convergence is achieved*.[2]

While the NPL algorithm provides an attractive apparatus for empirical researchers, little is known about its convergence properties. AM07 have obtained convergence in their simulations and illustrate that the limiting estimator performs very well relative to the two-step PML estimator. However, they neither provide the conditions under which the NPL algorithm converges nor analyze how fast the convergence occurs. On the other hand, Pesendorfer and Schmidt-Dengler (2008) provided simulation evidence that the NPL algorithm may not necessarily converge. Collard-Wexler (2006) used the NPL algorithm to estimate a model of entry and exit for the ready-mix concrete industry and found that $\tilde{P}_j$'s "cycle around several values without converging." In view of this mixed evidence and its practical importance, it is imperative that we understand the convergence properties of the NPL algorithm.

In the first of our two main contributions, this paper derives the condition under which the NPL algorithm converges. 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 operator achieves contraction, the closer the value obtained after one iteration is to the fixed point, and, therefore, we expect that the NPL algorithm works well if the operator has a good contraction property. We show that the NPL algorithm has a good contraction property if the modulus of the dominant eigenvalue of the Jacobian matrix $\partial\Psi(\theta, P)/\partial P$ evaluated at the fixed point P_θ is sufficiently smaller than 1.

As AM07 (p. 19) recognized, the possibility of non-convergence of the NPL algorithm is a concern. Using the dynamic game model of AM07, we find in our simulations that, when the degree of strategic substitutability is high, the smallest eigenvalue of the Jacobian matrix of the policy iteration mapping is less than -1, and the NPL algorithm fails to converge. In such cases, various two-step estimators can be used, but they may suffer from a large finite sample bias.

As our second contribution, we propose alternative sequential algorithms that are implementable even when the original NPL algorithm does not converge. The first estimator replaces the fixed point mapping $\Psi(\theta, P)$ in the NPL algorithm with $\Lambda(\theta, P) = [\Psi(\theta, P)]^\alpha P^{1-\alpha}$, which shares the same fixed point as Ψ. With an appropriate choice of α and under some conditions on Ψ, the mapping Λ has a better contraction property than Ψ.

The second algorithm requires more computation than the first algorithm but converges under general conditions. It builds upon the idea of the Recursive Projection Method (henceforth

[2]Two-step estimators can also be applied to models with unobserved heterogeneity when an initial consistent estimator of the type-specific conditional choice probabilities are available. Kasahara and Shimotsu (2006, 2008b) derived sufficient conditions for nonparametric identification of a finite mixture model of dynamic discrete choices and developed a series logit estimator which can be used as a consistent initial estimator for two-step estimators.

RPM) of Shroff and Keller (1993). The divergence of the fixed point mapping Ψ is often caused by a small number of eigenvalues of $\partial\Psi(\theta, P_\theta)/\partial P$ lying outside the unit circle. The key idea behind the RPM is to find the eigenvectors corresponding to the unstable modes and to decompose the space into the unstable subspace and its orthogonal complement. Then, it modifies the fixed point mapping Ψ by taking a Newton step on the unstable subspace while using the original fixed point iteration on the stable subspace. The modified mapping is contractive.

The third estimator uses a pseudo-likelihood objective function that is defined in terms of multiple iterations of the mapping as opposed to one iteration. Since such a modification increases computational cost substantially, we introduce an approximation method that requires evaluating the mapping and its Jacobian with respect to the parameter θ only once outside of the optimization routine. This algorithm converges faster than the original NPL algorithm and, upon convergence, the proposed estimator is more efficient than the estimator generated by the NPL algorithm.

The fourth algorithm we propose directly approximates a fixed point of the mapping but with additional computational cost. This sequential algorithm has an advantage over others in that it generates a sequence of estimators that approaches the maximum likelihood estimator (henceforth MLE) and, upon convergence, we obtain the MLE which is more efficient than the other proposed estimators.

Recently, Su and Judd (2008) advocate numerically solving a constrained optimization problem for estimating a structural model using a large-scale, state-of-the-art computing facility available via the internet. We do not know, however, how their method performs when it is applied to models with a very large state space, such as the models of dynamic games of AM07.

The rest of the paper is organized as follows. Section 2 introduces a class of models with fixed point constraints. Section 3 establishes the convergence properties of the NPL algorithm. In Section 4, we develop alternative sequential algorithms. Section 5 reports some simulation results. Section 6 concludes the paper.

2 Maximum likelihood estimation of models with a fixed point constraint

We consider a class of parametric discrete choice models in which restrictions are characterized by fixed point problems. Let $a_i \in A$ and $x_i \in X$ denote the choice variable and the conditioning variable, respectively. Upon estimating such models, researchers may consider the (conditional) MLE with a fixed point constraint:

$$\hat{\theta}_{MLE} = \arg\max_{\theta\in\Theta} \left\{ \max_{P\in\mathcal{M}_\theta} \; n^{-1} \sum_{i=1}^{n} \ln P(a_i|x_i) \right\}, \tag{2}$$

where $P(a_i|x_i)$ denotes the conditional choice probability of the ith observation, $P = \{P(a|x) : (a,x) \in A \times X\}$, and

$$\mathcal{M}_\theta \equiv \{P \in B_P : \ P = \Psi(\theta, P)\} \tag{3}$$

is the set of fixed points of $\Psi(\theta, \cdot)$ given the value of $\theta \in \Theta \subset \mathbb{R}^K$. Here, B_P represents the space of conditional choice probabilities while Θ is the set of possible parameter values. The model space—the set of conditional choice probabilities that are consistent with the parametric fixed point restrictions—is then defined as a union of $\mathcal{M}_\theta$ over Θ: $\mathcal{M} \equiv \cup_{\theta\in\Theta}\mathcal{M}_\theta = \{P \in B_P : \ P = \Psi(\theta, P), \ \theta \in \Theta\}$. The data is generated from the population conditional probability, denoted by P^0, which belongs to the model space $\mathcal{M}$, i.e., $P^0 \in \mathcal{M}$.

The fixed point constraint $P = \Psi(\theta, P)$ in (3) summarizes the restrictions of the model that is parametrized with a K-dimensional vector θ. For each θ, the operator $\Psi(\cdot, \theta)$ maps the space of conditional choice probabilities into itself. The true conditional choice probability P^0 is a fixed point of the operator $\Psi(\cdot, \theta)$ evaluated at the true parameter value θ^0.

The computation of the MLE in (2) requires repeatedly solving all the fixed points of $P = \Psi(\theta, P)$ at each parameter value to maximize the objective function with respect to θ. If evaluating the mapping Ψ is costly, the MLE could be extremely computationally intensive. Further, when there are multiple fixed points, finding all of the fixed points of $P = \Psi(\theta, P)$ may be infeasible. One of the major econometric issues in estimating models with a fixed point constraint is to develop an estimator that is computationally simple and has good finite sample properties as an alternative to the MLE.

3 The nested pseudo likelihood (NPL) algorithm

3.1 Asymptotic properties of the NPL estimator

This section briefly reviews the properties of the two-step pseudo maximum likelihood (PML) estimator and the estimator generated by the nested pseudo likelihood (NPL) algorithm as discussed in Aguirregabiria and Mira (2002, 2007). They are feasible alternatives to the MLE.

We assume that the support of (a_i, x_i) is finite, $A \times X = \{a^1, a^2, \ldots, a^{|A|}\} \times \{x^1, x^2, \ldots, x^{|X|}\}$. Accordingly, P is represented by an $L \times 1$ vector, where $L = |A||X|$. Given θ, the Jacobian $\nabla_{P'}\Psi(\theta, P)$ is an $L \times L$ matrix, where $\nabla_{P'} \equiv (\partial/\partial P')$. Define $\Psi_P \equiv \nabla_{P'}\Psi(\theta^0, P^0)$ and $\Psi_\theta \equiv \nabla_{\theta'}\Psi(\theta^0, P^0)$. Let $\nabla^{(s)} f$ denote the sth order derivative of a function f with respect to all of its parameters. Let $\mathcal{N}$ denote a closed neighborhood of (θ^0, P^0), and let $\mathcal{N}_{\theta^0}$ denote a closed neighborhood of θ^0.

We collect the assumptions employed in AM07. As in AM07, define $Q_0(\theta, P) \equiv E \ln \Psi(\theta, P)(a_i|x_i)$, $\tilde{\theta}_0(P) \equiv \arg\max_{\theta\in\Theta} Q_0(\theta, P)$, and $\phi_0(P) \equiv \Psi(\tilde{\theta}_0(P), P)$. Define the set of population NPL fixed points as $\mathcal{Y}_0 \equiv \{(\theta, P) \in \Theta \times B_P : \theta = \tilde{\theta}_0(P) \text{ and } P = \phi_0(P)\}$. See AM07 for details.

Assumption 1 *(a) The observations* $\{a_i, x_i : i = 1, \ldots, n\}$ *are independent and identically distributed, and* $dF(x) > 0$ *for any* $x \in X$*, where* $F(x)$ *is the distribution function of* x_i*. (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 *and* B_P *are compact. (e) There is a unique* $\theta^0 \in int(\Theta)$ *such that* $P^0 = \Psi(\theta^0, P^0)$*. (f) For any* $\theta \neq \theta^0$ *and* P *that solves* $P = \Psi(\theta, P)$*, it is the case that* $P \neq P^0$*. (g)* (θ^0, P^0) *is an isolated population NPL fixed point. (h)* $\tilde{\theta}_0(P)$ *is a single-valued and continuous function of* P *in a neighborhood of* P^0*. (i) the operator* $\phi_0(P) - P$ *has a nonsingular Jacobian matrix at* P^0*.*

Assumption 1(b)(c) implies that $\max_{(a,x)\in A\times X} \sup_{(\theta,P)\in\Theta\times B_P} ||\nabla^{(2)} \ln \Psi(\theta, P)(a|x)|| < \infty$ and hence $E \sup_{(\theta,P)\in\Theta\times B_P} ||\nabla^{(2)} \ln \Psi(\theta, P)(a_i|x_i)||^r < \infty$ for any positive integer r. Assumption 1(h) corresponds to assumption (iv) in Proposition 2 of AM07. A sufficient condition for Assumption 1(h), which holds in a class of models that AM07 estimated, is that Q_0 is globally concave in θ in a neighborhood of P^0 and $\nabla_{\theta\theta'} Q_0(\theta, P^0)$ is a nonsingular matrix.

Define $\Omega_{\theta\theta} \equiv E[\nabla_\theta \ln \Psi(\theta^0, P^0)(a_i|x_i) \nabla_{\theta'} \ln \Psi(\theta^0, P^0)(a_i|x_i)]$, and $\Omega_{\theta P} \equiv E[\nabla_\theta \ln \Psi(\theta^0, P^0)(a_i|x_i) \times \nabla_{P'} \ln \Psi(\theta^0, P^0)(a_i|x_i)]$. The two-step PML estimator is $\hat{\theta}_{PML} = \arg\max_{\theta\in\Theta} n^{-1} \sum_{i=1}^n \ln \Psi(\theta, \hat{P}_0)(a_i|x_i)$, where $\hat{P}_0$ is an initial consistent estimator of P^0. Proposition 1 of AM07 showed that the two-step PML estimator is consistent under Assumption 1, and, when $\hat{P}_0$ satisfies $\sqrt{n}(\hat{P}_0 - P^0) \to_d N(0, \Sigma)$, the estimator is asymptotically normal with asymptotic variance $V_{PML} = (\Omega_{\theta\theta})^{-1} + (\Omega_{\theta\theta})^{-1}\Omega_{\theta P}\Sigma(\Omega_{\theta P})'(\Omega_{\theta\theta})^{-1}$. The second term of the variance expression, $(\Omega_{\theta\theta})^{-1}\Omega_{\theta P}\Sigma(\Omega_{\theta P})'(\Omega_{\theta\theta})^{-1}$, captures the effect of the first step estimator $\hat{P}_0$ on $\hat{\theta}_{PML}$, and the two-step PML estimator may perform poorly when $\hat{P}_0$ is imprecisely estimated.

As discussed in the introduction, Aguirregabiria and Mira (2002, 2007) developed a recursive extension of the two-step PML estimator, called the NPL algorithm. Starting from an initial estimator of P^0, their algorithm generates a sequence of estimators $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^k$. If this sequence converges, its limit satisfies the following conditions:

$$\check{\theta} = \arg\max_{\theta\in\Theta} n^{-1} \sum_{i=1}^n \ln \Psi(\theta, \check{P})(a_i|x_i) \quad \text{and} \quad \check{P} = \Psi(\check{\theta}, \check{P}). \tag{4}$$

Any pair $(\check{\theta}, \check{P})$ that satisfies these two conditions in (4) is called an *NPL fixed point*. The *NPL estimator*, denoted by $(\hat{\theta}_{NPL}, \hat{P}_{NPL})$, is defined as the NPL fixed point with the highest value of the pseudo likelihood among all the NPL fixed points.

Proposition 2 of Aguirregabiria and Mira (2007) established the consistency of $\hat{\theta}_{NPL}$ under Assumption 1 and derived its asymptotic distribution: $\sqrt{n}(\hat{\theta}_{NPL} - \theta^0) \to_d N(0, V_{NPL})$, where $V_{NPL} = [\Omega_{\theta\theta} + \Omega_{\theta P}(I - \Psi_P)^{-1}\Psi_\theta]^{-1}\Omega_{\theta\theta}\{[\Omega_{\theta\theta} + \Omega_{\theta P}(I - \Psi_P)^{-1}\Psi_\theta]^{-1}\}'$. The NPL estimator does not depend on the initial estimator of P^0 and reduces the finite small sample relative to the PML estimator especially when the initial estimator of P^0 is imprecise.

While AM07 have obtained convergence in their simulations and show that the NPL esti-

mator substantially outperforms the PML estimator, they neither provide the conditions under which the NPL algorithm converges nor analyze how fast the convergence occurs. On the other hand, some other studies find potential problems with the convergence of the NPL algorithm (see Pesendorfer and Schmidt-Dengler, 2008; Collard-Wexler, 2006). To date, little is known about the convergence properties of the NPL algorithm.

3.2 Convergence properties of the NPL algorithm

We now analyze the conditions under which the NPL algorithm achieves convergence and derive its convergence rate when the algorithm is started from an initial consistent estimate of P^0. First, we state the regularity conditions. For matrix and nonnegative scalar sequences of random variables $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$, respectively, we write $X_n = O_p(Y_n)(o_p(Y_n))$ if $||X_n|| \leq CY_n$ for some (all) $C > 0$ with probability arbitrarily close to one for sufficiently large n.

Assumption 2 *Assumption 1 holds. Further,* $\tilde{P}_0 - P^0 = o_p(1)$, $\Psi(\theta, P)$ *is three times continuously differentiable, and* $\Omega_{\theta\theta}$ *is nonsingular.*

Define $f_x(x^s) \equiv \Pr(x = x^s)$ for $s = 1, \ldots, |X|$, and let f_x be an $L \times 1$ vector of $\Pr(x = x^s)$ whose elements are arranged conformably with $P_{\theta^0}(a^j|x^s)$. Let $\Delta_P \equiv diag(P^0)^{-1} diag(f_x)$. With this notation, we may write $\Omega_{\theta\theta} = \Psi_\theta' \Delta_P \Psi_\theta$ and $\Omega_{\theta P} = \Psi_\theta' \Delta_P \Psi_P$. The following lemma states the local convergence rate of the NPL algorithm and is one of the main results of this paper.

Lemma 1 *Suppose Assumption 2 holds. Then, for* $j = 1, \ldots, k$,

$$\begin{aligned} \tilde{\theta}_j - \hat{\theta}_{NPL} &= O_p(||\tilde{P}_{j-1} - \hat{P}_{NPL}||), \\ \tilde{P}_j - \hat{P}_{NPL} &= M_{\Psi_\theta} \Psi_P (\tilde{P}_{j-1} - \hat{P}_{NPL}) + O_p(n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{NPL}|| + ||\tilde{P}_{j-1} - \hat{P}_{NPL}||^2), \end{aligned}$$

where $M_{\Psi_\theta} \equiv I - \Psi_\theta (\Psi_\theta' \Delta_P \Psi_\theta)^{-1} \Psi_\theta' \Delta_P$.

Write the updating equation of $\tilde{P}_j$ as $\tilde{P}_j - \hat{P}_{NPL} = [M_{\Psi_\theta}\Psi_P + O_p(n^{-1/2} + ||\tilde{P}_{j-1} - \hat{P}_{NPL}||)](\tilde{P}_{j-1} - \hat{P}_{NPL})$, then recursive substitution gives $\tilde{P}_k - \hat{P}_{NPL} = (M_{\Psi_\theta}\Psi_P + o_p(1))^k (\tilde{P}_0 - \hat{P}_{NPL})$. If all the eigenvalues of $M_{\Psi_\theta}\Psi_P$ are inside the unit circle, then $(M_{\Psi_\theta}\Psi_P)^k \to 0$ as $k \to \infty$, and iterations move $\tilde{P}_j$ toward $\hat{P}_{NPL}$. Consequently, by choosing k sufficiently large, $(\tilde{\theta}_k, \tilde{P}_k)$ becomes arbitrary close to $(\hat{\theta}_{NPL}, \hat{P}_{NPL})$. In contrast, if some eigenvalues of $M_{\Psi_\theta}\Psi_P$ are outside the unit circle, then iterations move some elements of $\tilde{P}_j$ further away from $\hat{P}_{NPL}$, and iterations may not converge even when the initial estimate $\tilde{P}_0$ is in a neighborhood of $\hat{P}_{NPL}$. As we discuss in the next section, the convergence of $(M_{\Psi_\theta}\Psi_P)^k$ is primarily determined by the dominant eigenvalues of Ψ_P. If all the eigenvalues of Ψ_P are sufficiently smaller than 1 in absolute value, then $(M_{\Psi_\theta}\Psi_P)^k \to 0$ as $k \to \infty$.

Remark 1 *When* $\Psi_P = 0$, *the convergence rate is faster than linear:* $\tilde{P}_j - \hat{P}_{NPL} = O_p(n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{NPL}|| + ||\tilde{P}_{j-1} - \hat{P}_{NPL}||^2)$ *(cf. Kasahara and Shimotsu, 2008a).*

3.3 Convergence of $(M_{\Psi_\theta}\Psi_P)^k$

Define the spectral radius of A as $\rho(A) \equiv \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$. Then $A^k \to 0$ as $k \to \infty$ if and only if $\rho(A) < 1$ (Horn and Johnson, 1985, Theorem 5.6.12). Hence, the NPL algorithm converges if and only if $\rho(M_{\Psi_\theta}\Psi_P) < 1$. Because Ψ_P is often closely related to the property of the economic model, we want to find a bound of $\rho(M_{\Psi_\theta}\Psi_P)$ in terms of $\rho(\Psi_P)$.[3] In the following, we give two discussions on the relation between $\rho(M_{\Psi_\theta}\Psi_P)$ and $\rho(\Psi_P)$.[4]

3.3.1 Projection by M_{Ψ_θ} and eigenvalues of $M_{\Psi_\theta}\Psi_P$

Define $P_{\Psi_\theta} \equiv \Psi_\theta(\Psi_\theta'\Delta_P\Psi_\theta)^{-1}\Psi_\theta'\Delta_P$. P_{Ψ_θ} is a GLS projection matrix, whereas $M_{\Psi_\theta} = I - P_{\Psi_\theta}$ is the projection matrix that generates the "residuals". Since P_{Ψ_θ} is a projection matrix, we may decompose any $L-$vector x into two: $x = x_1 + x_2$, where $x_1 = P_{\Psi_\theta}x \in \mathcal{S}(\Psi_\theta)$ (the column space of Ψ_θ) and $x_2 = (I - P_{\Psi_\theta})x = M_{\Psi_\theta}x \in \mathcal{S}^\perp(\Delta_P\Psi_\theta)$ (the orthogonal complement of $\Delta_P\Psi_\theta$).

Suppose y is an eigenvector of Ψ_P with non-zero eigenvalue ν so that $\Psi_P y = \nu y$ and $\nu \neq 0$. Consider two extreme cases. First, suppose the GLS regression of y on Ψ_θ gives no fit. In this case, $M_{\Psi_\theta}\Psi_P y = \nu y$, and $M_{\Psi_\theta}\Psi_P$ and Ψ_P share the same eigenvector y with eigenvalue ν. Second, suppose the GLS regression of y on Ψ_θ gives a perfect fit. In this case, $M_{\Psi_\theta}\Psi_P y = 0$, and y is an eigenvector of $M_{\Psi_\theta}\Psi_P$ with eigenvalue 0.

Now we place the above discussion in the context of our model. Recall that y is an $L \times 1$ vector and Ψ_θ is a $K \times L$ matrix, and typically $L \gg K$ because the dimension of the state variable is much larger than the number of parameters. Then, for many y, regressing y on K regressors gives a poor fit, and the eigenvalues of Ψ_P and $M_{\Psi_\theta}\Psi_P$ are likely to be close. For some y, we may have a good fit, so the eigenvalue of $M_{\Psi_\theta}\Psi_P$ associated with such a y is close to zero and is not likely to be the dominant eigenvalue. Hence, we expect that the dominant eigenvalues of Ψ_P and $M_{\Psi_\theta}\Psi_P$ are close to each other. In our simulation based on a model of a dynamic game with $L = 72$ and $K = 2$, we find either one of the above two cases hold for most of the eigenvectors, and the spectral radius of $M_{\Psi_\theta}\Psi_P$ is very similar to the spectral radius of Ψ_P (see Table 1).

3.3.2 The case when Ψ_P is diagonalizable

We can obtain a bound of $\rho(M_{\Psi_\theta}\Psi_P)$ if we assume Ψ_P is diagonalizable, i.e., $\Psi_P = SDS^{-1}$ for a diagonal matrix D. A matrix A is diagonalizable if all the eigenvectors are linearly independent (Horn and Johnson, 1985, Theorem 1.3.7). A sufficient condition for the diagonalizability of A is that the eigenvalues of A are distinct (Horn and Johnson, 1985, Theorem 1.3.9). Although

[3] The contraction property of Ψ may or may not be related to the stability of equilibria in the economic model. Given a model, there are often multiple ways of formulating a fixed point mapping (e.g., Hotz and Miller, 1993; Arcidiacono and Miller, 2008) and its contraction property depends on which mapping a researcher chooses.

[4] The spectral radius is not submultiplicative; i.e., $\rho(AB) > \rho(A)\rho(B)$ is possible.

economic models do not give implications for the diagonalizability of Ψ_P, we expect that A is diagonalizable in some, and possibly many, cases.

For a matrix A, let $||A||_s$ denote its spectral norm: $||A||_s \equiv \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } A'A\}$, which satisfies $||AB||_s \leq ||A||_s||B||_s$. $\rho(\cdot)$ and $||\cdot||_s$ satisfies $\rho(S^{-1}AS) = \rho(A)$, $\rho(A) \leq ||A||_s$, and $||D||_s = \rho(D)$ if D is diagonal. It follows that, if Ψ_P is diagonalizable, $\rho(M_{\Psi_\theta}\Psi_P) = \rho(M_{\Psi_\theta}SDS^{-1}) = \rho(S^{-1}M_{\Psi_\theta}SD) \leq ||S^{-1}M_{\Psi_\theta}S||_s||D||_s = ||S^{-1}M_{\Psi_\theta}S||_s\rho(\Psi_P)$. Consequently, $(M_{\Psi_\theta}\Psi_P)^k$ converges to 0 if Ψ_P is diagonalizable and $\rho(\Psi_P)$ is sufficiently smaller than 1.

4 Alternative sequential likelihood-based estimators

When $\Psi(\theta, P)$ is not a contraction in a neighborhood of (θ^0, P^0), the NPL algorithm has a convergence problem and may not be implemented. This section discusses alternative estimation algorithms that are implementable even when the NPL algorithm encounters a convergence problem. Some of our proposed algorithms produce more efficient estimators than the NPL estimator.

4.1 Locally contractive mapping with the relaxation method

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

$$[\Lambda(\theta, P)](a|x) \equiv \{[\Psi(\theta, P)](a|x)\}^\alpha P(a|x)^{1-\alpha}, \tag{5}$$

for all $(a, x) \in A \times X$, where $\alpha \in [0, 1]$. This is called the relaxation method in numerical analysis.[5] Since P is a fixed point of $\Psi(\theta, P)$ if and only if it is a fixed point of $\Lambda(\theta, P)$, the fixed points of $\Psi(\theta, P)$ is obtained by solving the fixed points of $\Lambda(\theta, P)$.

Denote the largest and the smallest eigenvalues of Ψ_P by $\lambda_{\max}$ and $\lambda_{\min}$, respectively. As discussed in Judd (1998, pp. 78-80), when $\lambda_{\max} < 1$, we may choose the value of α so that $\Lambda(\theta, P)$ becomes locally contractive even when $\Psi(\theta, P)$ is not locally contractive. Define $\Lambda_P \equiv \nabla_{P'}\Lambda(\theta^0, P^0)$, and let α^* denote the value of α that minimizes the spectral radius of Λ_P.

Proposition 1 *If $\lambda_{\max} \geq 1 \geq \lambda_{\min}$, then there is no value of α such that $\rho(\Lambda_P)$ is less than 1. If $\lambda_{\max} < 1$, then $\alpha^* = 2/(2 - \lambda_{\max} - \lambda_{\min})$ and $\rho(\Lambda_P) = (\lambda_{\max} - \lambda_{\min})/(2 - \lambda_{\max} - \lambda_{\min}) < 1$.*

Consider the NPL algorithm that uses $\Lambda(\theta, P)$ in place of $\Psi(\theta, P)$. When the condition that $\lambda_{\max} < 1$ is satisfied, the NPL algorithm with $\Lambda(\theta, P)$ may converge even if the NPL algorithm with $\Psi(\theta, P)$ does not converge. Since $\ln \Lambda(\theta, P) = \alpha \ln \Psi(\theta, P) + (1 - \alpha) \ln P$, the objective function of the NPL estimator with Ψ and that of the NPL estimator with Λ are maximized at

[5] Başar (1987) and Krawczyk and Uryasev (2000) apply the relaxation method to find a Nash equilibrium of a game. Ljungqvist and Sargent (2004, p. 574) also suggest using the relaxation method to solve the fixed point problem for the model of Aiyagari (1994).

the same value of θ for a given P. The NPL estimator with Ψ and the NPL estimator with Λ are, therefore, numerically equivalent. The advantage of this method is its simplicity. Once an appropriate value of α is determined, it achieves convergence under weaker conditions than the original NPL algorithm without adding computational burden.[6]

4.2 Recursive Projection Method

In this subsection, we construct a mapping that has a better local contraction property than Ψ building upon the Recursive Projection Method (RPM) of Shroff and Keller (1993) (henceforth SK).

First, fix θ. If some eigenvalues of $\nabla_{P'}\Psi(\theta, P_\theta)$ are outside the unit circle, the iteration $P_j = \Psi(P_{j-1}, \theta)$ does not converge to P_θ. Suppose that a small number, m, of the eigenvalues of $\nabla_{P'}\Psi(\theta, P_\theta)$ are larger than $\delta \in (0,1)$ in absolute value:

$$|\lambda_1| \geq \cdots \geq |\lambda_m| > \delta \geq |\lambda_{m+1}| \geq \cdots \geq |\lambda_L|. \tag{6}$$

Define $\mathbb{P} \subseteq \mathbb{R}^L$ as the space spanned by the eigenvectors of $\nabla_{P'}\Psi(\theta, P_\theta)$ associated with $\{\lambda_k\}_{k=1}^m$, and let $\mathbb{Q} \equiv \mathbb{R}^L - \mathbb{P}$ be the orthogonal complement of $\mathbb{P}$. Let Π_θ denote the orthogonal projector from $\mathbb{R}^L$ on $\mathbb{P}$. We may write $\Pi_\theta = Z_\theta Z_\theta'$, where $Z_\theta \in \mathbb{R}^{L\times m}$ is an orthonormal basis of $\mathbb{P}$. Then, for each $P \in \mathbb{R}^L$, we have the unique decomposition $P = u + v$, where $u \equiv \Pi_\theta P \in \mathbb{P}$ and $v \equiv (I - \Pi_\theta)P \in \mathbb{Q}$.

Now apply Π_θ and $I - \Pi_\theta$ to $P = \Psi(\theta, P)$, and decompose the system as follows:

$$\begin{aligned} u &= f(u, v, \theta) \equiv \Pi_\theta \Psi(\theta, P), \\ v &= g(u, v, \theta) \equiv (I - \Pi_\theta)\Psi(\theta, P). \end{aligned}$$

For a given P_{j-1}, decompose it into $u_{j-1} = \Pi_\theta P_{j-1}$ and $v_{j-1} = (I - \Pi_\theta)P_{j-1}$. Since $g(u, v, \theta)$ is contractive in v (see Lemma 2.10 of SK), we can update v_{j-1} by the recursion $v_j = g(u, v_{j-1}, \theta)$. On the other hand, when the dominant eigenvalue of Ψ_P is outside the unit circle, the recursion $u_j = f(u_{j-1}, v, \theta)$ cannot be used to update u_{j-1} because $f(u, v, \theta)$ is not a contraction in u. Instead, the RPM performs a single Newton step on the system $u = f(u, v, \theta)$, leading to the following updating procedure:

$$\begin{aligned} u_j &= u_{j-1} + (I - \Pi_\theta \nabla_{P'}\Psi(\theta, P_{j-1})\Pi_\theta)^{-1}(f(u_{j-1}, v_{j-1}, \theta) - u_{j-1}) \equiv h(u_{j-1}, v_{j-1}, \theta), \\ v_j &= g(u_{j-1}, v_{j-1}, \theta). \end{aligned} \tag{7}$$

[6] We may estimate $\alpha^* = 2/(2 - \lambda_{\max} - \lambda_{\min})$, by first applying the PML estimator and then evaluating the eigenvalues of $\nabla_{P'}\Psi(\hat{\theta}_{PML}, \hat{P}_0)$, where $\hat{P}_0$ is an initial consistent estimator. Alternatively, we may simulate a sequence $\{P^j\}_{j=0}^k$ by iterating $P^j = \Psi(\hat{\theta}_{PML}, P^{j-1})$ and compute the mean of $||P^{j+1} - P^k||/||P^j - P^k||$ across $j = 1, \ldots, k-1$, which gives an estimate of the dominant eigenvalue. Repeating this procedure for different values of α, we may estimate α^* by the value of α that leads to the smallest value of the mean of $||P^{j+1} - P^k||/||P^j - P^k||$'s.

Lemma 3.11 of SK shows that the spectral radius of the Jacobian of the stabilized iteration (7) is no larger than δ, and thus the iteration $P_j = h(\Pi_\theta P_{j-1}, (I - \Pi_\theta)P_{j-1}, \theta) + g(\Pi_\theta P_{j-1}, (I - \Pi_\theta)P_{j-1}, \theta)$ is locally converging. In the following, we develop a sequential algorithm building upon the updating procedure (7) by replacing Π_θ with its consistent estimator.

Let $\Pi(\theta, P)$ be the orthogonal projector from $\mathbb{R}^L$ on the subspace of $\mathbb{R}^L$ spanned by the eigenvectors of $\nabla_{P'}\Psi(\theta, P)$ associated with its m largest (in absolute value) eigenvalues. Define $h^*(u, v, \theta)$ and $g^*(u, v, \theta)$ by replacing Π_θ in $h(u, v, \theta)$ and $g(u, v, \theta)$ with $\Pi(\theta, P)$, and define

$$\begin{aligned}
\Gamma(\theta, P) &\equiv h^*(u, v, \theta) + g^*(u, v, \theta) \\
&= \Pi(\theta, P)P + (I - \Pi(\theta, P)\nabla_{P'}\Psi(\theta, P)\Pi(\theta, P))^{-1}(\Pi(\theta, P)\Psi(\theta, P) - \Pi(\theta, P)P) \\
&\quad +(I - \Pi(\theta, P))\Psi(\theta, P) \\
&= \Psi(\theta, P) + [(I - \Pi(\theta, P)\nabla_{P'}\Psi(\theta, P)\Pi(\theta, P))^{-1} - I]\Pi(\theta, P)(\Psi(\theta, P) - P). \qquad (8)
\end{aligned}$$

P^0 is a fixed point of $\Gamma(\theta^0, \cdot)$, i.e., $P^0 = \Gamma(\theta^0, P^0)$, because all the fixed points of $\Psi(\theta, \cdot)$ are also fixed points of $\Gamma(\theta, \cdot)$. The following proposition shows two important properties of $\Gamma(\theta, P)$: local contraction and equivalence of fixed points of $\Gamma(\theta, P)$ and $\Psi(\theta, P)$.

Proposition 2 *(a) Suppose $I - \Pi(\theta, P)\nabla_{P'}\Psi(\theta, P)\Pi(\theta, P)$ is nonsingular and hence $\Gamma(\theta, P)$ is well-defined. Then $\Gamma(\theta, P)$ and $\Psi(\theta, P)$ have the same fixed points; i.e., $\Gamma(\theta, P) = P$ if and only if $\Psi(\theta, P) = P$. (b) $\rho(\nabla_{P'}\Gamma(\theta^0, P^0)) \leq \delta^0$, where δ^0 is defined by (6) in terms of the eigenvalues of $\nabla_{P'}\Psi(\theta^0, P^0)$. Hence, $\Gamma(\theta, P)$ is locally contractive.*

The matrix $I - \Pi(\theta, P)\nabla_{P'}\Psi(\theta, P)\Pi(\theta, P)$ is nonsingular if any of the eigenvalues of $\Pi(\theta, P)\nabla_{P'}\Psi(\theta, P)\Pi(\theta, P)$ is not unity.

Define an *RPM fixed point* as any pair $(\check{\theta}, \check{P})$ that satisfies $\check{\theta} = \arg\max_{\theta\in\Theta} n^{-1}\sum_{i=1}^n \ln\Gamma(\theta, \check{P})(a_i|x_i)$ and $\check{P} = \Gamma(\check{\theta}, \check{P})$. The *RPM estimator*, denoted by $(\hat{\theta}_{RPM}, \hat{P}_{RPM})$, is defined as the RPM fixed point with the highest value of the pseudo likelihood among all the RPM fixed points. The RPM estimator is consistent and asymptotically normally distributed under assumptions analogous to Assumption 1, where $\Psi(\theta, P)$ is replaced with $\Gamma(\theta, P)$. Define the *RPM algorithm* by the same sequential algorithm as the NPL algorithm except that it uses $\Gamma(\theta, P)$ in place of $\Psi(\theta, P)$. Since the mapping $\Gamma(\theta, \cdot)$ is locally contractive, the RPM algorithm will converge.

Assumption 3 *(a) Assumption 1 holds, and conditions (b)–(i) of Assumption 1 hold when $\Psi(\theta, P)$ is replaced with $\Gamma(\theta, P)$. (b) $\Gamma(\theta, P)$ is three times continuously differentiable in $\mathcal{N}$. (c) $\Omega^\Gamma_{\theta\theta} \equiv$
$E\nabla_\theta \ln\Gamma(\theta^0, P^0)(a_i|x_i)\nabla_{\theta'} \ln\Gamma(\theta^0, P^0)(a_i|x_i)$ is nonsingular. (d) $\tilde{P}_0 - P^0 = o_p(1)$, and $\tilde{\theta}_0 - \theta^0 = o_p(1)$.*

Proposition 3 *Suppose Assumption 3 holds. Suppose we obtain $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^k$ by the RPM algorithm. Then, for $j = 1, \ldots, k$, $\tilde{\theta}_j - \hat{\theta}_{RPM} = O_p(||\tilde{P}_{j-1} - \hat{P}_{RPM}||)$ and $\tilde{P}_j - \hat{P}_{RPM} = M_{\Gamma_\theta}\Gamma_P(\tilde{P}_{j-1} -$*

$\hat{P}_{RPM}) + O_p(n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{RPM}|| + ||\tilde{P}_{j-1} - \hat{P}_{RPM}||^2)$, *where* $M_{\Gamma_\theta} \equiv I - \Gamma_\theta(\Gamma_\theta'\Delta_P\Gamma_\theta)^{-1}\Gamma_\theta'\Delta_P$, $\Gamma_P \equiv \nabla_{P'}\Gamma(\theta^0, P^0)$, *and* $\Gamma_\theta \equiv \nabla_{\theta'}\Gamma(\theta^0, P^0)$.

We omit the proof of Proposition 3 because it is essentially the same as the proof of Lemma 1. Note that, from Assumption 3(a), the RPM estimator satisfies $\hat{P} - P^0 = O_p(n^{-1/2})$ and $n^{-1/2}(\hat{\theta} - \theta^0) \to_d V_{RPM}$, where $V_{RPM} = [\Omega^\Gamma_{\theta\theta} + \Omega^\Gamma_{\theta P}(I - \Gamma_P)^{-1}\Gamma_\theta]^{-1}\Omega^\Gamma_{\theta\theta}\{[\Omega^\Gamma_{\theta\theta} + \Omega^\Gamma_{\theta P}(I - \Gamma_P)^{-1}\Gamma_\theta]^{-1}\}'$.

Implementing the RPM algorithm is very costly because it requires evaluating $\Pi(\theta, P)$ and $\nabla_{P'}\Psi(\theta, P)$ for all the trial values of θ. We reduce the computational burden by evaluating $\Pi(\theta, P)$ and $\nabla_{P'}\Psi(\theta, P)$ outside the optimization routine by using a preliminary estimate of θ. This modification has only a second-order effect on the convergence of the algorithm because the derivatives of $\Gamma(\theta, P)$ with respect to $\Pi(\theta, P)$ and $\nabla_{P'}\Psi(\theta, P)$ are zero when evaluated at $P = \Psi(\theta, P)$; see the second term in (8). Let η be a preliminary estimate of θ. Replacing θ in $\Pi(\theta, P)$ and $\nabla_{P'}\Psi(\theta, P)$ with η, we define the following mapping

$$\Gamma(\theta, P, \eta) \equiv \Psi(\theta, P) + [(I - \Pi(\eta, P)\nabla_{P'}\Psi(\eta, P)\Pi(\eta, P))^{-1} - I]\Pi(\eta, P)(\Psi(\theta, P) - P).$$

Once $\Pi(\eta, P)$ and $\nabla_{P'}\Psi(\eta, P)$ are computed, a large part of computational cost of evaluating $\Gamma(\theta, P, \eta)$ comes from evaluating $\Psi(\theta, P)$, and the computational cost of evaluating $\Gamma(\theta, P, \eta)$ across different values of θ would be of a magnitude similar to that of evaluating $\Psi(\theta, P)$.

Let $(\tilde{\theta}_0, \tilde{P}_0)$ be an initial consistent estimator of (θ^0, P^0). For instance, $\tilde{\theta}_0$ can be the PML estimator. The *modified RPM algorithm* iterates the following steps until $j = k$:

Step 1: Given $(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$, update θ by $\tilde{\theta}_j = \arg\max_{\theta \in \bar{\Theta}_j} n^{-1} \sum_{i=1}^n \ln \Gamma(\theta, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(a_i|x_i)$, where $\bar{\Theta}_j \equiv \{\theta \in \Theta : \Gamma(\theta, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(a|x) \in [\epsilon, 1-\epsilon] \text{ for all } (a, x) \in A \times X\}$ for an arbitrary small $\epsilon > 0$. We impose this restriction in order to avoid computing $\ln(0)$.[7]

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

The following proposition shows that the modified RPM algorithm achieves the same convergence rate as the original RPM algorithm in the first order.

Proposition 4 *Suppose Assumption 3 holds. Suppose we obtain* $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^k$ *by the modified RPM algorithm. Then, for* $j = 1, \ldots, k$,

$$\begin{aligned}
\tilde{\theta}_j - \hat{\theta}_{RPM} &= O_p(||\tilde{P}_{j-1} - \hat{P}_{RPM}|| + n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}||^2), \\
\tilde{P}_j - \hat{P}_{RPM} &= M_{\Gamma_\theta}\Gamma_P(\tilde{P}_{j-1} - \hat{P}_{RPM}) + O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}|| \\
&\quad + ||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}||^2 + n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{RPM}|| + ||\tilde{P}_{j-1} - \hat{P}_{RPM}||^2).
\end{aligned}$$

[7] In practice, we may consider a penalized objective function by truncating $\Gamma(\theta, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ so that its value takes between ϵ and $1 - \epsilon$, and adding a penalty term that penalizes θ such that $\Gamma(\theta, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) \notin [\epsilon, 1 - \epsilon]$.

By choosing m sufficiently large, the dominant eigenvalue of Γ_P lies inside the unit circle, and the modified RPM algorithm can converge even when the NPL algorithm does not.

If an alternate preliminary consistent estimator, (θ^*, P^*), is used in forming $\Pi(\theta, P)$ and $\nabla_{P'}\Psi(\theta, P)$, it only affects the reminder terms in Proposition 4 as the following corollary shows. Therefore, if we use a root-n consistent (θ^*, P^*) to evaluate $\Pi(\theta, P)$ and $\nabla_{P'}\Psi(\theta, P)$ and keep these estimates unchanged throughout iterations, the resulting sequence of estimators is only $O_p(n^{-1})$ away from the corresponding estimators generated by the modified RPM algorithm.

Corollary 1 *Suppose Assumption 3 holds. Let (θ^*, P^*) be a consistent estimator of (θ^0, P^0), and suppose we obtain $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^k$ by the modified RPM algorithm with $\Pi(\theta^*, P^*)$ and $\nabla_{P'}\Psi(\theta^*, P^*)$ in place of $\Pi(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$ and $\nabla_{P'}\Psi(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$. Then, $\tilde{\theta}_j - \hat{\theta}_{RPM} = O_p(||\tilde{P}_{j-1} - \hat{P}_{RPM}|| + n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}||^2 + r^*_{nj})$ and $\tilde{P}_j - \hat{P}_{RPM} = M_{\Gamma_\theta}\Gamma_P(\tilde{P}_{j-1} - \hat{P}_{RPM}) + O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}_{RPM}||^2 + n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{RPM}|| + ||\tilde{P}_{j-1} - \hat{P}_{RPM}||^2 + r^*_{nj})$, where $r^*_{nj} = n^{-1/2}||\theta^* - \hat{\theta}_{RPM}|| + ||\theta^* - \hat{\theta}_{RPM}||^2 + n^{-1/2}||P^* - \hat{P}_{RPM}|| + ||P^* - \hat{P}_{RPM}||^2$.*

The supplementary appendix discusses how to implement the sequential RPM algorithm in details, including how to reduce the computational burden further by applying Corollary 1.

4.3 The q-NPL algorithm

When the spectral radius of Λ_P or Ψ_P is smaller than but close to 1, the convergence of the NPL algorithm could be very slow, and a sequence generated by the algorithm could behave erratically.[8] Furthermore, in such a case, the efficiency loss of the NPL estimator relative to that of the MLE can be substantial.

To improve the convergence of the NPL algorithm and to obtain a more efficient estimator, consider a q-fold operator of Λ as

$$\Lambda^q(\theta, P) \equiv \underbrace{\Lambda(\theta, (\Lambda(\theta, \ldots \Lambda(\theta, \Lambda}_{q \text{ times}}(\theta, P)) \ldots).$$

We may define $\Gamma^q(\theta, P)$ and $\Psi^q(\theta, P)$ analogously. Define the *q-NPL (q-RPM) algorithm* by using a q-fold operator Λ^q, Γ^q, and Ψ^q in place of Λ, Γ, or Ψ in the original NPL (RPM) algorithm. In the following, we focus on Λ^q but the same argument applies to Γ^q and Ψ^q.

If q-NPL iterations converge, its limit satisfies $\check{\theta} = \arg\max_{\theta\in\Theta} n^{-1}\sum_{i=1}^n \ln \Lambda^q(\theta, \check{P})(a_i|x_i)$ and $\check{\theta} = \Lambda^q(\check{\theta}, \check{P})$. Among the pairs $(\hat{\theta}, \hat{P})$ that satisfy these two conditions, the one that maximizes the value of the pseudo likelihood is called the *q-NPL estimator* and denoted by $(\hat{\theta}_{qNPL}, \hat{P}_{qNPL})$.

[8] As AM07 (pp. 20-21) discussed, if some eigenvalues of Λ_P or Ψ_P are equal to 1, then there could exist a continuum of NPL fixed points at (θ^0, P^0).

Since the result of Lemma 1 also applies here by replacing Ψ with Λ^q, the conditions under which the q-NPL algorithm converges is primarily determined by the spectral radius of $\Lambda^q_P \equiv \nabla_{P'}\Lambda^q(\theta^0, P^0)$. When $\rho(\Lambda_P)$ is less than 1, the q-NPL algorithm converges faster than the NPL algorithm because $\rho(\Lambda^q_P) = (\rho(\Lambda_P))^q$. Moreover, the variance of the q-NPL estimator approaches that of the MLE at the exponential rate of $(\rho(\Lambda_P))^q$ as $q \to \infty$.

Applying the q-NPL algorithm is computationally intensive because its Step 1 requires evaluating Λ^q at many different values of θ, where each evaluation of Λ^q is very costly. We reduce the computational burden by introducing a linear approximation of $\Lambda^q(\theta, P)$ around (η, P), where η is a preliminary estimate of θ: $\tilde{\Lambda}^q(\theta, P, \eta) \equiv \Lambda^q(\eta, P) + \nabla_{\theta'}\Lambda^q(\eta, P)(\theta - \eta)$.

Given a consistent estimator $(\tilde{\theta}_0, \tilde{P}_0)$, the *approximate q-NPL algorithm* iterates the following steps until $j = k$:

Step 1: Given $(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$, update θ by $\tilde{\theta}_j = \arg\max_{\theta \in \Theta^q_j} n^{-1} \sum_{i=1}^n \ln \tilde{\Lambda}^q(\theta, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(a_i|x_i)$, where $\Theta^q_j \equiv \{\theta \in \Theta : \tilde{\Lambda}^q(\theta, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(a|x) \in [\epsilon, 1-\epsilon] \text{ for all } (a,x) \in A \times X\}$ for an arbitrary small $\epsilon > 0$.

Step 2: Given $(\tilde{\theta}_j, \tilde{P}_{j-1})$, update P using the obtained estimate $\tilde{\theta}_j$ by $\tilde{P}_j = \Lambda^q(\tilde{\theta}_j, \tilde{P}_{j-1})$.

Implementing Step 1 requires evaluating $\Lambda^q(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$ and $\nabla_{\theta'}\Lambda^q(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$ only once outside of the optimization routine for θ and, thus, it involves much fewer evaluations of $\Lambda(\theta, P)$ across different values of P and θ than the original q-NPL algorithm.[9]

To establish the consistency of a sequence of estimators generated by the approximate q-NPL algorithm, we need the following assumptions.

Assumption 4 *(a) Assumption 1 holds, and conditions (b)–(i) of Assumption 1 hold when $\Psi(\theta, P)$ is replaced with $\Lambda^q(\theta, P)$. (b) $\Lambda^q(\theta, P)$ is three times continuously differentiable in $\mathcal{N}$. (c) $\Omega^q_{\theta\theta} \equiv$*
$E\nabla_\theta \ln \Lambda^q(\theta^0, P^0)(a_i|x_i)\nabla_{\theta'} \ln \Lambda^q(\theta^0, P^0)(a_i|x_i)$ is nonsingular. (d) For any $\nu \in \mathbb{R}^K$ such that $\nu \neq 0$, $\nabla_{\theta'}\Lambda^q(\theta^0, P^0)(a_i|x_i)\nu \neq 0$ with positive probability. (e) $\tilde{P}_0 - P^0 = o_p(1)$, and $\tilde{\theta}_0 - \theta^0 = o_p(1)$.

Assumption 4(d) is an identification condition for the probability limit of our objective function. It is required because we use an approximation of $\Lambda^q(\theta, P)(a|x)$ in the objective function.

Under these assumptions, we establish consistency:

Proposition 5 *Suppose that Assumption 4 holds. Suppose we obtain $\tilde{\theta}_k$ by the approximate q-NPL algorithm. Then $\tilde{\theta}_j - \theta^0 = o_p(1)$ for $j = 1, \ldots, k$.*

The following proposition establishes that the approximate q-NPL algorithm has the same convergence property as the original q-NPL algorithm.

[9] Using one-sided numerical derivatives, evaluating $\nabla_{\theta'}\Lambda^q(\tilde{\theta}_j, \tilde{P}_j)$ requires $(K+1)q$ function evaluations of $\Psi(\theta, P)$.

Assumption 5 *$\tilde{\Lambda}^q(\theta, P, \eta)$ is three times continuously differentiable in $\mathcal{N}_{\theta^0} \times \mathcal{N}$.*

Proposition 6 *Suppose Assumptions 4-5 hold. Suppose we obtain $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^k$ by the approximate q-NPL algorithm. Then, for $j = 1, \ldots, k$, $\tilde{\theta}_j - \hat{\theta}_{qNPL} = O_p(||\tilde{P}_{j-1} - \hat{P}_{qNPL}|| + n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{qNPL}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}_{qNPL}||^2)$ and $\tilde{P}_j - \hat{P}_{qNPL} = M_{\Lambda_\theta^q}\Lambda_P^q(\tilde{P}_{j-1} - \hat{P}_{qNPL}) + O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{qNPL}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}_{qNPL}||^2 + n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{qNPL}|| + ||\tilde{P}_{j-1} - \hat{P}_{qNPL}||^2)$, where $M_{\Lambda_\theta^q} \equiv I - \Lambda_\theta^q((\Lambda_\theta^q)'\Delta_P\Lambda_\theta^q)^{-1}(\Lambda_\theta^q)'\Delta_P$ with $\Lambda_\theta^q \equiv \nabla_{\theta'}\Lambda^q(\theta^0, P^0)$.*

Upon convergence, this approximate algorithm generates the q-NPL estimator, $\hat{\theta}_{qNPL}$, which is more efficient than the NPL estimator.

4.4 Approximate fixed point algorithm

It is possible to apply the idea of the approximate q-NPL algorithm to the fixed point, $P_\theta = \Psi(\theta, P_\theta)$, to approximate the MLE. From the Taylor expansion and the relation $\nabla_{\theta'}P_\theta = (I - \nabla_{P'}\Psi(\theta, P_\theta))^{-1}\nabla_{\theta'}\Psi(\theta, P_\theta)$, we can approximate P_θ as $P_\theta = P_{\theta^0} + (I - \nabla_{P'}\Psi(\theta^0, P_{\theta^0}))^{-1}\nabla_{\theta'}\Psi(\theta^0, P_{\theta^0})(\theta - \theta^0) + O(||\theta - \theta^0||^2)$, where $\nabla_{\theta'}P_{\theta^0}$ denotes the derivative of P_θ evaluated at $\theta = \theta^0$. Therefore, if we have a consistent estimate of θ^0 and P^0, we may approximate P_θ by a linear function of θ with the mappings $\nabla_{P'}\Psi(\theta, P)$ and $\nabla_{\theta'}\Psi(\theta, P)$.

We consider an estimation algorithm, called the *Approximate Fixed Point (AFXP) algorithm*, based on the following objective function: $Q_n(\theta, P, \eta) \equiv n^{-1}\sum_{i=1}^n \ln \Phi(\theta, P, \eta)(a_i|x_i)$, where

$$\Phi(\theta, P, \eta) \equiv P + (I - \nabla_{P'}\Psi(\eta, P))^{-1}\nabla_{\theta'}\Psi(\eta, P)(\theta - \eta).$$

Let $\tilde{\theta}_0$ be an initial estimator of θ^0, such as the PML estimator. The AFXP algorithm iterates the following steps until $j = k$:

Step 1: Given $\tilde{\theta}_{j-1}$, update P by solving the fixed point: $\tilde{P}_j = P_{\tilde{\theta}_{j-1}}$. If there are multiple fixed points, choose the one that maximizes the likelihood function: $\tilde{P}_j = \arg\max_{P \in \mathcal{M}_{\tilde{\theta}_{j-1}}} n^{-1}\sum_{i=1}^n \ln P(a_i|x_i)$, where $\mathcal{M}_\theta$ is defined in (3).

Step 2: Given $(\tilde{P}_j, \tilde{\theta}_{j-1})$, update θ by $\tilde{\theta}_j = \arg\max_{\theta \in \Theta_j} Q_n(\theta, \tilde{P}_j, \tilde{\theta}_{j-1})$, where $\Theta_j \equiv \{\theta \in \Theta : \Phi(\theta, \tilde{P}_j, \tilde{\theta}_{j-1})(a|x) \in [\epsilon, 1 - \epsilon]$ for all $(a, x) \in A \times X\}$ for an arbitrary small $\epsilon > 0$.

To establish the consistency of the sequential estimators generated by the AFXP algorithm, we impose the following assumptions. Assumption 6 is the standard regularity conditions for the consistency of the MLE. Assumption 7 is required for the consistency of the AFXP estimator.

Assumption 6 *(a) Θ is compact and, for any $\theta \in \Theta$, $\mathcal{M}_\theta$ is compact. (b) (a_i, x_i) for $i = 1, \ldots, M$, are independently and identically distributed, and $\Pr(x_i = x) > 0$ for any $x \in X$. (c) There is a unique $\theta^0 \in int(\Theta)$ and a unique $P_{\theta^0} \in \mathcal{M}_{\theta^0}$ such that, for any $(a, x) \in A \times$*

X, $P_{\theta^0}(a|x) = P^0(a|x)$. *(d) For any* $P_\theta \in \mathcal{M}_\theta$ *given any* $\theta \neq \theta^0$, $\Pr_{P^0}(\{(a,x) : P_\theta(a|x) \neq P^0(a|x)\}) > 0$. *(e)* $\ln P_\theta$ *is continuous in* θ. *(f)* $E \sup_{\theta\in\Theta} |\ln P_\theta(a_i|x_i)| < \infty$.

Assumption 7 *(a) For any* $\nu \in \mathbb{R}^K$ *such that* $\nu \neq 0$, $\nabla_{\theta'} P_{\theta^0}(a_i|x_i)\nu \neq 0$ *with positive probability. (b)* $\Phi(\theta, P, \eta)$ *is continuous in* $(\theta, P, \eta) \in \Theta \times \mathcal{N}$. *(c)* $E \sup_{\theta\in\Theta,(P,\eta)\in\mathcal{N}} |\ln \Phi(\theta, P, \eta)(a_i|x_i)| < \infty$.

Assumption 7(a) is similar to Assumption 4 and is an identification condition for the probability limit of our objective function. Assumption 7(b)(c) are regularity conditions required for the uniform convergence of the objective function.

Under these assumptions, the estimators generated by the AFXP algorithm are consistent:

Proposition 7 *Suppose that Assumptions 6-7 hold and* $\tilde{\theta}_0$ *is consistent. Suppose we obtain* $\tilde{\theta}_k$ *by the AFXP algorithm. Then* $\tilde{\theta}_j - \theta^0 = o_p(1)$ *for* $j = 1, \ldots, k$.

If a sequence of estimators generated by the AFXP algorithm converges, it converges to the MLE. To analyze the convergence properties of the AFXP algorithm, we introduce the following additional regularity conditions. Assumption 8(a)-(d) are required for the asymptotic normality of the MLE; see Theorem 3.3 of Newey and McFadden (1994).

Assumption 8 *(a) For* $\theta \in \mathcal{N}_{\theta^0}$, $\ln P_\theta$ *is twice continuously differentiable and* $P_\theta > 0$. *(b)* $E \sup_{\theta\in\mathcal{N}_{\theta^0}} ||\nabla_{\theta'} P_\theta(a_i|x_i)|| < \infty$, *and* $E \sup_{\theta\in\mathcal{N}_{\theta^0}} ||\nabla_{\theta\theta'} P_\theta(a_i|x_i)|| < \infty$. *(c)* $\mathcal{I}^0 \equiv E[\nabla_\theta \ln P_{\theta^0}(a_i|x_i) \times \nabla_{\theta'} \ln P_{\theta^0}(a_i|x_i)]$ *exists and is nonsingular. (d)* $E \sup_{\theta\in\mathcal{N}_{\theta^0}} ||\nabla_{\theta\theta'} \ln P_\theta(a_i|x_i)|| < \infty$. *(e)* $\Psi(\theta, P)$ *is twice continuously differentiable in* $(\theta, P) \in \mathcal{N}$. *(f)* $\Phi(\theta, P, \eta)$ *is three times continuously differentiable in* $\mathcal{N}_{\theta^0} \times \mathcal{N}$.

The following proposition establishes the convergence rate of the AFXP algorithm.

Proposition 8 *Suppose that Assumptions 6-8 hold and* $\tilde{\theta}_0$ *is consistent. Suppose we obtain* $\{\tilde{\theta}_j, \tilde{P}_j\}_{j=1}^k$ *by the AFXP algorithm. Then, for* $j = 1, \ldots, k$, $\tilde{P}_j - \hat{P}_{MLE} = O_p(||\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}||)$ *and* $\tilde{\theta}_j - \hat{\theta}_{MLE} = O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}||) + O_p(||\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}||^2)$.

Thus, the estimator generated by the AFXP algorithm is first-order equivalent to the MLE for all $k \geq 1$. This algorithm can be used to obtain the MLE because, upon convergence, its limit is identical to the MLE.

Implementing Step 1 of the AFXP algorithm may be impractical when finding all the fixed points is computationally infeasible. In such cases, we may replace the solution to the fixed point in Step 1 with its consistent estimator. Define the *q-AFXP algorithm* by the same sequential algorithm as the AFXP algorithm except that, starting from an initial consistent estimate $(\tilde{\theta}_0, \tilde{P}_0)$, Step 1 updates P by $\tilde{P}_j = \Lambda^q(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$ or $\tilde{P}_j = \Gamma^q(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$. In the following, we focus on the case in which P is updated using Λ^q but a similar argument applies to Γ^q.

The following propositions establish the consistency and the convergence properties of the estimators generated by the q-AFXP algorithm. Define a $K \times L$ matrix $\mathcal{J}$ as $\mathcal{J} \equiv E[\nabla_\theta \ln P_{\theta^0}(a_i|x_i) I(a_i|x_i)/P^0(a_i|x_i)]$, where $I(a_i|x_i)$ is the row of an $L \times L$ identity matrix that corresponds to $(a_i|x_i)$.

Proposition 9 *Suppose that Assumptions 6-7 hold and $(\tilde{\theta}_0, \tilde{P}_0)$ is consistent. Suppose we obtain $\tilde{\theta}_k$ by the q-AFXP algorithm. Then $\tilde{\theta}_j - \theta^0 = o_p(1)$ for $j = 1, \ldots, k$.*

Proposition 10 *Suppose that Assumptions 6-8 hold and $(\tilde{\theta}_0, \tilde{P}_0)$ is consistent. Suppose we obtain $\tilde{\theta}_k$ by the q-AFXP algorithm. Then, for $j = 1, \ldots, k$,*

$$\begin{aligned} \tilde{P}_j - \hat{P}_{MLE} &= \Lambda_P^q(\tilde{P}_{j-1} - \hat{P}_{MLE}) + \Lambda_\theta^q(\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}) + r_{nj}, \\ \tilde{\theta}_j - \hat{\theta}_{MLE} &= (\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}) - (\mathcal{I}^0)^{-1}\mathcal{J}(\tilde{P}_j - \hat{P}_{MLE}) + r_{nj}, \end{aligned}$$

where r_{nj} denotes a reminder term satisfying $r_{nj} = O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}_{MLE}||^2 + n^{-1/2}||\tilde{P}_{j-1} - \hat{P}_{MLE}|| + ||\tilde{P}_{j-1} - \hat{P}_{MLE}||^2)$.

Ignoring r_{nj}, arranging the two updating relations into a system of equations, solving for $\tilde{P}_j - \hat{P}_{MLE}$ and $\tilde{\theta}_j - \hat{\theta}_{MLE}$, and using $\Lambda_P^q = (\Lambda_P)^q$, $\Lambda_\theta^q = (I + \Lambda_P + \cdots + (\Lambda_P)^{q-1})\Lambda_\theta = (I - (\Lambda_P)^q)(I - \Lambda_P)^{-1}\Lambda_\theta = (I - (\Lambda_P)^q)\nabla_{\theta'} P_{\theta^0}$, and $\mathcal{J}\nabla_{\theta'} P_{\theta^0} = \mathcal{I}^0$, we obtain

$$\begin{pmatrix} \tilde{P}_j - \hat{P}_{MLE} \\ \tilde{\theta}_j - \hat{\theta}_{MLE} \end{pmatrix} = Q \begin{pmatrix} \tilde{P}_{j-1} - \hat{P}_{MLE} \\ \tilde{\theta}_{j-1} - \hat{\theta}_{MLE} \end{pmatrix}, \text{ where } Q = \begin{pmatrix} (\Lambda_P)^q & \Lambda_\theta^q \\ -(\mathcal{I}^0)^{-1}\mathcal{J}(\Lambda_P)^q & (\mathcal{I}^0)^{-1}\mathcal{J}(\Lambda_P)^q \nabla_{\theta'} P_{\theta^0} \end{pmatrix}.$$

Suppose $\rho(\Lambda_P) < 1$. Then, as q increases, $(\Lambda_P)^q$ approaches zero, and all the eigenvalues of Q approach zero. Therefore, all of the eigenvalues of Q are inside the unit circle for sufficiently large q, and iterating the q-AFXP algorithm converges to the MLE.

5 Monte Carlo experiments

We consider a dynamic game of market entry and exit. The model's setup is identical to that of Section 4 in AM07, and the reader is referred to AM07. The profit of firm i operating in market m in period t is equal to

$$\theta_{RS} \ln S_{mt} - \theta_{RN} \ln(1 + \sum_{j \neq i} a_{jmt}) - \theta_{FC,i} - \theta_{EC}(1 - a_{im,t-1}) + \epsilon_{imt}(1),$$

whereas its profit is $\epsilon_{imt}(0)$ if the firm is not operating. We assume that $\{\epsilon_{imt}(0), \epsilon_{imt}(1)\}$ follow i.i.d. type I extreme value distribution with zero mean and unit variance, and S_{mt} follows an exogenous first-order Markov process $f_S(S_{m,t+1}|S_{mt})$. We set the number of firms $N = 3$. The

state space for the market size S_{mt} is $\{2, 6, 10\}$.[10] The discount factor is set to $\beta = 0.96$. We normalize θ_{RS} to 1 and fix θ_{EC} to 1. Fixed operating costs are $\theta_{FC,1} = 1.0$, $\theta_{FC,2} = 0.9$, and $\theta_{FC,3} = 0.8$.

The value of parameter θ_{RN} determines the degree of strategic substitutabilities among firms and is the main determinant of the dominant eigenvalue of Ψ_P. We therefore vary the value of θ_{RN} to 2 and 4 across experiments and examine the performance of different estimators. As reported in Table 1, all of the eigenvalues of Ψ_P 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 estimate θ_{RS} and θ_{RN} while the other parameters are not estimated but fixed at the true values.

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, and then draw the choices at $t = 1$, $\{a_{1m1}, a_{2m1}, a_{3m1}\}$, given x_m randomly from the equilibrium choice probabilities. For $\theta_{RN} = 1$ and 2, the fixed point of $\Psi(\theta, P)$ is obtained by iterating the mapping $\Psi(\theta, P)$ starting from an initial vector of choice probabilities that are uniformly equal to 0.5. For $\theta_{RN} = 4$ and 6, the fixed point is obtained by iterating the mapping $[\Lambda(\theta, P)](a = 1|x) \equiv \{[\Psi(\theta, P)](a = 1|x)\}^{\alpha^*}\{P(a = 1|x)\}^{1-\alpha^*}$. We replicate 500 simulated samples, each of which contains $n = 500$, 2000, and 8000 observations.

As shown in Table 1, the absolute value of the dominant eigenvalue of $M_{\Psi_\theta}\Psi_P$ and $M_{\Lambda_\theta}\Lambda_P$ is similar to the corresponding eigenvalue of Ψ_P and Λ_P. Thus, in view of Lemma 1, the convergence rate of the NPL algorithm is primarily determined by the dominant eigenvalue of Ψ_P and Λ_P.

Table 2 compares the bias and the root mean squared error (RMSE) across different estimators for $\theta_{RN} = 2$ or 4. The maximum number of iterations for sequential estimators is set to $k = 50$. For $\theta_{RN} = 2$, the NPL estimator with Ψ (henceforth Ψ–NPL estimator) substantially improves the performance of the two-step PML estimator across different sample sizes, and the Ψ– and Λ–NPL estimator converge to the same estimate.

For $\theta_{RN} = 4$, however, reflecting its non-convergence, the estimator generated by 50 iterations of the NPL algorithm with Ψ (henceforth Ψ–NPL algorithm) performs substantially worse than the Λ–NPL estimator. With the sample size $n = 500$, the RMSE of the estimates of $\hat{P}$ generated by the Ψ–NPL algorithm is more than thirty times larger than those of the Λ–NPL estimator. Further, as the sample size increases from $n = 500$ to $n = 2000$, then to $n = 8000$, the RMSE of the Λ–NPL estimator decreases approximately at the rate of $n^{1/2}$, but the RMSE of the Ψ–NPL estimator decreases at a much slower rate. For $\theta_{RN} = 4$ and $n = 2000$ or 8000,

[10] The transition probability matrix of S_{mt} is given by

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

the performance of the Ψ–NPL estimator is worse than that of the PML estimator.

The fourth and the fifth rows of each panel of Table 2 report the performance of the estimator generated by the modified RPM algorithm with $\delta = 0.5$ and 0.8, respectively. See the supplementary appendix for our implementation of the modified RPM algorithm. Both estimators perform better than the Ψ–NPL estimator, especially when $\theta_{RN} = 4$, and their performance is comparable to that of the Λ–NPL estimator. Note also that the modified RPM algorithm performs better with $\delta = 0.5$ than with $\delta = 0.8$ as the former achieves faster contraction.

The sixth and the seventh rows of each panel of Table 2 report the performance of the q-NPL estimator with Λ^q and the q-AFXP estimator that uses Λ^q to update P, respectively, where q is set to 4. For both $\theta_{RN} = 2$ and $\theta_{RN} = 4$, they perform better than the Ψ– and Λ–NPL estimator, suggesting their efficiency gain over the NPL estimator.

Table 3 compares the RMSE across the estimators generated by different sequential algorithms after $j = 5, 10, \ldots, 25$ iterations with the sample size $n = 8000$. For $\theta_{RN} = 2$, the RMSE does not change after $j = 5$ iterations for any of the algorithms. Thus, they either converge or are close to convergence after 5 iterations. For $\theta_{RN} = 4$, the RMSE of the estimators generated by the NPL algorithm with Ψ increases with the number of iterations, suggesting its divergence. On the other hand, our proposed alternative algorithms are convergent.

6 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 show that, when the fixed point mapping has a local contraction property, the NPL algorithm achieves convergence in a neighborhood of the true value.

In practice, the convergence condition may be violated. In such a case, the NPL algorithm will not converge even when an initial estimate is in a small 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 Monte Carlo experiments illustrate, these alternative estimators work well even when the NPL algorithm has a convergence problem, and their performance can be substantially better than that of the two-step estimator.

In the presence of (a finite number of) multiple equilibria, the limit of a sequence of estimators generated by the NPL algorithm is still consistent if the NPL algorithm is locally converging and the initial estimator is asymptotically in a neighborhood of the true equilibrium choice probabilities. We emphasize, however, that our convergence result is local. When there are multiple NPL fixed points and the initial point is far away from the NPL estimator, there is no guarantee that the NPL algorithm converges to the NPL estimator. This is analogous to the situation often encountered by a researcher when using Newton's method to solve the optimization problem with multiple local maxima. When a reliable initial estimate is not available,

it is recommended to repeatedly apply the NPL algorithm with different initial values.

In the supplementary appendix, we also show that convergence properties similar to that of the NPL algorithm hold for models with permanent unobserved heterogeneity. Furthermore, we develop a recursive extension of two-step generalized method of moment estimators and derive its convergence properties.

7 Appendix

7.1 Proof of Lemma 1

We suppress the subscript NPL from $\hat{P}_{NPL}$ and $\hat{\theta}_{NPL}$. Define $\overline{\psi}(\theta, P) \equiv n^{-1} \sum_{i=1}^{n} \ln \Psi(\theta, P)(a_i|x_i)$. First, Proposition 1 of AM07 implies that $\tilde{\theta}_j$ is consistent if $\tilde{P}_{j-1}$ is consistent, and the continuity of $\Psi(\theta, P)$ implies $\tilde{P}_j \to_p P^0$ if $(\tilde{\theta}_j, \tilde{P}_{j-1}) \to_p (\theta^0, P^0)$. Then, since $\tilde{P}_0$ is consistent, the consistency of $(\tilde{\theta}_j, \tilde{P}_j)$ for $j = 1, \ldots, k$ follows from induction.

We proceed to derive the stated representation of $\tilde{\theta}_j - \hat{\theta}$ and $\tilde{P}_j - \hat{P}$. First, $\tilde{\theta}_j$ satisfies the first order condition $\nabla_\theta \overline{\psi}(\tilde{\theta}_j, \tilde{P}_{j-1}) = 0$. Expanding this around $(\hat{\theta}, \hat{P})$ and using $\nabla_\theta \overline{\psi}(\hat{\theta}, \hat{P}) = 0$ gives

$$0 = \nabla_{\theta\theta'} \overline{\psi}(\bar{\theta}, \bar{P})(\tilde{\theta}_j - \hat{\theta}) + \nabla_{\theta P'} \overline{\psi}(\bar{\theta}, \bar{P})(\tilde{P}_{j-1} - \hat{P}), \tag{9}$$

where $(\bar{\theta}, \bar{P})$ lie between $(\tilde{\theta}_j, \tilde{P}_{j-1})$ and $(\hat{\theta}, \hat{P})$. Since $\nabla_{\theta\theta'} \overline{\psi}(\bar{\theta}, \bar{P}) = -\Omega_{\theta\theta} + o_p(1)$ and $\nabla_{\theta P'} \overline{\psi}(\bar{\theta}, \bar{P}) = -\Omega_{\theta P} + o_p(1)$ follow from the consistency of $(\bar{\theta}, \bar{P})$, positive definiteness of $\Omega_{\theta\theta}$ allows us to obtain $\tilde{\theta}_j - \hat{\theta} = O_p(||\tilde{P}_{j-1} - \hat{P}||)$, giving the first result.

For the second result, note that the second derivatives of $\Psi(\theta, P)$ are uniformly bounded in $(\theta, P) \in \Theta \times B_P$ from Assumption 1(c). Hence, expanding the right hand side of $\tilde{P}_j = \Psi(\tilde{\theta}_j, \tilde{P}_{j-1})$ twice around $(\hat{\theta}, \hat{P})$ and using $\Psi(\hat{\theta}, \hat{P}) = \hat{P}$, root-$n$ consistency of $(\hat{\theta}, \hat{P})$, and $\tilde{\theta}_j - \hat{\theta} = O_p(||\tilde{P}_{j-1} - \hat{P}||)$, we obtain

$$\tilde{P}_j - \hat{P} = \Psi_\theta(\tilde{\theta}_j - \hat{\theta}) + \Psi_P(\tilde{P}_{j-1} - \hat{P}) + O_p(n^{-1/2}||\tilde{P}_{j-1} - \hat{P}|| + ||\tilde{P}_{j-1} - \hat{P}||^2). \tag{10}$$

Refine (9) as $\tilde{\theta}_j - \hat{\theta} = -\Omega_{\theta\theta}^{-1}\Omega_{\theta P}(\tilde{P}_{j-1} - \hat{P}) + O_p(n^{-1/2}||\tilde{P}_{j-1} - \hat{P}|| + ||\tilde{P}_{j-1} - \hat{P}||^2)$ by using $\nabla_{\theta P'} \overline{\psi}(\bar{\theta}, \bar{P}) = -\Omega_{\theta P} + O_p(||\tilde{P}_{j-1} - \hat{P}||) + O_p(n^{-1/2})$ and $\nabla_{\theta\theta'} \overline{\psi}(\bar{\theta}, \bar{P}) = -\Omega_{\theta\theta} + O_p(||\tilde{P}_{j-1} - \hat{P}||) + O_p(n^{-1/2})$. Substituting this into (10) in conjunction with $\Omega_{\theta\theta}^{-1}\Omega_{\theta P} = (\Psi_\theta' \Delta_P \Psi_\theta)^{-1} \Psi_\theta' \Delta_P \Psi_P$ gives the stated result. □

7.2 Proof of Proposition 1

For any eigenvalue λ of Ψ_P, the corresponding eigenvalue of Λ_P is $\alpha\lambda + (1 - \alpha) = \alpha(\lambda - 1) + 1$. Suppose $\lambda_{\max} \geq 1 \geq \lambda_{\min}$. If $\alpha \geq 0$, then $\alpha(\lambda_{\max} - 1) + 1 \geq 1$. If $\alpha < 0$, then $\alpha(\lambda_{\min} - 1) + 1 \geq 1$. Therefore, there is no value of α such that $\alpha(\lambda - 1) + 1 < 1$ for both $\lambda = \lambda_{\max}$ and $\lambda_{\min}$, giving the first result. Now, assume that $\lambda_{\max} < 1$. We derive the value of α that minimizes the

spectral radius of Λ_P. First, such α needs to be positive because $\alpha(\lambda-1)+1 \geq 1$ if $\alpha \leq 0$. When $\alpha > 0$, we have $1 > \alpha(\lambda_{\max}-1)+1 \geq \alpha(\lambda_{\min}-1)+1$. Therefore, the optimal α satisfies $\alpha^*(\lambda_{\max}-1)+1 = -\alpha^*(\lambda_{\min}-1)-1$, giving $\alpha^* = 2/(2-\lambda_{\max}-\lambda_{\min})$. □

7.3 Proof of Proposition 2

For part (a), write $\Gamma(\theta,P)-P$ as $\Gamma(\theta,P)-P = A(\theta,P)(\Psi(\theta,P)-P)$, where $A(\theta,P) \equiv (I-\Pi(\theta,P)\nabla_{P'}\Psi(\theta,P)\Pi(\theta,P))^{-1}\Pi(\theta,P)+(I-\Pi(\theta,P))$. Let $Z(\theta,P)$ denote an orthonormal basis of the column space of $\Pi(\theta,P)$, so that $Z(\theta,P)Z(\theta,P)' = \Pi(\theta,P)$ and $Z(\theta,P)'Z(\theta,P) = I_m$. Suppress (θ,P) from $\Pi(\theta,P)$, $Z(\theta,P)$, and $\nabla_{P'}\Psi(\theta,P)$. A direct calculation gives $(I-\Pi\nabla_{P'}\Psi\Pi)^{-1}\Pi = Z(I-Z'\nabla_{P'}\Psi Z)^{-1}Z'$, so we can write $A(\theta,P)$ as $A(\theta,P) = Z(I-Z'\nabla_{P'}\Psi Z)^{-1}Z'+(I-\Pi)$. The stated result follows since $A(\theta,P)$ is nonsingular because $\text{rank}[Z(I-Z'\nabla_{P'}\Psi Z)^{-1}Z'] = m$, $\text{rank}(I-\Pi) = N-m$, and $Z(I-Z'\nabla_{P'}\Psi Z)^{-1}Z'$ and $I-\Pi$ are orthogonal to each other.

For part (b), define $\Gamma_P \equiv \nabla_{P'}\Gamma(\theta^0,P^0)$ and $\Pi^0 \equiv \Pi(\theta^0,P^0)$. Define $\mathbb{P}$ with respect to $\Psi_P \equiv \nabla_{P'}\Psi(\theta^0,P^0)$. Computing $\nabla_{P'}\Gamma(\theta,P)$ and noting that $\Psi(\theta^0,P^0) = P^0$, we find $\Gamma_P = \Pi^0 + (I-\Pi^0\Psi_P\Pi^0)^{-1}\Pi^0(\Psi_P - I) + (I-\Pi^0)\Psi_P$. Observe that $\Gamma_P\Pi^0 = (I-\Pi^0)\Psi_P\Pi^0 = 0$, where the last equality follows because $\Psi_P\Pi^0 P \in \mathbb{P}$ for any $P \in \mathbb{R}^L$ by the definition of Π^0. Hence, $\Gamma_P = \Gamma_P(I-\Pi^0)$. We also have $(I-\Pi^0)\Gamma_P = (I-\Pi^0)\Psi_P$ because a direct calculation gives $(I-\Pi^0\Psi_P\Pi^0)^{-1}\Pi^0 = Z^0(I-(Z^0)'\Psi_P Z^0)^{-1}(Z^0)'$ where $Z^0 = Z(\theta^0,P^0)$, and hence $(I-\Pi^0)(I-\Pi^0\Psi_P\Pi^0)^{-1}\Pi^0 = 0$. Then, in conjunction with $\Gamma_P = \Gamma_P(I-\Pi^0)$, we obtain $(I-\Pi^0)\Gamma_P = (I-\Pi^0)\Psi_P(I-\Pi^0)$. Since $\Gamma_P(I-\Pi^0)$ has the same eigenvalues as $(I-\Pi^0)\Gamma_P$ (see Theorem 1.3.20 of Horn and Johnson, 1985), we have $\rho(\Gamma_P) = \rho(\Gamma_P(I-\Pi^0)) = \rho((I-\Pi^0)\Gamma_P) = \rho[(I-\Pi^0)\Psi_P(I-\Pi^0)] \leq \delta^0$, where the last inequality follows from Lemma 2.10 of SK: P, Q, and F_u^* in SK correspond to our Π^0, $I-\Pi^0$, and Ψ_P. □

7.4 Proof of Proposition 4

Write the objective function as $\bar{\gamma}(\theta,P,\eta) \equiv n^{-1}\sum_{i=1}^n \ln\Gamma(\theta,P,\eta)(a_i|x_i)$, and define $\gamma(\theta,P,\eta) \equiv E\ln\Gamma(\theta,P,\eta)(a_i|x_i)$. Define $\Omega_{\theta P}^{\Gamma} \equiv E\nabla_\theta \ln\Gamma(\theta^0,P^0)(a_i|x_i)\nabla_{P'}\ln\Gamma(\theta^0,P^0)(a_i|x_i)$.

We use induction. First, we prove the consistency, i.e., $(\tilde{\theta}_j,\tilde{P}_j) \to_p (\theta^0,P^0)$ if $(\tilde{\theta}_{j-1},\tilde{P}_{j-1}) \to_p (\theta^0,P^0)$. To show the consistency of $\tilde{\theta}_j$, we show that $\bar{\Theta}_j$ is compact and

$$\sup_{(\theta,P,\eta)\in\bar{\Theta}_j\times\mathcal{N}} |\bar{\gamma}(\theta,P,\eta)-\gamma(\theta,P,\eta)| = o_p(1), \tag{11}$$

$$\gamma(\theta,P^0,\theta^0) \text{ is continuous in } \theta, \text{ and } \gamma(\theta,P^0,\theta^0) \text{ is uniquely maximized at } \theta^0. \tag{12}$$

Then the consistency of $\tilde{\theta}_j$ follows from Theorem 2.1 of Newey and McFadden (1994) because (11) in conjunction with the consistency of $(\tilde{\theta}_{j-1},\tilde{P}_{j-1})$ and the triangle inequality implies $\sup_{\theta\in\bar{\Theta}_j}|\bar{\gamma}(\theta,\tilde{P}_{j-1},\tilde{\theta}_{j-1})-\gamma(\theta,P^0,\theta^0)| = o_p(1)$.

$\bar{\Theta}_j$ is compact because $\bar{\Theta}_j$ is an intersection of the compact set Θ and $|A||X|$ closed sets.

Take $\mathcal{N}$ sufficiently small, then it follows from the consistency of $(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$ and the continuity of $\Gamma(\theta, P, \eta)$ that $\Gamma(\theta, P, \eta)(a|x) \in [\epsilon/2, 1-\epsilon/2]$ for all $(a, x) \in A \times X$ and $(\theta, P, \eta) \in \bar{\Theta}_j \times \mathcal{N}$ with probability approaching one (henceforth wpa1). Observe that (i) $\bar{\Theta}_j \times \mathcal{N}$ is compact, (ii) $\ln \Gamma(\theta, P, \eta)$ is continuous in $(\theta, P, \eta) \in \bar{\Theta}_j \times \mathcal{N}$, and (iii) $E \sup_{(\theta,P,\eta)\in\bar{\Theta}_j\times\mathcal{N}} |\ln \Gamma(\theta, P, \eta)(a_i|x_i)| \leq (|\ln(\epsilon/2)| + |\ln(1-\epsilon/2)|) < \infty$ because of the way we choose $\mathcal{N}$. Therefore, (11) follows from Lemma 2.4 of Newey and McFadden (1994). Lemma 2.4 of Newey and McFadden (1994) also implies that $\gamma(\theta, P, \eta)$ is continuous, giving the first part of (12). Finally, the second part of (12) holds because θ^0 is the only parameter such that $P^0 = \Gamma(\theta, P^0, \theta^0)$, and we prove the consistency of $\tilde{\theta}_j$. The consistency of $\tilde{P}_j$ then follows from the continuity of $\Gamma(\theta, P, \eta)$ and the consistency of $(\tilde{\theta}_j, \tilde{P}_{j-1})$, and we establish the consistency of $(\tilde{\theta}_j, \tilde{P}_j)$.

We proceed to derive the stated representation of $\tilde{\theta}_j - \hat{\theta}_{RPM}$ and $\tilde{P}_j - \hat{P}_{RPM}$. Henceforth, we suppress the subscript RPM from $\hat{\theta}_{RPM}$ and $\hat{P}_{RPM}$. $\tilde{\theta}_j$ satisfies the first order condition $\nabla_\theta \bar{\gamma}(\tilde{\theta}_j, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) = 0$. Expanding it twice around $(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ gives

$$0 = \nabla_\theta \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) + \nabla_{\theta\theta'} \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(\tilde{\theta}_j - \hat{\theta}) + O_p(||\tilde{\theta}_j - \hat{\theta}||^2). \tag{13}$$

We analyze $\nabla_\theta \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ on the right of (13) first. Expanding $\nabla_\theta \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ twice around $(\hat{\theta}, \hat{P}, \hat{\theta})$ gives $\nabla_\theta \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) = \nabla_\theta \bar{\gamma}(\hat{\theta}, \hat{P}, \hat{\theta}) + \nabla_{\theta P'} \bar{\gamma}(\hat{\theta}, \hat{P}, \hat{\theta})(\tilde{P}_{j-1} - \hat{P}) + \nabla_{\theta\eta'} \bar{\gamma}(\hat{\theta}, \hat{P}, \hat{\theta})(\tilde{\theta}_{j-1} - \hat{\theta}) + O_p(||\tilde{\theta}_{j-1} - \hat{\theta}||^2 + ||\tilde{P}_{j-1} - \hat{P}||^2)$. First, the RPM estimator satisfies $\nabla_\theta \bar{\gamma}(\hat{\theta}, \hat{P}, \hat{\theta}) = 0$ wpa1 because $\nabla_{\theta'} \bar{\gamma}(\hat{\theta}, \hat{P}) = 0$ from the first order condition, and Proposition 2(a) implies $\Psi(\hat{\theta}, \hat{P}) = \hat{P}$ wpa1 and hence $\nabla_{\theta'} \Gamma(\hat{\theta}, \hat{P}, \hat{\theta}) = \nabla_{\theta'} \Gamma(\hat{\theta}, \hat{P})$ wpa1. Second, the information matrices such as $\Omega^\Gamma_{\theta\theta}$ are defined equivalently in terms of either by $\Gamma(\theta, P, \eta)$ or $\Gamma(\theta, P)$ because $\Gamma(\theta^0, P^0, \theta^0) = \Gamma(\theta^0, P^0)$, $\nabla_{\theta'} \Gamma(\theta^0, P^0, \theta^0) = \nabla_{\theta'} \Gamma(\theta^0, P^0)$, and $\nabla_{P'} \Gamma(\theta^0, P^0, \theta^0) = \nabla_{P'} \Gamma(\theta^0, P^0)$ from $P^0 = \Psi(\theta^0, P^0)$. Third, the information matrix equality and $\nabla_{\eta'} \Gamma(\theta^0, P^0, \theta^0) = 0$ imply $E \nabla_{\theta\eta'} \ln \Gamma(\theta^0, P^0, \theta^0)(a_i|x_i) = 0$. Therefore, in conjunction with the root-n consistency of $(\hat{\theta}, \hat{P})$, we have

$$\nabla_\theta \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) = -\Omega^\Gamma_{\theta P}(\tilde{P}_{j-1} - \hat{P}) + r_{nj}, \tag{14}$$

where r_{nj} denotes a generic reminder term of $O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}||^2 + n^{-1/2}||\tilde{P}_{j-1} - \hat{P}|| + ||\tilde{P}_{j-1} - \hat{P}||^2)$. The stated bound, $\tilde{\theta}_j - \hat{\theta} = O_p(||\tilde{P}_{j-1} - \hat{P}||) + r_{nj}$, follows from writing the second and third terms on the right of (13) together as $(-\Omega^\Gamma_{\theta\theta} + o_p(1))(\tilde{\theta}_j - \hat{\theta})$ and using the positive definiteness of $\Omega^\Gamma_{\theta\theta}$.

For the representation of $\tilde{P}_j - \hat{P}$, first we have

$$\tilde{P}_j = \hat{P} + \Gamma_\theta(\tilde{\theta}_j - \hat{\theta}) + \Gamma_P(\tilde{P}_{j-1} - \hat{P}) + r_{nj}, \tag{15}$$

by expanding $\tilde{P}_j = \Gamma(\tilde{\theta}_j, \tilde{P}_{j-1}, \tilde{\theta}_j)$ around $(\hat{\theta}, \hat{P}, \hat{\theta})$ and using $\Gamma(\hat{\theta}, \hat{P}, \hat{\theta}) = \hat{P}$. Next, refine (13) as $0 = \nabla_\theta \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) - \Omega^\Gamma_{\theta\theta}(\tilde{\theta}_j - \hat{\theta}) + r_{nj}$ by expanding $\nabla_{\theta\theta'} \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ in (13) around $(\hat{\theta}, \hat{P}, \hat{\theta})$ to write it as $\nabla_{\theta\theta'} \bar{\gamma}(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) = -\Omega^\Gamma_{\theta\theta} + O_p(n^{-1/2}) + O_p(||\tilde{\theta}_{j-1} - \hat{\theta}||) + O_p(||\tilde{P}_{j-1} - \hat{P}||)$

and using the bound of $\tilde{\theta}_j - \hat{\theta}$ obtained above. Substituting this into (14) gives $\tilde{\theta}_j - \hat{\theta} = -(\Omega^{\Gamma}_{\theta\theta})^{-1}\Omega^{\Gamma}_{\theta P}(\tilde{P}_{j-1} - \hat{P}) + r_{nj}$. The stated result follows from substituting this into (15) in conjunction with $(\Omega^{\Gamma}_{\theta\theta})^{-1}\Omega^{\Gamma}_{\theta P} = (\Gamma'_\theta \Delta_P \Gamma_\theta)^{-1}\Gamma'_\theta \Delta_P \Gamma_P$. □

7.5 Proof of Corollary 1

The proof of the consistency of $(\tilde{\theta}_j, \tilde{P}_j)$ is the same as the proof of Proposition 4. For the bound of $\tilde{\theta}_j$ and $\tilde{P}_j$, define $\Gamma(\theta, P, \eta, Q) \equiv \Psi(\theta, P) + [(I - \Pi(\eta, Q)\nabla_{P'}\Psi(\eta, Q)\Pi(\eta, Q))^{-1} - I]\Pi(\eta, Q)(\Psi(\theta, P) - P)$, and write the objective function in Step 1 as $\bar{\gamma}(\theta, \tilde{P}_{j-1}, \theta^*, P^*)$. Since $\nabla_{Q'}\Gamma(\theta^0, P^0, \theta^0, P^0) = 0$, the stated result follows from starting from the first order condition $\nabla_{\theta'}\bar{\gamma}(\tilde{\theta}_j, \tilde{P}_{j-1}, \theta^*, P^*) = 0$, and following the proof of Proposition 4. □

7.6 Proof of Proposition 5

We use induction. Assume $(\tilde{\theta}_{j-1}, \tilde{P}_{j-1}) \to_p (\theta^0, P^0)$. Define $Q^q_n(\theta, P, \eta) \equiv n^{-1}\sum_{i=1}^n \ln \tilde{\Lambda}^q(\theta, P, \eta)(a_i|x_i)$ and $Q^q(\theta, P, \eta) \equiv E \ln \tilde{\Lambda}^q(\theta, P, \eta)(a_i|x_i)$. In order to show $\tilde{\theta}_j \to_p \theta^0$, it suffices to show that (11)–(12) in the proof of Proposition 4 hold if we replace $\bar{\gamma}(\theta, P, \eta)$ and $\gamma(\theta, P, \eta)$ with $Q^q_n(\theta, P, \eta)$ and $Q^q(\theta, P, \eta)$. Take $\mathcal{N}$ sufficiently small, then (i) $\Theta^q_j \times \mathcal{N}$ is compact, (ii) $\ln \tilde{\Lambda}^q(\theta, P, \eta)$ is continuous in $(\theta, P, \eta) \in \Theta^q_j \times \mathcal{N}$, and (iii) $E \sup_{(\theta,P,\eta)\in\Theta^q_j\times\mathcal{N}} |\ln \tilde{\Lambda}^q(\theta, P, \eta)(a_i|x_i)| < \infty$. Therefore, (11) and the first result of (12) hold for $Q^q_n(\theta, P, \eta)$ and $Q^q(\theta, P, \eta)$.

We proceed to show that θ^0 uniquely maximizes $Q^q(\theta, P^0, \theta^0)$. Note that

$$\begin{aligned} Q^q(\theta, P^0, \theta^0) - Q^q(\theta^0, P^0, \theta^0) &= E \ln(\nabla_{\theta'}\Lambda^q(\theta^0, P^0)(\theta - \theta^0) + P^0)(a_i|x_i) - E \ln P^0(a_i|x_i) \\ &= E \ln\left(\frac{\nabla_{\theta'}\Lambda^q(\theta^0, P^0)(a_i|x_i)(\theta - \theta^0)}{P^0(a_i|x_i)} + 1\right). \end{aligned} \tag{16}$$

Recall that $\ln(y + 1) \leq y$ for all $y > -1$ where the inequality is strict if $y \neq 0$, and that Assumption 4(d) implies $\nabla_{\theta'}\Lambda^q(\theta^0, P^0)(a_i|x_i)(\theta - \theta^0)/P^0(a_i|x_i) \neq 0$ with positive probability for all $\theta \neq \theta^0$. Therefore, the right hand side of (16) is strictly smaller than

$$E\left[\frac{\nabla_{\theta'}\Lambda^q(\theta^0, P^0)(a_i|x_i)(\theta - \theta^0)}{P^0(a_i|x_i)}\right] \quad \text{for all } \theta \neq \theta^0. \tag{17}$$

Because $E[\nabla_{\theta'}\Lambda^q(\theta^0, P^0)(a_i|x_i)/P^0(a_i|x_i)] = 0$, we have $Q^q(\theta, P^0, \theta^0) - Q^q(\theta^0, P^0, \theta^0) < 0$ for all $\theta \neq \theta^0$, and θ^0 uniquely maximizes $Q(\theta, P^0, \theta^0)$. Therefore, $\tilde{\theta}_j \to_p \theta^0$. Finally, $\tilde{P}_j \to_p P^0$ follows from $\Lambda^q(\tilde{\theta}_j, \tilde{P}_{j-1}) \to_p \Lambda^q(\theta^0, P^0) = P^0$. □

7.7 Proof of Proposition 6

The proof is similar to the proof of the updating formula of Proposition 4. We suppress the subscript qNPL from $\hat{\theta}_{qNPL}$ and $\hat{P}_{qNPL}$. $(\hat{\theta}, \hat{P})$ is root-n consistent from applying the

proof of Proposition 2 of Aguirregabiria and Mira (2007) with replacing $\Psi(\theta, P)$ by $\Lambda^q(\theta, P)$. Define $Q_n^q(\theta, P, \eta) \equiv n^{-1}\sum_{i=1}^n \ln \tilde{\Lambda}^q(\theta, P, \eta)(a_i|x_i)$. First, expanding the first order condition $0 = \nabla_\theta Q_n^q(\tilde{\theta}_j, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ twice around $(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ gives $0 = \nabla_\theta Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) + \nabla_{\theta\theta'} Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(\tilde{\theta}_j - \hat{\theta}) + O_p(||\tilde{\theta}_j - \hat{\theta}||^2)$, which corresponds to (13) in the proof of Proposition 4. Second, note that the q-NPL estimator satisfies $\nabla_\theta Q_n^q(\hat{\theta}, \hat{P}, \hat{\theta}) = 0$, and that $\tilde{\Lambda}^q(\theta^0, P^0, \theta^0) = \Lambda^q(\theta^0, P^0)$, $\nabla_{\theta'}\tilde{\Lambda}^q(\theta^0, P^0, \theta^0) = \nabla_{\theta'}\Lambda^q(\theta^0, P^0)$, $\nabla_{P'}\tilde{\Lambda}^q(\theta^0, P^0, \theta^0) = \nabla_{P'}\Lambda^q(\theta^0, P^0)$, and $\nabla_{\eta'}\tilde{\Lambda}^q(\theta^0, P^0, \theta^0) = 0$. Therefore, expanding $\nabla_\theta Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ twice around $(\hat{\theta}, \hat{P}, \hat{\theta})$ and using the root-n consistency of $(\hat{\theta}, \hat{P})$ and the information matrix equality, we obtain $\nabla_\theta Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) = -\Omega_{\theta P}^q(\tilde{P}_{j-1} - \hat{P}) + r_{nj}$, where r_{nj} denotes a reminder term of $O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}||^2 + n^{-1/2}||\tilde{P}_{j-1} - \hat{P}|| + ||\tilde{P}_{j-1} - \hat{P}||^2)$. This corresponds to (14) in the proof of Proposition 4. The stated bound of $\tilde{\theta}_j - \hat{\theta}$ follows from noting that $\nabla_{\theta\theta'} Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1}) = -\Omega_{\theta\theta}^q + o_p(1)$ and repeating the argument of the proof of Proposition 4.

The proof of the representation of $\tilde{P}_j - \hat{P}$ follows from the proof of Proposition 4, because (i) $\tilde{P}_j = \hat{P} + \Lambda_\theta^q(\tilde{\theta}_j - \hat{\theta}) + \Lambda_P^q(\tilde{P}_{j-1} - \hat{P}) + r_{nj}$, which corresponds to (15) in the proof of Proposition 4, from expanding $\Lambda^q(\tilde{\theta}_j, \tilde{P}_{j-1})$ twice around $(\hat{\theta}, \hat{P})$ and using $\hat{P} = \Lambda^q(\hat{\theta}, \hat{P})$, (ii) $\nabla_{\theta\theta'} Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})(\tilde{\theta}_j - \hat{\theta}) = -\Omega_{\theta\theta}^q(\tilde{\theta}_j - \hat{\theta}) + r_{nj}$ from expanding $\nabla_{\theta\theta'} Q_n^q(\hat{\theta}, \tilde{P}_{j-1}, \tilde{\theta}_{j-1})$ around $(\hat{\theta}, \hat{P}, \hat{\theta})$ and using the bound of $\tilde{\theta}_j - \hat{\theta}$ obtained above, and (iii) $(\Omega_{\theta\theta}^q)^{-1}\Omega_{\theta P}^q = ((\Lambda_\theta^q)'\Delta_P\Lambda_\theta^q)^{-1}(\Lambda_\theta^q)'\Delta_P\Lambda_P^q$. □

7.8 Proof of Proposition 7

The proof is essentially the same as the proof of Proposition 5. The argument of the proof of Proposition 5 carries through if we replace $\tilde{\Lambda}^q(\theta, P, \eta)$ and $\Lambda^q(\theta^0, P^0)$ with $\Phi(\theta, P, \eta)$ and P_{θ^0}. □

7.9 Proof of Proposition 8

We suppress the subscript MLE from $\hat{\theta}_{MLE}$ and $\hat{P}_{MLE}$. First, $\tilde{P}_j - \hat{P} = O_p(||\theta_{j-1} - \hat{\theta}||)$ follows easily from Taylor expansion. To show the bound of $\tilde{\theta}_j - \hat{\theta}$, define $\Phi(\theta, \eta) \equiv \Phi(\theta, P_\eta, \eta) = P_\eta + \nabla_{\theta'} P_\eta(\theta - \eta)$ and $Q_n(\theta, \eta) \equiv Q_n(\theta, P_\eta, \eta) = n^{-1}\sum_{i=1}^n \ln \Phi(\theta, \eta)(a_i|x_i)$, so that $\tilde{\theta}_j = \arg\max_{\Theta_j} Q_n(\theta, \tilde{\theta}_{j-1})$. We expand the first order condition $\nabla_\theta Q_n(\tilde{\theta}_j, \tilde{\theta}_{j-1}) = 0$ twice around $(\hat{\theta}, \tilde{\theta}_{j-1})$ as

$$\begin{aligned} 0 &= \nabla_\theta Q_n(\hat{\theta}, \tilde{\theta}_{j-1}) + \nabla_{\theta\theta'} Q_n(\hat{\theta}, \tilde{\theta}_{j-1})(\tilde{\theta}_j - \hat{\theta}) + O_p(||\tilde{\theta}_j - \hat{\theta}||^2) \\ &= \nabla_\theta Q_n(\hat{\theta}, \tilde{\theta}_{j-1}) + (-\mathcal{I}^0 + o_p(1))(\tilde{\theta}_j - \hat{\theta}), \end{aligned} \tag{18}$$

where the second equality follows from $E[\nabla_{\theta\theta'} Q_n(\theta^0, \theta^0)] = -\mathcal{I}^0$ and the consistency of $(\hat{\theta}, \tilde{\theta}_{j-1})$. Since the MLE satisfies $\nabla_\theta Q_n(\hat{\theta}, \hat{\theta}) = 0$, expanding $\nabla_\theta Q_n(\hat{\theta}, \tilde{\theta}_{j-1})$ around $(\hat{\theta}, \hat{\theta})$ gives $\nabla_\theta Q_n(\hat{\theta}, \tilde{\theta}_{j-1}) = \nabla_{\theta\eta'} Q_n(\hat{\theta}, \hat{\theta})(\tilde{\theta}_{j-1} - \hat{\theta}) + O_p(||\tilde{\theta}_{j-1} - \hat{\theta}||^2)$. Now, $\nabla_{\theta\eta'} Q_n(\hat{\theta}, \hat{\theta}) = n^{-1}\sum_{i=1}^n \nabla_{\theta\theta'} P_{\hat{\theta}}(a_i|x_i)/P_{\hat{\theta}}(a_i|x_i) = O_p(n^{-1/2})$, where the last equality follows from the root-n consistency of $\hat{\theta}$ because the infor-

mation matrix equality implies $E[\nabla_{\theta\theta'} P_{\theta^0}(a_i|x_i)/P_{\theta^0}(a_i|x_i)] = 0$. Therefore, $\nabla_\theta Q_n(\hat{\theta}, \tilde{\theta}_{j-1}) = O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}||^2)$, and the stated bound of $\tilde{\theta}_j - \hat{\theta}$ follows from (18). □

7.10 Proof of Proposition 9

The proof is the same as that of Proposition 7 and is omitted. □

7.11 Proof of Proposition 10

We suppress the subscript MLE from $\hat{\theta}_{MLE}$ and $\hat{P}_{MLE}$. The updating formula of $\tilde{P}_j$ follows from expanding $\tilde{P}_j = \Lambda^q(\tilde{\theta}_{j-1}, \tilde{P}_{j-1})$ around $(\hat{\theta}, \hat{P})$ and using the root-n consistency of $(\hat{\theta}, \hat{P})$.

For the bound of $\tilde{\theta}_j - \hat{\theta}$, expanding the first order condition $\nabla_\theta Q_n(\tilde{\theta}_j, \tilde{P}_j, \tilde{\theta}_{j-1}) = 0$ twice around $(\hat{\theta}, \tilde{P}_j, \tilde{\theta}_{j-1})$, we have

$$0 = \nabla_\theta Q_n(\hat{\theta}, \tilde{P}_j, \tilde{\theta}_{j-1}) + \left[-\mathcal{I}^0 + O_p\left(n^{-1/2} + ||\tilde{P}_j - \hat{P}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}||\right)\right](\tilde{\theta}_j - \hat{\theta}) + O_p(||\tilde{\theta}_j - \hat{\theta}||^2), \tag{19}$$

where the second term on the right follows from expanding $\nabla_{\theta\theta'} Q_n(\hat{\theta}, \tilde{P}_j, \tilde{\theta}_{j-1})$ around $(\hat{\theta}, \hat{P}, \hat{\theta})$ and using the root-n consistency of $(\hat{\theta}, \hat{P})$ and the information matrix equality.

Since $\nabla_\theta Q_n(\hat{\theta}, \hat{P}, \hat{\theta}) = 0$ and $E\left[\nabla^{(j)}\Phi(\theta, P, \eta)(a_i|x_i)/\Phi(\theta, P, \eta)(a_i|x_i)\right]$ evaluated at $(\theta, P, \eta) = (\theta^0, P^0, \theta^0)$ is equal to 0 for $j \geq 1$, expanding $\nabla_\theta Q_n(\hat{\theta}, \tilde{P}_j, \tilde{\theta}_{j-1})$ on the right of (19) twice around $(\hat{\theta}, \hat{P}, \hat{\theta})$ in conjunction with the root-n consistency of $(\hat{\theta}, \hat{P})$ and the information matrix equality gives

$$\nabla_\theta Q_n(\hat{\theta}, \tilde{P}_j, \tilde{\theta}_{j-1}) = -E[\nabla_\theta \ln P_{\theta^0}(a_i|x_i) I(a_i|x_i)/P^0(a_i|x_i)](\tilde{P}_j - \hat{P}) + \mathcal{I}^0(\tilde{\theta}_{j-1} - \hat{\theta}) + r_{nj}, \tag{20}$$

where $I(a_i|x_i)$ is the row of an $L \times L$ identity matrix corresponding to $(a_i|x_i)$, and r_{nj} is a reminder term of $O_p(n^{-1/2}||\tilde{\theta}_{j-1} - \hat{\theta}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}||^2 + n^{-1/2}||\tilde{P}_j - \hat{P}|| + ||\tilde{P}_j - \hat{P}||^2)$. Hence, we have $\tilde{\theta}_j - \hat{\theta} = O_p(||\tilde{\theta}_{j-1} - \hat{\theta}|| + ||\tilde{P}_j - \hat{P}||)$ from (19) and (20). Substituting this bound of $\tilde{\theta}_j - \hat{\theta}$ into the $O_p(||\tilde{\theta}_j - \hat{\theta}||^2)$ term in (19) and using $\tilde{P}_j - \hat{P} = O_p(||\tilde{P}_{j-1} - \hat{P}|| + ||\tilde{\theta}_{j-1} - \hat{\theta}||)$ from the updating formula of $\tilde{P}_j$, we obtain the stated updating formula of $\tilde{\theta}_j$ from (19) and (20). □

References

Aiyagari, S. Rao (1994). "Uninsured idiosyncratic risk and aggregate saving." *Quarterly Journal of Economics* 109(3): 659-684.

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.

Aguirregabiria, V. and P. Mira (2007). "Sequential estimation of dynamic discrete games." *Econometrica* 75(1): 1-53.

Arcidiacono, P. and R. A. Miller (2008). CCP estimation of dynamic discrete choice models with unobserved heterogeneity. Mimeographed, Duke university.

Bajari, P., Benkard, C. L., and Levin, J. (2007). "Estimating dynamic models of imperfect competition." *Econometrica* 75(5): 1331-1370.

Bajari, P. and H. Hong (2006). Semiparametric estimation of a dynamic game of incomplete information. NBER Technical Working Paper 320.

Başar, T. (1987). "Relaxation Techniques and Asynchronous Algorithms for On-line Computation of Noncooperative Equilibria." *Journal of Economic Dynamics and Controls*, 11: 531-549.

Collard-Wexler, A. (2006) Demand fluctuations and plant turnover in the Ready-Mix concrete industry. Mimeographed, NYU.

Horn R. A. and C. R. Johnson (1985) *Matrix Analysis*. Cambridge University Press.

Hotz, J. and R. A. Miller (1993). "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies* 60: 497-529.

Judd, L. J. (1998) *Numerical Methods in Economics*. Cambridge, Massachusetts: The MIT Press.

Kasahara, H. and K. Shimotsu (2006). Nonparametric Identification and Estimation of Finite Mixture Models of Dynamic Discrete Choices. Mimeographed, Queen's University.

Kasahara, H. and K. Shimotsu (2008a) "Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, 146: 92-106.

Kasahara, H. and K. Shimotsu (2008b) "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, forthcoming.

Krawczyk, J. B. and S. Uryasev (2000) "Relaxation Algorithms to Find Nash Equilibria with Economic Applications," *Environmental Modeling and Assessment*, 5: 63-73.

Krusell, P. and A. Smith Jr. (1998) "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(5): 867-896

Ljungqvist, L. and T. J. Sargent (2004) *Recursive Macroeconomic Theory*, 2nd ed., MIT Press.

Newey, W. K. and D. McFadden (1994). "Large Sample Estimation and Hypothesis Testing," in R. F. Engle and D. L. McFadden (eds.) Handbook of Econometrics, Vol. 4, Elsevier.

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.

Pesendorfer, M. and P. Schmidt-Dengler (2008). "Asymptotic least squares estimators for dynamic games," *Review of Economic Studies*, 75, 901-928.

Rust, J. (1987). "Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher." *Econometrica* 55(5): 999-1033.

Shroff, G. M. and H. B. Keller (1993) "Stabilization of unstable procedures: the recursive projection method," *SIAM Journal of Numerical Analysis*, 30(4): 1099-1120.

Su, Che-Lin, and Kenneth L. Judd (2008) Constrained optimization approaches to estimation of structural models. Mimeographed, University of Chicago.

Table 1: The Largest and Smallest Eigenvalues of Ψ_P and Λ_P

	Eig(Ψ_P)		Eig(Λ_P)			
θ_{RN}	λ_{max}	λ_{min}	λ_{max}	λ_{min}	$\rho(M_{\Psi_\theta}\Psi_P)$	$\rho(M_{\Lambda_\theta}\Lambda_P)$
1	0.2104	-0.3365	0.2572	-0.2572	0.2922	0.2555
2	0.4275	-0.6925	0.4945	-0.4945	0.5996	0.4937
4	0.7596	-1.1839	0.8017	-0.8017	1.1788	0.8056
6	0.8914	-1.4788	0.9161	-0.9161	1.4775	0.9150

A pair $(\lambda_{max}, \lambda_{min})$ represents the largest and the smallest eigenvalues of Ψ_P or Λ_P. The last two columns report the absolute value of the dominant eigenvalue of $M_{\Psi_\theta}\Psi_P$ and $M_{\Lambda_\theta}\Lambda_P$.

Table 2: Bias and RMSE

	Estimator	$\theta_{RN}=2$						$\theta_{RN}=4$					
		$n=500$		$n=2000$		$n=8000$		$n=500$		$n=2000$		$n=8000$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\hat{\theta}_{RS}$	PML with Ψ	-0.2277	0.2703	-0.0752	0.1125	-0.0258	0.0502	-0.1162	0.1438	-0.0323	0.0508	-0.0065	0.0196
	NPL with Ψ	-0.0147	0.1415	-0.0038	0.0646	-0.0037	0.0335	-0.0098	0.0685	-0.0056	0.0472	-0.0019	0.0403
	NPL with Λ	-0.0147	0.1415	-0.0038	0.0646	-0.0037	0.0335	0.0036	0.0593	-0.0015	0.0296	0.0011	0.0144
	RPM ($\delta=0.5$)	-0.0162	0.1399	-0.0063	0.0636	-0.0041	0.0325	0.0033	0.0586	-0.0019	0.0280	0.0008	0.0140
	RPM ($\delta=0.8$)	-0.0150	0.1410	-0.0038	0.0645	-0.0038	0.0334	0.0016	0.0617	-0.0027	0.0299	0.0010	0.0143
	q-NPL with Λ^q	-0.0135	0.1296	-0.0046	0.0595	-0.0023	0.0301	0.0024	0.0569	-0.0016	0.0278	0.0009	0.0139
	q-AFXP with Λ^q	-0.0131	0.1299	-0.0045	0.0596	-0.0023	0.0302	0.0021	0.0561	-0.0018	0.0276	0.0007	0.0137
$\hat{\theta}_{RN}$	PML with Ψ	-0.8116	0.9555	-0.2681	0.3988	-0.0935	0.1789	-0.7167	0.8270	-0.1798	0.2447	-0.0403	0.0871
	NPL with Ψ	-0.0450	0.4840	-0.0131	0.2285	-0.0144	0.1180	-0.1569	0.2753	-0.1168	0.1956	-0.0982	0.1624
	NPL with Λ	-0.0450	0.4840	-0.0131	0.2285	-0.0144	0.1180	0.0187	0.1346	0.0055	0.0678	0.0043	0.0350
	RPM ($\delta=0.5$)	-0.0502	0.4798	-0.0223	0.2242	-0.0161	0.1144	0.0196	0.1462	0.0042	0.0688	0.0038	0.0350
	RPM ($\delta=0.8$)	-0.0451	0.4843	-0.0132	0.2285	-0.0144	0.1181	-0.0099	0.1657	-0.0008	0.0727	0.0043	0.0357
	q-NPL with Λ^q	-0.0413	0.4411	-0.0165	0.2090	-0.0094	0.1052	0.0196	0.1267	0.0049	0.0651	0.0038	0.0330
	q-AFXP with Λ^q	-0.0403	0.4418	-0.0164	0.2094	-0.0092	0.1052	0.0184	0.1221	0.0046	0.0643	0.0034	0.0326
$\hat{P}$ ($\times 100$)	PML with Ψ	-0.0654	2.1491	-0.0103	0.5553	0.0237	0.1877	-0.0967	5.7026	-0.0831	1.9414	-0.0183	0.4722
	NPL with Ψ	0.0211	0.1625	0.0175	0.0392	0.0157	0.0363	-0.5544	3.4606	-0.1975	3.0148	-0.0150	2.8906
	NPL with Λ	0.0211	0.1625	0.0175	0.0392	0.0157	0.0363	0.0009	0.1113	-0.0453	0.0531	0.0048	0.0392
	RPM ($\delta=0.5$)	0.0209	0.1649	0.0200	0.0542	0.0169	0.0408	-0.0017	0.1774	-0.0464	0.0630	0.0011	0.0313
	RPM ($\delta=0.8$)	0.0165	0.1637	0.0167	0.0390	0.0135	0.0357	-0.2429	1.0161	-0.0938	0.3194	0.0028	0.0363
	q-NPL with Λ^q	0.0150	0.1397	0.0194	0.0424	0.0130	0.0245	-0.0211	0.1045	-0.0451	0.0523	0.0028	0.0324
	q-AFXP with Λ^q	0.0157	0.1390	0.0192	0.0421	0.0130	0.0240	-0.0232	0.0978	-0.0479	0.0556	0.0012	0.0285

The result is based on 500 simulated samples. The maximum number of iterations is set to 50. For the q-NPL and q-AFXP, we set $q=4$.

Table 3: RMSE for $j=5,10,\ldots,25$ with $n=8000$

	$\theta_{RN}=2$									
	RMSE of $\hat{\theta}_{RS}$					RMSE of $\hat{\theta}_{RN}$				
	j=5	j=10	j=15	j=20	j=25	j=5	j=10	j=15	j=20	j=25
NPL with Ψ	0.0335	0.0335	0.0335	0.0335	0.0335	0.1181	0.1180	0.1180	0.1180	0.1180
NPL with Λ	0.0335	0.0335	0.0335	0.0335	0.0335	0.1181	0.1180	0.1180	0.1180	0.1180
RPM ($\delta=0.5$)	0.0328	0.0326	0.0326	0.0325	0.0325	0.1153	0.1150	0.1148	0.1146	0.1145
RPM ($\delta=0.8$)	0.0334	0.0334	0.0334	0.0334	0.0334	0.1183	0.1181	0.1181	0.1181	0.1181
q-NPL with Λ^q	0.0302	0.0301	0.0301	0.0301	0.0301	0.1052	0.1052	0.1052	0.1052	0.1052
q-AFXP with Λ^q	0.0302	0.0302	0.0302	0.0302	0.0302	0.1052	0.1052	0.1052	0.1052	0.1052
	$\theta_{RN}=4$									
	RMSE of $\hat{\theta}_{RS}$					RMSE of $\hat{\theta}_{RN}$				
	j=5	j=10	j=15	j=20	j=25	j=5	j=10	j=15	j=20	j=25
NPL with Ψ	0.0173	0.0223	0.0265	0.0324	0.0350	0.0682	0.0777	0.1195	0.1271	0.1534
NPL with Λ	0.0144	0.0145	0.0144	0.0144	0.0144	0.0364	0.0351	0.0350	0.0350	0.0350
RPM ($\delta=0.5$)	0.0142	0.0140	0.0139	0.0140	0.0139	0.0379	0.0350	0.0350	0.0350	0.0351
RPM ($\delta=0.8$)	0.0148	0.0154	0.0144	0.0148	0.0143	0.0392	0.0394	0.0365	0.0409	0.0360
q-NPL with Λ^q	0.0139	0.0139	0.0139	0.0139	0.0139	0.0330	0.0330	0.0330	0.0330	0.0330
q-AFXP with Λ^q	0.0137	0.0137	0.0137	0.0137	0.0137	0.0325	0.0326	0.0326	0.0326	0.0326

The result is based on 500 simulated samples. The maximum number of iterations is set to 50. For the q-NPL and q-AFXP, we set $q=4$.