# SOCIAL INTERACTIONS: A GAME THEORETIC APPROACH

HAIQING XU

ABSTRACT. This paper uses a game theoretic model to capture the interactions among individuals within a social network, and establishes nonparametric identification and inference on the game structural model. Consider observations from a single equilibrium of a network game in which each player chooses an action from a finite set and is subject to interactions that are local — the interactions only occur among friends. All observations are potentially dependent on each other because they are interpreted as arising from a single equilibrium of settings where players interact directly or indirectly. Simple assumptions about the structure are made that ensure that the game has a unique equilibrium and the equilibrium has a stability property. The formulation of this stability property is new and serves as the basis for statistical inference. I establish the identification of the structural model and introduce an estimation procedure called (sieve) maximum approximated likelihood.

**Keywords:** Local interaction, social networks, incomplete information games, sieve maximum likelihood estimation, maximum approximated likelihood estimation
**JEL**: C14, C35, C62 and C72

---

*Date*: February 28, 2012.
Department of Economics, The University of Texas at Austin, BRB 3.160, Mailcode C3100, Austin, TX, 78712
h.xu@austin.utexas.edu.
I gratefully acknowledge Joris Pinkse and Sung Jae Jun for their guidance and advice. I am very indebted to Herman Bierens, Isabelle Perrigne, Quang Vuong and Neil Wallace for long and fruitful discussions about this research project. I also thank Victor Aguirregabiria, Kalyan Chatterjee, Xiaohong Chen, Edward Coulson, Frank Erickson, Paul Grieco, Han Hong, Hiroyuki Kasahara, Vijay Krishna, Margaret Slade, James Tybout, Yuanyuan Wan, Halbert White, Daniel Xu, and the participants of seminars at UBC, Toronto, UT Austin, Boston College, 2011 Cornell-PSU Macro Seminar, the 2011 North American Econometric Society Summer Meeting, and Texas Econometrics Camp for providing helpful comments.

## 1. INTRODUCTION

In many social science research areas, observations are usually dependent on each other because of the interactions among agents directly and indirectly. Common features of these social interactions are that agents are embedded in a social network and the interactions spread through the network. In this paper, I propose a simple discrete game model to describe the social interactions that are 'local' — for example, teenagers' smoking behaviors are affected by their closest friends (Nakajima (2007)). The social network is assumed to be exogenously given and captures all the local relationships among agents. I establish the existence and uniqueness of equilibrium. Simple assumptions about the structure are made to ensure that the game has a unique equilibrium and the equilibrium has a stability property. I also give identification and estimation results for the structural model in a semiparametric setup.

The structure of the model is as follows: there are $N$ players with exogenously determined locations in a social network. Each player makes a choice from a finite set, $\mathcal{A} = \{0, 1, 2, \cdots, K\}$. Player $i$'s payoff depends on not only her private information $\epsilon_i$, commonly observed characteristics $X_i$ and her own choice, but also the choices of her direct neighbors on the network. The players move simultaneously and the equilibrium concept is the pure strategy Bayesian Nash Equilibrium (BNE). Simple assumptions about the structure are made to ensure that the game has a unique equilibrium. The BNE gives rise to a conditional distribution over actions induced by the distribution of private information, where the conditioning variables are the commonly observed characteristics of all players. This game is simple enough to allow straightforward analysis, yet rich enough for us to illustrate the central issues for the social–interaction analysis. Early examples of local interactions network game include Carlsson and Van Damme (1993) and Morris and Shin (2002), etc.

Conceivably, in this network game, all observations are potentially dependent on each other because they are interpreted as arising from a single equilibrium of settings where players interact directly or indirectly based on the network configuration. The dependence structure, in contrast to the time series and spatial econometrics, is derived from the equilibrium of the network game. I characterize such a dependence structure with a stability property — the dependence between any two players' choices will vanish at an exponential rate, as the distance between them increases. The formulation of such a property is the basis for statistical inference. Such a dependence structure is different from the existing social interaction literature, e.g., Manski (1993, 2000), Brock and Durlauf (2001a,b) and others, where each individual reacts to the average behavior of the group that she belongs to, and individuals in the same group are equally affected by the averages.[1]

Several applications can be entertained in social interactions, such as obesity, drug addiction and criminality among teenagers but also the impact of social networks on job search. Most of the existing social network literature assumes that each player's payoff is not affected by her friends' current choices but the choices in the previous stage (see, e.g., Maxwell (2002) and Ennett and Bauman (1993)), which effectively assumes away the presence of 'endogenous interactions' (see Manski (2000)).

An important feature of the empirical social network game is that the data observed are usually from one (or a few) social network and the number of agents on the network is large. Most of the existing empirical game literature, e.g. Aguirregabiria and Mira (2002, 2007), Bajari, Hong, Krainer, and Nekipelov (2010), Bjorn and Vuong (1984), Bresnahan and Reiss (1991a,b), Pesendorfer and Schmidt-Dengler (2003) and Tamer (2003), assumes that the number of players is fixed and that the same game is played repeatedly in a sequence of independent local markets, or at different points in time. The number of markets is moreover assumed to grow large. In contrast, I assume that there is only a single market and that the number of players increases in the asymptotic analysis. Therefore, I consider a sequence of games indexed by the number of players. This asymptotics approach is new and more applicable in the case that observations come from one or several games and a large number of players are involved in each game.

To estimate the network game structure, I develop a new estimation method rather than rely on the two–step method pioneered by Aguirregabiria and Mira (2002) and used in much of the literature on such games. My method starts with the (pseudo) loglikelihood function, which obtains by approximating each player's equilibrium strategy by solving a much

[1]As a consequence of the react–to–average models, the dependence between any two individuals decreases with the number of members in the same group. As the number of members increases, the equilibrium solution converges to a competitive equilibrium.

smaller–sized game — player's '$h$–neighborhood' game. In player $i$'s '$h$–neighborhood' game, she and all rivals that are within her $h$ distance interact strategically with each other and all the other players outside of the distance $h$ have not been taken into account. When $h$ is small, obtaining a solution to each player's $h$–neighborhood game is computationally less expensive than the original one. By choosing appropriately the distance $h$, namely $h$ to grow with $N$ at a polynomial rate, I show that the approximated likelihood behaves asymptotically as well as the underlying likelihood.

In the existing dependent–data econometrics literature, there are few papers dealing with discrete choice models; exceptions include Pinkse and Slade (1998, 2007) and Klier and McMillen (2008). When the choice variables are discrete, the dependence pattern is inherently nonlinear in nature. In this paper, I derive this nonlinear dependence pattern from the best responses of all individuals, instead of assuming a predefined pattern. The pioneering work by Bresnahan (1987) and Seim (2006) also studied local interactions among firms with differentiated products, but focused on the endogenous product–location choices in the product space.

The organization of the paper is as follows. In the next section, I specify a static discrete choice game of incomplete information. The solution concept adopted is standard: the Bayesian Nash Equilibrium. In Section 3, I establish the identification in a semiparametric setup. In Section 4, I propose the MAL estimation approach in both parametric and semiparametric setups. Asymptotics properties of MAL estimators are established.

## 2. THE MODEL

2.1. **Game structure.** I consider a simultaneous–move game of incomplete information. There are $N$ players indexed by $i \in \mathcal{I}_N = \{1, \cdots, N\}$, with exogenously determined locations. Players simultaneously choose their actions $Y_1, \cdots, Y_N \in \mathcal{A} = \{0, 1, 2, ..., K\}$.[2]

I further assume that the payoff of player $i$ from choosing $k \in \mathcal{A}$ given other players' choices equals

$$u_{ik} = \beta_0(X_i, k) + \sum_{j \neq i} \alpha_0(k, Y_j) g_{ij} + \epsilon_i(k) \tag{1}$$

[2]Here I assume that the set of actions is identical across players. This assumption is only for notational convenience and could be relaxed.

where $X_i \in \mathcal{X} \subseteq \mathbb{R}^p$, a vector of exogenous variables, is publicly observed by all players and $\beta_0(\cdot, k) : \mathcal{X} \to \mathbb{R}$ is a choice–specific function. For each $\ell \in \mathcal{A}$, $\alpha_0(k, \ell) \in \mathbb{R}$ is the strategic effect coefficient if another player chooses $\ell$. Moreover, $g_{ij} \in \mathbb{R}^+$, for $j \neq i$, is an exogenous variable which describes the strength of the strategic effect; $g_{ij}$ will be specified later in terms of players' network locations. Let $G_i = (g_{i1}, \cdots, g_{iN})$ where $g_{ii} = 0$. $G_i$ is assumed to be public information. Finally, $\epsilon_i = (\epsilon_i(0), \cdots, \epsilon_i(K))$ is a vector of player $i$'s action–dependent payoff shocks, which is privately observed by player $i$ before actions are taken. Other players cannot observe $\epsilon_i$; however, they know how the $\epsilon_i$'s are distributed.

As is standard for discrete choice models, only the differences of the choice–specific payoff functions matter to decision–makers. It is therefore necessary to impose some normalization. Without loss of generality, I normalize the mean payoff of action 0 to zero by assuming $\beta_0(x, 0) = \alpha_0(0, \ell) = 0$ for all $x \in \mathcal{X}$ and $\ell \in \mathcal{A}$. Thereafter, let $\alpha_{0k} = (\alpha_0(k, 0), \cdots, \alpha_0(k, K))^T$ for $k \in \mathcal{A} \backslash \{0\}$. Let further $\alpha_0 = \left(\alpha_{01}^T, \alpha_{02}^T, \cdots, \alpha_{0K}^T\right)^T \in \mathbb{R}^{K(K+1)}$ and $\beta_0 = (\beta_0(\cdot, 1), \beta_0(\cdot, 2), \cdots, \beta_0(\cdot, K))^T$. Hence, $\alpha_0$ is a finite–dimensional parameter and $\beta_0$ is a vector of functions. Further, let $\theta_0 = (\alpha_0^T, \beta_0^T)^T \in \Theta = \mathbb{A} \times \mathcal{B}$, where $\mathbb{A} \subseteq \mathbb{R}^{K(K+1)}$ and $\mathcal{B}$ is a vector–valued function space.

Let $W_i = (X_i^T, G_i^T)^T$ and $S_N = (W_1^T, \cdots, W_N^T)^T$. Let further $\mathscr{S}_N$ and $\mathbb{R}^{K+1}$ be the support of $S_N$ and $\epsilon_i$ respectively. Given the structural parameter value $\theta_0$, a strategy for player $i$ is a function $r_i(S_N, \epsilon_i; \theta_0)$ which maps from player $i$'s private information $\epsilon_i$ and the public signal $S_N$ to a discrete choice $Y_i = r_i(S_N, \epsilon_i; \theta_0)$. In BNE, each player's strategy maximizes her (conditional) expected payoff given all the information available to her. Let $\{r_i^*\}_{i=1}^N$ be a BNE strategy profile. Thus player $i$'s equilibrium strategy satisfies

$$
\begin{aligned}
r_i^*(S_N, \epsilon_i; \theta_0) &= \underset{k \in \mathcal{A}}{\operatorname{argmax}} \, \mathbb{E}\left(u_{ik} | S_N, \epsilon_i\right) \\
&= \underset{k \in \mathcal{A}}{\operatorname{argmax}} \left[\beta_0(X_i, k) + \sum_{j \neq i} \sum_{\ell=0}^{K} \left\{\alpha_0(k, \ell) g_{ij} \times \mathbb{P}\left(r_j^*(S_N, \epsilon_j; \theta_0) = \ell \middle| S_N, \epsilon_i\right)\right\} + \epsilon_i(k)\right]. \quad (2)
\end{aligned}
$$

2.2. **Equilibrium characterization.** I first make an assumption on the distribution of private signals.

**Assumption A.** *The private shocks $\epsilon_i(k)$ are distributed i.i.d. across both actions and players, and conform to an extreme value distribution with density $f(t) = \exp(-t)\exp[-\exp(-t)]$.*

Let $\sigma^*_{ik}(S_N;\theta_0) = \mathbb{P}[r^*_i(S_N,\epsilon_i;\theta_0) = k|S_N]$ be the conditional choice probability of player $i$ choosing $k$ in equilibrium. Note that $\mathbb{P}\{r^*_j(S_N,\epsilon_j;\theta_0) = \ell|S_N,\epsilon_i\} = \sigma^*_{j\ell}(S_N;\theta_0)$ for $j \neq i$ when $\epsilon_i$ and $\epsilon_j$ are independent of each other. Hence, under Assumption A, obtaining an expression for $\sigma^*_{ik}(S_N;\theta_0)$ in terms of $X_i$, $G_i$ and $\sigma^*_{j\ell}(S_N;\theta_0)$ ($j \neq i$, $\ell \in \mathcal{A}$) is straightforward. Now I arrive at the following lemma

**Lemma 1.** *Let $\sigma^*_i(S_N;\theta_0) = (\sigma^*_{i0}(S_N;\theta_0),\cdots,\sigma^*_{iK}(S_N;\theta_0))$ be player i's equilibrium conditional choice probability. Then the profile $\{\sigma^*_i(S_N;\theta_0)\}_{i=1}^N$ is a one to one mapping of $\{r^*_i(\cdot,S_N;\theta_0)\}_{i=1}^N$, and a BNE solution can be obtained by solving: for all $i \in \mathcal{I}_N$ and $k \in \mathcal{A}$*

$$\sigma^*_{ik}(S_N;\theta_0) = \frac{\exp\left[\beta_0(X_i,k) + \sum_{l=0}^K \left\{\alpha_0(k,\ell)\sum_{j\neq i} g_{ij}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}{1 + \sum_{q=1}^K \exp\left[\beta_0(X_i,q) + \sum_{l=0}^K \left\{\alpha_0(q,\ell)\sum_{j\neq i} g_{ij}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}. \tag{3}$$

*Proof.* See Appendix A.1. □

Equation (3) is the common logit functional form, except that there are choice probabilities of player $i$'s neighbors on the right hand side. It is routine to apply Brouwer's fixed point theorem to prove the existence of a solution in (3), which therefore gives me the existence of a BNE solution in this game. Note that multiple solutions to equation (3) could exist and each of them would correspond to a BNE in the game. Later, I will impose further conditions to obtain a unique solution.

2.3. **Neighbors.** I assume that all players are exogenously located on a social network, and the network locations are public information. Observables $g_{ij}$ depend on player $i$'s and $j$'s locations. In this paper, I take $g_{ij} = 1$ if $j$ is $i$'s 'neighbor' on the social network and zero otherwise.[3] Note that $g_{ii} = 0$ by definition, as mentioned at the beginning of this section. A similar setup can be found in the spatial econometrics literature (see e.g. Case (1991) and Pinkse and Slade (1998)). Although restricting $g_{ij}$ to be binary is a limitation, using this

[3]The concept of neighbor can be formally defined in many ways, for instance, player $j$ is $i$'s neighbor on the social network if $j$ is one of $i$'s best friends. Note that the symmetry of the neighborship, however, is not necessarily required, i.e., $g_{ij} \neq g_{ji}$ is allowed.

makes my model more tractable. In principle, it is possible to allow for $g_{ij}$ that depend on the network distance between $i$ and $j$; I intend to pursue this possibility in future work.

Under the above specification of $g_{ij}$, equation (3) can be simplified as follows. Let $\mathcal{N}_i = \{j \in \mathcal{I}_N : g_{ij} = 1\}$ be the collection of player $i$'s neighbors. For all $i \in \mathcal{I}_N$ and all $k \in \mathcal{A}$,

$$\sigma^*_{ik}(S_N;\theta_0) = \frac{\exp\left[\beta_0(X_i,k) + \sum_{\ell=0}^{K}\left\{\alpha_0(k,\ell)\sum_{j\in\mathcal{N}_i}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}{1+\sum_{q=1}^{K}\exp\left[\beta_0(X_i,q) + \sum_{\ell=0}^{K}\left\{\alpha_0(q,\ell)\sum_{j\in\mathcal{N}_i}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}. \tag{4}$$

Hereafter, I focus on equation (4) instead of (3) to discuss the equilibrium solution.

Note that the strategic interaction between a pair of players could occur directly, if they are neighbors on the social network, or indirectly through their neighbors, or neighbors' neighbors, etc.

2.4. **Interaction strength.** To establish my results, I now restrict the interaction strength between players in my structural model, which ensures that the BNE solution is unique and has a spatial stability property. Similar to other spatial models, with too much interaction, there could be multiple equilibria, or a counterfactual change of one player's characteristics could cause the equilibrium to change radically (see Pinkse and Slade (2010)). Hence, I make the following primitive assumptions.

**Assumption B.** *There exists a constant $M \in \mathbb{N}$ (independent of $N$) such that* $\max_{i\in\mathcal{I}_N} C(\mathcal{N}_i) \leq M$ *with probability one, where $C(\mathcal{N}_i)$ is the cardinality of the set $\mathcal{N}_i$.*

Assumption B restricts the number of neighbors of any player to be less than a constant $M$. This assumption can also be found in Morris (2000) for the contagion analysis in local interaction games. In the one–dimensional spatial competition model, e.g. Salop (1979), $M$ can equal two. Note that $M$ does not depend on $N$, which restricts the pattern of neighborship formation of new players in a growing spatial structure.

**Assumption C.** $\lambda_0 = \max_{k,\ell,m\in\mathcal{A}}|\alpha_0(k,\ell) - \alpha_0(m,\ell)| \times \frac{MK}{K+1} < 1$.

Given the number of choices $(K+1)$ and the upper–bound of the number of neighbors $(M)$, Assumption C restricts the scale of the strategic effect coefficient differences. In the

Salop (1979) circle model, if players compete with each other in terms of choosing an output level and $\alpha_0(k,\ell) \leq \alpha_0(k',\ell') \leq 0$ for all $k' \leq k$ and $\ell' \leq \ell$, then Assumption C means that $\lambda_0 = -\alpha_0(K,K) \times \frac{MK}{K+1} < 1$. Assumption C plays a similar role as the requirement in autoregression models that all roots lie outside the unit circle. It should be noted that Assumption A has already implicitly imposed a normalization restriction on the scale of $\alpha_0$ since the standard error of the private signals has been assumed to be one.

Under Assumptions B and C, I establish the properties of the BNE in the next lemma. To proceed, I first introduce some notation and a definition of spatial stability condition.

For any integer $h \geq 0$, let $\mathcal{N}_{(i,h)}$ be the $h$–neighborhood of $i$, which is defined inductively,

$$\mathcal{N}_{(i,0)} = \{i\} \quad \text{and} \quad \mathcal{N}_{(i,h)} = \mathcal{N}_{(i,h-1)} \bigcup \left( \bigcup_{j \in \mathcal{N}_{(i,h-1)}} \mathcal{N}_j \right).$$

By definition, $\mathcal{N}_{(i,1)} = \{i\} \cup \mathcal{N}_i$. Let $S_N^{(i,h)}$ be all the public information within $i$'s $h$–neighborhood, i.e.

$$S_N^{(i,h)} = \left( \{X_j\}_{j \in \mathcal{N}_{(i,h)}}, \{g_{nj}\}_{n,j \in \mathcal{N}_{(i,h)}} \right). \tag{5}$$

Now I define a choice probability profile $\left\{ \sigma_j^{(i,h)} \left( S_N^{(i,h)}; \theta_0 \right) \right\}_{j \in \mathcal{N}_{(i,h)}}$ only for those players in $i$'s $h$–neighborhood, which depends only on the public information within $i$'s $h$–neighborhood, i.e. for any $j \in \mathcal{N}_{(i,h)}$

$$\sigma_{jk}^{(i,h)} \left( S_N^{(i,h)}; \theta_0 \right) = \frac{\exp\left[ \beta_0(X_j,k) + \sum_{\ell=0}^K \left\{ \alpha_0(k,\ell) \sum_{n \in \mathcal{N}_j \cap \mathcal{N}_{(i,h)}} \sigma_{n\ell}^{(i,h)}(S_N^{(i,h)}; \theta_0) \right\} \right]}{1 + \sum_{q=1}^K \exp\left[ \beta_0(X_j,q) + \sum_{\ell=0}^K \left\{ \alpha_0(q,\ell) \sum_{n \in \mathcal{N}_j \cap \mathcal{N}_{(i,h)}} \sigma_{n\ell}^{(i,h)}(S_N^{(i,h)}; \theta_0) \right\} \right]}. \tag{6}$$

By definition, the solution of equation (6) can be viewed as a BNE solution to a smaller–sized game than the original one. Players outside player $i$'s $h$–neighborhood are not taken into account. I can now formally define the spatial stability condition.

**Definition 1** (Spatial stability condition)**.** *Suppose $\{(\sigma_1^*, \cdots, \sigma_N^*)\}_{N=1}^{\infty}$ is a sequence of equilibria indexed by the number of players in the game structures described in Section 2.1. This equilibria sequence satisfies the spatial stability condition if there exists a determinant sequence, $\xi_h \downarrow 0$, such* *that for any* $h, N \in \mathbb{N}$*,*

$$\max_{i\in\mathcal{I}_N}\left\|\sigma_i^*(S_N;\theta_0)-\sigma_i^{(i,h)}\left(S_N^{(i,h)};\theta_0\right)\right\|_1\le\xi_h,\quad a.s. \tag{7}$$

The spatial stability condition implies that if one player's characteristics change, the counterfactual impact on another player's equilibrium strategy will decrease with the distance between them.

**Lemma 2.** *Suppose that Assumptions A through C hold. Then, for any* $N \in \mathbb{N}$*, (i) there exists a unique BNE; (ii) the game structure satisfies the spatial stability condition, in particular* $\xi_h = 2\lambda_0^h$*.*

*Proof.* See Appendix A.2 □

The proof of the uniqueness of the BNE involves two conditions related to the private information term: additivity in the payoff function, and independence of the private signals across players. These two conditions allow for a contraction mapping analysis being conducted in choice probability space, instead of in strategy space (see e.g. Mason and Valentinyi (2010)). Equation (7) shows that the spatial stability condition is satisfied in this (unique) BNE, where the dependence of a player's equilibrium strategy on the characteristics of other players vanishes with their distances at an exponential rate.

Spatial stability condition serves as the basis for empirical inference, when all observations come from the equilibrium of a single draw of the game, instead of a repetition of the same game. Albeit complicated, the distribution of observables can be consistently estimated under spatial stability property. Equation (7) implies that, conditioning on a player's $h$–neighborhood state variable $S_N^{(i,h)}$, the heterogeneity of the choice probability can be bounded above by a function of $h$ that decreases to zero at an exponential rate. The following result summarizes above analysis.

**Theorem 1.** *Suppose that Assumptions A through C hold. Then, for any* $N \in \mathbb{N}$ *and* $k \in \mathcal{A}$

$$\max_{i\in\mathcal{I}_N}\left|\mathbb{P}(Y_i=k|S_N)-\mathbb{P}\left(Y_i=k|S_N^{(i,h)}\right)\right|\le 4\lambda_0^h,\ \ a.s. \tag{8}$$

*Proof.* See Appendix A.3

Hence, $\mathbb{P}(Y_i = k|S_N)$ can be consistently estimated by a kernel-based nonparametric estimator of $\mathbb{P}\left(Y_i = k|S_N^{(i,h)}\right)$ with $h$ growing to infinity with $N$.[4]

## 3. Identification

I now discuss identification in the sense of Hurwicz (1950) and Koopmans and Reiersol (1950). As the limit of estimation, identification analysis is concerned with the possibility of getting a unique value $\theta_0 \in \Theta$ to rationalize the joint distribution of observables $\mathbb{P}_{Y_1,\cdots,Y_N;S_N}$. My approach is constructive, i.e., for the given $\mathbb{P}_{Y_1,\cdots,Y_N;S_N}$, I provide explicit formulas for both $\alpha_0$ and $\beta_0$ under sufficient conditions. My identification results are related to those of Bajari, Hong, Krainer, and Nekipelov (2010).

Note that $\sigma^*_{ik}(S_N;\theta_0) = \mathbb{P}(Y_i = k|S_N)$ when the equilibrium is unique under Assumptions A through C (see Lemma 2). Let $\Delta_{ik}(S_N) = \ln \mathbb{P}(Y_i = k|S_N) - \ln \mathbb{P}(Y_i = 0|S_N)$ for $k \in \mathcal{A}$. Both $\sigma^*_{ik}(S_N;\theta_0)$ and $\Delta_{ik}(S_N)$ can be derived from $\mathbb{P}_{Y_1,\cdots,Y_N|S_N}$, and hence they are identified. Note also that equation (4) gives me

$$\Delta_{ik}(S_N) = \beta_0(X_i, k) + \sum_{\ell=0}^{K} \left\{ \alpha_0(k,\ell) \sum_{j\in\mathcal{N}_i} \mathbb{P}(Y_j = \ell|S_N) \right\}, \; i \in \mathcal{I}_N, \; k \in \mathcal{A}. \tag{9}$$

I will derive an expression for $\alpha_0$ and $\beta_0$, respectively, in terms of $\{\mathbb{P}_{Y_j|S_N}\}_{j\in\mathcal{N}_i}$ and $\{\Delta_{ik}(S_N)\}_{k\in\mathcal{A}}$ from equation (9).

Recall that $S_N$ consists of all public information in the game. In equation (9), if I hold $X_i$ constant and vary $\sum_{j\in\mathcal{N}_i} \mathbb{P}(Y_j = \ell|S_N)$ by changing $S_N$, then $\alpha_0$ can be identified if an additional rank condition is satisfied. Essentially, equation (9) can, for the purpose of identification, be thought of as a partial linear model (see Robinson (1988)) and the rank condition in Assumption D originates in that literature.

[4]Here, $\mathbb{P}(Y_i = k|S_N)$ also varies with sample size $N$ and the consistency of its estimator, $\mathbb{P}(\widehat{Y_i = k}|S_N)$, is defined as follows,

$$\lim_{N\to\infty} \left| \mathbb{P}(\widehat{Y_i = k}|S_N) - \mathbb{P}(Y_i = k|S_N) \right| = 0.$$

**Assumption D.** *For* $\phi_{i\ell}(S_N) = \sum_{j \in \mathcal{N}_i} \mathbb{P}(Y_j = \ell | S_N)$ *and* $\Phi_i(S_N) = (\phi_{i0}(S_N), \cdots, \phi_{iK}(S_N))^T$,

$$\max_{i \in \mathcal{I}_N} \left| \det \left( \mathbb{E} \left[ \operatorname{Var} \left\{ \Phi_i(S_N) | X_i \right\} \right] \right) \right| > 0.^5 \tag{10}$$

Assumption D is not primitive, but testable if the conditional choice probabilities can be consistently estimated. Assumption D fails to hold when the number of neighbors is a constant for all players as in e.g. the Salop (1979) model. Indeed, then $(\phi_{i0}, \cdots, \phi_{iK})$ would be collinear, since $\sum_{k=0}^{K} \phi_{ik}$ would be a constant. In this case, a further normalization of the coefficients will be necessary, e.g. $\alpha_0(k, 0) = 0$ for all $k \in \mathcal{A}$. On the other hand, if $\mathbb{E} \left[ \operatorname{Var} \left\{ C(\mathcal{N}_i) | X_i \right\} \right] > 0$, then $\alpha_0(k, 0)$ can be identified using the variation in $C(\mathcal{N}_i)$ while $X_i$ is held fixed.

Under Assumption D, I obtain an expression for $(\alpha_0, \beta_0)$ from equation (9), which gives me the identification results. Before that, I introduce some notation. Let $U_i(S_N) = \Phi_i(S_N) - \mathbb{E} \left\{ \Phi_i(S_N) | X_i \right\}$ and $V_{ik}(S_N) = \Delta_{ik}(S_N) - \mathbb{E} \left\{ \Delta_{ik}(S_N) | X_i \right\}$ for $k \in \mathcal{A}$.

**Theorem 2.** *Suppose that Assumptions A through D hold. Then the structural parameter* $\theta_0$ *is identified, i.e.* $\overline{\theta} \neq \theta_0 \Longrightarrow \mathbb{P}_{Y_1, \cdots, Y_I; S_N}(\overline{\theta}) \neq \mathbb{P}_{Y_1, \cdots, Y_I; S_N}(\theta_0)$. *Moreover, for all* $k \in \mathcal{A}$ *and* $x \in \mathcal{X}$,

$$\alpha_{0k} = \left[ \mathbb{E} \left\{ U_i(S_N) U_i^T(S_N) \right\} \right]^{-1} \mathbb{E} \left\{ U_i(S_N) V_{ik}(S_N) \right\}, \tag{11}$$

$$\beta_0(x, k) = \mathbb{E} \left\{ \Delta_{ik}(S_N) | X_i = x \right\} - \sum_{\ell=0}^{K} \alpha_0(k, \ell) \mathbb{E} \left\{ \phi_{i\ell}(S_N) | X_i = x \right\}. \tag{12}$$

*Proof.* See Appendix A.4. □

It should be noted that, the above identification results are established under fixed $N$. However, for estimation I need $N$ to increase to infinity and I must hence establish identification in the limit, also.

**Assumption E** (Rank Condition)**.**

$$\liminf_{N \to \infty} \max_{i \in \mathcal{I}_N} \left| \det \left( \mathbb{E} \left[ \operatorname{Var} \left\{ \Phi_i(S_N) | X_i \right\} \right] \right) \right| > 0.$$

**Theorem 3.** *Suppose that Assumptions A through C, and E hold, then* $\theta_0$ *is identified when N is sufficiently large.*

[5] The determinant of a matrix A is denoted det(A).

*Proof.* See Appendix A.5. □

Analogous identification conditions can be formulated in fully parametrized models, and such conditions are more straightforward than those used in Theorem 3. If one assumes that $\beta_0(X_i, k) = X_i^T \beta_{0k}$ for $\beta_{0k} \in \mathbb{R}^p$ and $k \in \mathcal{A}\backslash\{0\}$. Let $\overline{X}_{Ni} = \left(\Phi_i^T(S_N); X_i^T\right)^T$. Then the rank condition can be directly derived from equation (9).

**Assumption F** (Rank Condition for linear–index setup)**.**

$$\liminf_{N\to\infty} \max_{i\in\mathcal{I}_N} \left|\det\left\{\mathbb{E}\left(\overline{X}_{Ni}\overline{X}_{Ni}^T\right)\right\}\right| > 0.$$

Replace Assumption E with F in Theorem 3, then the identification of $\alpha_0$ and $(\beta_{01}, \cdots, \beta_{0K})$ is straightforward for the sufficiently large $N$; moreover, for all $k \in \mathcal{A}\backslash\{0\}$,

$$\left(\alpha_{0k}^T, \beta_{0k}^T\right)^T = \left[\mathbb{E}\left\{\overline{X}_{Ni}\overline{X}_{Ni}^T\right\}\right]^{-1} \mathbb{E}\left\{\overline{X}_{Ni}\Delta_{ik}(S_N)\right\}. \tag{13}$$

## 4. Estimation

In this section, I discuss estimation of the structural parameters. The identification arguments suggest an estimator of $\theta_0$ based on nonparametrically estimated $\sigma_{ik}^*(S_N; \theta_0)$'s. However, I do not pursue this approach because the domain of the equilibrium choice probability functions increases with $N$ in the asymptotic analysis. Consequently, one would be estimating a nonparametric function whose domain is increasing with the sample size. In contrast, my MAL estimation procedure solves $\sigma_{ik}^*(S_N; \theta)$ as a fixed point of equation (4).

From hereon, I add a subscript $N$ to $\sigma_i^*$ for emphasizing the fact that the equilibrium solution depends on the number of players. As in the classical multinomial logit model, the likelihood function depends on the equilibrium choice probability of each action of each player. Formally, the likelihood function is

$$p_N(S_N) \prod_{i=1}^{N} \prod_{k=0}^{K} \left\{\sigma_{Nik}^*(S_N; \theta)\right\}^{\mathbf{1}(Y_i = k)}, \tag{14}$$

where $p_N$ is the density function of $S_N$. Because $\sigma_{Nik}^*$ does not have an analytic expression and its numerical calculation becomes costly as $N$ increases, the classical ML estimator is infeasible to compute. Instead, the basic idea of my estimation procedure is to approximate $\sigma_{Nik}^*$ by $\sigma_{ik}^{(i,h)}$, where $h$ is an integer that depends on $N$ and will be specified later. Namely, the approximated likelihood function is given by

$$p_N(S_N)\prod_{i=1}^{N}\prod_{k=0}^{K}\left\{\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)\right\}^{\mathbf{1}(Y_i=k)}. \tag{15}$$

Therefore, I define my MAL estimator $\widehat{\theta}\in\Theta$ as the maximizer of the approximated loglikelihood function

$$\widehat{L}_N^h(\theta)=\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{k=0}^{K}\left\{\mathbf{1}\left(Y_i=k\right)\ln\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)\right\}\right].$$

The choice of $h$ depends on the sample size, namely $h=h_0\times[N^\omega]$ for some constant $h_0\in\mathbb{N}$ and $w>0$, where $[a]$, for arbitrary $a\in\mathbb{R}$, is the largest interger which is no larger than $a$. The consistency of estimation requires that $h$ should increase to infinity with the sample size $N$, such that the approximation error of the likelihood function would vanish. Moreover, by choosing $h$ at a polynomial rate of $N$, I show in Theorem 5 that the MAL estimator performs asymptotically as same as the infeasible ML estimator.

It should also be noted that if $h=0$, then $\mathcal{N}_{(i,0)}=\{i\}$ and equation (6) becomes

$$\sigma_{ik}^{(i,0)}(S_N^{(i,0)},\theta)=\frac{\exp\{\beta(X_i,k)\}}{1+\sum_{q=1}^{K}\exp\{\beta(X_i,q)\}}.$$

which is the choice probability in the classical multinomial logit model.

Now, I will first illustrate my method in a parametric setup and then in a semiparametric framework. For the sake of notational simplicity, I assume that all players' choices are observed. An extension to the situation of missing observations is discussed in Section 5.

4.1. **Parametric Approach.** Here I consider a case where the payoff functions are known up to a finite–dimensional vector of parameters. In particular, I assume that $\beta_0(x,k)=x^T\beta_{0k}$ for all $k\in\mathcal{A}\backslash\{0\}$, where $\beta_{0k}\in\mathbb{R}^p$. Let $(\beta_{01}^T,\cdots,\beta_{0K}^T)^T\in\mathbb{B}\subseteq\mathbb{R}^{Kp}$. I denote $\beta_0=(\beta_{01}^T,\cdots,\beta_{0K}^T)^T\in\mathbb{R}^{Kp}$, $\theta_0=(\alpha_0^T,\beta_0^T)^T\in\mathbb{R}^L$ and $\Theta=\mathbb{A}\times\mathbb{B}\subseteq\mathbb{R}^L$, where $L=Kp+K(K+1)$.

The following assumptions are also made for the consistency of $\widehat{\theta}$.

**Assumption G.** *$\mathcal{X}$ is bounded.*

**Assumption H.** *$\Theta$ is compact.*

**Assumption I.** *There exists a* $\lambda \in (0,1)$ *such that*

$$\sup_{\alpha\in\mathbb{A}} \max_{k,\ell,m\in\mathcal{A}} |\alpha(k,\ell) - \alpha(m,\ell)| \times \frac{MK}{K+1} \leq \lambda.$$

Assumption G ensures that choice probabilities are bounded away from zero so that the likelihood function is bounded. Assumption H is standard. Assumption I strengthens Assumption C.

**Assumption J** (identical distribution)**.** *For each* $N \in \mathbb{N}$*,* $S_N$ *conforms to a probability distribution* $\mathbb{P}_{S_N}$*. For any permutation* $\{i_1, \cdots, i_N\}$ *of the player's index set* $\{1, \cdots, N\}$*, let* $S'_N$ *and* $\mathbb{P}_{S'_N}$ *be the state variable* $(W^T_{i_1}, \cdots, W^T_{i_N})$ *and its probability distribution respectively. Then* $\mathbb{P}_{S_N} = \mathbb{P}_{S'_N}$*.*

Assumption J means that the joint distribution $\mathbb{P}_{S_N}$ is symmetric across players.

**Theorem 4.** *Suppose that Assumptions A, B, and F through J hold. Then* $\widehat{\theta} \xrightarrow{p} \theta_0$*.*

*Proof.* See Appendix B.4. □

Now that the consistency of $\widehat{\theta}$ is established, I discuss its asymptotic normality. First, I introduce some notation. Let $Z_{Ni} = (Y_i, S_N)$ and $f_{Ni}(Z_{Ni}, \theta) = \prod_{k=0}^{K} \sigma^*_{Nik}(S_N; \theta)^{\mathbf{1}(Y_i=k)}$. Let further $J_{Ni}(\theta_0) = \mathbb{E}\left\{ \frac{\partial}{\partial\theta} \ln f_{Ni}(Z_{Ni}, \theta_0) \frac{\partial}{\partial\theta^T} \ln f_{Ni}(Z_{Ni}, \theta_0) \middle| S_N \right\}$ and $J_N(\theta_0) = N^{-1} \sum_{i=1}^{N} J_{Ni}(\theta_0)$.

**Assumption K.** $\theta_0$ *belongs to the interior of* $\Theta$*.*

**Assumption L.** *There exists a non–singular L by L matrix* $J_0$*, such that* $J_N(\theta_0) \xrightarrow{p} J_0$*.*

Assumption K is standard in the literature. Assumption L imposes restrictions on the sequence of game models indexed by the number of players in the asymptotic analysis. It could be derived from primitive restrictions on the probability measure of $(S_1, \cdots, S_\infty)$, for instance, fix any $h \in \mathbb{N}$, the probability distribution $\mathbb{P}_{S_N^{(i,h)}}$ converges to a limit distribution as $N \to \infty$.

**Theorem 5.** *Suppose that Assumptions A, B, F through L hold. Then* $\sqrt{N}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, J_0^{-1}\right)$*.*

*Proof.* See Appendix B.5.

Because $J_0$ is the Fisher information matrix when the number of players goes to infinity, Theorem 5 implies that the MAL estimator behaves asymptotically as well as the maximum likelihood estimator. Note that $J_0$ can be consistently estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\partial}{\partial\theta}\ln f_{Ni}^{h}(Z_{Ni},\widehat{\theta})\frac{\partial}{\partial\theta^{T}}\ln f_{Ni}^{h}(Z_{Ni},\widehat{\theta})\right\},$$

where $f_{Ni}^{h}(Z_{Ni},\theta)=\prod_{k=0}^{K}\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)^{\mathbf{1}(Y_i=k)}$.

4.2. **A Semiparametric Approach.** I now consider a semiparametric setup in which the payoff function $\beta_0$ is an unknown element of an infinite–dimensional function space $\mathcal{B}$. Below I propose a sieve estimator which I show to be consistent and asymptotically normal.

There is a variety of function spaces one may consider in nonparametric estimation. In particular, I assume that $\mathcal{B}$ is a Hölder class of functions, which is known to be well–approximated by linear sieves (see Chen (2007)).

Further, I assume that $X$ is scalar–valued. This assumption is for notational convenience only and could be relaxed. In addition, the analysis below for scalar–valued $X$ can be easily extended to single–index specifications, e.g. $\beta_0(x,k)=F_{0k}(x^T\gamma_{0k})$ with some identification restrictions (see e.g. Bierens (2008)). To avoid the curse of dimensionality in the case of vector–valued $X$, the following analysis with modifications according to a particular single–index specification can be useful in practice.

**Assumption M.** *For $q\in\mathbb{N}$, $K_0\in\mathbb{R}_+$ and $q+m>1/2$,*

$$\mathcal{B}=\Big\{\beta=(\beta_1,\cdots,\beta_K)^T:\ \beta_k:\mathcal{X}\to\mathbb{R};\quad \|\beta^{(s)}\|_{\sup}<\infty,\ s=0,\cdots,q;\quad \sup_{x_1,x_2\in\mathcal{X};\ x_1\neq x_2}\left|\beta_k^{(q)}(x_1)-\beta_k^{(q)}(x_2)\right|\leq K_0\,|x_1-x_2|^m\Big\},$$

*where $\beta_k^{(s)}$ is the s–th derivative of $\beta_k$.*

Let $\rho(\cdot,\cdot)$ be some pseudo–distance on $\Theta$. Like Shen and Wong (1994), I use trigonometric polynomials to approximate $\Theta$. Let

$$\mathcal{B}_N = \Big\{\beta:\ \beta_k(x) = a_{k0} + \sum_{j=1}^{r_N}\left\{a_{kj}\cos(2\pi jx) + b_{kj}\sin(2\pi jx)\right\},$$
$$a_{k0}^2 + \sum_{j=1}^{r_N} j^{2(q+m)}\left(a_{kj}^2 + b_{kj}^2\right) \le A_0 \ln N;\ a_{kj}, b_{kj} \in \mathbb{R} \text{ and } k \in \mathcal{A}\backslash\{0\}\Big\},$$

for some $A_0 \in \mathbb{R}_+$ and some positive integer $r_N$. Let $\Theta_N = \mathbb{A} \times \mathcal{B}_N$. Note that other linear sieves could also be used in my context, for instance, polynomials, B–splines (see Chen (2007) for more discussions).

**Assumption N.** *$\mathbb{A}$ is compact.*

**Assumption O.** $r_N = O\left(N^{\frac{1}{2q+2m+1}}\right)$.

Under Assumption O, it is known that $\rho(\pi_N\theta,\theta) = O\left(N^{-\frac{q+m}{2q+2m+1}}\right)$ (see Lorentz (1966)), where $\pi_N\theta$ is the projection of $\theta$ on the sieve space $\Theta_N$. Assumption N restricts the space of parametric part to be compact in this semiparametric setup. In contrast, compactness is not required for the space of the nonparametric part.

The estimator $\tilde{\theta} = (\tilde{\alpha},\tilde{\beta})$ is defined as follows:

$$\tilde{\theta} = \underset{\theta\in\Theta_N}{\operatorname{argmax}}\, \tilde{L}_N^h(\theta) = \underset{\theta\in\Theta_N}{\operatorname{argmax}} \sum_{i=1}^{N}\left[\sum_{k=0}^{K}\left\{\mathbf{1}\left(Y_i = k\right)\ln\tilde{\sigma}_{Nik}^{(i,h)}(S_N;\theta)\right\}\right],$$

where $\tilde{\sigma}_{Ni}^{(i,h)}(S_N;\theta)$ is defined in equation (6).

**Theorem 6.** *Suppose Assumptions A, B, E, I, and K through O hold, then*

$$\rho(\tilde{\theta},\theta_0) = O_p\left(N^{-\frac{q+m}{2q+2m+1}}\right).$$

*Proof.* See Appendix B.6 □

To derive the limiting distribution of a random process, I need to define the directional derivative in the functional space. Let $m(\cdot): \mathbb{R}^{d_Z}\times\Theta \to \mathbb{R}$, where $d_Z \in \mathbb{N}_+$. Then, for any

$\nu \in \Theta - \{\theta\}$,[6] let further

$$m_{\theta(\nu)}(Z,\theta) = \frac{\partial m(Z,\theta)}{\partial \theta(\nu)} = \lim_{t\to 0} \frac{m(Z,\theta+t\nu) - m(Z,\theta)}{t}.$$

Take $m^*(Z_{Ni}, f_{Ni}, \theta) = \ln f_{Ni}(Z_{Ni}, \theta)$. Let further

$$H_0(\tau,\nu) = \lim_{N\to\infty} \mathbb{E}\left[\frac{\partial m^*(Z_{Ni}, f_{Ni}, \theta_0)}{\partial \theta(\tau)} \times \frac{\partial m^*(Z_{Ni}, f_{Ni}, \theta_0)}{\partial \theta(\nu)}\right].$$

For any $\theta_1, \theta_2 \in \overline{\Theta}$, let $\langle \theta_1, \theta_2 \rangle$ be defined by

$$\langle \theta_1, \theta_2 \rangle = \sum_{k,\ell \in \mathcal{A}} \{\alpha_1(k,\ell)\alpha_2(k,\ell)\} + \int_{\mathcal{X}} \beta_1(x)\beta_2(x) dF_X.$$

For any $\nu \in \Theta - \{\theta_0\}$, let $\xi(\nu) \in \Theta$ be defined by

$$H_0(\xi,\tau) = \langle \nu, \tau \rangle \text{, for all } \tau \in \Theta - \{\theta_0\}.$$

In the parametric case, where $\theta_0$ is a finite–dimensional vector, $\xi(\nu) = \nu \times J^{-1}(\theta_0)$.

**Theorem 7.** *Suppose Assumptions A, B, E, I, and K through O hold. Let $\nu \in \Theta - \{\theta_0\}$. Then*

$$\sqrt{N} \langle \tilde{\theta} - \theta_0, \nu \rangle \xrightarrow{d} \mathcal{N}\Big(0, H_0\left(\xi(\nu), \xi(\nu)\right)\Big).$$

*Proof.* See Appendix B.7. □

When $\nu \in (\mathbb{A} - \alpha_0) \times \mathbf{0}_{\mathcal{B}}$, Theorem 7 provides the limiting distribution of $\tilde{\alpha}$ in sieve estimates.

## References


AGUIRREGABIRIA, V., AND P. MIRA (2002): "Swapping the nested fixed point algorithm: a class of estimators for discrete Markov decision models," *Econometrica*, 70(4), 1519–1543.

——— (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.

BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2010): "Estimating static models of strategic interactions," *Journal of Business and Economic Statistics*, 28(4), 469–482.


[6] For any $\bar{\theta} \in \Theta$, let $\Theta - \{\bar{\theta}\} = \{\theta - \bar{\theta}\}_{\theta \in \Theta}$.

BIERENS, H. (2008): "Semi-nonparametric interval-censored mixed proportional hazard models: Identification and consistency results," *Econometric Theory*, 24(03), 749–794.

BJORN, P. A., AND Q. H. VUONG (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," Working Papers 537, California Institute of Technology, Division of the Humanities and Social Sciences.

BRESNAHAN, T. (1987): "Competition and collusion in the American automobile industry: The 1955 price war," *The Journal of Industrial Economics*, 35(4), 457–482.

BRESNAHAN, T. F., AND P. C. REISS (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48(1-2), 57–81.

——— (1991b): "Entry and Competition in Concentrated Markets," *The Journal of Political Economy*, 99(5), 977–1009.

BROCK, W., AND S. DURLAUF (2001a): "Discrete choice with social interactions," *Review of Economic Studies*, 68(2), 235–260.

——— (2001b): "Interactions-based models," *Handbook of econometrics*, 5, 3297–3380.

CARLSSON, H., AND E. VAN DAMME (1993): "Global games and equilibrium selection," *Econometrica: Journal of the Econometric Society*, pp. 989–1018.

CASE, A. C. (1991): "Spatial Patterns in Household Demand," *Econometrica*, 59(4), 953–65.

CHEN, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6, 5549–5632.

ENNETT, S., AND K. BAUMAN (1993): "Peer group structure and adolescent cigarette smoking: a social network analysis," *Journal of Health and Social Behavior*, 34(3), 226–236.

HURWICZ, L. (1950): "Generalization of the Concept of Identification," *Statistical inference in dynamic economic models*, 10.

KLIER, T., AND D. MCMILLEN (2008): "Clustering of Auto Supplier Plants in the United States," *Journal of Business and Economic Statistics*, 26(4), 460–471.

KOOPMANS, T., AND O. REIERSOL (1950): "The identification of structural characteristics," *The Annals of Mathematical Statistics*, pp. 165–181.

LORENTZ, G. (1966): "Approximation of Functions.–Holt, Rinehart and Wilson," *Inc., New York*.

MANSKI, C. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60(3), 531–542.

——— (2000): "Economic analysis of social interactions," *The Journal of Economic Perspectives*, 14(3), 115–136.

MASON, R., AND A. VALENTINYI (2010): "The existence and uniqueness of monotone pure strategy equilibrium in Bayesian games," *Discussion papers in Economics and Econometrics, University of Southampton*, 710.

MAXWELL, K. (2002): "Friends: The role of peer influence across adolescent risk behaviors," *Journal of Youth and Adolescence*, 31(4), 267–277.

MORRIS, S. (2000): "Contagion," *Review of Economic Studies*, 67(1), 57–78.

MORRIS, S., AND H. SHIN (2002): "Global Games: Theory and Applications, forthcoming in Advances in Economics and Econometrics," in *the Eighth World Congress (M. Dewatripont, L. Hansen and S. Turnovsky, eds.), Cambridge University Press*.

NAKAJIMA, R. (2007): "Measuring peer effects on youth smoking behaviour," *Review of Economic Studies*, 74(3), 897–935.

NEWEY, W., AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2003): "Identification and estimation of dynamic games," *NBER working paper*.

PINKSE, J., AND M. SLADE (2007): "Semi-structural models of advertising competition," *Journal of Applied Econometrics*, 22(7), 1227–1246.

PINKSE, J., AND M. SLADE (2010): "THE FUTURE OF SPATIAL ECONOMETRICS*," *Journal of Regional Science*, 50(1), 103–117.

PINKSE, J., AND M. E. SLADE (1998): "Contracting in space: An application of spatial statistics to discrete-choice models," *Journal of Econometrics*, 85(1), 125–154.

POLLARD, D. (1990): *Empirical processes: theory and applications*. Inst of Mathematical Statistic.

ROBINSON, P. (1988): "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, 56(4), 931–954.

SALOP, S. (1979): "Monopolistic competition with outside goods," *The Bell Journal of Economics*, 10(1), 141–156.

SEIM, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *The RAND Journal of Economics*, 37(3), 619–640.

SHEN, X. (1997): "On methods of sieves and penalization," *The Annals of Statistics*, pp. 2555–2591.

SHEN, X., AND W. WONG (1994): "Convergence rate of sieve estimates," *The Annals of Statistics*, 22(2), 580–615.

TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70(1), 147–165.

VAN DER VAART, A. (2000): *Asymptotic statistics*. Cambridge Univ Pr.

## APPENDIX A.

A.1. **proof of Lemma 1.** By definition, $\{\sigma_i^*(S_N;\theta_0)\}_{i=1}^N$ can be derived from the equilibrium strategy profile $\{r_i^*(\cdot,S_N;\theta_0)\}_{i=1}^N$. So it suffices to show that $\{r_i^*(\cdot,S_N;\theta_0)\}_{i=1}^N$ can also be induced from $\{\sigma_i^*(S_N;\theta_0)\}_{i=1}^N$. To see this, note that equation (2) can also be written as

$$r_i^*(\epsilon_i,S_N;\theta_0)=\operatorname*{argmax}_{k\in\mathcal{A}}\left[\beta_0(X_i,k)+\sum_{j\neq i}\sum_{\ell=0}^{K}\left\{\alpha_0(k,\ell)g_{ij}\sigma_{jl}^*(S_N;\theta_0)\right\}+\epsilon_i(k)\right].$$

□

A.2. **proof of Lemma 2.** I first establish (ii) by mathematical induction. Fix $i,h,s$. I show that for all $q=1,\cdots,h$, there is

$$\max_{j\in\mathcal{N}_{(i,h-q)}}\left\|\sigma_j^*(s;\theta_0)-\sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1\leq 2\lambda_0^q. \tag{16}$$

Lemma A.6 implies that (16) holds for $q=1$. Moreover, if for any $q\leq q_0\in\{1,\cdots,h-1\}$, (16) is satisfied. Then, I need to show that (16) also holds for $q=q_0+1$. Lemma A.6 implies that for any $j\in\mathcal{N}_{(i,h-q_0-1)}\subset\mathcal{N}_{(i,h-1)}$, there is

$$\left\|\sigma_j^*(s;\theta_0)-\sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1\leq\lambda_0\times\max_{n\in\mathcal{N}_j}\left\|\sigma_n^*(s;\theta_0)-\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1. \tag{17}$$

Note that $n\in\mathcal{N}_j$ and $j\in\mathcal{N}_{(i,h-q_0-1)}$ implies that $n\in\mathcal{N}_{(i,h-q_0)}$. Thus, by equation (16)

$$\begin{aligned}&\max_{n\in\mathcal{N}_j}\left\|\sigma_n^*(s;\theta_0)-\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1\\ &\qquad\leq\max_{n\in\mathcal{N}_{(i,h-q_0)}}\left\|\sigma_n^*(s;\theta_0)-\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1\leq 2\lambda_0^{q_0}.\end{aligned} \tag{18}$$

Thus, equations (17) and (18) imply that (16) also holds for $q_0+1$. Under Assumption C, $2\lambda_0^h\downarrow 0$ as $h\to\infty$.

I then establish (i) by contradiction. Suppose that there are two equilibria $\{\sigma_n^*(s;\theta_0)\}_{n=1}^N$ and $\{\overline{\sigma}_n(s;\theta_0)\}_{n=1}^N$ for some $N\in\mathbb{N}$ and $s\in\mathscr{S}_N$. Using a similar argument as that in Lemma A.6, for any $i\in\mathcal{I}_N$

$$\|\sigma_i^*(s;\theta_0)-\overline{\sigma}_i(s;\theta_0)\|_1\leq\lambda_0\times\max_{j\in\mathcal{N}_i}\left\|\sigma_j^*(s;\theta_0)-\overline{\sigma}_j(s;\theta_0)\right\|_1\leq\lambda_0\times\max_{j\in\mathcal{I}_N}\left\|\sigma_j^*(s;\theta_0)-\overline{\sigma}_j(s;\theta_0)\right\|_1.$$

Hence,

$$\max_{i\in\mathcal{I}_N} \|\sigma_i^*(s;\theta_0) - \overline{\sigma}_i(s;\theta_0)\|_1 \le \lambda_0 \times \max_{j\in\mathcal{I}_N} \left\|\sigma_j^*(s;\theta_0) - \overline{\sigma}_j(s;\theta_0)\right\|_1.$$

Since $0 < \lambda_0 < 1$, contradiction. □

A.3. **Proof of Theorem 1.** From Lemma 2, for any $N \in \mathbb{N}$ and $k \in \mathcal{A}$

$$\max_{i\in\mathcal{I}_N} \left|\mathbb{P}(Y_i = k|S_N) - \sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta_0\right)\right| \le 2\lambda_0^h, \;\; a.s.$$

and

$$\max_{i\in\mathcal{I}_N} \left|\mathbb{P}\left(Y_i = k|S_N^{(i,h)}\right) - \sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta_0\right)\right| \le 2\lambda_0^h, \;\; a.s..$$

Hence, equation (8) can be derived from above two equations. □

A.4. **Proof of Theorem 2.** My proof of identification is constructive. By definition,

$$V_{ik}(S_N) = \Delta_{ik}(S_N) - \mathbb{E}\left\{\Delta_{ik}(S_N)|X_i\right\} = \sum_{\ell=0}^{K}\left\{\alpha_0(k,\ell)U_{i\ell}(S_N)\right\}.$$

Under Assumption D, there exists an $i \in \mathcal{I}_N$ such that $\mathbb{E}\left\{U_i(S_N)U_i^T(S_N)\right\}$ is invertible. Hence

$$\alpha_{0k} = \left[\mathbb{E}\left\{U_i(S_N)U_i^T(S_N)\right\}\right]^{-1}\left[\mathbb{E}\left\{U_i(S_N)V_{ik}(S_N)\right\}\right].$$

Moreover, $\beta_0(\cdot,k)$ is also identified by

$$\beta_0(X_i,k) = \Delta_{ik}(S_N) - \sum_{\ell=0}^{K}\alpha_0(k,\ell)\phi_{i\ell}(S_N).$$

□

A.5. **Proof of Theorem 3.** Note that when $N$ is sufficient large, there is

$$\max_{i\in\mathcal{I}_N}\left|\det\left(\mathbb{E}\left[\,\mathrm{Var}\left\{\Phi_i(S_N)|X_i\right\}\right]\right)\right| > 0.$$

Then the identification result follows Theorem 2. □

A.6. **Lemma A.6.** Suppose that Assumptions A through C hold. Thus, for any $h, N \in \mathbb{N}$, $s \in \mathscr{S}_N$, $i \in \mathcal{I}$ and $j \in \mathcal{N}_{(i,h-1)}$,

$$\left\| \sigma_j^*(s;\theta_0) - \sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1 \leq \lambda_0 \times \max_{n\in\mathcal{N}_j} \left\| \sigma_n^*(s;\theta_0) - \sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1.$$

In particular

$$\left\| \sigma_j^*(s;\theta_0) - \sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1 \leq 2\lambda_0.$$

*Proof.* To begin with, let me first introduce some notation. Let $\overline{\Delta}(\alpha) = \max_{k\ell,m\in\mathcal{A}} |\alpha(k,\ell) - \alpha(m,\ell)|$. For any $i \in \mathcal{I}_N$ and $k \in \mathcal{A}$, let

$$\Gamma_{ik}\left(W_i, \{\sigma_j\}_{j\in\mathcal{N}_i}, \theta\right) = \frac{\exp\left[\beta(X_i,k) + \sum_{\ell=0}^K \left\{\alpha(k,\ell)\sum_{n\in\mathcal{N}_i}\sigma_{j\ell}\right\}\right]}{1+\sum_{q=1}^K \exp\left[\beta(X_i,q) + \sum_{\ell=0}^K \left\{\alpha(q,\ell)\sum_{n\in\mathcal{N}_i}\sigma_{j\ell}\right\}\right]}. \tag{19}$$

Then, equation (4) can be written as

$$\sigma_{ik}^*(S_N;\theta_0) = \Gamma_{ik}\left(W_i, \{\sigma_j^*(S_N;\theta_0)\}_{j\in\mathcal{N}_i}, \theta_0\right).$$

Now, fix $h, N, s$. For any $i \in \mathcal{I}$ and $j \in \mathcal{N}_{(i,h-1)}$,

$$\begin{aligned}
&\left\| \sigma_j^*(s;\theta_0) - \sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1 \\
&\quad = \left\| \Gamma_j\left(w_j, \{\sigma_n^*(s;\theta_0)\}_{n\in\mathcal{N}_j};\theta_0\right) - \Gamma_j\left(w_j, \left\{\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\}_{n\in\mathcal{N}_j};\theta_0\right) \right\|_1 \\
&\quad = \sum_{k=0}^K \left| \sum_{n\in\mathcal{N}_j}\sum_{\ell=0}^K \frac{\partial}{\partial\sigma_{n\ell}}\Gamma_{jk}(w_j, \{\sigma_n^\dagger\}_{n\in\mathcal{N}_j};\theta_0)\left\{\sigma_{n\ell}^*(s;\theta_0) - \sigma_{n\ell}^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\} \right|,
\end{aligned}$$

where $\{\sigma_n^\dagger\}_{n\in\mathcal{N}_j}$ is a choice probability profile between $\{\sigma_n^*(s;\theta_0)\}_{n\in\mathcal{N}_j}$ and $\left\{\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\}_{n\in\mathcal{N}_j}$. Because of the definition of $\Gamma_{jk}(w_j, \Sigma_N;\theta)$, for any $n \in \mathcal{N}_j$

$$\frac{\partial\Gamma_{jk}}{\partial\sigma_{n\ell}} = \Gamma_{jk}\sum_{q\neq k}\left[\Gamma_{jq}\left\{\alpha_0(k,\ell) - \alpha_0(q,\ell)\right\}\right].$$

Moreover, (i) $0 \leq \Gamma_{jk} \leq 1$; (ii) $\sum_{k=0}^K \Gamma_{jk} = 1$. Thus

$$\sum_{k=0}^K \left|\frac{\partial\Gamma_{jk}}{\partial\sigma_{n\ell}}\right| \leq \bar{\Delta}(\alpha_0) \times \sum_{k=0}^K \Gamma_{jk}\sum_{q\neq k}\Gamma_{jq} = \bar{\Delta}(\alpha_0) \times \sum_{k=0}^K \left\{\Gamma_{jk}\left(1-\Gamma_{jk}\right)\right\} \leq \frac{\bar{\Delta}(\alpha_0)K}{K+1}.$$

The last step comes from the fact that $\sum_{k=0}^{K}\left\{\Gamma_{jk}\left(1-\Gamma_{jk}\right)\right\} \leq K/(K+1)$ for any $\Gamma_j$ satisfying (i) and (ii). Hence,

$$\begin{aligned}
&\left\|\sigma_j^*(s;\theta_0)-\sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1 \\
&\leq \sum_{n\in\mathcal{N}_j}\sum_{\ell=0}^{K}\sum_{k=0}^{K}\left|\frac{\partial}{\partial\sigma_{n\ell}}\Gamma_{jk}(w_j,\Sigma_N^{\dagger};\theta_0)\right| \times \left|\sigma_{n\ell}^*(s;\theta_0)-\sigma_{n\ell}^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right| \\
&\leq \frac{\bar{\Delta}(\alpha_0)MK}{K+1} \times \max_{n\in\mathcal{N}_j}\left\|\sigma_n^*(s;\theta_0)-\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1 \\
&= \lambda_0 \times \max_{n\in\mathcal{N}_j}\left\|\sigma_n^*(s;\theta_0)-\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1 .
\end{aligned}$$

In particular, for any $n \in \mathcal{I}_N$

$$\left\|\sigma_n^*(s;\theta_0)-\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1 \leq \left\|\sigma_n^*(s;\theta_0)\right\|_1 + \left\|\sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right)\right\|_1 = 2.$$

□

## APPENDIX B.

In this section, I provide proofs for asymptotic analysis. As mentioned in Section 4.1, I add subscript N to $\sigma_i^*(S_N;\theta)$. Let $\widehat{L}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}\ln f_{Ni}(Z_{Ni},\theta)$ and $L_N(\theta) = \mathbb{E}\ln f_{N1}(Z_{N1},\theta)$. Moreover, let $f_{Ni}^h(Z_{Ni},\theta) = \prod_{k=0}^{K}\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)^{\mathbf{1}(Y_i=k)}$ and $\widehat{L}_N^h(\theta) = \frac{1}{N}\sum_{i=1}^{N}\ln f_{Ni}^h(Z_{Ni},\theta)$. Let further $\widehat{G}_N(\theta) = \partial\widehat{L}_N(\theta)/\partial\theta$ and $\widehat{G}_N^h(\theta) = \partial\widehat{L}_N^h(\theta)/\partial\theta$.

As convention, for arbitrary $\epsilon > 0$, let $B_\epsilon(\theta_0)$ be an open ball centered at $\theta_0$ with $\epsilon$ radius in the space $\Theta$.

B.1. **Lemma B.1.** Assume (i) For any $\epsilon > 0$, $\limsup_N \sup_{\theta\in\Theta\cap B_\epsilon^c(\theta_0)}\left\{L_N(\theta)-L_N(\theta_0)\right\} < 0$; (ii) $\widehat{L}_N(\theta)$ converges uniformly in probability to $L_N(\theta)$, i.e.

$$\sup_{\theta\in\Theta}\left|\widehat{L}_N(\theta)-L_N(\theta)\right| \xrightarrow{p} 0;$$

(iii) $\widehat{L}_N(\widehat{\theta}) \geq \sup_{\theta\in\Theta}\widehat{L}_N(\theta) - o_p(1)$. Then $\widehat{\theta} \xrightarrow{p} \theta_0$.

*Proof.* The following proofs constructed are basically based on the proofs in Newey and McFadden (1994), Theorem 2.1. First, by (ii) and (iii), with propability approaching one (w.p.a.1),

$$L_N(\widehat{\theta}) > \widehat{L}_N(\widehat{\theta}) - \eta/3 > \widehat{L}_N(\theta_0) - 2\eta/3 > L_N(\theta_0) - \eta$$

Thus for any $\eta > 0$, $L_N(\widehat{\theta}) > L_N(\theta_0) - \eta$ w.p.a.1.

Next, for any $\epsilon > 0$, choose $\eta = -\frac{1}{2}\limsup_N \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\}$. It follows that w.p.a.1,

$$L_N(\widehat{\theta}) - L_N(\theta_0) > \frac{1}{2}\limsup_N \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\}.$$

Because for sufficient large $N$,

$$\sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\} - \limsup_N \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\} \leq -\frac{1}{2}\limsup_N \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\},$$

hence, for sufficient large $N$,

$$\frac{1}{2}\limsup_N \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\} \geq \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\}.$$

Hence, w.p.a.1,

$$L_N(\widehat{\theta}) - L_N(\theta_0) > \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\},$$

which implies that $\widehat{\theta} \in B_\epsilon(\theta_0)$ w.p.a.1. Because $\epsilon$ can be arbitrarily small, $\widehat{\theta} \xrightarrow{p} \theta_0$. □

**Remark 1.** *In the standard case for ML estimation with $L_N(\theta) = L(\theta)$ (see e.g. Newey and McFadden (1994), Theorem 2.1), condition (i) can be satisfied if one assume (1) $\Theta$ is compact; (2) $L(\theta)$ is continuous in $\theta$; and (3) $\theta_0$ is a unique maximizer of $L(\theta)$. It should also be noted that condition (iii) is a non–trivial statement here, since $\hat{\theta}$ is a maximizer of $\widehat{L}_N^h$, instead of $\widehat{L}_N$.*

B.2. **Lemma 3.** Suppose that Assumptions A, B, and I hold. Then,

$$\widehat{L}_N(\widehat{\theta}) \geq \sup_{\theta \in \Theta} \widehat{L}_N(\theta) - o_p(1).$$

*Proof.* Because $\widehat{\theta}$ is a maximizer of $\widehat{L}_N^h(\theta)$, it is sufficient to show that

$$\sup_{\theta}\left|\widehat{L}_N^h(\theta)-\widehat{L}_N(\theta)\right| \to 0.$$

Note that

$$\sup_{\theta}\left|\widehat{L}_N^h(\theta)-\widehat{L}_N(\theta)\right| \le \sup_{\theta}\frac{1}{N}\sum_{i=1}^{N}\left|\ln f_{Ni}^h(Z_{Ni},\theta)-\ln f_{Ni}(Z_{Ni},\theta)\right|$$
$$\le \sup_{\theta}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K}\left|\ln \sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)-\ln \sigma_{Nik}^{*}(S_N;\theta)\right|.$$

By Taylor expansion, for any $k \in \mathcal{A}$ and $i \in \mathcal{I}_N$

$$\left|\ln \sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)-\ln \sigma_{Nik}^{*}(S_N;\theta)\right| = \left|\frac{1}{\widetilde{\sigma}}\left\{\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)-\sigma_{Nik}^{*}(S_N;\theta)\right\}\right|$$
$$\le \frac{1}{\sigma_L}\left|\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)-\sigma_{Nik}^{*}(S_N;\theta)\right| \le \frac{2\lambda^h}{\sigma_L}, \ \ a.s.$$

where $\widetilde{\sigma}$ is some real value between $\sigma_{ik}^{(i,h)}\left(S_N^{(i,h)};\theta\right)$ and $\sigma_{Nik}^{*}(S_N;\theta)$, and the last two steps come from Lemma C.1 and Lemma 2, respectively.

Hence

$$\sup_{\theta}\left|\widehat{L}_N^h(\theta)-\widehat{L}_N(\theta)\right| \le \frac{2(K+1)\lambda^h}{\sigma_L}, \ \ a.s.$$

By the choice of $h$ and the fact that $\lambda < 1$,

$$\sup_{\theta}\left|\widehat{L}_N^h(\theta)-\widehat{L}_N(\theta)\right| \xrightarrow{p} 0.$$

□

Note that for the standard ML estimator, (iii) is a trivial statement and conditions (i) and (ii) are sufficient for its consistency. Given Lemma 3, the following argument for the consistency of MAL estimator can also apply to the ML estimator.

B.3. **Lemma 4.** Suppose that Assumptions A, B, and F through I hold. Then, For any $\epsilon > 0$, $\limsup_N \sup_{\theta\in\Theta\cap B_\epsilon^c(\theta_0)}\{L_N(\theta)-L_N(\theta_0)\} < 0$.

*Proof.* Proof by contradiction. Because $\theta_0$ is identified (see Theorem 2), for any fixed $N$, there is $L_N(\theta)-L_N(\theta_0) < 0$ for all $\theta$. Hence, $\limsup_N \sup_{\theta\in\Theta\cap B_\epsilon^c(\theta_0)}\{L_N(\theta)-L_N(\theta_0)\} \le 0$.

Suppose $\limsup_N \sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_N(\theta) - L_N(\theta_0)\} = 0$ for some $\epsilon > 0$. Then there exists a sequence $\{N_t\}_{t=1}^\infty$, such that

$$\sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_{N_t}(\theta) - L_{N_t}(\theta_0)\} \to 0.$$

Because $\Theta$ is compact and by Lemma C.2, $L_{N_t}(\cdot)$ is continuous for every $N_t$, there exists a sequence $\{\theta_{N_t}\}_{t=1}^\infty$ in $\Theta \cap B_\epsilon^c(\theta_0)$, such that $\sup_{\theta \in \Theta \cap B_\epsilon^c(\theta_0)} \{L_{N_t}(\theta) - L_{N_t}(\theta_0)\} = L_{N_t}(\theta_{N_t}) - L_{N_t}(\theta_0)$. Note that $L_N(\theta) = \mathbb{E}\left[\sum_{k=0}^K \left\{\sigma_{N1k}^*(S_N;\theta_0) \ln \sigma_{N1k}^*(S_N;\theta)\right\}\right]$. Therefore,

$$\mathbb{E}\left[\sum_{k=0}^K \sigma_{N1k}^*(S_N;\theta_0) \left\{\ln \sigma_{N1k}^*(S_N;\theta_{N_t}) - \ln \sigma_{N1k}^*(S_N;\theta_0)\right\}\right] \to 0.$$

Since $\sum_{k=0}^K \sigma_{Nik}^*(S_N;\theta_0) \left\{\ln \sigma_{N1k}^*(S_N;\theta_{N_t}) - \ln \sigma_{N1k}^*(S_N;\theta_0)\right\} \leq 0$ a.s. under the standard argument, then

$$\sum_{k=0}^K \sigma_{N1k}^*(S_N;\theta_0) \left\{\ln \sigma_{N1k}^*(S_N;\theta_{N_t}) - \ln \sigma_{N1k}^*(S_N;\theta_0)\right\} \xrightarrow{p} 0.$$

By Taylor expansion,

$$\begin{aligned}
&\sum_{k=0}^K \sigma_{N1k}^*(S_N;\theta_0) \left\{\ln \sigma_{N1k}^*(S_N;\theta_{N_t}) - \ln \sigma_{N1k}^*(S_N;\theta_0)\right\} \\
&\qquad = -\sum_{k=0}^K \frac{\sigma_{N1k}^*(S_N;\theta_0)}{\overline{\sigma}^2} \left\{\sigma_{N1k}^*(S_N;\theta_{N_t}) - \sigma_{N1k}^*(S_N;\theta_0)\right\}^2 \\
&\qquad\qquad \leq -\sigma_L \sum_{k=0}^K \left\{\sigma_{N1k}^*(S_N;\theta_{N_t}) - \sigma_{N1k}^*(S_N;\theta_0)\right\}^2 \leq 0
\end{aligned}$$

for some $\overline{\sigma}$ between $\sigma_{N1k}^*(S_N;\theta_{N_t})$ and $\sigma_{N1k}^*(S_N;\theta_0)$, where the last step is because of Lemma C.1. Hence,

$$-\sigma_L \sum_{k=0}^K \left\{\sigma_{N1k}^*(S_N;\theta_{N_t}) - \sigma_{N1k}^*(S_N;\theta_0)\right\}^2 \xrightarrow{p} 0.$$

So $\sigma_{N1k}^*(S_N;\theta_{N_t}) - \sigma_{N1k}^*(S_N;\theta_0) \xrightarrow{p} 0$ for all $k \in \mathcal{A}$.

Because of Assumption J, the distribution of $S_N$ is symmetric over players, then for any $\eta > 0$, $\mathbb{P}\left[\left|\sigma_{Nik}^*(S_N;\theta_{N_t}) - \sigma_{Nik}^*(S_N;\theta_0)\right| > \eta\right] = \mathbb{P}\left[\left|\sigma_{N1k}^*(S_N;\theta_{N_t}) - \sigma_{N1k}^*(S_N;\theta_0)\right| > \eta\right] \to 0.$

Thus, by the identification analysis in Theorem 2, there is $\|\theta_{N_t} - \theta_0\| \to 0$, which is a contradiction to the definition of $\theta_{N_t}$. □

B.4. **Proof of Theorem 4.** In this proof, I'll check the conditions in Lemma B.1. By Lemma 3 and 4, it suffices to show the uniform convergence of $\widehat{L}_N(\theta)$ to $L_N(\theta)$, i.e.

$$\sup_{\theta \in \Theta} \left| \widehat{L}_N(\theta) - L_N(\theta) \right| \xrightarrow{p} 0.$$

By Lemma C.1 and C.2, $\ln f_{Ni}(Z_{Ni}, \theta)$ is a bounded continuous function in $\theta$. Since $\Theta$ is compact, then $\mathcal{F}_N = \{\ln f_{N1}(Z_{N1}, \theta) : \theta \in \Theta\}$ can be covered by a finite number of $\epsilon$–brackets. To apply the classical Glivenko-Cantelli argument, it suffices to show the pointwise LLN, i.e. for any $\theta \in \Theta$

$$\widehat{L}_N(\theta) - L_N(\theta) \xrightarrow{p} 0.$$

Because

$$\begin{aligned}
\mathbb{E}\left\{\widehat{L}_N(\theta) - L_N(\theta)\right\}^2 &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\Big[\ln f_{Ni}(Z_{Ni}, \theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni}, \theta)\right\}\Big]\right)^2 \\
&= \frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\ln f_{Ni}(Z_{Ni}, \theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni}, \theta)\right\}\Big]\right)^2 \Big| S_N\right\}\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\ln f_{Ni}(Z_{Ni}, \theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni}, \theta)|S_N\right\}\Big]\right)^2 \Big| S_N\right\}\right] \\
&+ \frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\mathbb{E}\left[\ln f_{Ni}(Z_{Ni}, \theta)|S_N\right] - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni}, \theta)\right\}\Big]\right)^2 \Big| S_N\right\}\right]. \quad (20)
\end{aligned}$$

Note that I suppress the zero terms in RHS of (20). Conditional on $S_N$, $\{Y_i\}_{i=1}^N$ is independent among each other. Then $\{\ln f_{Ni}(Z_{Ni}, \theta)\}_{i=1}^N$ is also conditionally independent, so

$$\begin{aligned}
&\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\ln f_{Ni}(Z_{Ni}, \theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni}, \theta)|S_N\right\}\Big]\right)^2 \Big| S_N\right\} \\
&\qquad = \sum_{i=1}^{N}\mathbb{E}\left(\Big[\ln f_{Ni}(Z_{Ni}, \theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni}, \theta)|S_N\right\}\Big]^2 \Big| S_N\right).
\end{aligned}$$

By Lemma C.1, $\ln f_{Ni}(\cdot, \theta)$ is a bounded function uniformly in $N$, $i$ and $\theta$. Thus

$$\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\left\{\mathbb{E}\left(\left[\ln f_{Ni}(Z_{Ni},\theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}\right]^2 \middle| S_N\right)\right\} = o(1).$$

Moreover, by Lemma C.3

$$\frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\} - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2 \middle| S_N\right\}\right] = o(1).$$

Then $\mathbb{E}\left\{\widehat{L}_N(\theta) - L_N(\theta)\right\}^2 \to 0$, so pointwise LLN obtains. □

B.5. **Proof of Theorem 5.**

*Proof.* First, by Lemma D.1 and the definition of $\widehat{\theta}$, there is $\widehat{G}_N(\widehat{\theta}) = o_p(1/\sqrt{N})$. Hence,

$$o_p\left(\frac{1}{\sqrt{N}}\right) = \widehat{G}_N(\theta_0) + \frac{\partial \widehat{G}_N(\theta^\dagger)}{\partial \theta^T}(\widehat{\theta} - \theta_0), \tag{21}$$

for some $\theta^\dagger$ between $\theta_0$ and $\widehat{\theta}$. Then it suffices to show: (i) $\sqrt{N} \times \widehat{G}_N(\theta_0) \xrightarrow{d} N(0, J_0)$; and (ii) $\frac{\partial \widehat{G}_N(\theta^\dagger)}{\partial \theta^T} \xrightarrow{p} -J_0$.

Proofs of (i). Let $\varphi_{Ni} = \frac{\partial}{\partial \theta}\ln f_{Ni}(Z_{Ni},\theta_0)$, then $\widehat{G}_N(\theta_0) = \sum_{i=1}^{N}\varphi_{Ni}$. Because for any $N$ and $i$, $\theta_0$ maximizes the smooth function $\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}$ almost surely, then $\mathbb{E}\left(\varphi_{Ni}|S_N\right) = 0$. Since $\varphi_{Ni}$ is conditionally independent across $i$. Then

$$\mathbb{E}\left[\left\{\sum_{i=1}^{N}\varphi_{Ni}\right\} \times \left\{\sum_{i=1}^{N}\varphi_{Ni}^T\right\} \middle| S_N\right] = \sum_{i=1}^{N}\mathbb{E}\left(\varphi_{Ni} \times \varphi_{Ni}^T \middle| S_N\right) = NJ_N(\theta_0).$$

For any $\kappa \in \mathbb{R}^L$, let $\psi_N(\kappa, S_N) = \kappa^T \times \left\{J_N(\theta_0)\right\}^{-1/2}$, then there is

$$\begin{aligned}
N^{-1}\mathbb{E}&\left[\psi_N(\kappa,S_N)\left\{\sum_{i=1}^{N}\varphi_{Ni}\right\} \times \left\{\sum_{i=1}^{N}\varphi_{Ni}^T\right\}\psi_N^T(\kappa,S_N)\middle| S_N\right] \\
&= N^{-1}\psi_N(\kappa,S_N)\sum_{i=1}^{N}\mathbb{E}\left(\varphi_{Ni} \times \varphi_{Ni}^T \middle| S_N\right)\psi_N^T(\kappa,S_N) \\
&= \psi_N(\kappa,S_N)J_N(\theta_0)\psi_N^T(\kappa,S_N) = \kappa^T \times \kappa.
\end{aligned}$$

Moreover, $\|\varphi_{Ni}\|_1$ is bounded almost surely by Lemma D.2. Hence, by the Lindeberg-Feller Theorem (see Van der Vaart (2000), page 20),

$$\left(\kappa^T \kappa N\right)^{-1/2} \psi_N(\kappa, S_N) \sum_{i=1}^{N} \varphi_{Ni} \to \mathcal{N}(0,1).$$

Since $\kappa$ is arbitrary in $\mathbb{R}^L$, by Cramer–Wold device,

$$\{N J_N(\theta_0)\}^{-1/2} \sum_{i=1}^{N} \varphi_{Ni} \to \mathcal{N}(0, \mathbf{1}_L).$$

where $\mathbf{1}_L$ is the $L$ by $L$ identity matrix. Moreover, by Assumption L, (i) obtains.

To prove (ii), it is sufficient to show $\partial \widehat{G}_N(\theta^\dagger)/\partial\theta^T + J_N(\theta_0) \xrightarrow{p} 0$. By Lemma D.3, there is $\left|\partial^2 \ln f_{Ni}(Z_{Ni}, \theta)/\partial\theta_m \partial\theta_{m'}\right| < \delta_3$ almost surely for all $N, i, m, m'$. Moreover, by Lemma C.2 and D.3, $\partial^2 \ln f_{Ni}(Z_{Ni}, \theta)/\partial\theta_m \partial\theta_{m'}$ is continuous and uniformly bounded (in $\theta$). Hence by a similar argument as the proofs in Theorem 4

$$\sup_{\theta} \left[ \frac{\widehat{G}_N(\theta)}{\partial\theta^T} - \mathbb{E}\left\{ \frac{\partial^2}{\partial\theta\partial\theta^T} \ln f_{N1}(Z_{N1}, \theta) \right\} \right] \xrightarrow{p} 0.$$

Moreover, by $\theta^\dagger \xrightarrow{p} \theta_0$, it follows that $\partial \widehat{G}_N(\theta^\dagger)/\partial\theta^T - \mathbb{E}\left\{ \frac{\partial^2}{\partial\theta\partial\theta^T} \ln f_{N1}(Z_{N1}, \theta_0) \right\} \xrightarrow{p} 0$.

By information matrix equality, $J_N(\theta_0) = -\mathbb{E}\left\{ \frac{\partial^2}{\partial\theta\partial\theta^T} \ln f_{N1}(Z_{N1}, \theta_0) \right\}$, then (ii) is proved. □

B.6. **Proof of Theorem 6.**

*Proof.* Without causing any confusion in notation, I still denote my objective function as $L_N(\theta)$ in this semiparametric setup. Similarly, $L_0(\theta) = \limsup_{N\to\infty} L_N(\theta)$. First, similarly as in the proof of consistency in parametric part, $\tilde{L}_N(\tilde{\theta}) \geq \sup_{\theta\in\Theta_N} \tilde{L}_N(\theta) - o_p(1)$.

Then the consistency part can be proved by checking the conditions in Chen (2007), Theorem 3.1, where the conditions 3.1∼ 3.4 can be easily verified by the properties of the sieve I choose and similar arguments as in the proof of consistency in parametric setup. Hence, it suffices to verify condition 3.5, the uniform convergence over sieves.

The Uniform LLN can obtain by using an empirical process argument, i.e. a class $\mathcal{F}_N$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ is said to be P–Glivenko–Cantelli class if the sample path of $\mathbb{P}_N f$ get uniformly closer to $Pf$ as $N \to \infty$. For the analysis of empirical process, a key step of constructing probabilistic bounds for the maximal deviation of a sum of independent stochastic process is called symmetrization, which requires the independence of the process. In my case, $\left\{\sum_{k=0}^{K}\left\{\mathbf{1}\left(Y_i=k\right)\ln\tilde{\sigma}^*_{Nik}(S_N;\theta)\right\}:i=1,\cdots,N\right\}\subseteq\mathcal{F}_N$ is a dependent sequence. However, the conditional independence obtains by conditioning on $S_\infty$, the distribution of which does not affect the function class $\mathcal{F}_N$. The symmetrization idea still go through after taking conditioning probability first and then reexpressing as unconditional after the symmetrization. It could be verified that all the results for the bounds of the RHS of the symmetrization inequality still hold in empirical process theory (see Pollard (1990) for more details). Hence, it suffices to examine the class of functions $\mathcal{F}_N$. Similarly as in parametric setup, it could be verified that $\mathcal{F}_N$ can be Hölder class of functions with $(q+m)$–th smoothness. Hence ULLN obtains.

The proof of convergence rate follows Shen and Wong (1994), Theorem 1. The conditions C1 and C2 in the theorem can be verified by choosing $\alpha=\beta=1$ (in their notation), similarly as in Example 2, Shen and Wong (1994). For condition C3, similarly as in Example 3, pick $r_0=\frac{1}{2(2q+2m+1)}$ and $r=0^+$. Thus by their Theorem 1, the converge rate is $N^{-\frac{q+m}{2q+2m+1}}$. Note that the extra $\ln n$ factor in their Theorem 1 for the case $r=r_0^+$ can be removed when the criterion difference is continuous in the Remark 4, which is exactly the case here. □

B.7. **Proof of Theorem 7.**

*Proof.* The proof follows the Theorem 1, Shen (1997). Similarly as in parametric setup, there exists a constant $\delta_7\in\mathcal{R}_+$, such that

$$\left\|\ln f_{Ni}(Z_{Ni},\theta)-\ln f_{Ni}(Z_{Ni},\theta_0)-\frac{\partial\ln f_{Ni}(Z_{Ni},\theta_0)}{\partial\theta(\nu)}\right\|\le\delta_7\left(\theta-\theta_0\right)^2,$$

where $\nu=\theta-\theta_0$. Then, to check the conditions in Theorem 1 of Shen (1997), a similar argument follows as their Example 1(b). Note that, to satisfy the Condition B in Shen (1997), I need impose a normalization on the parameter space, i.e.

$$\widetilde{\Theta}=\left\{\theta':\theta'=\xi(\theta);\theta\in\Theta\right\}.$$

Then the objective function defined on $\widetilde{\Theta}$ satisfies all the conditions for Theorem 1 of Shen (1997). □

## APPENDIX C.

C.1. **Lemma C.1.** Suppose Assumptions A, B, G and H hold, then there exists $\sigma_L \in (0,1)$ such that

$$\sigma^*_{Nik}(S_N;\theta) \geq \sigma_L \quad a.s.$$

for all $N$, $i$, $k$ and $\theta$.

*Proof.* Fix arbitrary $N \in \mathbb{N}$, $i \in \mathcal{I}_N$, $k \in \mathcal{A}$ and $\theta \in \Theta$,

$$\sigma^*_{Nik}(S_N;\theta) = \Gamma_{ik}\left(W_i, \{\sigma^*_{Nj}(S_N;\theta)\}_{j\in\mathcal{N}_i};\theta\right).$$

where the operator $\Gamma_{ik}$ is defined by equation (19). By definition, $\Gamma_{ik}(x, \{\sigma_j\}_{j\in\mathcal{N}_i};\theta)$ is a continuous and strictly positive function for any $x \in \mathbb{R}^p$, choice probability profile $\{\sigma_j\}_{j\in\mathcal{N}_i}$, and $\theta \in \Theta$. Since $\mathcal{X}$ is bounded, $0 \leq \sum_{j\in\mathcal{N}_i} \sigma^*_{j\ell}(S_N;\theta) \leq M$ for all $i$ and $\ell \in \mathcal{A}$, and $\theta$ is in a compact space $\Theta$. Then $\Gamma_{ik}\left(x, \{\sigma_j\}_{j\in\mathcal{N}_i}, \theta\right)$ is greater than some constant $0 < c_k < 1$, which is independent of $N$, $i$, $x$ and $\theta$. Hence $\Gamma_{ik}\left(W_i, \{\sigma^*_{Nj}(S_N;\theta)\}_{j\in\mathcal{N}_i};\theta\right) \geq c_k > 0$, a.s. Thus, take $\sigma_L = \min_{k\in\mathcal{A}} c_k$. □

C.2. **Lemma C.2.** Suppose that Assumptions A through C hold. Hence, for all $N \in \mathbb{N}$, $i \in \mathcal{I}$ and $z \in \mathcal{A} \times \mathscr{S}_N$, $f_{Ni}(z,\cdot) \in \mathcal{C}^\infty(\Theta)$.

*Proof.* By the definition of $f_{Ni}$, it is sufficient to show that for arbitrarily fixed $N \in \mathbb{N}$ and $s \in \mathscr{S}_N$, $\sigma^*_{Ni}(s;\theta) \in \mathcal{C}^\infty(\Theta)$ for all $i \in \mathcal{I}_N$. Note that $\{\sigma^*_{Ni}(s;\theta)\}_{i=1}^N$ is a solution in equation system (4) with $S_N = s$, and the solution is unique by Lemma 2. Let $\Sigma$ is the space of choice probability profile for all $N$ players. Define a mapping $\Gamma : \Sigma \times \Theta \to \Sigma$ such that

$$\Gamma(\Sigma;\theta) = \left(\Gamma_1\left(s, \{\sigma_j\}_{j\in\mathcal{N}_1};\theta\right), \cdots, \Gamma_N\left(s, \{\sigma_j\}_{j\in\mathcal{N}_N};\theta\right)\right)^T, \tag{22}$$

where $\Gamma_i\left(s, \{\sigma_j\}_{j\in\mathcal{N}_i};\theta\right) = \left(\Gamma_{i0}\left(s, \{\sigma_j\}_{j\in\mathcal{N}_i};\theta\right), \cdots, \Gamma_{iK}\left(s, \{\sigma_j\}_{j\in\mathcal{N}_i};\theta\right)\right)$. It is straightforward that $\Gamma(\Sigma;\theta) \in \mathcal{C}^\infty\left(\mathbb{R}^{N(K+1)} \times \Theta; \mathbb{R}^{N(K+1)}\right)$, then by implicit function theorem $\sigma^*_{Ni}(s;\theta) \in \mathcal{C}^\infty(\Theta)$ for all $i \in \mathcal{I}_N$. □

C.3. **Lemma C.3.** Suppose that Assumptions A, B, G, H and I hold, then

$$\frac{1}{N^2}\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\Big]\right)^2\right\}=o(1).$$

*Proof.* Let $h^*\in\mathbb{N}$. Then by a similar argument as in Lemma 2, for all $i$,

$$\sup_{\theta}\left|\sigma^*_{Nik}(S_N;\theta)-\sigma^{(i,h^*)}_{ik}\left(S_N^{(i,h^*)};\theta\right)\right|\leq 2\lambda^{h^*},\ \ a.s.$$

Thus by Taylor expansion,

$$\sup_{\theta}\left|\sum_{k=0}^{K}\sigma^*_{Nik}(S_N;\theta_0)\ln\sigma^*_{Nik}(S_N;\theta)-\sum_{k=0}^{K}\sigma^{(i,h^*)}_{ik}\left(S_N^{(i,h^*)};\theta_0\right)\ln\sigma^{(i,h^*)}_{ik}\left(S_N^{(i,h^*)};\theta\right)\right|$$
$$\leq 2(1-\ln\sigma_L)(K+1)\lambda^{h^*},\ \ a.s.$$

Thus,

$$\begin{aligned}&\sup_{\theta}\left|\mathbb{E}\left\{\ln f_{Ni}(Z_{ni}),\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{ni}),\theta)|S_N^{(i,h^*)}\right\}\right|\\&\quad\leq\sup_{\theta}\left|\sum_{k=0}^{K}\sigma^*_{Nik}(S_N;\theta_0)\ln\sigma^*_{Nik}(S_N;\theta)-\sum_{k=0}^{K}\sigma^{(i,h^*)}_{ik}\left(S_N^{(i,h^*)};\theta_0\right)\ln\sigma^{(i,h^*)}_{ik}\left(S_N^{(i,h^*)};\theta\right)\right|\\&+\sup_{\theta}\int\left|\sum_{k=0}^{K}\sigma^*_{Nik}(S_N;\theta_0)\ln\sigma^*_{Nik}(S_N;\theta)-\sum_{k=0}^{K}\sigma^{(i,h^*)}_{ik}\left(S_N^{(i,h^*)};\theta_0\right)\ln\sigma^{(i,h^*)}_{ik}(S_N^{(i,h^*)};\theta)\right|d\mathbb{P}_{S_N|S_N^{(i,h^*)}}\\&\leq 4(1-\ln\sigma_L)(K+1)\lambda^{h^*},\ \ a.s..\end{aligned}\tag{23}$$

Because

$$\begin{aligned}&\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\Big]\right)^2\right\}\\&\quad=\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h^*)}\right\}\Big]\right)^2\right\}\\&\qquad+\mathbb{E}\left\{\left(\sum_{i=1}^{N}\Big[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h^*)}\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\Big]\right)^2\right\}.\end{aligned}$$

Note that I suppress zero terms in the RHS of above equation.

By equation (23), and by Lemma C.1, $\ln f_{Ni}(\cdot, \theta)$ is a bounded function uniformly in $N$, $i$ and $\theta$. Then,

$$\mathbb{E}\left\{ \left( \sum_{i=1}^{N} \left[ \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) | S_N \right\} - \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) | S_N^{(i,h^*)} \right\} \right] \right)^2 \right\} = O\left( N^2 \lambda^{2h^*} \right),$$

and

$$\mathbb{E}\left\{ \left( \sum_{i=1}^{N} \left[ \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) | S_N^{(i,h^*)} \right\} - \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) \right\} \right] \right)^2 \right\}$$

$$= \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{(i,h)}} Cov\left[ \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) | S_N^{(i,h^*)} \right\}, \mathbb{E}\left\{ \ln f_{Nj}(Z_{Nj}, \theta) | S_N^{(i,h^*)} \right\} \right]$$

$$+ \sum_{i=1}^{N} Var\left[ \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) | S_N^{(i,h^*)} \right\} \right] = O\left( N M^{h^*} \right) + O(N)$$

Choose $h^* = \frac{b \ln N}{\ln M}$ for some $b \in (0,1)$. Then $h^* \to \infty$ as $N \to \infty$ and $M^{h^*} = o(N)$. Note that $h^*$ is different from the $h$ in the MAL estimator and only serves for this proof. Hence,

$$\frac{1}{N^2} \mathbb{E}\left\{ \left( \sum_{i=1}^{N} \left[ \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) | S_N \right\} - \mathbb{E}\left\{ \ln f_{Ni}(Z_{Ni}, \theta) \right\} \right] \right)^2 \right\} = o(1).$$

□

## Appendix D.

D.1. **Lemma D.1.** Suppose that Assumptions A through C, G and H hold.

$$\sup_{\theta \in \Theta} \left\| \widehat{G}_N(\theta) - \widehat{G}_N^h(\theta) \right\| = o_p\left( \frac{1}{\sqrt{N}} \right)$$

*Proof.* First, because Lemma E.3,

$$\sup_{\theta \in \Theta} \left\| \widehat{G}_N(\theta) - \widehat{G}_N^h(\theta) \right\|_1 \leq \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{K} \left\| \frac{\partial \ln \sigma_{Nik}^*(S_N; \theta)}{\partial \theta} - \frac{\partial \ln \sigma_{ik}^{(i,h)}(S_N^{(i,h)}; \theta)}{\partial \theta} \right\|_1 = O_p(h \lambda^h).$$

Since $\lambda < 1$ and $h \propto N^\omega$ for some $\omega > 0$, then

$$h \lambda^h = o\left( \frac{1}{\sqrt{N}} \right).$$

$\square$

D.2. **Lemma D.2.** Suppose that Assumptions A, B, G H and I hold. Then there exists a $C_1 \in \mathbb{R}_+$, such that

$$\sup_N \max_{i\in\mathcal{I}_N} \sup_{\theta\in\Theta} \left\| \frac{\partial \ln \sigma^*_{Ni}(S_N;\theta)}{\partial\theta} \right\|_1 \le C_1, \quad a.s.$$

and

$$\sup_N \max_{i\in\mathcal{I}_N} \sup_{\theta\in\Theta} \left\| \frac{\partial \sigma^*_{Ni}(S_N;\theta)}{\partial\theta} \right\|_1 \le C_1, \quad a.s.$$

*Proof.* First, I derive an expression for $\frac{\partial}{\partial\theta}\sigma^*_{Ni}(S_N;\theta)$, from which I obtain a uniform bound over $N$, $i$ and $\theta$. For all $i \in \mathcal{I}_N$, $k \in \mathcal{A}$ and $m = 1, \cdots, L$,

$$\frac{\partial \sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m} = \frac{\partial\Gamma_{ik}\left(W_i, \{\sigma^*_{Nj}(S_N,\theta)\}_{j\in\mathcal{N}_i};\theta\right)}{\partial\theta_m} + \sum_{j\in\mathcal{N}_i}\sum_{\ell=0}^{K}\left\{\frac{\partial\Gamma_{ik}\left(W_i, \{\sigma^*_{Nj}(S_N,\theta)\}_{j\in\mathcal{N}_i};\theta\right)}{\partial\sigma^*_{Nj\ell}} \times \frac{\partial\sigma^*_{Nj\ell}(S_N;\theta)}{\partial\theta_m}\right\}. \quad (24)$$

Let $\sigma^*_N(S_N;\theta) = \left(\sigma^*_{N1}(S_N;\theta)^\top, \cdots, \sigma^*_{NN}(S_N;\theta)^\top\right)^\top$ and $\zeta(S_N,\theta) = \partial\sigma^*_N(S_N;\theta)/\partial\theta_m$. Let further $\Lambda(S_N,\theta) = \partial\Gamma(\sigma^*_N(S_N;\theta);\theta)/\partial\theta_m$ and $D(S_N,\theta) = \partial\Gamma(\sigma^*_N(S_N;\theta);\theta)/\partial\Sigma$, where $\Gamma$ is defined in equation (22). Thus, equation (24) becomes

$$\left\{\mathbf{1}_{N(K+1)} - D(S_N,\sigma^*_N;\theta)\right\}\zeta(S_N,\theta) = \Lambda(S_N,\theta).$$

Thus

$$\zeta(S_N,\theta) = \left\{\mathbf{1}_{N(K+1)} - D(S_N,\sigma^*_N;\theta)\right\}^{-1}\Lambda(S_N,\theta) = \left\{\sum_{t=0}^{\infty} D^t(S_N,\sigma^*_N(S_N;\theta);\theta)\right\}\Lambda(S_N,\theta),$$

where the last step comes from Lemma E.1.

For each $i \in \mathcal{I}_N$ and $k \in \mathcal{A}$, let $\iota_{ik}$ be an $N(K+1)$ dimensional vector with value one only at the $[(i-1)(K+1)+k+1]$–th component and zero elsewhere. Hence

$$\sum_{k=0}^{K}\left|\frac{\partial\sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m}\right| \le \sum_{k=0}^{K}\left|\sum_{t=0}^{\infty}\iota'_{ik}D^t(S_N,\sigma^*_N;\theta)\Lambda(S_N,\theta)\right| \le \delta_2\sum_{k=0}^{K}\sum_{t=0}^{\infty}\lambda^t,$$

where the last step comes from Holder inequality, Lemma C.1 and Lemma E.2. Thus

$$\sum_{k=0}^{K} \left| \frac{\partial \ln \sigma^*_{Nik}(S_N;\theta)}{\partial \theta_m} \right| \leq \frac{\delta_2(K+1)}{1-\lambda}. \tag{25}$$

Moreover, by Lemma C.1,

$$\left| \frac{\partial \ln \sigma^*_{Nik}(S_N;\theta)}{\partial \theta_m} \right| \leq \frac{1}{\sigma_L} \left| \frac{\partial \sigma^*_{Nik}(S_N;\theta)}{\partial \theta_m} \right| \leq \frac{\delta_2(K+1)}{\sigma_L(1-\lambda)}. \tag{26}$$

To complete the proof, just note that the RHS's of (25) and (26) do not depend on $N, i$ or $\theta$. □

D.3. **Lemma D.3.** Suppose that Assumptions A, B , G, H and I hold. Then there exists a $\delta_4 \in \mathbb{R}_+$, such that

$$\sup_N \max_{i\in\mathcal{I}_N} \sup_{\theta\in\Theta} \sum_{k\in\mathcal{A}} \left\| \frac{\partial^2 \sigma^*_{Nik}(S_N;\theta)}{\partial\theta\partial\theta^T} \right\|_1 \leq \delta_4, \quad a.s.$$

*Proof.* It is sufficient to show that for any $m, m' = 1, \cdots, L$, then there exists a $\delta_4' \in \mathbb{R}_+$, such that

$$\sup_N \max_{i\in\mathcal{I}_N} \sup_{\theta\in\Theta} \sum_{k\in\mathcal{A}} \left| \frac{\partial^2 \sigma^*_{Ni}(S_N;\theta)}{\partial\theta_m\partial\theta_{m'}} \right| \leq \delta_4', \quad a.s.$$

Note that for any $m, m' = 1, \cdots, L$,

$$\begin{aligned}
\frac{\partial^2 \sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m\partial\theta_{m'}} = & \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\theta_m\partial\theta_{m'}} + \sum_{j\in\mathcal{N}_i} \sum_{\ell=0}^{K} \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\theta_m\partial\sigma^*_{j\ell}} \times \frac{\partial \sigma^*_{Nj\ell}}{\partial\theta_m} \\
& + \sum_{j\in\mathcal{N}_i} \sum_{\ell=0}^{K} \left\{ \frac{\partial \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\sigma^*_{Nj\ell}} \times \frac{\partial^2 \sigma^*_{Nj\ell}}{\partial\theta_m\partial\theta_{m'}} \right\} \\
& + \sum_{j\in\mathcal{N}_i} \sum_{\ell=0}^{K} \left\{ \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\sigma^*_{Nj\ell}\partial\theta_{m'}} \times \frac{\partial \sigma^*_{Nj\ell}}{\partial\theta_m} \right\} \\
& + \sum_{j,j'\in\mathcal{N}_i} \sum_{\ell,\ell'\in\mathcal{A}} \left\{ \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\sigma^*_{Nj\ell}\partial\sigma^*_{Nj'\ell'}} \frac{\partial \sigma^*_{Nj\ell}}{\partial\theta_m} \frac{\partial \sigma^*_{Nj'\ell'}}{\partial\theta_{m'}} \right\},
\end{aligned}$$

which can also be written as

$$\left\{ \mathbf{1}_{N(K+1)} - D(S_N, \sigma^*_N; \theta) \right\} \times \Psi_N(\theta) = T_N(\theta)$$

where $\Psi_N(\theta)$ and $T_N(\theta)$ are $N(K+1)$ dimensional vectors. $\Psi_N$'s $(i-1)(K+1)+k+1$–the component is given by follows

$$\Psi_N = \frac{\partial^2 \sigma^*_{ik}(S_N, \theta)}{\partial\theta_m \partial\theta_{m'}}.$$

$T_N$'s $(i-1)(K+1)+k+1$–th component is

$$\begin{aligned}
T_{Nik}(\theta) = & \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\theta_m \partial\theta_{m'}} + \sum_{j\in\mathcal{N}_i}\sum_{\ell=0}^{K} \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\theta_m \partial\sigma^*_{Nj\ell}} \times \frac{\partial \sigma^*_{Nj\ell}}{\partial\theta_m} \\
& + \sum_{j\in\mathcal{N}_i}\sum_{\ell=0}^{K} \left\{ \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\sigma^*_{Nj\ell} \partial\theta_{m'}} \times \frac{\partial \sigma^*_{Nj\ell}}{\partial\theta_m} \right\} \\
& + \sum_{j,j'\in\mathcal{N}_i}\sum_{\ell,\ell'\in\mathcal{A}} \left\{ \frac{\partial^2 \Gamma_{ik}(W_i, \Sigma^*_N(S_N;\theta), \theta)}{\partial\sigma^*_{Nj\ell} \partial\sigma^*_{Nj'\ell'}} \frac{\partial \sigma^*_{Nj\ell}}{\partial\theta_m} \frac{\partial \sigma^*_{Nj'\ell'}}{\partial\theta_{m'}} \right\} \\
& = T_{Nik1}(\theta) + T_{Nik2}(\theta) + T_{Nik3}(\theta) + T_{Nik4}(\theta),
\end{aligned}$$

Now I am going to show

$$\max_i \max_k |T_{Nik}(\theta)| \leq \delta_5$$

for some $\delta_5 \in \mathbb{R}_+$. Since for any $k, \ell \in \mathcal{A}$ and $q \in \mathcal{A}/\{0\}$,

$$\begin{aligned}
\frac{\partial^2 \Gamma_{ik}}{\partial\beta_q \partial\theta_{m'}} &= -\frac{\partial \sigma^*_{Niq}}{\partial\theta_{m'}} \sigma^*_{Nik} X_{iq} + \frac{\partial \sigma^*_{Nik}}{\partial\theta_{m'}} \left\{ \mathbf{1}(q=k) - \sigma^*_{Niq} \right\} X_{iq}, \\
\frac{\partial^2 \ln \Gamma_{ik}}{\partial\alpha(q,\ell) \partial\theta_{m'}} &= \left\{ \mathbf{1}(q=k) - \sigma^*_{Niq} \right\} \sigma^*_{Nik} \sum_{j\in\mathcal{N}_i} \frac{\sigma^*_{Nj\ell}}{\partial\theta_{m'}} - \frac{\partial \sigma^*_{Niq}}{\partial\theta_{m'}} \sigma^*_{Nik} \sum_{j\in\mathcal{N}_i} \sigma^*_{j\ell} \\
& \quad + \left\{ \mathbf{1}(q=k) - \sigma^*_{Niq} \right\} \frac{\partial \sigma^*_{Nik}}{\partial\theta_{m'}} \sum_{j\in\mathcal{N}_i} \sigma^*_{Nj\ell}.
\end{aligned}$$

By Lemma D.2 and Assumption G,

$$\sup_N \max_{i\in\mathcal{I}_N} \sum_{k\in\mathcal{A}} |T_{Nik1}(\theta)| \leq 2(K+1)\delta_2 C_1.$$

Similarly, it can be shown that

$$\begin{aligned}
\sup_N \max_{i\in\mathcal{I}_N} \sum_{k\in\mathcal{A}} |T_{Nik2}(\theta)| &\leq 2\lambda(K+1)\delta_2 C_1, \\
\sup_N \max_{i\in\mathcal{I}_N} \sum_{k\in\mathcal{A}} |T_{Nik3}|(\theta) &\leq 2\lambda(K+1)\delta_2 C_1.
\end{aligned}$$

For term $T_{Nik4}(\theta)$, because for any $j \in \mathcal{N}_i$,

$$\frac{\partial \Gamma_{ik}(W_i, \{\sigma^*_{Nj}(S_N;\theta)\}_{j\in\mathcal{N}_i}, \theta)}{\partial \sigma_{j\ell}} = \sigma^*_{Nik} \sum_{q\neq k} \left[\sigma^*_{Niq} \left\{\alpha(k,\ell) - \alpha(q,\ell)\right\}\right].$$

Then for all $j, j' \in \mathcal{N}_i$

$$\frac{\partial^2 \Gamma_{ik}(W_i, \{\sigma^*_{Nj}(S_N;\theta)\}_{j\in\mathcal{N}_i}, \theta)}{\partial \sigma_{j\ell}\partial \sigma_{j',\ell'}} = \sigma^*_{Nik} \left(\sum_{q\neq k} \left[\sigma^*_{Niq} \left\{\alpha(k,\ell) - \alpha(q,\ell)\right\}\right]\right)^2$$
$$+ \sigma^*_{Nik} \sum_{q\neq k} \left[\left(\sigma^*_{Niq} \sum_{q'\neq q} \left[\sigma^*_{Niq'} \left\{\alpha(q,\ell) - \alpha(q',\ell)\right\}\right]\right) \left\{\alpha(k,\ell) - \alpha(q,\ell)\right\}\right].$$

Thus

$$\sum_{k=0}^{K} \left| \frac{\partial^2 \Gamma_{ik}(W_i, \{\sigma^*_{Nj}(S_N;\theta)\}_{j\in\mathcal{N}_i}, \theta)}{\partial \sigma_{j\ell}\partial \sigma_{j',\ell'}} \right|$$
$$\leq \bar{\Delta}^2(\alpha) \sum_{k=0}^{K} \sigma^*_{Nik} \sum_{q\neq k} \left(\sigma^*_{Niq}\right)^2 + \bar{\Delta}^2(\alpha) \sum_{k=0}^{K} \left[\sigma^*_{Nik} \left\{\sum_{q\neq k} \sigma^*_{Niq} \left(1 - \sigma^*_{Niq}\right)\right\}\right]$$
$$= \bar{\Delta}^2(\alpha) \sum_{k=0}^{K} \left\{\sigma^*_{Nik} \left(\sum_{q\neq k} \sigma^*_{Niq}\right)\right\} \leq \bar{\Delta}^2(\alpha) \times \frac{K}{K+1}.$$

It follows that

$$\sum_{k\in\mathcal{A}} |T_{Nik4}(\theta)| \leq \bar{\Delta}^2(\alpha) \times \frac{K}{K+1} \sum_{k\in\mathcal{A}} \sum_{j,j'\in\mathcal{N}_i} \sum_{\ell,\ell'\in\mathcal{A}} \left| \frac{\partial \sigma^*_{Nj\ell}}{\partial \theta_m} \frac{\partial \sigma^*_{Nj'\ell'}}{\partial \theta_{m'}} \right|$$
$$\leq \bar{\Delta}^2(\alpha) \times \frac{K}{K+1}(K+1)^2 M^2 = \lambda^2 \times \frac{(K+1)^3}{K}.$$

From above analysis, there exists a constant $\delta_5 \in \mathbb{R}_+$, which does not depend on $N, i$ or $\theta$, such that

$$\max_i \max_k |T_{Nik}(\theta)| \leq \delta_5.$$

Therefore,

$$\sum_{k\in\mathcal{A}}\left|\frac{\partial^2\sigma^*_{ik}(S_N,\theta)}{\partial\theta_m\partial\theta_{m'}}\right| = \sum_{k\in\mathcal{A}}\left|\iota'_{ik}\left\{\mathbf{1}_{N(K+1)} - D(S_N,\sigma^*_N;\theta)\right\}^{-1}T_N(\theta)\right|$$
$$\leq \delta_5\sum_{k\in\mathcal{A}}\sum_{t=0}^{\infty}\left\|\iota'_{ik}D^t(S_N,\theta)\right\|_1 \leq \frac{\delta_5(K+1)}{1-\lambda}.$$

Since the RHS of above inequality does not depend on $N, i$ an $\theta$, so the lemma is proved. □

## Appendix E.

E.1. **Lemma E.1.** Suppose Assumptions A, B and I hold, then for any choice probability profile $\Sigma_N$, $s \in \mathscr{S}_N$, and arbitrary real vector $\mu \in \mathbb{R}^{N(K+1)}$,

$$\left\{\mathbf{1}_{N(K+1)} - T(s,\Sigma_N;\theta)\right\}^{-1}\mu = \sum_{t=0}^{\infty}\left\{T(s,\Sigma_N;\theta)\mu\right\},$$

and

$$\left\|\left\{\mathbf{1}_{N(K+1)} - T(s,\Sigma_N;\theta)\right\}^{-1}\mu\right\|_1 \leq \frac{\|\mu\|_1}{1-\lambda}.$$

where $\lambda = \bar{\Delta}(\alpha)MK/(1+K)$.

*Proof.* By the definition of $D(s,\Sigma_N;\theta)$,

$$\|D(s,\Sigma_N;\theta)\mu\|_1 \leq \sum_{i\in\mathcal{I}_N}\sum_{k=0}^{K}\sum_{j\in\mathcal{N}_i}\sum_{\ell=0}^{K}\left|\frac{\partial\Gamma_{ik}(w_i,\Sigma_N;\theta)}{\partial\sigma_{j\ell}}\right| \times \left|\mu_{(j-1)(K+1)+\ell+1}\right|,$$

and from the proof in Lemma A.6,

$$\frac{\partial}{\partial\sigma_{j\ell}}\Gamma_{ik}(w_i,\Sigma_N;\theta) = \Gamma_{ik}\sum_{q\neq k}\left[\Gamma_{iq}\left\{\alpha(k,\ell)-\alpha(q,\ell)\right\}\right].$$

Thus

$$\|D(s,\Sigma_N;\theta)\mu\|_1 \leq \bar{\Delta}(\alpha)\sum_{i\in\mathcal{I}_N}\sum_{k=0}^{K}\sum_{j\in\mathcal{N}_i}\left\{\left|\Gamma_{ik}\sum_{q\neq k}\Gamma_{iq}\right| \times \sum_{\ell=0}^{K}\left|\mu_{(j-1)(K+1)+\ell+1}\right|\right\}$$
$$\leq \frac{\bar{\Delta}(\alpha)K}{K+1}\sum_{i\in\mathcal{I}_N}\sum_{j\in\mathcal{N}_i}\sum_{\ell=0}^{K}\left|\mu_{(j-1)(K+1)+\ell+1}\right| \leq \frac{\bar{\Delta}(\alpha)MK}{K+1}|\mu|_1 = \lambda\,\|\mu\|_1. \quad (27)$$

Next, I'll show $\left\|D^{T+1}(s,\Sigma_N;\theta)\mu\right\|_1 \to 0$ using equation (27). Because

$$\left\|D^{T+1}(s,\Sigma_N;\theta)\mu\right\|_1 \leq \lambda \left\|D^{T}(s,\Sigma_N;\theta)\mu\right\|_1 \leq \lambda^{T+1}\|\mu\|_1,$$

then $\lim_{T\to\infty}\left\|D^{T+1}(s,\Sigma_N;\theta)\mu\right\|_1 = 0$. Hence, for any $\mu \in \mathbb{R}^{N(K+1)}$, there is

$$\begin{aligned}&\left\{\mathbf{1}_{N(K+1)} - D(s,\Sigma_N;\theta)\right\} \times \sum_{t=0}^{\infty} D^t(s,\Sigma_N;\theta) \times \mu \\ &\qquad = \left\{\mathbf{1}_{N(K+1)} - \lim_{T\to\infty} D^{T+1}(s,\Sigma_N;\theta)\right\} \times \mu = \mu,\end{aligned}$$

which implies that

$$\sum_{t=0}^{\infty} D^t(s,\Sigma_N;\theta) \times \mu = \left\{\mathbf{1}_{N(K+1)} - D(s,\Sigma_N;\theta)\right\}^{-1} \times \mu.$$

Moreover

$$\begin{aligned}\left\|\left\{\mathbf{1}_{N(K+1)} - D(s,\Sigma_N;\theta)\right\}^{-1} \times \mu\right\|_1 &= \left\|\sum_{t=0}^{\infty} D^t(s,\Sigma_N;\theta) \times \mu\right\|_1 \\ &\leq \sum_{t=0}^{\infty}\left\|D^t(s,\Sigma_N;\theta) \times \mu\right\|_1 \leq \sum_{t=0}^{\infty}\lambda^t\|\mu\|_1 = \frac{\|\mu\|_1}{1-\lambda}.\end{aligned}$$

□

E.2. **Lemma E.2.** Suppose Assumptions A, B, and G hold, then there exists $\delta_2 \in \mathbb{R}_+$ such that for all $\theta$ and $m = 1, \cdots, L$,

$$\sup_{s\in\mathscr{S}_N}\left|\frac{\partial \ln \Gamma_{ik}(w_i,\Sigma_N;\theta)}{\partial \theta_m}\Big|_{\Sigma_N=\Sigma_N^*(s;\theta)}\right| \leq \delta_2.$$

*Proof.* Since for arbitrary $s \in \mathscr{S}_N$, choice probability profile $\Sigma_N$ and $\theta \in \Theta$, consider $\Gamma_{ik}(w_i,\Sigma_N;\theta)$ for $k \in \mathcal{A}$ and $i \in \mathcal{I}$.

$$\frac{\partial}{\partial \beta_q}\ln\Gamma_{ik}(w_i,\Sigma_N;\theta) = \left\{\mathbf{1}(q=k) - \Gamma_{iq}\right\} x_i,$$

$$\frac{\partial}{\partial \alpha(q,\ell)}\ln\Gamma_{ik}(w_i,\Sigma_N;\theta) = \left\{\mathbf{1}(q=k) - \Gamma_{iq}\right\} \sum_{j\in\mathcal{N}_i}\sigma_{j\ell},$$

and $q \in \mathcal{A}/\{0\}$. Furthermore, by Assumption G, there exists $C_0 \in \mathbb{R}_+$ such that $\|x\| \leq C_0$ for all $x \in \mathcal{X}$. Hence

$$\left|\frac{\partial \ln \Gamma_{ik}(w_i, \Sigma_N; \theta)}{\partial \beta_q}\right| \leq C_0, \quad \left|\frac{\partial \ln \Gamma_{ik}(w_i, \Sigma_N; \theta)}{\partial \alpha(q, \ell)}\right| \leq M.$$

Take $\delta_2 = C_0 + M$, thus, for any $m = 1, \cdots, L$

$$\sup_{s \in \mathscr{S}_N} \sup_{\Sigma_N} \left|\frac{\partial \ln \Gamma_{ik}(w_i, \Sigma_N; \theta)}{\partial \theta_m}\right| \leq \delta_2,$$

which implies that

$$\sup_{s \in \mathscr{S}_N} \left|\frac{\partial \ln \Gamma_{ik}(w_i, \Sigma_N; \theta)}{\partial \theta_m}\Big|_{\Sigma_N = \Sigma_N^*(s;\theta)}\right| \leq \delta_2.$$

□

E.3. **Lemma E.3.** Suppose that Assumptions A through C and G hold. Then

$$\sup_{\theta \in \Theta} \sup_{i \in \mathcal{I}_N} \sum_{k=0}^{K} \left\|\frac{\partial \sigma^*_{Nik}(S_N; \theta)}{\partial \theta} - \frac{\partial \sigma^{(i,h)}_{ik}(S^{(i,h)}_N; \theta)}{\partial \theta}\right\|_1 = O_p(h\lambda^h),$$

and

$$\sup_{\theta \in \Theta} \sup_{i \in \mathcal{I}_N} \sum_{k=0}^{K} \left\|\frac{\partial \ln \sigma^*_{Nik}(S_N; \theta)}{\partial \theta} - \frac{\partial \ln \sigma^{(i,h)}_{ik}(S^{(i,h)}_N; \theta)}{\partial \theta}\right\|_1 = O_p(h\lambda^h).$$

*Proof.* First, for any $q \in \mathcal{A} \backslash \{0\}$ and $\ell \in \mathcal{A}$, consider any $j \in \mathcal{N}_{(i,h)}$

$$\begin{aligned}
\left\|\frac{\partial}{\partial \theta}\sigma^*_{Njk}(S_N;\theta) - \frac{\partial}{\partial \theta}\sigma^{(i,h)}_{Njk}(S_N;\theta)\right\|_1 &\leq \left\|\frac{\partial}{\partial \theta}\Gamma_{jk}(W_j, \Sigma^*_N;\theta) - \frac{\partial}{\partial \theta}\Gamma_{jk}(W_j, \Sigma^{(i,h)}_N;\theta)\right\|_1 \\
&+ \left\|\sum_{n \in \mathcal{N}_j}\sum_{\ell \in \mathcal{A}}\left\{\frac{\partial}{\partial \sigma_{Nn\ell}}\Gamma_{jk}(W_j, \Sigma^*_N;\theta) \times \frac{\partial \sigma^*_{Nn\ell}}{\partial \theta} - \frac{\partial}{\partial \sigma_{Nn\ell}}\Gamma_{jk}(W_j, \Sigma^{(i,h)}_N;\theta) \times \frac{\partial \sigma^{(i,h)}_{Nn\ell}}{\partial \theta}\right\}\right\|_1 \\
&\leq \left\|\frac{\partial}{\partial \theta}\Gamma_{jk}(W_j, \Sigma^*_N;\theta) - \frac{\partial}{\partial \theta}\Gamma_{jk}(W_j, \Sigma^{(i,h)}_N;\theta)\right\|_1 \\
&+ \left\|\sum_{n \in \mathcal{N}_j}\sum_{\ell \in \mathcal{A}}\left\{\frac{\partial}{\partial \sigma_{Nn\ell}}\Gamma_{jk}(W_j, \Sigma^*_N;\theta) - \frac{\partial}{\partial \sigma_{Nn\ell}}\Gamma_{jk}(W_j, \Sigma^{(i,h)}_N;\theta)\right\} \times \frac{\partial \sigma^*_{Nn\ell}}{\partial \theta}\right\|_1 \\
&+ \left\|\sum_{n \in \mathcal{N}_j}\sum_{\ell \in \mathcal{A}}\frac{\partial}{\partial \sigma_{Nn\ell}}\Gamma_{jk}(W_j, \Sigma^{(i,h)}_N;\theta) \times \left\{\frac{\partial \sigma^*_{Nn\ell}}{\partial \theta} - \frac{\partial \sigma^{(i,h)}_{Nn\ell}}{\partial \theta}\right\}\right\|_1, \quad a.s.
\end{aligned}$$

where $\Sigma_N^{(i,h)} = \Sigma_N^{(i,h)}(S_N^{(i,h)};\theta)$. Furthermore, by Lemma D.2 and E.4,

$$\begin{aligned}
&\sum_{k=0}^{K}\left\|\frac{\partial}{\partial\theta}\sigma^*_{Njk}(S_N;\theta) - \frac{\partial}{\partial\theta}\sigma^{(i,h)}_{Njk}(S_N;\theta)\right\|_1 \\
&\quad\le (K+1)C_2 \times \sup_{\theta\in\Theta}\max_{n\in\mathcal{N}_{(j,1)}}\left\|\sigma^*_{Nn} - \sigma^{(i,h)}_{Nn}\right\|_1 + (K+1)C_1C_3 \times \sup_{\theta\in\Theta}\left\|\sigma^*_{Nj} - \sigma^{(i,h)}_{Nj}\right\|_1 \\
&\qquad + \max_{n\in\mathcal{N}_j}\sum_{\ell=0}^{K}\left\|\frac{\partial\sigma^*_{Nn\ell}}{\partial\theta} - \frac{\partial\sigma^{(i,h)}_{Nn\ell}}{\partial\theta}\right\|_1 \times \sum_{k=0}^{K}\sum_{n\in\mathcal{N}_j}\left|\frac{\partial}{\partial\sigma_{Nn\ell}}\Gamma_{jk}(W_j,\Sigma_N^{(i,h)};\theta)\right|, \ \ a.s.
\end{aligned}$$

By the proof in Lemma A.6, the last term is bounded by

$$\lambda \times \max_{n\in\mathcal{N}_j}\sum_{\ell=0}^{K}\left\|\frac{\partial\sigma^*_{Nn\ell}}{\partial\theta} - \frac{\partial\sigma^{(i,h)}_{Nn\ell}}{\partial\theta}\right\|_1, \ \ a.s.$$

Hence,

$$\begin{aligned}
&\sum_{k=0}^{K}\left\|\frac{\partial}{\partial\theta}\sigma^*_{Njk}(S_N;\theta) - \frac{\partial}{\partial\theta}\sigma^{(i,h)}_{Njk}(S_N;\theta)\right\|_1 \\
&\qquad\le C_4 \times \sup_{\theta\in\Theta}\max_{n\in\mathcal{N}_{(j,1)}}\left\|\sigma^*_{Nn} - \sigma^{(i,h)}_{Nn}\right\|_1 + \lambda \times \max_{n\in\mathcal{N}_j}\sum_{\ell=0}^{K}\left\|\frac{\partial\sigma^*_{Nn\ell}}{\partial\theta} - \frac{\partial\sigma^{(i,h)}_{Nn\ell}}{\partial\theta}\right\|_1, \ \ a.s.
\end{aligned}$$

where $C_4 = (K+1)\times(C_2 + C_1C_3)$. Since $\left\|\sigma^*_{Nn} - \sigma^{(i,h)}_{Nn}\right\|_1 \le 2$ almost surely for all $N, n$ and $\theta$, and by Lemma D.2, $\max_{n\in\mathcal{N}_j}\left\|\partial\sigma^*_{Nn}/\partial\theta - \partial\sigma^{(i,h)}_{Nn}/\partial\theta\right\|_1 \le 2C_1$, almost surely. Then for all $j \in \mathcal{N}_{(i,h)}$,

$$\sum_{k=0}^{K}\left\|\frac{\partial\sigma^*_{Njk}(S_N;\theta)}{\partial\theta} - \frac{\partial\sigma^{(i,h)}_{Njk}(S_N;\theta)}{\partial\theta}\right\|_1 \le 2C_4 + 2\lambda C_1, \ \ a.s.$$

From the proof of A.2, for all $j \in \mathcal{N}_{(i,h-1)}$, $\sup_{\theta\in\Theta}\max_{n\in\mathcal{N}_{(j,1)}}\left\|\sigma^*_{Nn} - \sigma^{(i,h)}_{Nn}\right\|_1 \le 2\lambda$ almost surely, then

$$\sum_{k=0}^{K}\left\|\frac{\partial\sigma^*_{Njk}(S_N;\theta)}{\partial\theta} - \frac{\partial\sigma^{(i,h)}_{Njk}(S_N;\theta)}{\partial\theta}\right\|_1 \le 2C_4\lambda + \lambda \times (2C_4 + 2\lambda C_1), \ \ a.s.$$

By induction method, for all $j \in \mathcal{N}_{(i,h-d)}$ ($d \in \mathbb{N}; d \leq h.$), there is

$$\sum_{k=0}^{K} \left\| \frac{\partial \sigma^*_{Njk}(S_N;\theta)}{\partial \theta} - \frac{\partial \sigma^{(i,h)}_{Njk}(S_N;\theta)}{\partial \theta} \right\|_1 \leq 2\lambda^d \left(\lambda C_1 + (d+1)C_4\right), a.s.$$

Note that $i \in \mathcal{N}_{(i,0)}$, then

$$\sum_{k=0}^{K} \left\| \frac{\partial \sigma^*_{Nik}(S_N;\theta)}{\partial \theta} - \frac{\partial \sigma^{(i,h)}_{ik}(S^{(i,h)}_N;\theta)}{\partial \theta} \right\|_1 \leq 2\lambda^h \left(\lambda C_1 + (h+1)C_4\right), a.s.$$

$$\begin{aligned}
&\sum_{k=0}^{K} \left\| \frac{\partial \ln \sigma^*_{Nik}(S_N;\theta)}{\partial \theta} - \frac{\partial \ln \sigma^{(i,h)}_{ik}(S^{(i,h)}_N;\theta)}{\partial \theta} \right\|_1 \\
&\qquad \leq \sum_{k=0}^{K} \left\{ \left\| \frac{\partial \sigma^*_{Nik}(S_N;\theta)}{\partial \theta} - \frac{\partial \sigma^{(i,h)}_{ik}(S^{(i,h)}_N;\theta)}{\partial \theta} \right\|_1 \times \left| \frac{1}{\sigma^*_{Nik}(S_N;\theta)} \right| \right\} \\
&\qquad + \sum_{k=0}^{K} \left\{ \left\| \frac{\partial \sigma^{(i,h)}_{ik}(S^{(i,h)}_N;\theta)}{\partial \theta} \right\|_1 \times \left| \frac{\sigma^*_{Nik}(S_N;\theta) - \sigma^{(i,h)}_{ik}(S^{(i,h)}_N;\theta)}{\sigma^*_{Nik}(S_N;\theta)\sigma^{(i,h)}_{ik}(S^{(i,h)}_N;\theta)} \right| \right\} \\
&\qquad\qquad \leq \frac{2\lambda^h \left(\lambda C_1 + (h+1)C_4\right)}{\sigma_L} + \frac{2C_1\lambda^{h+1}}{\sigma_L^2}, a.s.
\end{aligned}$$

□

E.4. **Lemma E.4.** Suppose that Assumptions A through C and G hold. Then there exist $C_2, C_3 \in \mathbb{R}_+$, such that for all $N, i, k$ and arbitrary two strategy profile $\Sigma_N = \{\sigma_n\}_{n \in \mathcal{I}}$, $\Sigma'_N = \{\sigma'_n\}_{n \in \mathcal{I}}$.

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \Gamma_{ik}(W_i, \Sigma_N;\theta)}{\partial \theta} - \frac{\partial \Gamma_{ik}(W_i, \Sigma'_N;\theta)}{\partial \theta} \right\|_1 \leq C_2 \times \sup_{\theta \in \Theta} \max_{j \in \mathcal{N}_{(i,1)}} \left\| \Gamma_j - \Gamma'_{Nj} \right\|_1, \quad a.s.$$

and

$$\sup_{\theta \in \Theta} \left| \sum_{j \in \mathcal{N}_i} \sum_{\ell \in \mathcal{A}} \left\{ \frac{\partial \Gamma_{ik}(W_i, \Sigma_N;\theta)}{\partial \sigma_{Nj\ell}} - \frac{\partial \Gamma_{ik}(W_i, \Sigma'_N;\theta)}{\partial \sigma_{Nj\ell}} \right\} \right| \leq C_3 \times \sup_{\theta \in \Theta} \left\| \Gamma_i - \Gamma'_{Ni} \right\|_1, \quad a.s.$$

where $\Gamma'_{Ni} = \Gamma_i(W_i, \Sigma'_N;\theta)$.

*Proof.* First, for any $q \in \mathcal{A}\backslash\{0\}$ and $\ell \in \mathcal{A}$,

$$\left|\frac{\partial\Gamma_{ik}(W_i, \Sigma_N; \theta)}{\partial\alpha(q, \ell)} - \frac{\partial\Gamma_{ik}(W_i, \Sigma'_N; \theta)}{\partial\alpha(q, \ell)}\right|$$

$$= \left|\Gamma_{ik}\left\{\mathbf{1}(q = k) - \Gamma_{iq}\right\}\sum_{j\in\mathcal{N}_i}\Gamma_{j\ell} - \Gamma'_{Nik}\left\{\mathbf{1}(q = k) - \Gamma'_{Niq}\right\}\sum_{j\in\mathcal{N}_i}\Gamma'_{Nj\ell}\right|$$

$$\leq M \times \left|\Gamma_{ik} - \Gamma'_{Nik}\right| + M \times \left|\Gamma_{iq} - \Gamma'_{Niq}\right| + M \times \max_{j\in\mathcal{N}_i}\left|\Gamma_{j\ell} - \Gamma'_{Nj\ell}\right|,$$

almost surely, where the last step is because $0 \leq \Gamma_{jk} \leq 1$ for all $N, i, k$ and Assumption B. Thus

$$\left|\frac{\partial\Gamma_{ik}(W_i, \Sigma_N; \theta)}{\partial\alpha(q, \ell)} - \frac{\partial\Gamma_{ik}(W_i, \Sigma'_N; \theta)}{\partial\alpha(q, \ell)}\right| \leq 3M \times \max_{j\in\mathcal{N}_{(i,1)}}\left\|\Gamma_j - \Gamma'_{Nj}\right\|_1, \quad a.s.$$

Second, for any $q = 1, \cdots, p$

$$\left\|\frac{\partial\Gamma_{ik}(W_i, \Sigma_N; \theta)}{\partial\beta_{\ell'}} - \frac{\partial\Gamma_{ik}(W_i, \Sigma'_N; \theta)}{\partial\beta_q}\right\|_1$$

$$\leq \left|\Gamma_{ik}\left\{\mathbf{1}(q = k) - \Gamma_{iq}\right\} - \Gamma'_{Nik}\left\{\mathbf{1}(q = k) - \Gamma'_{Niq}\right\}\right| \times \|X_i\|_1$$

$$\leq 2C_0 \times \max_i \|\Gamma_i - \Gamma_i\|_1,$$

almost surely. Then let $C_2 = L \times \max\{3M, 2C_0\}$, there is

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\Gamma_{ik}(W_i, \Sigma_N; \theta)}{\partial\theta} - \frac{\partial\Gamma_{ik}(W_i, \Sigma'_N; \theta)}{\partial\theta}\right\|_1 \leq C_2 \times \sup_{\theta\in\Theta}\max_{j\in\mathcal{N}_{(i,1)}}\left\|\Gamma_j - \Gamma'_{Nj}\right\|_1, \quad a.s.$$

For the second part of the lemma,

$$\sup_{\theta\in\Theta}\left|\sum_{j\in\mathcal{N}_i}\sum_{\ell\in\mathcal{A}}\left\{\frac{\partial\Gamma_{ik}(W_i, \Sigma_N; \theta)}{\partial\sigma_{Nj\ell}} - \frac{\partial\Gamma_{ik}(W_i, \Sigma'_N; \theta)}{\partial\sigma_{Nj\ell}}\right\}\right|$$

$$= \sup_{\theta\in\Theta}\left|\sum_{j\in\mathcal{N}_i}\sum_{\ell\in\mathcal{A}}\sum_{q\neq k}\left[\left(\Gamma_{ik}\Gamma_{iq} - \Gamma'_{Nik}\Gamma'_{Niq}\right)\left\{\alpha(k, \ell) - \alpha(q, \ell)\right\}\right]\right|$$

$$\leq MK(K+1)\bar{\Delta}(\alpha) \times \sup_{\theta\in\Theta}\|\Gamma_i - \Gamma'_{Ni}\|_1,$$

almost surely. Then take $C_3 = MK(K+1)\bar{\Delta}(\alpha)$. □