Econometric Theory, **26**, 2010, 863–881.
doi:10.1017/S0266466609990132

PANEL DATA MODELS WITH FINITE NUMBER OF MULTIPLE EQUILIBRIA

JINYONG HAHN
UCLA

HYUNGSIK ROGER MOON
USC

We study a nonlinear panel data model in which the fixed effects are assumed to have finite support. The fixed effects estimator is known to have the incidental parameters problem. We contribute to the literature by making a qualitative observation that the incidental parameters problem in this model may not be not as severe as in the conventional case. Because fixed effects have finite support, the probability of correctly identifying the fixed effect converges to one even when the cross sectional dimension grows as fast as some exponential function of the time dimension. As a consequence, the finite sample bias of the fixed effects estimator is expected to be small.

1. INTRODUCTION

The empirical industrial organization literature has now begun to pay attention to the problem of multiple equilibria. In many interesting game theoretic models, multiplicity of Nash equilibria is unavoidable.[1] The statistical difficulties associated with such multiplicity of equilibria usually have been dealt with by imposing some auxiliary assumptions on equilibrium selection. Sometimes, it is assumed that all markets in the data choose the same equilibrium, as is done in, e.g., Aguirregabiria and Mira (2004). Another possibility is to assume that the equilibria are selected independent of the observed market characteristics, which is the approach taken by, e.g., Ackerberg and Gowrisankaran (2006) and Sweeting (2004). A slightly more general approach is to assume that equilibria are selected randomly, with probability depending on some *observed* market characteristics; see Bajari, Hong, and Ryan (2004). All these approaches rule out the possibility that the equilibrium selection may depend on some unobserved market characteristic, which may be correlated with observed characteristics.

In this paper, we propose yet another approach. Our approach is applicable when the econometrician has access to a panel data set, where each market is observed over several periods of time. We will impose a restriction that each market

We are thankful to Guido Kuersteiner and two referees for helpful comments. We also thank Timothy Derdenger for proofreading. Address correspondence to Jinyong Hahn, Department of Economics, UCLA, 8283 Bunche Hall, Mail Stop 147703, Los Angeles, CA 90095; e-mail: hahn@econ.ucla.edu; Hyungsik Roger Moon, Department of Economics, KAP 300, USC, Los Angeles, CA 90089; e-mail: moonr@usc.edu.

 0266-4666/10 $15.00

chooses the *same* equilibrium over the period of observation, yet at the same time, we allow for the possibility that equilibria are selected depending on *unobserved* market characteristics. When each market chooses the same equilibrium over time, the equilibrium selection plays the same role as the *fixed effects* in the standard panel data models. The "fixed effects assumption" is often imposed in the literature; see Ackerberg and Gowrisankaran (2006) and Pakes, Ostrovsky, and Berry (2005).

In standard nonlinear panel data analysis, it is well-known that the standard fixed effects estimator will be subject to the incidental parameters problem; see Neyman and Scott (1948) or Nickell (1981). The recent panel literature shows that, for typical panel data models, the bias due to the incidental parameter problem can be significant even when the cross-sectional dimension (N) is of the same order of magnitude as the time series dimension (T). In order to rule out the incidental parameters problem in the models considered in this literature, it is theoretically necessary for $N = o(T)$; see Hahn and Kuersteiner (2002, 2004), Woutersen (2002), Carro (2007), Hahn and Newey (2004), Fernandez-Val (2005), or Arellano and Hahn (2007).

Our contribution is the qualitative prediction that the incidental parameters problem is not as severe in many game theoretic models. Unlike panel data models discussed in the recent literature, many interesting game theoretic models predict that the number of possible equilibria is *finite*. For these models, estimation of the fixed effects is equivalent to estimation of selected equilibrium out of the finite set of equilibria. We show that, under such circumstances, the incidental parameters problem is negligible even when N grows as some *exponential* function of T. Our result is a pointwise result, not a uniform result. It is therefore unclear if our result would apply when identification of equilibria may be problematic. For given data and econometric models, it is not clear whether we can ignore the error in estimation of selected equilibrium for the inference of structural parameters.

Our result is predicated on the assumption that the number of possible equilibria is finite, which is equivalent to the assumption that the support of the "fixed effects" is a finite set. We establish that the probability of correctly identifying equilibrium over the entire market in the data converges to one even when N grows as fast as some exponential function of T.

Although our result is applicable to any nonlinear panel data model where the support of the fixed effects is known to be finite, game theoretic models are the only class of models where finiteness can be theoretically proved, at least to our knowledge. We do understand that the finiteness assumption may not be applicable to all game theoretic models, but our contribution is deemed relevant for many empirically interesting models, in light of the literature such as Ackerberg and Gowrisankaran (2006), Sweeting (2004), Bajari et al. (2004).

This paper is organized as follows: In Section 2 we derive our main result under a high-level assumption on the rate at which N and T grow to infinity. In Section 3 we show that the required rate is such that N grows as fast as some exponential function of T. In Section 4 we present Monte Carlo simulations with an illustrative example of a simple entry-exit game.

2. MAIN RESULT

In this section we derive our main result under some high-level assumptions. We discuss how the degree of incidental parameters problem may be understood in terms of the relative magnitude of T and N that ensures bias-free estimation of common parameters of interest. In the next section, we use the large deviation principle and show that the relative magnitude requirement is rather mild if the number of fixed effects is finite.

We first describe our basic model. Suppose that each market i is characterized by a finite number J of equilibria. We denote the likelihood of equilibrium j by $f_{(j)}(x_{it};\theta_0)$. Here, x_{it} denotes the vector of observed outcomes and characteristics. Without loss of generality, we write the likelihood as $f(x_{it};\theta_0,\gamma_{i0})$. Our object of interest is the common parameter θ_0. Here, γ_{i0} denotes the equilibrium selected by market i. Written this way, the equilibrium selection "acts" as a fixed effect. As discussed in the introductory section, we assume that, once an equilibrium is selected, the market continues to choose the same equilibrium over time.

We consider the fixed effects maximum likelihood estimator, which solves

$$\max_{\theta,\gamma_1,\ldots,\gamma_N}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi(x_{it};\theta,\gamma_i),$$

where $\psi(x_{it};\theta,\gamma_i)=\log f(x_{it};\theta,\gamma_i)$. Hahn and Kuersteiner (2002) and Hahn and Newey (2004), among others, considered such an estimator and showed how the incidental parameters problem can be understood in the asymptotic framework where N and T grow to infinity at the same rate. Their results imply that the incidental parameters problem disappears as long as T grows to infinity sufficiently fast, i.e., $N=o(T)$.

Our purpose in this paper is to show that in the case where γ_i has finite support and satisfies a certain regularity condition discussed below, the incidental parameters problem disappears even when the time series dimension T grows at a much slower rate. In particular, we show that the fixed effects estimator does not suffer from the problem even when the cross-sectional dimension N grows at an exponential function of T.

We impose the following conditions.

Condition 1.

(i) $\varepsilon^*\equiv\inf_i\left[G_{(i)}(\theta_0,\gamma_{i0})-\sup_{\{\gamma_i\neq\gamma_{i0}\}}G_{(i)}(\theta_0,\gamma_i)\right]>0$, where $G_{(i)}(\theta,\gamma)\equiv\mathrm{E}_{(\theta_0,\gamma_{i0})}[\psi(x_t;\theta,\gamma)]$;

(ii) for all $\eta>0$, $\inf_i\left[G_{(i)}(\theta_0,\gamma_{i0})-\sup_{\{|\theta-\theta_0|>\eta,\gamma\}}G_{(i)}(\theta,\gamma)\right]>0$;

(iii) the parameter space Θ is compact; and

(iv) there exists some $M(x)$ such that $\sup_{\theta,\gamma}\left|\partial\psi(x;\theta,\gamma)/\partial\theta^k\right|\leq M(x)$ for $k=0,1$ and $\max_i\mathrm{E}\left[M(x_{it})^2\right]<\infty$.

Remark 1. Condition 1(i) is the crucial condition for our result. It requires that each equilibrium $f_{(j)}(\cdot;\cdot)$ is well separated from each other. Our proof is based on the idea that $\widehat{G}_{(i)}(\theta,\gamma) \equiv T^{-1}\sum_{t=1}^{T}\psi(x_{it};\theta,\gamma)$ uniformly converges to $G_{(i)}(\theta,\gamma)$, and that $G_{(i)}(\theta,\gamma)$ are well separated from each other.

Condition 2. For each i, $\{x_{it}, t=1,2,\ldots\}$ is strictly stationary. Furthermore, we assume that the difference, if any, of the joint distribution of $\{x_{i1}, x_{i2}, x_{i3},\ldots\}$ across i is completely characterized by the difference of γ_{i0}.

Remark 2. Condition 2 assumes that the joint density of $\{x_{it}\}_{t=1}^{T}$ can be written as $f_T(x_{i1},\ldots,x_{iT};\theta_0,\gamma_{i0})$. Equivalently put, it assumes that, for each i, the distribution of $\{x_{i1}, x_{i2}, x_{i3},\ldots\}$ is determined as one of the finite number of distributions determined by the value of γ. This condition is typically satisfied when the time series of the observed data is a time homogeneous Markov process, where $x_{it} = (y_{it}, y_{it-1})'$ and $p(y_{it}|y_{it-1};\theta_0)$ denotes the transition density of y_{it}. In this case, the stationary Markov process $\{y_{i1}, y_{i2}, y_{i3},\ldots\}$ should allow for a unique invariant measure. Meyn and Tweedie (1993, Thm. 10.1), for example, provides a sufficient condition.

Condition 3. Let θ, $\varepsilon > 0$, and $\eta > 0$ be given. There is some $h(T)$ strictly increasing in T such that, for all (γ,γ') combinations, we have

$$\Pr_{(\theta_0,\gamma')}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(\psi(x_t;\theta,\gamma) - \mathrm{E}_{(\theta_0,\gamma')}[\psi(x_t;\theta,\gamma)]\right)\right| > \frac{\eta}{3}\right] = o\left(\frac{1}{h(T)}\right), \quad \textbf{(1)}$$

$$\Pr_{(\theta_0,\gamma')}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M(x_t) - \mathrm{E}_{(\theta_0,\gamma')}[M(x_t)]\right)\right| > \frac{\eta}{3\varepsilon}\right] = o\left(\frac{1}{h(T)}\right), \quad \textbf{(2)}$$

where the (θ_0,γ') subscript above denotes the probability, and the expectation of $\{x_t\}_{t=1}^{T}$ is taken with respect to the density $f_T(x_{i1},\ldots,x_{iT};\theta_0,\gamma')$.

Remark 3. The $h(T)$ function in Condition 3 determines the required rate of growth for N and T that guarantees consistent identification of every γ_{i0}. We provide the detail on the $h(T)$ function in Section 3.

Our main result is the following theorem.

THEOREM 1. *Let*

$$\widehat{\gamma}_i(\theta) \equiv \underset{j=1,\ldots,J}{\operatorname{argmax}}\left\{\sum_{t=1}^{T}\log f_{(1)}(x_{it};\theta),\ldots,\sum_{t=1}^{T}\log f_{(J)}(x_{it};\theta)\right\},$$

$$\widehat{\theta} \equiv \underset{\theta}{\operatorname{argmax}}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi\left(x_{it};\theta,\widehat{\gamma}_i(\theta)\right),$$

$$\widetilde{\theta} \equiv \underset{\theta}{\operatorname{argmax}}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi(x_{it};\theta,\gamma_{i0}).$$

Suppose that $\sqrt{NT}\left(\widetilde{\theta}-\theta_0\right) \to N(0,\Omega)$ *for some* Ω*. Under Conditions 1, 2, and 3, we have* $\sqrt{NT}\left(\widehat{\theta}-\theta_0\right) \to N(0,\Omega)$ *if* $N \to \infty$ *and* $T \to \infty$ *such that* $N = O(h(T))$.

Proof. See Appendix, Section A.1. ■

Here $\widetilde{\theta}$ requires that the econometrician can identify the correct equilibrium for every market in the data set and, therefore, it is an infeasible estimator. Also, $\widehat{\theta}$ is a feasible estimator requiring the econometrician to estimate the selected equilibrium for each market. Therefore, $\widehat{\theta}$ can be understood to be the fixed effects estimator. In models often encountered in panel data analysis, the fixed effects are estimated from a continuum, and the incidental parameters problem disappears only when $N = o(T)$. Theorem 1 implies that the condition on N can be relaxed to $N = O\left(h(T)\right)$ when the fixed effects are assumed to be in a finite set.

Remark 4. Under the asymptotics where N grows to infinity while T is fixed, we would have $\Pr\left[\widehat{\gamma}_i \neq \gamma_{i0}\right] > 0$, even asymptotically, which would create a problem similar to the one discussed by, e.g., Leeb and Pötscher (2005). It is interesting to note how Neyman and Scott's (1948) classical incidental parameters problem is related to Leeb and Pötscher (2005) in this particular context.

In the next section, we show that $h(T)$ is typically exponential in T. Our paper therefore shows that the incidental parameter problem is far less severe if the support of the fixed effects can be assumed to be finite in models.

3. CONDITION 3 AND THE LARGE DEVIATION PRINCIPLE

In Condition 3 it is assumed that the tail probabilities tend to zero at the speed of $h(T)$. We argue in this section that the tail probabilities are typically exponential; that is, $h(T)$ is an increasing exponential function of T. By Theorem 1 it implies that $\widehat{\gamma}_i = \gamma_{i0}$ for all i with probability approaching one if N is some exponential function of T.

For simplicity of notation, we omit index i in x_{it}. Without loss of generality, we let ξ_t denote either $\psi\left(x_t;\theta_j,\gamma\right) - \mathrm{E}\left[\psi\left(x_t;\theta_j,\gamma\right)\right]$ or $M(x_t) - \mathrm{E}[M(x_t)]$, so that $\mathrm{E}[\xi_t] = 0$. We also denote $S_T = \frac{1}{T}\sum_{t=1}^{T}\xi_t$ to be the sample average of $\{\xi_1,\xi_2,\ldots,\xi_T\}$. Let P_T be the probability measure of S_T. We show that the $h(T)$ in Condition 3 is some exponential function by characterizing the bound on the tail probability of S_T.

We consider three cases: (i) when $\{\xi_1,\xi_2,\ldots\}$ is i.i.d.; (ii) when $\{\xi_1,\xi_2,\ldots\}$ may allow some serial correlation but is an α-mixing process; and (iii) when $\{\xi_1,\xi_2,\ldots\}$ may allow some serial correlation but is an ϕ-mixing process.

3.1. I.I.D.

We first consider the case where ξ_t is i.i.d. In general, we establish the exponential bound of the tail probability of S_T using the large deviation principle (LDP).[2]

We denote $\Lambda(\lambda) = \log \mathrm{E}\left[e^{\lambda \xi_t}\right]$, the logarithmic moment generating function of ξ_t and $\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\}$, the Fenchel-Legendre transform of $\Lambda(\lambda)$. The following is Cramer's theorem, which provides the LDP of P_T.

THEOREM 2 (Cramer's theorem, Dembo and Zeitouni, 1998, Thm. 2.3.3). *The sequence of measures $\{P_T\}$ satisfies the LDP with the convex rate function $\Lambda^*(x)$; that is, for any closed set $F \subset \mathbb{R}$,*

$$\limsup_T \frac{1}{T} \log P_T(F) \leq -\inf_{x \in F} \Lambda^*(x), \tag{3}$$

and for any open set $G \subset \mathbb{R}$,

$$\liminf_T \frac{1}{T} \log P_T(G) \geq -\inf_{x \in G} \Lambda^*(x).$$

The $\Lambda^*(x)$ in Theorem 2 is called a rate function.[3] A desired exponential bound can be derived if we assume

$$\Lambda^*(x) > 0 \quad \text{if } x \neq 0. \tag{4}$$

(In the Appendix, Section A.3, we provide sufficient conditions under which restriction (4) is satisfied.) Then, in view of (3) in Theorem 2 together with (4), for any $\eta > 0$, we can choose a small $\epsilon > 0$ such that $-\inf_{|x| \geq \eta} \Lambda^*(x) + \epsilon = -\inf(\Lambda^*(-\eta), \Lambda^*(\eta)) + \epsilon < 0$. Then, equation (3) implies that if T is large,

$$P\{|S_T| \geq \eta\} < \exp\left\{-T\left[\inf\left(\Lambda^*(-\eta);\ \Lambda^*(\eta)\right) - \epsilon\right]\right\} \to 0.$$

Therefore, if we set

$$h(T) = \exp\left\{T\left[\inf\left(\Lambda^*(-\eta),\ \Lambda^*(\eta)\right) - 2\epsilon\right]\right\},$$

then Condition 3 is satisfied.

3.2. α-Mixing

Now consider the case when ξ_t may be serially correlated.[4] We first look at the case when ξ_t are stationary α-mixing random vectors. Let $\|\xi\|_r$, $1 \leq r \leq \infty$ denote the L_r-norm of random variable ξ. Assume there exist constants m and M such that

$$0 < mk \leq \|\xi_{t+1} + \cdots + \xi_{t+k}\|_\infty \leq Mk \quad \text{for any } t \text{ and } k. \tag{5}$$

Notice that whenever ξ_t are bounded and there exists an $m > 0$ such that $P\{\xi_{t+1} \geq m, \ldots, \xi_{t+k} \geq m\} > 0$, then condition (5) is satisfied. Define

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^t,\ B \in \mathcal{F}_{t+k}^\infty} |P(A \cap B) - P(A)P(B)|$$

to be α-mixing coefficients of ξ_t, where $\mathcal{F}_{t_2}^{t_1}$ denotes the sigma fields generated by $\{\xi_s : t_2 \leq s \leq t_1\}$.

THEOREM 3 (Bosq, 1993, Cors. 4.1 and 4.2). *Suppose that ξ_t are stationary α-mixing random vectors that satisfy the restriction in (5).*

(a) If $\{\xi_t\}$ is k_0-dependent, that is, $\alpha(k) = 0$ if $k > k_0$, then

$$\mathbb{P}\{S_T \geq \eta\} \leq 8 \exp\left(-\frac{\eta^2}{25M}\frac{T}{k_0+1}\right), \quad \text{where } T \geq 2(k_0+1).$$

(b) If $\alpha(k) = a\rho^k$, $a > 0$, $0 < \rho < 1$, $k \geq 1$, then

$$\mathbb{P}\{S_T \geq \eta\} \leq c_1 \exp\left(-c_2\sqrt{T}\right),$$

where $n \geq 2$, and c_1 and c_2 are strictly positive constants that depend on m, M, a, ρ, η.

By Theorem 3, if $\{\xi_t\}$ is k_0-dependent, then we can take

$$h(T) = \exp\left(\left(\frac{\eta^2}{25M}\frac{1}{k_0+1} - \epsilon\right) T\right)$$

to have Condition 3 satisfied, where $\epsilon > 0$ is an arbitrarily small number. If $\alpha(k) = a\rho^k$, then it suffices to take $h(T) = \exp\left((c_2 - \epsilon)\sqrt{T}\right)$.

3.3. ϕ-Mixing

We now discuss how to use the LDP in deriving an exponential bound for the tail probabilities when ξ_t may be serially dependent stationary ϕ-mixing random vectors:

$$\phi(k) = \sup_{A \in \mathcal{F}_{-\infty}^{t},\, B \in \mathcal{F}_{t+k}^{\infty}} \{P(B|A) - P(A) : P(A) > 0\} \to 0 \quad \text{as } k \to \infty.$$

We impose the following regularity conditions: First, we assume that the mixing coefficients $\phi(k)$ tend to zero in a hypergeometric rate, i.e., $e^{ck}\phi(k) \to 0$ as $k \to \infty$ for each $c \geq 0$. We also assume that ξ_t are bounded. These two conditions lead the exponential bound to be qualitatively identical to the i.i.d. case:

THEOREM 4 (Bryc, 1992, Thm. 1). *Suppose that ξ_t are stationary hypergeometric ϕ-mixing random vectors whose support is bounded. Then S_{TT} satisfies the LDP in $\mathbb{R}$; that is, there exists a convex lower semicontinuous rate function $\Lambda^* : \mathbb{R}^d \to [0, \infty]$ with compact level sets $\Lambda^{*-1}[0, a]$, $a \geq 0$, such that*

$$\limsup_T \frac{1}{T} \log P_T(F) \leq -\inf_{x \in F} \Lambda^*(x)$$

for any closed set $F \subset \mathbb{R}$*;*

$$\liminf_T \frac{1}{T} \log P_T(G) \geq -\inf_{x \in G} \Lambda^*(x)$$

and for any open set $G \subset \mathbb{R}$*. Moreover, the limit*

$$\lim_T \frac{1}{T} \log \mathrm{E}\left[e^{T\lambda S_T}\right] = \Lambda(\lambda)$$

exists for each $\lambda \in \mathbb{R}$*, and the rate function is given by*

$$\Lambda^*(x) = \sup_\lambda (\lambda x - \Lambda(\lambda)).$$

As discussed earlier in the i.i.d. case, if $\Lambda^*(x) > 0$ for $x \neq 0$, the exponential bound for the tail probability of S_T follows as $P\{|S_T| \geq \eta\} < \exp\left(-T\left(\inf_{|x|\geq\eta} \Lambda^*(x) - \epsilon\right)\right)$ for some small $0 < \epsilon < \inf_{|x|\geq\eta} \Lambda^*(x)$. Therefore, it suffices to take

$$h(T) = \exp\left(-T\left(\inf_{|x|\geq\eta} \Lambda^*(x) - 2\epsilon\right)\right)$$

to have Condition 3 satisfied.

4. MONTE CARLO: A SIMPLE ENTRY-EXIT GAME

An illustrative example of the model that we discuss here is a simple yet extensively discussed entry-exit game (e.g., Ciliberto and Tamer, 2004) of market i at time t whose payoff matrix is summarized as follows:

		Player b	
		Exit	Enter
Player a	Exit	$(0,0)$	$\left(0, \varepsilon_{it}^b\right)$
	Enter	$\left(\varepsilon_{it}^a, 0\right)$	$\left(\theta_{a0} + \varepsilon_{it}^a, \theta_{b0} + \varepsilon_{it}^b\right)$

In the game, $\varepsilon_{it} = \left(\varepsilon_{it}^a, \varepsilon_{it}^b\right)$ is observed by players a and b but unobservable to the econometrician. We assume that $\varepsilon_{it} = \left(\varepsilon_{it}^a, \varepsilon_{it}^b\right)$ are i.i.d. across i over t, with a known distribution $F_\varepsilon(\cdot,\cdot)$. The parameter of interest is $\theta_0 = (\theta_{a0}, \theta_{b0})'$. Denote by $x_{it}^k = 0$ if player k in market i at time t chooses "Exit" and otherwise, $x_{it}^k = 1$, where $k = a, b$. Let $x_{it} = \left(x_{it}^a, x_{it}^b\right)'$. We make three additional assumptions.

Assumption 1. The parameter set for $\theta_0 = (\theta_{a0}, \theta_{b0})$ is $[-M_l, -M_u] \times [-M_l, -M_u]$, where $0 < M_u < M_l < \infty$, and the true parameter θ_0 exists in an interior of the parameter set $[-M_l, -M_u] \times [-M_l, -M_u]$.

Assumption 2. ε_{it} is a continuous random vector whose support is $\mathbb{R}^2$ and the cdf function of ε_{it} is $F_\varepsilon(\cdot,\cdot)$ and is continuously differentiable.

Assumption 3. The observed outcome x_{it} is a Nash equilibrium.

An economic justification of Assumption (i) is that a firm's profit as a monopolist is greater than as a duopolist. Examination of $f_{(1)}(x_{it};\theta)$ and $f_{(2)}(x_{it};\theta)$ derived below reveals that they are different from each other only when $\theta_0 \neq 0$. With $\theta_0 \neq 0$, the two equilibria are distinct from each other and can be correctly identified with a large enough T. On the other hand, if $\theta_0 = 0$, then there exists a unique equilibrium since the two equilibria are identical. This suggests that, when θ_0 is close to zero, the two equilibria are very similar to each other. Consequently, identification of equilibria would be very difficult. Although identification of equilibria may be difficult, it may not lead to a disastrously incorrect identification of parameter θ_0. Suppose that the econometrician incorrectly identifies the equilibrium to be 1, even though the true equilibrium is 2. When the two likelihoods are very similar, the pseudo-parameter maximizing the incorrect likelihood would tend to be very close to the true parameter. As a consequence, the difficulties of identification of equilibria and parameter seem to be inversely related when θ_0 is close to 0. We are yet to articulate and generalize this intuition in a rigorous manner.

Under these assumptions, we have

$$x_{it} = (0,0) \Leftrightarrow \varepsilon_{it}^a \leq 0,\ \varepsilon_{it}^b \leq 0,$$
$$x_{it} = (1,1) \Leftrightarrow \varepsilon_{it}^a \geq -\theta_{a0},\ \varepsilon_{it}^b \geq -\theta_{b0},$$

while

$$x_{it} = (0,1) \Rightarrow \varepsilon_{it}^a \leq -\theta_{a0},\ \varepsilon_{it}^b \geq 0,$$
$$x_{it} = (1,0) \Rightarrow \varepsilon_{it}^a \geq 0,\ \varepsilon_{it}^b \leq -\theta_{b0}.$$

The reason is that when $0 \leq \varepsilon_{it}^a \leq -\theta_{a0}$, $0 \leq \varepsilon_{it}^b \leq -\theta_{b0}$, the game has two pure strategy equilibria, $x_{it} = (0,1)$ and $(1,0)$. (That is, the equilibrium identifies the number of firms in the market, in this case, only one firm, but not which firm enters.)

We denote

$$G_{00}(\theta) = \Pr\left\{\varepsilon_{it}^a \leq 0,\ \varepsilon_{it}^b \leq 0\right\},$$
$$G_{11}(\theta) = \Pr\left\{\varepsilon_{it}^a \geq -\theta_{a0},\ \varepsilon_{it}^b \geq -\theta_{b0}\right\},$$
$$G_{01}(\theta) = \Pr\left\{\varepsilon_{it}^a \leq -\theta_{a0},\ \varepsilon_{it}^b \geq 0\right\},$$
$$G_{10}(\theta) = \Pr\left\{\varepsilon_{it}^a \geq 0,\ \varepsilon_{it}^b \leq -\theta_{b0}\right\},$$
$$G_{10,01}(\theta) = \Pr\left\{0 \leq \varepsilon_{it}^a \leq -\theta_{a0}, 0 \leq \varepsilon_{it}^b \leq -\theta_{b0}\right\}.$$

Then,

$$\Pr\{x_{it} = (0,0)\} = G_{00}(\theta_0),$$

$$\Pr\{x_{it} = (1,1)\} = G_{11}(\theta_0),$$

$$G_{01}(\theta_0) - G_{10,01}(\theta_0) \leq \Pr\{x_{it} = (0,1)\} \leq G_{01}(\theta_0),$$

$$G_{10}(\theta_0) - G_{10,01}(\theta_0) \leq \Pr\{x_{it} = (1,0)\} \leq G_{10}(\theta_0).$$

When the game has multiple equilibria, in the case of $0 \leq \varepsilon_{it}^a \leq -\theta_{a0}$, $0 \leq \varepsilon_{it}^b \leq -\theta_{b0}$, we assume players in each market select one of the pure strategy equilibria $\{(0,1), (1,0)\}$ and maintain the same selection over time. Write $\gamma_i = 1$ if market i selects $(1,0)$ and $\gamma_i = 2$ otherwise. The log likelihood is then

$$\begin{aligned}
&\log f_{(1)}(x_{it};\theta) \\
&\quad = 1\{x_{it} = (0,0)\}\log G_{00}(\theta) + 1\{x_{it} = (1,1)\}\log G_{11}(\theta) \\
&\qquad + 1\{x_{it} = (0,1)\}\log\left\{G_{01}(\theta) - G_{10,01}(\theta)\right\} + 1\{x_{it} = (1,0)\}\log\{G_{10}(\theta)\},
\end{aligned}$$

and

$$\begin{aligned}
&\log f_{(2)}(x_{it};\theta) \\
&\quad = 1\{x_{it} = (0,0)\}\log G_{00}(\theta) + 1\{x_{it} = (1,1)\}\log G_{11}(\theta) \\
&\qquad + 1\{x_{it} = (0,1)\}\log G_{01}(\theta) + 1\{x_{it} = (1,0)\}\log\left\{G_{10}(\theta) - G_{10,01}(\theta)\right\}.
\end{aligned}$$

This game satisfies conditions in our paper; see the Appendix, Section A.4.

Using the illustrative game model above, we investigated finite sample properties of the estimators $\widehat{\theta}$ and $\widetilde{\theta}$ in Theorem 1 with small-scale Monte Carlo simulations. For computational simplicity, we assume that $\theta_{a0} = \theta_{b0} = \theta_0$ and $\theta_a = \theta_b = \theta$. The data generating process of the Monte Carlo simulations is $\varepsilon_{it} = \left(\varepsilon_{it}^a, \varepsilon_{it}^b\right) \overset{iid}{\sim} N(0, I_2)$, and the ratio of the type 1 equilibrium and the type 2 equilibrium is 0.5. The sample sizes we consider are $N = \{100,\ 250\}$ and $T = \{5,\ 10,\ 20\}$. The true parameters we consider are $\theta_0 \in \{0.01,\ 0.05,\ 0.1,\ 0.5,\ 1\}$.[5]

The results of our Monte Carlo simulations are summarized in Table 1. Even with $T = 5$, the bias of $\widehat{\theta}$ due to the incidental parameters is quite small over the whole parameter set that we consider, $\{0.01,\ 0.05,\ 0.1,\ 0.5,\ 1\}$. When $T = 20$, it is almost negligible.

Although the bias is small in absolute terms, it can be large in relative terms, especially when $T = 5$. The RMSE of $\widehat{\theta}$ is about five times as large as that of $\widetilde{\theta}$ when $\theta_0 = -0.01$ and $T = 5$, although they become almost equal when $T = 20$, as is predicted by our theory. The large relative RMSE does indicate that our result should be taken with a grain of salt when it comes to parameter inference in a given finite sample.

In the beginning of this section, we mentioned that if the two equilibria are difficult to decipher, it may actually help identification of the parameter of interest θ_0. This intuition is confirmed in the Monte Carlo simulation. The absolute

TABLE 1. Monte Carlo results

θ_0	N	T	$\mathrm{E}\left[\widetilde{\theta}-\theta_0\right]$	$\sqrt{\mathrm{E}\left[\left(\widetilde{\theta}-\theta_0\right)^2\right]}$	$\mathrm{E}\left[\widehat{\theta}-\theta_0\right]$	$\sqrt{\mathrm{E}\left[\left(\widehat{\theta}-\theta_0\right)^2\right]}$	$\mathrm{E}\left[\left\lvert\widetilde{\theta}-\widehat{\theta}\right\rvert\right]$	$\sqrt{\mathrm{E}\left[\left(\widetilde{\theta}-\widehat{\theta}\right)^2\right]}$
–0.01	100	5	–0.00042	0.04685	0.01659	0.20226	0.01701	0.18459
–0.01	100	10	–0.00293	0.03350	–0.00464	0.03808	0.00171	0.00522
–0.01	100	20	0.00102	0.02292	0.00025	0.02493	0.00077	0.00299
–0.01	250	5	0.00078	0.02896	0.00069	0.06698	0.00009	0.05458
–0.01	250	10	–0.00086	0.02076	–0.00219	0.02361	0.00133	0.00366
–0.01	250	20	–0.00018	0.01472	–0.00107	0.01598	0.00089	0.00250
–0.05	100	5	–0.00121	0.04805	–0.01124	0.05862	0.01004	0.01409
–0.05	100	10	–0.00321	0.03342	–0.01029	0.03890	0.00709	0.00860
–0.05	100	20	0.00056	0.02316	–0.00393	0.02562	0.00448	0.00541
–0.05	250	5	0.00031	0.02923	–0.00941	0.03615	0.00972	0.01147
–0.05	250	10	–0.00086	0.02116	–0.00743	0.02508	0.00657	0.00745
–0.05	250	20	–0.00066	0.01509	–0.00534	0.01716	0.00469	0.00521
–0.1	100	5	–0.00109	0.04916	–0.02067	0.06224	0.01959	0.02211
–0.1	100	10	–0.00418	0.03428	–0.01770	0.04214	0.01352	0.01448
–0.1	100	20	–0.00016	0.02367	–0.00896	0.02715	0.00880	0.00935
–0.1	250	5	–0.00058	0.02992	–0.01995	0.04083	0.01937	0.02038
–0.1	250	10	–0.00155	0.02158	–0.01475	0.02837	0.01320	0.01369
–0.1	250	20	–0.00085	0.01542	–0.01002	0.01944	0.00918	0.00952

(continued)

TABLE 1. *continued.*

θ_0	N	T	$\mathrm{E}\left[\tilde{\theta}-\theta_0\right]$	$\sqrt{\mathrm{E}\left[\left(\tilde{\theta}-\theta_0\right)^2\right]}$	$\mathrm{E}\left[\hat{\theta}-\theta_0\right]$	$\sqrt{\mathrm{E}\left[\left(\hat{\theta}-\theta_0\right)^2\right]}$	$\mathrm{E}\left[\left\lvert\tilde{\theta}-\hat{\theta}\right\rvert\right]$	$\sqrt{\mathrm{E}\left[\left(\tilde{\theta}-\hat{\theta}\right)^2\right]}$
–0.5	100	5	0.00329	0.06109	–0.08655	0.11323	0.08984	0.09182
–0.5	100	10	–0.00184	0.04030	–0.05958	0.07452	0.05775	0.05870
–0.5	100	20	0.00140	0.02904	–0.03493	0.04671	0.03633	0.03699
–0.5	250	5	–0.00157	0.03735	–0.09125	0.10156	0.08968	0.09044
–0.5	250	10	–0.00123	0.02691	–0.05915	0.06644	0.05792	0.05835
–0.5	250	20	–0.00043	0.01877	–0.03738	0.04255	0.03695	0.03725
–1	100	5	–0.00043	0.08738	–0.17162	0.20503	0.17120	0.17632
–1	100	10	–0.00224	0.05929	–0.09812	0.11857	0.09588	0.09832
–1	100	20	–0.00342	0.04212	–0.05294	0.06884	0.04953	0.05085
–1	250	5	–0.00385	0.05588	–0.17250	0.18628	0.16865	0.17057
–1	250	10	–0.00164	0.03948	–0.09776	0.10755	0.09612	0.09707
–1	250	20	–0.00111	0.02775	–0.05144	0.05921	0.05033	0.05096

Note: All results are based on 1,000 Monte Carlo runs. The $\tilde{\theta}$ denotes the infeasible MLE that maximizes the likelihood with correct equilibria. The $\hat{\theta}$ denotes our estimator that maximizes the likelihood with estimated equilibria.

bias of $\widehat{\theta}$ is, in fact, larger with $\theta_0 = -1$ than with $\theta_0 = -0.01$ for most values of T.

NOTES

1. Most game theoretic econometric models are plagued with computational difficulty in addition to the problem of multiplicity of equilibria. In order to focus on the latter problem, we assume away the computational issue in this note.

2. If ξ has a bounded support, then the exponential bound can be proved by using the Hoeffding's inequality as well. See Pollard (2002), for example.

3. In the Appendix, we provide examples of $\Lambda^*(x)$ associated with some distribution functions.

4. When ξ_t is bounded, that is, $|\xi_t| \le M < \infty$, it is possible to extend the Hoeffding's inequality for a mixing process. One can find exponential inequalities for various mixing processes in Doukhan (1995), for example.

5. To estimate parameter θ_0, we employed a grid search with grid size 0.005 over the parameter set $[-2, 2]$. All the estimates ended up in an interior of the parameter set.

REFERENCES

Ackerberg, D.A. & G. Gowrisankaran (2006) Quantifying equilibrium network externalities in the ACH banking industry. *Rand Journal of Economics* 37, 738–761.

Aguirregabiria, V. & P. Mira (2004) Sequential Estimation of Dynamic Discrete Games. Working paper, University of Toronto.

Arellano, M. & J. Hahn (2007) Understanding bias in nonlinear panel models: Some recent developments. In R. Blundell, W.K. Newey, and T. Persson (eds.), *Advances in Economics and Econometrics*. Cambridge University Press.

Bajari, P., H. Hong, & S. Ryan (2004) Identification and Estimation of Discrete Games of Complete Information. NBER Working paper no. T0301.

Bosq, D. (1993) Bernstein-type large deviations inequality for partial sums of strong mixing processes. *Statistics* 24, 59–70.

Bryc, W. (1992) On large deviations for uniformly strong mixing sequences. *Stochastic Processes and their Applications* 41, 191–202.

Carro, J. (2007) Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics* 140, 503–528.

Ciliberto, F. & E. Tamer (2004) Market Structure and Multiple Equilibria in Airline Markets. Manuscript, University of Virginia and Northwestern University.

Dembo, A. & O. Zeitouni (1998) *Large Deviations Techniques and Applications.* Springer-Verlag.

Deuschel, J.D. & D.W. Stroock (1989) *Large Deviations.* Academic Press.

Doukhan, P. (1995) *Mixing: Properties and Examples.* Lecture Notes in Statistics 85, Springer-Verlag.

Fernandez-Val, I. (2005) Bias Correction in Panel Data Models with Individual Specific Parameters. Working paper, Boston University.

Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.

Hahn, J. & G. Kuersteiner (2004) Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. Working paper, UCLA.

Hahn, J. & W.K. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.

Leeb, H. & B.M. Pötscher (2005) Model selection and inference: Facts and fiction. *Econometric Theory* 21, 21–59.

Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability.* Springer.

Neyman, J. & E. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–31.

Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1426.

Pakes, A., M. Ostrovsky, & S. Berry (2005) Simple Estimators for the Parameters of Discrete Dynamic Games. Harvard Institute of Economic Research Discussion Paper no. 2036.
Pollard, D. (2002) *A User's Guide to Measure Theoretic Probability.* Cambridge University Press.
Sweeting, A. (2004) Coordination Games, Multiple Equilibria and the Timing of Radio Commercials. Working paper, Northwestern University.
Woutersen, T. (2002) Robustness Against Incidental Parameters. Working paper, University of Western Ontario.

APPENDIX

A.1. Proofs.

Proof of Theorem 1. In what follows (see Lemma 3) we show $\Pr[\exists i$, such that $\widehat{\gamma}_i \neq \gamma_{i0}] = o(1)$.

Then, since

$$\widehat{\theta} \equiv \operatorname*{argmax}_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi\left(x_{it}; \theta, \widehat{\gamma}_i\right) \quad \text{and} \quad \widetilde{\theta} \equiv \operatorname*{argmax}_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi\left(x_{it}; \theta, \gamma_{i0}\right),$$

we have

$$\begin{aligned}
\Pr\left\{\hat{\theta} \neq \tilde{\theta}\right\} &\leq \Pr\left\{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(\psi\left(x_{it}; \theta, \widehat{\gamma}_i\right) - \psi\left(x_{it}; \theta, \gamma_{i0}\right)\right) \neq 0\right\} \\
&\leq \Pr\left[\exists i, \quad \text{such that} \quad \widehat{\gamma}_i \neq \gamma_{i0}\right] \\
&= o(1).
\end{aligned}$$

The required result follows since

$$\begin{aligned}
&\left|\Pr\left\{\sqrt{NT}\left(\hat{\theta} - \theta_0\right) \leq c\right\} - \Pr\left\{\sqrt{NT}\left(\tilde{\theta} - \theta_0\right) \leq c\right\}\right| \\
&\quad \leq \left|\Pr\left\{\sqrt{NT}\left(\hat{\theta} - \theta_0\right) \leq c, \sqrt{NT}\left(\tilde{\theta} - \theta_0\right) > c\right\}\right| \\
&\quad\quad + \left|\Pr\left\{\sqrt{NT}\left(\hat{\theta} - \theta_0\right) > c, \sqrt{NT}\left(\tilde{\theta} - \theta_0\right) \leq c\right\}\right| \\
&\quad \leq \Pr\left\{\hat{\theta} \neq \tilde{\theta}\right\} + \Pr\left\{\hat{\theta} \neq \tilde{\theta}\right\}
\end{aligned}$$

for any c. ■

LEMMA 1. *Suppose that Conditions 1, 2, and 3 hold. For all $\eta > 0$, it follows that*

$$\Pr\left[\max_{1 \leq i \leq N} \sup_{(\theta,\gamma)} \left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \geq \eta\right] = o\left(\frac{N}{h(T)}\right)$$

Proof. Let $\eta > 0$ be given. By Condition 1 (iv), we can choose $\varepsilon > 0$ such that $2\varepsilon \max_i \mathrm{E}[M(x_{it})] < \eta/3$. By Condition 1 (iii), we can divide Θ into subsets $\Theta_1, \Theta_2, \ldots, \Theta_{C(\varepsilon)}$, such that $\left|\theta-\theta'\right| < \varepsilon$ whenever θ and θ' are in the same subset. Let θ_j denote *some* point in Θ_j for each j. Then,

$$\max_{1\le i\le N}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| = \max_{1\le i\le N}\max_{\gamma}\max_{j}\sup_{\Theta_j}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right|,$$

and therefore

$$\Pr\left[\max_{1\le i\le N}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| > \eta\right]$$
$$\le \sum_{j=1}^{C(\varepsilon)}\Pr\left[\max_{1\le i\le N}\max_{\gamma}\sup_{\Theta_j}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| > \eta\right]. \tag{A.1}$$

For $\theta\in\Theta_j$, we have

$$\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| \le \left|\widehat{G}_{(i)}\left(\theta_j,\gamma\right)-G_{(i)}\left(\theta_j,\gamma\right)\right|$$
$$+\frac{\varepsilon}{T}\left|\sum_{t=1}^{T}\left(M(x_{it})-\mathrm{E}[M(x_{it})]\right)\right| + 2\varepsilon\mathrm{E}[M(x_{it})].$$

Then,

$$\Pr\left[\max_{1\le i\le N}\max_{\gamma}\sup_{\Theta_j}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| > \eta\right]$$
$$\le \Pr\left[\max_{1\le i\le N}\max_{\gamma}\left|\widehat{G}_{(i)}\left(\theta_j,\gamma\right)-G_{(i)}\left(\theta_j,\gamma\right)\right| > \frac{\eta}{3}\right]$$
$$+\Pr\left[\max_{1\le i\le N}\frac{1}{T}\left|\sum_{t=1}^{T}\left(M(x_{it})-\mathrm{E}[M(x_{it})]\right)\right| > \frac{\eta}{3\varepsilon}\right]. \tag{A.2}$$

We will bound the two terms on the right-hand side of (A.2). For this purpose, we note that

$$\Pr\left[\max_{1\le i\le N}\max_{\gamma}\left|\widehat{G}_{(i)}\left(\theta_j,\gamma\right)-G_{(i)}\left(\theta_j,\gamma\right)\right| > \frac{\eta}{3}\right]$$
$$\le \sum_{\gamma=1}^{J}\sum_{i=1}^{N}\Pr_{(\theta_0,\gamma_{i0})}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(\psi\left(x_{it};\theta_j,\gamma\right)-\mathrm{E}_{(\theta_0,\gamma_{i0})}\left[\psi\left(x_{it};\theta_j,\gamma\right)\right]\right)\right| > \frac{\eta}{3}\right] \tag{A.3}$$

and

$$\Pr\left[\max_{1\le i\le N}\frac{1}{T}\left|\sum_{t=1}^{T}\left(M(x_{it})-\mathrm{E}[M(x_{it})]\right)\right| > \frac{\eta}{3\varepsilon}\right]$$
$$\le \sum_{i=1}^{N}\Pr_{(\theta_0,\gamma_{i0})}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M(x_{it})-\mathrm{E}_{(\theta_0,\gamma_{i0})}[M(x_{it})]\right)\right| > \frac{\eta}{3\varepsilon}\right]. \tag{A.4}$$

Note that the probabilities and expectations on the right-hand side in (A.3) and (A.4) are with respect to the true distribution of x_{it}, which explains the subscript (θ_0, γ_{i0}). Because there are only J possible values of γ_{i0}, we have

$$\Pr_{(\theta_0,\gamma_{i0})}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(\psi\left(x_{it};\theta_j,\gamma\right)-\mathrm{E}_{(\theta_0,\gamma_{i0})}\left[\psi\left(x_{it};\theta_j,\gamma\right)\right]\right)\right|>\frac{\eta}{3}\right]$$
$$\leq\sum_{\gamma'=1}^{J}\Pr_{(\theta_0,\gamma')}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(\psi\left(x_t;\theta_j,\gamma\right)-\mathrm{E}_{(\theta_0,\gamma')}\left[\psi\left(x_t;\theta_j,\gamma\right)\right]\right)\right|>\frac{\eta}{3}\right]$$

and

$$\Pr_{(\theta_0,\gamma_{i0})}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(x_{it}\right)-\mathrm{E}_{(\theta_0,\gamma_{i0})}\left[M\left(x_{it}\right)\right]\right)\right|>\frac{\eta}{3\varepsilon}\right]$$
$$\leq\sum_{\gamma'=1}^{J}\Pr_{(\theta_0,\gamma')}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(x_t\right)-\mathrm{E}_{(\theta_0,\gamma')}\left[M\left(x_t\right)\right]\right)\right|>\frac{\eta}{3\varepsilon}\right].$$

Note that the x_t on the right-hand side no longer has the i subscript by Condition 2. That is because x_t there simply denotes a generic random variable following the density characterized by $\left(\theta_0, \gamma'\right)$. We therefore obtain

$$\Pr\left[\max_{1\leq i\leq N}\max_{\gamma}\left|\widehat{G}_{(i)}\left(\theta_j,\gamma\right)-G_{(i)}\left(\theta_j,\gamma\right)\right|>\frac{\eta}{3}\right]$$
$$\leq\sum_{i=1}^{N}\sum_{\gamma=1}^{J}\sum_{\gamma'=1}^{J}\Pr_{(\theta_0,\gamma')}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(\psi\left(x_{it};\theta_j,\gamma\right)-\mathrm{E}_{(\theta_0,\gamma')}\left[\psi\left(x_{it};\theta_j,\gamma\right)\right]\right)\right|>\frac{\eta}{3}\right]$$

and

$$\Pr\left[\max_{1\leq i\leq N}\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(x_{it}\right)-\mathrm{E}\left[M\left(x_{it}\right)\right]\right)\right|>\frac{\eta}{3\varepsilon}\right]$$
$$\leq\sum_{i=1}^{N}\sum_{\gamma'=1}^{J}\Pr_{(\theta_0,\gamma')}\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(x_{it}\right)-\mathrm{E}_{(\theta_0,\gamma')}\left[M\left(x_{it}\right)\right]\right)\right|>\frac{\eta}{3\varepsilon}\right].$$

By Condition 3, we obtain

$$\Pr\left[\max_{1\leq i\leq N}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}\left(\theta,\gamma\right)-G_{(i)}\left(\theta,\gamma\right)\right|>\eta\right]$$
$$\leq C\left(\varepsilon\right)NJ^2o\left(\frac{1}{h(T)}\right)+C\left(\varepsilon\right)NJo\left(\frac{1}{h(T)}\right)$$
$$=o\left(\frac{N}{h(T)}\right). \qquad \blacksquare$$

LEMMA 2. *Suppose that Conditions 1, 2, and 3 hold. Then,* $\Pr\left[\left|\widehat{\theta}-\theta_0\right| \geq \eta\right] = o(N/h(T))$ *for every* $\eta > 0$.

Proof. Let η be given, and let $\varepsilon \equiv \inf_i \left[G_{(i)}(\theta_0, \gamma_{i0}) - \sup_{\{|\theta-\theta_0|>\eta,\gamma\}} G_{(i)}(\theta,\gamma)\right]$. Note that $\varepsilon > 0$ by Condition 1 (ii). With probability equal to $1-o(N/h(T))$, we have

$$\begin{aligned}\max_{|\theta-\theta_0|>\eta,\gamma_1,\ldots,\gamma_N} N^{-1}\sum_{i=1}^N \widehat{G}_{(i)}(\theta,\gamma_i) &\leq \max_{|\theta-\theta_0|>\eta,\gamma_1,\ldots,\gamma_N} N^{-1}\sum_{i=1}^N G_{(i)}(\theta,\gamma_i)+\frac{1}{3}\varepsilon\\ &< N^{-1}\sum_{i=1}^N G_{(i)}(\theta_0,\gamma_{i0})-\frac{2}{3}\varepsilon\\ &< N^{-1}\sum_{i=1}^N \widehat{G}_{(i)}(\theta_0,\gamma_{i0})-\frac{1}{3}\varepsilon,\end{aligned}$$

where the first and third inequalities are based on Lemma 1, and the second inequality is based on the definition of ε. Because

$$\max_{\theta,\gamma_1,\ldots,\gamma_N} N^{-1}\sum_{i=1}^N \widehat{G}_{(i)}(\theta,\gamma_i) \geq N^{-1}\sum_{i=1}^N \widehat{G}_{(i)}(\theta_0,\gamma_{i0})$$

by definition, we can conclude that $\Pr\left[\left|\widehat{\theta}-\theta_0\right| \geq \eta\right] = o(N/h(T))$. ■

LEMMA 3. *Suppose that Conditions 1, 2, and 3 hold. Then,* $\Pr\left[\exists i \text{ such that } \widehat{\gamma}_i \neq \gamma_{i0}\right] = o(N/h(T))$

Proof. We first prove that

$$\Pr\left[\max_{1\leq i\leq N}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \geq \eta\right] = o\left(\frac{N}{h(T)}\right), \tag{A.5}$$

for every $\eta > 0$. Note that

$$\begin{aligned}&\max_{1\leq i\leq N}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right|\\ &\quad\leq \max_{1\leq i\leq N}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\gamma\right)-G_{(i)}\left(\widehat{\theta},\gamma\right)\right|+\max_{1\leq i\leq N}\sup_{\gamma}\left|G_{(i)}\left(\widehat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right|\\ &\quad\leq \max_{1\leq i\leq N}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right|+\max_{1\leq i\leq N}\mathrm{E}\left[M(x_{it})\right]\cdot\left|\widehat{\theta}-\theta_0\right|,\end{aligned}$$

where we note that $\mathrm{E}[M(x_{it})] < \infty$ by Condition 1 (iv). Therefore,

$$\begin{aligned}&\Pr\left[\max_{1\leq i\leq N}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\gamma\right)-G_{(i)}(\theta_0,\gamma)\right| \geq \eta\right]\\ &\quad\leq \Pr\left[\max_{1\leq i\leq N}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma)-G_{(i)}(\theta,\gamma)\right| \geq \frac{\eta}{2}\right]\end{aligned}$$

$$+\Pr\left[\left|\widehat{\theta}-\theta_0\right| \geq \frac{\eta}{2\left(1+\max_i \mathrm{E}\left[M\left(x_{it}\right)\right]\right)}\right]$$

$$= o\left(\frac{N}{h(T)}\right),$$

by Lemmas 1 and 2.

Now recall that $\varepsilon^* \equiv \inf_i \left[G_{(i)}(\theta_0, \gamma_{i0}) - \sup_{\{\gamma_i \neq \gamma_{i0}\}} G_{(i)}\left(\theta_0, \gamma_i\right)\right] > 0$ by Condition 1(i). Conditional on the event $\left\{\max_{1\leq i\leq N} \sup_\gamma \left|\widehat{G}_{(i)}\left(\widehat{\theta}, \gamma\right) - G_{(i)}\left(\theta_0, \gamma\right)\right| \leq \frac{1}{3}\varepsilon^*\right\}$, which has a probability equal to $1 - o(N/h(T))$ by (A.5), we then have

$$\max_{\gamma_i \neq \gamma_{i0}} \widehat{G}_{(i)}\left(\widehat{\theta}, \gamma_i\right) < \max_{\gamma_i \neq \gamma_{i0}} G_{(i)}(\theta_0, \gamma_i) + \frac{1}{3}\varepsilon^* < G_{(i)}(\theta_0, \gamma_{i0}) - \frac{2}{3}\varepsilon^* < \widehat{G}_{(i)}\left(\widehat{\theta}, \gamma_{i0}\right) - \frac{1}{3}\varepsilon^*.$$

Since $\widehat{G}_{(i)}\left(\widehat{\theta}, \widehat{\gamma}_i\right) \geq \widehat{G}_{(i)}\left(\widehat{\theta}, \gamma_{i0}\right)$, it follows that $\hat{\gamma}_i \neq \gamma_{i0}$ for every i. ■

A.2. Examples of the Rate Functions in Theorem 2. In this section we list the Fenchel-Legendre transform of $\Lambda(\lambda)$, $\Lambda^*(x)$, of several different underlying distributions. These examples are found, for example, in Dembo and Zeitouni (1998) and Deuschel and Stroock (1989).

- For a Poisson distribution with parameter θ,

$$\Lambda^*(x) = \theta - x + x\log\left(\frac{x}{\theta}\right) \quad \text{for } x \geq 0$$
$$= \infty \quad \text{otherwise.}$$

- For a Bernoulli distribution that takes value a with probability p and b with probability $1-p$, where $a < b$,

$$\Lambda^*(x) = \frac{x-a}{b-a}\log\frac{x-a}{(1-p)(b-a)} + \frac{b-x}{b-a}\log\frac{b-x}{p(b-a)} \quad \text{if } x \in [a, b]$$
$$= \infty \quad \text{otherwise.}$$

- For an exponential distribution of parameter θ,

$$\Lambda^*(x) = \theta x - 1 - \log(\theta x) \quad \text{for } x > 0$$
$$= \infty \quad \text{otherwise.}$$

- For a normal distribution of mean θ_1 and variance θ_2,

$$\Lambda^*(x) = -\frac{(x-\theta_1)^2}{2\theta_2}.$$

- If $\Lambda_X^*(x)$ is the rate function of random variable X, and $Y = X - \mathrm{E}(X)$, then the rate function $\Lambda_Y^*(x)$ of the centered random variable Y is

$$\Lambda_Y^*(x) = \Lambda_X^*\left(x + \mathrm{E}(X)\right).$$

A.3. Sufficient Conditions for (4). A sufficient condition for (4) is that (i) $\mathrm{E}[|\xi|] < \infty$, and (ii) $\Lambda^*(x)$ is strictly convex around $\mathrm{E}[\xi]$ that is assumed to be zero in the paper. Then, at $x = \mathrm{E}[\xi] = 0$, $\Lambda^*(x) = 0$ by equation (2.2.8) of Dembo and Zeitouni (1998). Now, because $\Lambda^*(x)$ is strictly convex around zero and convex overall, $\Lambda^*(x) > 0$ if $x \neq 0$.

According to Exercise 2.2.24 in Dembo and Zeitouni (1998), $\Lambda^*(x)$ is strictly convex in the interior of $\left\{\Lambda'(\lambda) : \lambda \in \mathcal{D}_\Lambda^0\right\}$, where $\mathcal{D}_\Lambda^0$ is the interior of $\{\lambda : \Lambda(\lambda) < \infty\}$.

Notice that all the rate functions in the previous example sections satisfy restriction (4).

A.4. Entry-Exit Game and Regularity Conditions. Now we discuss how and why the game in Section 4 satisfies the conditions in our paper. For Condition 1(i) and (ii), we assume that $f_{(1)}(x_{it}; \theta_0)$ is the true model without loss of generality.

1. For Condition 1(i), notice that $0 < G_{01}(\theta_0) - G_{10,01}(\theta_0) \leq \Pr\{x_{it} = (0,1)\}$ and $0 < G_{10}(\theta_0) - G_{10,01}(\theta_0) \leq \Pr\{x_{it} = (1,0)\}$ by Assumptions 1 and 2. Also, we have $0 < G_{01}(\theta_0), G_{10}(\theta_0), G_{10,01}(\theta_0) < 1$. Then, $\log f_{(1)}(x_{it}; \theta_0) \neq \log f_{(2)}(x_{it}; \theta_0)$, and by the strict version of Jensen's inequality, it follows that $\mathrm{E} \log f_{(1)}(x_{it}; \theta_0) > \mathrm{E} \log f_{(2)}(x_{it}; \theta_0)$, where the expectation is taken by the true distribution function $f_{(1)}(x_{it}; \theta_0)$.
2. For Condition 1(ii), it is enough to show that (a) $\log f_{(1)}(x_{it}; \theta_0) \neq \log f_{(1)}(x_{it}; \theta)$ for any $\theta \neq \theta_0$, (b) $\log f_{(1)}(x_{it}; \theta_0) \neq \log f_{(2)}(x_{it}; \theta)$ for all θ, and (c) $\mathrm{E}[\log f_{(j)}(x_{it}; \theta)]$ is continuous in θ. Then, by the strict version of the Jensen's inequality, (a) and (b) imply that

$$\begin{aligned} \mathrm{E} \log f_{(1)}(x_{it}; \theta_0) > \mathrm{E} \log f_{(1)}(x_{it}; \theta) \quad & \text{for any } \theta \neq \theta_0 \\ > \mathrm{E} \log f_{(2)}(x_{it}; \theta) \quad & \text{for any } \theta. \end{aligned}$$

   Combining this with (c) and the compact parameter set assumption, the required result follows.

   Part (a) follows immediately by Assumptions 1 and 2. Part (b) holds, since $\Pr\{x_{it} = (1,1)\} = G_{11}(\theta_0) > 0$ and by Assumptions 1 and 2. Part (c) follows, since $\log f_{(j)}(x_{it}; \theta)$ is continuous in θ and by the dominated convergence theorem with $\sup_{k,\theta} \left|\log f_{(k)}(x_{it}; \theta)\right| \leq M$.
3. Condition 1(iii) follows by Assumption 1.
4. Notice that the functions $G_{\cdot\cdot}(\theta)$ and $G_{01,10}(\theta)$ are continuously differentiable functions over the compact parameter set. Therefore, it is possible to find a constant M such that

$$\sup_{k,\theta} \left|\log f_{(k)}(x_{it}; \theta)\right| \leq M \quad \text{and} \quad \sup_{k,\theta} \left|\frac{\partial \log f_{(k)}(x_{it}; \theta)}{\partial \theta}\right| \leq M,$$

   as required for Condition 1(iv).
5. Conditions 2 and 3 follow because x_{it} are i.i.d. random vectors with a finite number of states.