Estimation of a Binary Choice Game Model with Network Links*

Xingbai Xu and Lung-fei Lee

November 25, 2015

Department of Economics, The Ohio State University, Columbus, OH.

Abstract

This paper studies a binary choice game model with network links, where the network peer effects are non-negative, and there might be only one or few networks in the sample. The model might have multiple Nash equilibria. We assume that the maximum Nash equilibrium, which always exists and is strongly coalition-proof and Pareto optimal, is selected. We investigate a simulated moment method for estimation. The challenging econometric issues are the possible correlation among all dependent variables in a network setting and the discontinuous functional form of our simulated moments. We overcome these challenges via the empirical process theory and derive the spatial near-epoch dependence (NED) of the dependent variable. We establish a criterion for an NED random field to be stochastically equicontinuous and we apply it to develop the consistency and asymptotic normality of the estimator. We examine computational issues and finite sample properties of the simulated moment method by some Monte Carlo experiments.

JEL: C13, C21, C25, C57, C63

Keywords: spatial autoregressive model; binary choice; supermodular game; near-epoch dependence; method of simulated moments

*An earlier version of this paper was presented in the 5th Shanghai Econometrics Workshop at the Shanghai University of Finance and Economics. We thank the audiences for their comments. We also acknowledge valuable discussions with Professor Robert de Jong.

1. Motivation and Introduction

In this paper, we study a spatial autoregressive (SAR) binary choice model with network links based on a static complete information game. This model can be used in various fields in economics, including agricultural economics, IO, spatial econometrics, and social networks, etc, for example, the decision for Wal-Mart to enter a county or not (Jia, 2008), peer effects in education and sport activities for adolescents (Liu, Patacchini, Zenou, 2014), peer effects for students in smoking (Krauth, 2006, Hsieh and Lee, 2014), presidential election (Lacombe and LeSage, 2013), the decision of adopting an agricultural program (Holloway, Shankara, and Rahman, 2002), the decision on whether to convert land-use from agricultural to non-agricultural uses (LeSage and Pace, 2009), the decision of adopting the District Planning System in Japan (Hoshino, 2009).

By representing a network as a graph, there are several methods to model binary choices on a graph. For the Besag logistic auto-model (Besag, 1974, Gaetan and Guyon, 2010), statisticians treat the data as a Markov random field and generalize the Ising model in statistical mechanics to study binary choices on graphs. This model is widely used in epidemiology. But this model does not involve rational decision, and thus it is not quite relevant in economics. For empirical economists, some have used the latent SAR probit model: $y_{i,n} = 1(y^*_{i,n} > 0)$ and $y^*_{i,n} = \lambda \sum_{j=1}^{n} w_{ij,n} y^*_{j,n} + x_{i,n}\beta + \epsilon_{i,n}$, where $y^*_{i,n}$ is a latent dependent variable, $W_n = (w_{ij,n})$ is a specified spatial weight matrix or adjacency matrix in network analysis, and $\epsilon_{i,n}$'s are usually modeled as normally distributed random variables. In this latent SAR probit model, it is the latent variables that directly affect neighborhood's utility, and the observed binary variable is an indicator on signs of the latent variable. This model is often estimated by Bayesian methods. Another interesting model is the simultaneous SAR binary choice model, $y_{i,n} = 1(\lambda \sum_{j=1}^{n} w_{ij,n} y_{j,n} + x_{i,n}\beta + \epsilon_{i,n} > 0)$, which is is different from the latent SAR probit model, in that $y^*_{j,n}$ is replaced by $y_{j,n}$ on the right hand side (RHS) of the equation. In the simultaneous SAR binary choice model, it is the linked individuals' realizations, rather than their latent variables, that affect the dependent variable of an individual. Both the latent and simultaneous SAR models are game models with perfect information. In the latent model, an individual's utility is affected by his/her friends' utilities, but econometricians can only observe binary indicators y's. In the simultaneous model, his utility is influenced by his friends' actions, rather than their utilities. So the model is more related to an econometric game model on discrete choices.

For the simultaneous SAR binary model, there are some challenging issues that need to overcome. First, there may be multiple solutions, or multiple Nash equilibria (NE) in the language of game theory. Furthermore, the number of equilibria might increase as number of players n increases. There may be correlation between any two $y_{i,n}$'s. Specifically, we pay special attention to the case of a single network, not only many similar independent markets or groups as assumed in the existing literature on econometrics for games and social interactions. For the estimation of games with multiple equilibria, there are various approaches including equilibrium selection (see, e.g., Bajari, Hong and Ryan, 2010) and set estimation (see, e.g., Chernozhukov, Hong and Tamer, 2007). In this paper, we adopt the approach of equilibrium selection. Our main goal is to develop a rigorous large sample theory for the estimation of the simultaneous discrete choice model within a single network, even though the theory may also be applied to the situation with many independent networks. In order to have some law of large numbers (LLN) and central limit theorems (CLT) for asymptotic analysis, we take advantage of recent development on nonlinear spatial processes by establishing weak spatial dependence properties for relevant variables and functions in the model. As there might exist multiple solutions and the indicator function $1(\cdot > 0)$ is not continuous, the investigation of weak dependence for variables of the model is a research pursuit.

This paper contributes to the literature in the following aspects. (1) It develops a consistent and asymptotically normally distributed estimator for the model; so it enriches econometric tools to deal with binary data on networks, in particular, a single network. (2) We introduce a procedure to establish spatial near-epoch dependence (NED) of outcome variables of a nonlinear model without the need of a contraction mapping. (3) It extends the literature of spatial NED to the case of a network game. (4) It shows some criteria for the stochastic equicontinuity (SEC) for NED random fields. (5) It extends some simulation estimation theory to the case with spatial dependence as the estimation method under consideration is a simulated moment estimator.

More Related Literature. In addition to the literature mentioned above, this paper is related to some other publications. Some researchers apply simulation methods to discrete choices models, e.g., Pakes and Pollard (1989) and Krauth (2006). Simulation estimation is needed because correlation leads to complex integration. But their studies focus on many independent individuals or markets setting. Our paper differs from their studies in that: (i) we focus on games with many individuals linked in a network; (ii) simulation is needed to locate selected equilibria over the range of unobserved disturbances in order to calculate choice probabilities. Recently, the studies of estimation of games with many players have attracted more attention. Menzel (2014) investigates two-sided matching markets with a large number of participants and non-transferable utility. Menzel (2015) examines anonymous games where payoffs depend on an agent's own action and the empirical distribution of others' actions. He requires that the number of players in every group increases at the same rate. Our paper differs from Menzel (2015) in at least two aspects: (i) we do not need to exclude the existence of some small and independent markets or various increasing rates of market size; (ii) the predetermined relationship of individuals, reflected in W_n, indicates diverse importance of players to each specific individual.

The structure of this paper is as follows: In Section 2, we introduce an SAR binary choice model and discuss the strongly coalition-proof equilibrium of this model. For any equilibrium from a selection rule, it is shown to be a spatial NED process. In Section 3, we consider the simulated moment estimation of this model and establish consistency and asymptotic normality of the estimator. In Section 4, Monte Carlo studies are performed to examine the finite sample performance of the estimator. In Appendix A, we list conditions for SEC of NED random fields. In Appendix B, we generalize conditions for asymptotic distributions of simulated moment estimators to network data with spatial correlation. In Appendix C, we show some results on derivatives of the probability of the NE, which is needed for asymptotic analysis in the main text. All the proofs for theorems, propositions and lemmas in the main text are collected in Appendix D.

2. The Model

2.1. Model Setup and Equilibrium Concepts

Assume that there are n individuals or players, which may be consumers, firms, local governments, etc, living in a d-dimensional Euclidean space $\mathbb{R}^d$. We use $\vec{i} \in \mathbb{R}^d$ to represent individual i's location in $\mathbb{R}^d$. $\vec{i}$ may include both its geographic location and some socio-cultural or economic characteristics of individual i. Denote $d_{ij} \equiv d(\vec{i}, \vec{j})$, the distance between individuals i and j.

Assumption 1. For any $i \neq j$, d_{ij} is larger than or equal to a specific positive constant, without loss of generality, say, 1.

Assumption 1 implies the increasing domains asymptotic and excludes infilled asymptotic. Each individual is endowed with a vector $x_{i,n} \in \mathbb{R}^K$ of some exogenous characteristics, which is observable to all players and econometricians, and $\epsilon_{i,n} \in \mathbb{R}$ is public to all players but is unobservable to econometricians, i.e., we are considering a complete information static game. Player i chooses his action $y_{i,n}$ from the strategy set $\{0, 1\}$. At the alternative 0, his utility is normalized to be zero; and for alternative 1, his utility is $u(y_{i,n} = 1|y_{-i,n}, X_n) = \lambda_0 \sum_{j=1}^n w_{ij,n} y_{j,n} + x_{i,n}\beta_0 + \epsilon_{i,n}$, where $w_{ij,n} \geqslant 0$ is the (i, j)-entry of the exogenous spatial weight matrix (adjacency matrix) W_n,[1] the subscript 0 for a parameter represents its true value, X_n consists of all $x_{i,n}$, and $-i \equiv \{1, \cdots, n\}\backslash\{i\}$ is the set of all individuals but with i excluded. Individual i chooses $y_{i,n} = 1$ iff $u(y_{i,n} = 1|y_{-i,n}, X_n) \geqslant 0$. This model can be written as

$$y_{i,n} = 1(\lambda \sum_{j=1}^{n} w_{ij,n} y_{j,n} + x_{i,n}\beta + \epsilon_{i,n} > 0), \tag{1}$$

where $1(\cdot)$ is the set indicator. We treat the solution to Eq. (1) for all $i = 1, \cdots, n$ as the NE of the n-player game. So the game under study is a non-cooperative game.

Assumption 2. $\lambda_0 \geqslant 0$.

By Assumption 2, an individual's action has non-negative externalities on other players and the game is strategically complementary. Such a game is a supermodular game (Milgrom and Roberts, 1990). Supermodular games are used in oligopoly competition, macroeconomics (Diamond search model), arms races, technology adoption and diffusion, and many others (Milgrom and Roberts, 1990). The estimation of supermodular games is studied in Uetake and Watanabe (2013), Molinari and Rosen (2008), Jia (2008) where she investigates the expansion of Wal-Mart in counties, and Miyauchi (2014) where he examines the estimation of network formation games. If Assumption 2 fails, there might be no pure strategy NE. See Appendix B.2 in Jia (2008) for a numerical counter-example.

From Milgrom and Roberts (1990), a complete information static supermodular game always has at least one NE.[2] There might be multiple NE, and the set of NE is a complete sublattice in $\{0, 1\}^n$,

[1] Endogenous W_n is an interesting issue in network formation. But in this paper, our intention is to develop asymptotic theory for estimation of a discrete choice game with players connected in a network, so we treat W_n as given and binary choice decisions are not subject to selectivity in network formation.

[2] This is from Tarski's fixed point theorem, which applies for a complete lattice. For a linear SAR model, Tarski's fixed point theorem is not applicable, because $\mathbb{R}^n$, the range of y_n in a linear SAR model, is not a complete lattice.

which contains its supremum and infimum. Furthermore, the largest NE is Pareto optimal. There is a concept stronger than NE, namely, strongly coalition-proof equilibrium (SCPE) in Milgrom and Roberts (1995).

Definition 1. Let $\Gamma = \{\{1, 2, \cdots, n\}, \mathcal{S}, (\pi_1, \cdots, \pi_n)\}$ be a normal form game, where $\{1, 2, \cdots, n\}$ is the set of players, $\mathcal{S} = \prod_{i=1}^{n} \mathcal{S}_i$ is the strategy set defined to be the product of strategy spaces of players , where $\mathcal{S}_i$ is the strategy space of player i, and $\pi_i : \mathcal{S} \to \mathbb{R}$ is the payoff function of individual i. $s^* \in \mathcal{S}$ is called an *SCPE* if and only if for any proper nonempty subset $J \subsetneqq \{1, 2, \cdots, n\}$, there is no $s_J \in \prod_{j\in J} \mathcal{S}_j$ such that $\pi_j(s^*) \leqslant \pi_j(s_J, s^*_{-J})$ for all $j \in J$ and $\pi_j(s^*) < \pi_j(s_J, s^*_{-J})$ for some $j \in J$.

SCPE is immune to incentive-compatible deviations by coalitions. In an environment where players can freely discuss their strategies, but cannot make binding commitment, it is possible for coalitions of players to arrange plausible, mutually beneficial deviations for Nash agreements (Bernheim, Peleg and Whinston, 1987). From Theorem 2 in Milgrom and Roberts (1995), the maximum NE of our model is an SCPE. If we do not consider the critical case that some individual is indifferent for his choices, of which the probability is zero, the maximum NE of our game is the unique SCPE (Milgrom and Roberts, 1995). Thus, our model has a unique SCPE almost surely and we only consider the maximum NE as the chosen one in a sample in this paper. Jia (2008, pp. 1279-1280) proposed a simple and fast way to calculate the maximum NE: (1) let $y_n^0 = (1, \cdots, 1) \in \mathbb{R}^n$; (2) $y_n^{t+1} = 1(\lambda W_n y_n^t + X_n\beta + \epsilon_n > 0)$, where $1(u > 0) \equiv (1(u_1 > 0), \cdots, 1(u_n > 0))$; (3) the iteration process stops once $y_n^{t+1} = y_n^t$. The iteration has at most n steps, since y_n^t is nonincreasing in t.

2.2. Near-epoch Dependence

In this model, there are both heterogeneity among players and correlation among the decisions of players due to interactions. Thus, we need some type of weak dependence concept so that proper LLN and CLT can be valid in order to develop a rigorous asymptotic theory for estimators. We utilize the concept of spatial NED developed in Jenish and Prucha (2012). Intuitively, if a random field $\{z_{i,n}\}_{i=1}^n$ is NED on a base $\{\epsilon_{i,n}\}_{i=1}^n$, then the $\epsilon_{j,n}$'s, with j's near i, are able to give a good prediction of $z_{i,n}$. To establish NED, we need more assumptions.

Assumption 3. $W_n \neq 0$ is a non-stochastic $n \times n$ matrix with non-negative entries and zero diagonal elements. Furthermore, $B_W \equiv \sup_n ||W_n||_\infty < \infty$.

Assumption 4. $\{\epsilon_{i,n}\}_{i=1}^n$ are i.i.d. with support $\mathbb{R}$; $f(\cdot)$, the pdf of $\epsilon_{i,n}$, and its derivative $f'(\cdot)$ satisfy $B_f = \sup_\epsilon f(\epsilon) < \infty$ and $B_{f'} = \sup_\epsilon |f'(\epsilon)| < \infty$; $\{\epsilon_{i,n}\}_{i=1}^n$ are independent of $\{x_{i,n}\}_{i=1}^n$.

Assumption 5. The parameter space of $\theta = (\lambda, \beta')$ is $\Theta = [0, B_\lambda] \times \prod_{k=1}^K [-B_{\beta_k}, B_{\beta_k}] \subseteq \mathbb{R}^{K+1}$.

Under Assumptions 3 - 5, $\sup_{i,n,\theta,x_{i,n}} \Pr(-\lambda ||W_n||_\infty - x_{i,n}\beta \leqslant \epsilon_{i,n} < -x_{i,n}\beta | x_{i,n}) = \bar{\delta}$ for some $\bar{\delta}$ such that $0 < \bar{\delta} < 1$. Previously, we have defined a distance between i and j, d_{ij}, used in Assumption 1. Now we need another concept of distance in terms of network links between individuals. We call that individual j affects individual i directly, denoted $j \to i$, if and only if $w_{ij,n} \neq 0$.[3] A path $j_k \to j_{k-1} \to \cdots \to j_0$, is defined to satisfy two conditions: (1) any two individuals involved are different, and (2) $j_p \to j_{p-1}$ for all $1 \leqslant p \leqslant k$. Call the length of $j_k \to j_{k-1} \to \cdots \to j_0$ as k. For $j \neq i$, define d^{ij} as the smallest $k \in \mathbb{N} \equiv \{1, 2, \cdots\}$ such that there exists a path from j to i with length k. If there are no paths from j to i, then $d^{ij} \equiv \infty$. And define $d^{ii} \equiv 0$. Thus, $j \to i$ is equivalent to $d^{ij} = 1$. Notice that it is possible that $d^{ij} \neq d^{ji}$ in a directed graph. But for an undirected graph, $d^{ij} = d^{ji}$. Another equivalent definition for d^{ij} is that $d^{ij} \equiv \inf\{1 \leqslant k \in \mathbb{N} : (W_n^k)_{ij} \neq 0\}$ when $i \neq j$. For a set A, $|A|$ denotes its cardinality.

Assumption 6. There is an $m_0 \in \mathbb{N}$ such that $\bar{\delta}\bar{l}_p < 1$, where

$$\bar{l}_p \equiv \sup_{m \geqslant m_0} \sup_{i,n} |\{\text{path}\, j_m \to j_{m-1} \to \cdots \to j_1 \to i : d^{ij_m} = m\}|^{1/m}.$$

Assumption 6 includes the case of many independent groups, where the size of each group is bounded by a given natural number $\bar{n}$. Let $m_0 = \bar{n} + 1$, then $\bar{l}_p = 0$ and Assumption 6 holds trivially. For a single network or a network with large components, define a matrix W_n^* as follows: $w_{ij,n}^* = 1$ if and only if $w_{ij,n} > 0$; otherwise, $w_{ij,n} = 0$. $\sum_{j_m=1}^n (W_n^*)_{ij_m}^n = \sum_{j_m} \cdots \sum_{j_1} w_{ij_1,n}^* w_{j_1 j_2,n}^* \cdots w_{j_{m-1}j_m,n}^*$ is the number of walks[4] ending in i. Thus, $|\{\text{path}\, j_m \to \cdots \to j_1 \to i : d^{ij_m} = m\}| \leqslant \sum_{j_m=1}^n (W_n^*)_{ij_m}^m \leqslant ||(W_n^*)^m||_\infty$. By Gelfand's formula, $\lim_{m\to\infty} ||(W_n^*)^m||_\infty^{1/m} = \rho(W_n^*)$, where $\rho(W_n^*)$ is the spectral radius of W_n^*. Thus, for any $\epsilon > 0$, we can choose m_0 large

[3]In our model, we have $w_{ii,n} = 0$, so we do not care about direct self-influence. In fact, direct self-influence plays no role in our theory. We take this motivation from a named friend network. In such a network, $w_{ij,n}$=1 if individual i names j as his friend. Thus, j is i's role model who will influence i.

[4]$j_m j_{m-1} \cdots j_1$ is defined to be a walk iff $j_p \to j_{p-1}$ for all $2 \leqslant p \leqslant m$. Any two individuals on a path are different, but they can be the same on a walk.

enough such that $\bar{l}_p \leqslant \sup_n \rho(W_n^*) + \epsilon$. Since $\bar{\delta}$ is related to $\lambda||W_n||_\infty$, Assumption 6 adds some constraints on $B_\lambda||W_n||_\infty$ and the structure of the network (due to $\bar{l}_p$). More dense network (greater $\bar{l}_p$) the structure is, weaker interaction (less $\lambda||W_n||_\infty$) has to be assumed. For instance, when $\epsilon_{i,n}$ is standard normally distributed, then $\bar{\delta} = 2\Phi(B_\lambda \sup_n ||W_n||_\infty/2) - 1$. Assumption 6 implies that $\Phi(B_\lambda \sup_n ||W_n||_\infty/2) < \frac{\bar{l}_p+1}{2\bar{l}_p}$. If $||W_n||_\infty = 1$, $\bar{l}_p = 2$, then $B_\lambda < 1.349$; $||W_n||_\infty = 1$, $\bar{l}_p = 3$, then $B_\lambda < 0.86$. For the standard logistic distribution, if $||W_n||_\infty = 1$, $\bar{l}_p = 3$, then $B_\lambda < 1.38$.

Assumption 7. $d^{ij} = 1$ implies $d_{ij} \leqslant \bar{d}_0$ for some constant distance $\bar{d}_0 > 1$.

Assumptions 2-6 are sufficient to establish the NED of $y_{i,n}(\theta)$, if the metric used is d^{ij} (see Proposition 1). For Euclidean distance, Assumptions 1 and 7 are required. Assumption 7 implies that only individuals within distance $\bar{d}_0$ may directly affect each other. This assumption is widely used in spatial statistics and spatial econometrics, e.g., Xu and Lee (2015a). Consider the system $y_{i,n} = 1(\lambda_i \sum_{j=1}^n w_{ij,n} y_{j,n} + x_{i,n}\beta + \epsilon_{i,n} > 0)$, where $0 \leqslant \lambda_i \leqslant B_\lambda$. This system is more general in that each individual i may have a different interaction coefficient λ_i. This generalization is just for the purpose of having a general theoretical result in Lemmas C.2 and C.3.[5] Denote $\theta = (\lambda_1, \cdots, \lambda_n, \beta')'$ and $X_n = (x'_{1,n}, \cdots, x'_{n,n})'$. Then $y_{i,n} = y_{i,n}(\epsilon_n, X_n, \theta)$. For any natural number $m \in \mathbb{N}$ and any individual i, separate individuals into two sets: $\{j : d^{ij} \leqslant m\}$ and $\{j : d^{ij} > m\}$. Conformable to this partition, we have $X_n = (x_n^{(i,\leqslant m)}, x_n^{(i,>m)})$ and $\epsilon_n = (\epsilon_n^{(i,\leqslant m)}, \epsilon_n^{(i,>m)})$. The following proposition describes the conditional probability of difference for an individual's choices with $(x_n^{(i,>m)}, \epsilon_n^{(i,>m)})$ at two different values. The probability in part (1) of Proposition 1 is taken over $\epsilon_n^{(i,\leqslant m)}$, and is relevant for the spatial NED property. The probability in part (2) has an additional conditional argument, and will be useful for analyzing second order derivatives of choice probabilities.

Proposition 1. *Let $m \in \mathbb{N}$ such that $m_0 \leqslant m$. Denote $\bar{y}_{i,n} = y_{i,n}(\epsilon_n^{(i,\leqslant m)}, \bar{\epsilon}_n^{(i,>m)}, x_n^{(i,\leqslant m)}, \bar{x}_n^{(i,>m)}, \theta)$ and $\tilde{y}_{i,n} = y_{i,n}(\epsilon_n^{(i,\leqslant m)}, \tilde{\epsilon}_n^{(i,>m)}, x_n^{(i,\leqslant m)}, \tilde{x}_n^{(i,>m)}, \theta)$, where $(\bar{x}_n, \bar{\epsilon}_n)$ and $(\tilde{x}_n, \tilde{\epsilon}_n)$ are two different values of (x_n, ϵ_n). Then,*

(1) *for all i, n, θ, $\bar{x}_n^{(i,>m)}$, $\tilde{x}_n^{(i,>m)}$, $\bar{\epsilon}_n^{(i,>m)}$ and $\tilde{\epsilon}_n^{(i,>m)}$*

$$\Pr\left(\bar{y}_{i,n} \neq \tilde{y}_{i,n} | \, x_n^{(i,\leqslant m)}, \bar{x}_n^{(i,>m)}, \tilde{x}_n^{(i,>m)}, \bar{\epsilon}_n^{(i,>m)}, \tilde{\epsilon}_n^{(i,>m)}\right) \leqslant \bar{\delta}(\bar{\delta}\bar{l}_p)^m; \tag{2}$$

[5]For estimation, we can not allow λ_i's to be different. Otherwise, there are too many parameters to be estimable.

(2) *if* $d^{ik} \leqslant m$, *then for all* i, n, θ, $\epsilon_{k,n}$, $\bar{x}_n^{(i,>m)}$, $\tilde{x}_n^{(i,>m)}$, $\bar{\epsilon}_n^{(i,>m)}$ *and* $\tilde{\epsilon}_n^{(i,>m)}$,

$$\Pr\left(\bar{y}_{i,n} \neq \tilde{y}_{i,n} | \, \epsilon_{k,n}, x_n^{(i,\leqslant m)}, \bar{x}_n^{(i,>m)}, \tilde{x}_n^{(i,>m)}, \bar{\epsilon}_n^{(i,>m)}, \tilde{\epsilon}_n^{(i,>m)}\right) \leqslant (\bar{\delta}\bar{l}_p)^m. \quad (3)$$

Now we state the NED result for the dependent variable of our model.[6]

Corollary 1. *Let* $\mathcal{F}_{i,n}(s) \equiv \sigma(\{x_{j,n}, \epsilon_{j,n} : d_{ij} \leqslant s\})$ *and* $y_{i,n}(\theta) = y_{i,n}(\epsilon_n, X_n, \theta)$. *Let* $m_0 \leqslant m \in \mathbb{N}$.

(1) *Under Assumptions 2-6,* $\sup_{i,n,\theta\in\Theta} ||y_{i,n}(\theta) - \mathrm{E}[y_{i,n}(\theta)|x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m]||_{L^2} \leqslant (\bar{\delta}\bar{l}_p)^{m/2}$.

(2) *Under Assumptions 1-7,* $\sup_{i,n,\theta\in\Theta} ||y_{i,n}(\theta) - \mathrm{E}[y_{i,n}(\theta)|\mathcal{F}_{i,n}(m\bar{d}_0)]||_{L^2} \leqslant (\bar{\delta}\bar{l}_p)^{m/2}$.

Corollary 2. *Denote* $B_X = \sup_{i,k,n} ||x_{ik,n}||_{L^2}$. *Let* $m_0 \leqslant m \in \mathbb{N}$.

(1) *Under Assumptions 2-6,* $\sup_{i,k,n,\theta\in\Theta} ||y_{i,n}(\theta)x_{ik,n} - \mathrm{E}[y_{i,n}(\theta)x_{ik,n}|x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m]||_{L^2} \leqslant B_X(\bar{\delta}\bar{l}_p)^{m/2}$ *and* $\sup_{i,k,n,\theta\in\Theta} ||y_{i,n}(\theta)\sum_{l=1}^n w_{il,n}x_{lk,n} - \mathrm{E}[y_{i,n}(\theta)\sum_{l=1}^n w_{il,n}x_{lk,n}|x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m]||_{L^2} \leqslant B_X B_W(\bar{\delta}\bar{l}_p)^{m/2}$.

(2) *Under Assumptions 1-7,* $\sup_{i,n,k,\theta\in\Theta} ||y_{i,n}(\theta)x_{ik,n} - \mathrm{E}[y_{i,n}(\theta)x_{ik,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_{L^2} \leqslant B_X(\bar{\delta}\bar{l}_p)^{m/2}$ *and* $\sup_{i,k,n,\theta\in\Theta} ||y_{i,n}(\theta)\sum_{l=1}^n w_{il,n}x_{lk,n} - \mathrm{E}[y_{i,n}(\theta)\sum_{l=1}^n w_{il,n}x_{lk,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_{L^2} \leqslant B_X B_W(\bar{\delta}\bar{l}_p)^{m/2}$.

Notice that the NED property is uniform not only in i and n, but also in θ. This is needed when we apply some empirical process techniques to establish a large sample theory for our proposed estimator.

3. MSM and its Large Sample Properties

3.1. The MSM Estimator and Its Consistency

The likelihood of the model is $L_n(\theta|Y_n, X_n) = \Pr(\{\epsilon_n : \text{The maximum NE is } Y_n\}|\theta, X_n)$, where $\theta = (\lambda, \beta')'$. However, the method of ML estimation might be difficult to work with for this model. There are several disadvantages for the ML approach: (1) For large n, the closed form of the likelihood function is not available. (2) It is also hard to simulate. Due to the complex model structure, it seems natural to simulate outcomes by simulating disturbances and uses a frequency simulator to approximate the probability of an observed SCPE Y_n. However, because there are totally 2^n various and possible Y_n's, it needs an exponential number of simulations to obtain an accurate estimation of $L_n(\theta|Y_n)$. Furthermore, that is only for a specific θ. To maximize the

[6] The idea of proof is enlightened by de Jong and Woutersen (2011), even though the latter is for time series.

likelihood function, we need to evaluate it at different values of θ. (3) It seems complicated to obtain asymptotic properties from this likelihood function. It is difficult to establish pointwise convergence in probability, let alone uniform convergence in probability and asymptotic distribution.

Therefore, we decide to consider generalized method of moments (GMM), or more precisely, method of simulated moments (MSM). Since the closed form of moments is not available, simulation estimation is needed. Various moment conditions can be obtained from the conditional moment condition $\mathrm{E}[\Pr(y_{i,n}=1|X_n,\theta_0)-y_{i,n}|X_n]=0$, such as $\mathrm{E}\{[\Pr(y_{i,n}=1|X_n,\theta_0)-y_{i,n}]x'_{i,n}\}=0$ and $\mathrm{E}\{[\Pr(y_{i,n}=1|X_n,\theta_0)-y_{i,n}](w_{i\cdot,n}X_n)'\}=0$, where $w_{i\cdot,n}$ is the i-th row of W_n. The following proposition implies that this conditional moment condition is sufficient to identify the unknown parameters of our model.

Proposition 2. *Denote* $X_n=[X_{1n},X_{2n}]$ *where* X_{2n} *is an* n*-dimensional column vector. Assume* $\mathrm{support}(X_{2n}|X_{1n})=\mathbb{R}^n$ *and* $\beta_{20}\neq 0$, where β_{20} is the coefficient of of X_{2n}. *Under Assumptions 3 and 4, if* $\mathrm{E}(x'_{i,n}x_{i,n})$ *has full rank for every* i*, then* $\mathrm{E}[\Pr(y_{i,n}=1|X_n,\theta)-y_{i,n}|X_n]=0$ for all i implies $\theta=\theta_0$.

Let $q_{i,n}$ be a vector of IV variables for i, and $Q_n\equiv(q'_{1,n},\cdots,q'_{n,n})'$, which are functions of W_n and X_n. Let

$$g_{i,n}(\theta)\equiv[\Pr(y_{i,n}=1|X_n,\theta)-y_{i,n}]q'_{i,n}. \tag{4}$$

Then $\mathrm{E}\,g_n(\theta_0)=0$. By having R random draws $\epsilon_n^{(r)}$ $(r=1,\cdots,R)$ from the distribution of ϵ_n, where R does not need to depend on n, and then generating $y_{i,n}(\epsilon_n^{(r)},X_n,\theta)$'s as dependent variables from the model, $\hat{\Pr}(y_{i,n}=1|X_n,\theta)=\frac{1}{R}\sum_{r=1}^{R}y_{i,n}(\epsilon_n^{(r)},X_n,\theta)$ is an unbiased simulator of $\Pr(y_{i,n}=1|X_n,\theta)$. Denote

$$\hat{g}_{i,n}(\theta)=[\hat{\Pr}(y_{i,n}=1|X_n,\theta)-y_{i,n}]q'_{i,n}. \tag{5}$$

Notice that $\mathrm{E}\,\hat{g}_{i,n}(\theta_0)=0$, because $\hat{\Pr}(y_{i,n}=1|X_n,\theta)$ is an unbiased simulator. When the number of moments is greater than $K+1$, we choose a (possibly stochastic) positive definite matrix $\Omega_n(\theta)$ to be a GMM weighting matrix.

Assumption 8. The (stochastic) $\Omega_n(\theta)$ converges in probability uniformly to $\Omega(\theta)$, i.e., $\sup_{\theta\in\Theta}|\Omega_n(\theta)-\Omega(\theta)|=o_p(1)$. The $\Omega(\theta)$ is positive definite for any $\theta\in\Theta$ such that $\inf_{\theta\in\Theta}\min\mathrm{eig}\,\Omega(\theta)>0$ and $\sup_{\theta\in\Theta}\max\mathrm{eig}\,\Omega(\theta)<\infty$, where $\mathrm{eig}\,\Omega(\theta)$ denotes the set of eigenvalues of $\Omega(\theta)$.

Then, the MSM estimator is $\arg\min_{\theta\in\Theta} Q_n(\theta)$, where

$$Q_n(\theta) \equiv \left[\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{i,n}(\theta)\right]' \Omega_n(\theta) \left[\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{i,n}(\theta)\right] \equiv \left\|\Omega_n^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{i,n}(\theta)\right\|^2, \tag{6}$$

is a sample moment objective function with $||\cdot||$ being the Euclidean norm. Let

$$\bar{Q}_n(\theta) \equiv \left[\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\, g_{i,n}(\theta)\right]' \Omega(\theta) \left[\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\, g_{i,n}(\theta)\right] \equiv \left\|\Omega^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\, g_{i,n}(\theta)\right\|^2$$

be the norm of the population moment counterpart. We need some moment conditions in order to obtain large sample properties of the estimator. Formally, the requirements for the IVs are stated in Assumption 9:

Assumption 9. (1) There are two real numbers $2 < p_0 \leqslant q_0$ and an even number $w_0 > 2p_0(K+1)r_0^{-1}$, where $r_0 \equiv \frac{p_0^{-1}-q_0^{-1}}{2^{-1}-q_0^{-1}}$, such that $\frac{1}{2} = \frac{1}{p_0} + \frac{w_0-1}{q_0}$ and $B_Q \equiv \sup_{k,i,n}||q_{ik,n}||_{L^{q_0}} < \infty$;

(2) $q_{i,n}$ is measurable with respect to $\sigma(\{x_{j,n} : d_{ij} \leqslant m_0\bar{d}_0\})$.

(3) $\forall\, \theta \neq \theta_0$, $\liminf_{n\to\infty} \bar{Q}_n(\theta) > 0$.

Assumption 9(1) adds some moment requirements on $q_{i,n}$. If $x_{i,n}$ is of dimension K with $K \geqslant 2$, then we need $w_0 > 12$ and $q_0 \geqslant 2w_0 > 24$.[7] Thus, we require high moment conditions for $q_{i,n}$. For example, $(p_0, w_0, q_0, r_0) = (3, 10K + 10, 60K + 54, \frac{20K+17}{30K+26})$ satisfies Assumption 9(1). With Assumptions 9 (1) and (2), similarly to Corollary 2, $\sup_{i,k,n,\theta\in\Theta}||y_{i,n}(\theta)q_{ik,n} - \mathrm{E}[y_{i,n}(\theta)q_{ik,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_{L^2} \leqslant B_Q(\bar{\delta}\bar{l}_p)^{m/2}$, whenever $m \geqslant m_0$. Assumption 9(3) is an identification condition, which can be satisfied if the IVs extract enough identification information from the conditional expectation. Finally, some α-mixing conditions are required in order that the LLN in Jenish and Prucha (2012) can be applied.

Assumption 10. $\{x_{i,n}\}_{i=1}^{n}$is spatially α-mixing with coefficients $\alpha(u, v, r) \leqslant (u + v)^\tau\hat{\alpha}(r)$ for some $\tau \geqslant 0$ and $\sum_{r=1}^{\infty} r^{d-1}\hat{\alpha}(r) < \infty$.[8]

With all these assumptions, the consistency of the MSM estimator follows.

Theorem 1. *Under Assumptions 1-10,* $\hat{\theta}_n \xrightarrow{p} \theta_0$.

[7]Because $p_0 \leqslant q_0$, $\frac{1}{2} = \frac{1}{p_0} + \frac{w_0-1}{q_0} \geqslant \frac{1}{q_0} + \frac{w_0-1}{q_0} = \frac{w_0}{q_0}$ and $q_0 \geqslant 2w_0$.

[8]See Jenish and Prucha (2009) for the concept of spatially α-mixing random fields.

Remark. As contrary to a simulated likelihood approach, for the MSM estimation, we do not require that the number of random draws, R, increases as the sample size n increases in order to obtain consistency (Train, 2009).

3.2. Asymptotic Distribution of the Estimator

Theorem 3.3 in Pakes and Pollard (1989) is a general result for the asymptotic distribution of an estimator based on simulation of frequencies. We generalize their result to allow heterogeneity in Theorem B.1 and to allow that the true parameter is on the boundary of the parameter space in Theorem B.2 in the Appendix. Condition (iii) in Theorems B.1 and B.2 requires stochastic equicontinuity (SEC) of the empirical process of sample moments. Thus, we first apply Theorem A.1 to obtain the SEC of $g_{in}(\theta)$ in Eq. (4) and $\hat{g}_{in}(\theta)$ in Eq. (5) of our model.

Proposition 3. *Let $w \in (2p_0(K+1)r_0^{-1}, w_0]$ be an even number. Then, for any $\epsilon > 0$, there exists an $\eta > 0$, such that for any k,*

$$\limsup_{n\to\infty} \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \left| n^{-1/2} \sum_{i=1}^{n} [g_{ik,n}(\theta_1) - g_{ik,n}(\theta_2)] \right| \right\|_{L^w} \leqslant \epsilon.$$

Corollary 3. *Let $w \in (2p_0(K+1)r_0^{-1}, w_0]$ be an even number. Then, for any $\epsilon > 0$, there exists an $\eta > 0$, such that for any k,*

$$\limsup_{n\to\infty} \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \left| n^{-1/2} \sum_{i=1}^{n} [\hat{g}_{ik,n}(\theta_1) - \hat{g}_{ik,n}(\theta_2)] \right| \right\|_{L^w} \leqslant \epsilon.$$

We still need some additional regularity conditions in order to apply Theorem B.1 to establish the asymptotic distribution of the estimator.

Assumption 11. $\lambda_0 < B_\lambda$ and $\beta_0 \in \prod_{k=1}^{K}(-B_{\beta_k}, B_{\beta_k})$.

Assumption 12. The α-mixing coefficient $(u+v)^\tau \hat{\alpha}(r)$ of $\{x_{i,n}\}_{i=1}^n$ satisfies

$$\sum_{r=1}^{\infty} r^{d(\tau_*+1)-1} \hat{\alpha}(r)^{\delta_*/(4+2\delta_*)} < \infty$$

for some $\delta_* > 0$, where $\tau_* \equiv \delta_* \tau/(2+\delta_*)$.

Assumption 13. $\frac{1}{n}\sum_{i=1}^{n} \partial \operatorname{E} g_{i,n}(\theta_0)/\partial\theta' \to \Gamma$ and Γ has full column rank.

Assumption 14. var $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_{i,n}(\theta_0) \to V$ and V is nonsingular.

Theorem 2. *Under Assumptions 1-14,* $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (1+\frac{1}{R})(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega V\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1})$ *if* $\lambda_0 \in (0, B_\lambda)$. *On the other hand, if* $\lambda_0 = 0$, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *will converge to a mixture of distributions: with a half probability, a truncated* $N(0, (1+\frac{1}{R})(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega V\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1})$ *on* $(0,\infty)\times\mathbb{R}^K$*, and with a half probability, a multivariate distribution with its first component degenerated at 0 and the remaining components* $N(0, (1+\frac{1}{R})[(\Gamma'\Omega\Gamma)_{-1}]^{-1}[\Gamma'\Omega V\Omega\Gamma]_{-1}[(\Gamma'\Omega\Gamma)_{-1}]^{-1})$ *on* $\mathbb{R}^K$*, where, for a matrix, "-1" means the submatrix without the first row and column.*

To conduct statistical inference, we need to estimate the asymptotic distribution of $\hat{\theta}_n$. The analytical variance approximation is rather intractable, so we suggest constructing confidence intervals by bootstrap. Specifically, (1) generate B vectors of $\epsilon_n^{(b)}$'s, where $b = 1, \cdots, B$, of ϵ_n, and generate the corresponding $y_n(\epsilon_n^{(b)}, X_n, \hat{\theta}_n)$; (2) for each set of samples, estimate a $\hat{\theta}_n^{(b)}$; (3) construct confidence intervals according to $\hat{\theta}_n^{(b)}$'s. The validity of the Bootstrap method is based on the following proposition. Denote $\hat{g}_{i,n}^{(b)}(\theta) = [\hat{\Pr}(y_{i,n} = 1|X_n,\theta) - y_n(\epsilon_n^{(b)}, X_n, \hat{\theta}_n)]q_{i,n}'$, where $Q_n^{(b)}(\theta) \equiv \left\|\Omega_n^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{i,n}^{(b)}(\theta)\right\|^2$ and $\hat{\theta}_n^{(b)} \equiv \arg\min_{\theta\in\Theta} Q_n^{(b)}(\theta)$.

Proposition 4. *(1)* $\hat{\theta}_n^{(b)} = \theta_0 + o_p(1)$. *(2) When* $\lambda_0 \in [0, B_\lambda)$, $\sqrt{n}(\hat{\theta}_n^{(b)} - \theta_0)$ *has the same limiting distribution as that of* $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

4. Monte Carlo Simulation

In this section, we design some experiments to investigate finite sample properties of the MSM estimator. The data generating process for the experiments is $y_{i,n} = 1(\lambda_0\sum_{j=1}^{n} w_{ij,n}y_{j,n} + \beta_{10} + \beta_{20}x_{i,n} + \epsilon_{i,n} > 0)$, where $\beta_{10} = 0$, $\beta_{20} = 1$, but λ_0 is designed to be 0.6 or 0.3; $x_{i,n}$'s *i.i.d.* $N(0,1)$, $N(0,(\sqrt{2})^2)$ or $N(0,(\sqrt{3})^2)$, and $\epsilon_{i,n}$'s are *i.i.d.* standard normally distributed. $y_{i,n}$ is generated by the method mentioned in Section 2.1, namely, by iterating from $(1,\cdots,1)'$. Experiments show that two to six iterations, and on average fewer than 4 iterations, are sufficient to generate $y_n = (y_{1,n},\cdots,y_{n,n})'$ when sample size varies between 400 and 1600. Thus the algorithm is quite fast. We use 1, $x_{i,n}$ and $\sum_{j=1}^{n} w_{ij,n}x_{j,n}$ as the IVs for the orthogonality moment conditions. Because the number of IVs and the number of parameters are both 3, the parameters are exactly identified. The choice of Ω_n does not matter theoretically and thus the weighting matrix for moments can be the 3×3 identity matrix. For the simulated probability $\hat{\Pr}(y_{i,n} = 1|X_n,\theta) = \frac{1}{R}\sum_{r=1}^{R} y_{i,n}(\epsilon_n^{(r)}, X_n, \theta)$,

R is chosen to be 100 in order to balance efficiency, computational time and numerical precision, where the last issue will be discussed later in this section. For each θ_0 and n, 500 experiments are performed to obtain sample means and sample standard deviations of estimators.

Individuals are assumed to be located on a $\sqrt{n} \times \sqrt{n}$ square integer lattice (chess board). In the experiments, we try $\sqrt{n} = 20$, 30, 40. Thus, the corresponding sample sizes are 400, 900 and 1600. $W^*_{ij,n} = 1$ iff the Euclidean distance between i and j is 1 and W_n is row-normalized from W^*_n.

Because $y_{i,n}(\epsilon_n|X_n, \theta)$ is an index function of θ, the criterion function $Q_n(\theta)$ is a simple function, a linear combination of some index functions. As a result, almost every θ is a local minimum of $Q_n(\theta)$ and it is not easy to reach the global minimum. In order to get $\hat{\theta}_n$ as close as to the global minimum, on the one hand, we adopt a large R such that $Q_n(\theta)$ is close to be a smooth function; on the other hand, we choose carefully initial values for minimization. By some experiments, we find that a probit model of $y_{i,n}$ on 1 and $x_{i,n}$ produces a not too bad initial guess of β_{20}. Thus, we use it as the initial value for β_2. And we choose the initial values of (λ, β_1) from $\{(0.1a, 0.2b - 1.2) : a = 1, 2, \cdots, 13, b = 1, 2, \cdots, 11\}$ to minimize $Q_n(\theta)$. To do so, it is not guaranteed that we can obtain the global minimums, but experiments show that the estimators produced in this way have satisfactory finite sample performance. Another reason for such choices of initial values is to balance computational time and precision of estimators: a larger set will produce more initial values but require more computational time.

From Table 1, we observe the following phenomena: (1) For smaller λ_0, the bias of $\hat{\lambda}_n$ is larger, since the sample distribution of $\hat{\lambda}_n$ is skewed as $\hat{\lambda}_n \geqslant 0$. (2) As sample sizes increases, the bias of $\hat{\lambda}_n$ decrease fast in most cases. (3) As sample sizes increase, the standard deviations decrease. Overall, for larger sample sizes, the performance is closer to the asymptotic results.

To check the robustness of our estimator, which is based on the best NE being selected, we design some experiments with misspecification of equilibrium selection rule. We select the minimum NE to generate the dependent variable, but we estimate the model as if the maximum NE was selected. The results are summarized in Table 2. When the sample size is 400, the performance under misspecification is clearly worse, but not very bad. For example, when $\lambda_0 = 0.6$, the biases for $\hat{\lambda}_n$ and $\hat{\beta}_{1,n}$ under misspecification are respectively 0.0934 and -0.0941, but they are 0.0666 and -0.0534 when the specification is correct. When sample sizes are large, the difference is smaller, but overall, the biases and RMSE under misspecification are slightly larger. The differences when $\lambda_0 = 0.6$ are slightly greater than that when $\lambda_0 = 0.3$. The reason is that with greater λ_0, the probability for ϵ_n to be located in areas where multiple equilibria occur is larger.

Table 1: Monte Carlo Simulation Results

	n		$\lambda = 0.6$	$\beta_1 = 0$	$\beta_2 = 1$	$\lambda = 0.3$	$\beta_1 = 0$	$\beta_2 = 1$
x_{1n}	400	bias	0.0666	-0.0534	-0.0036	0.1354	-0.0926	0.0014
		std	0.4724	0.2958	0.1072	0.4263	0.2465	0.1074
		RMSE	0.4771	0.3006	0.1072	0.4473	0.2633	0.1075
	900	bias	-0.0088	0.0060	0.0121	0.0221	-0.0144	0.0072
		std	0.3361	0.2115	0.0773	0.2918	0.1673	0.0712
		RMSE	0.3363	0.2116	0.0782	0.2926	0.1679	0.0716
	1600	bias	0.0179	-0.0057	0.0072	0.0029	0.0042	0.0036
		std	0.2525	0.1601	0.0492	0.2303	0.1303	0.0489
		RMSE	0.2532	0.1602	0.0497	0.2303	0.1304	0.0490
x_{2n}	400	bias	0.0781	-0.0560	0.0159	0.0980	-0.0653	0.0093
		std	0.4952	0.2962	0.1054	0.4056	0.2319	0.0994
		RMSE	0.5013	0.3014	0.1066	0.4173	0.2410	0.0998
	900	bias	0.0161	-0.0066	0.0173	0.0355	-0.0176	0.0161
		std	0.3435	0.2104	0.0712	0.2969	0.1692	0.0668
		RMSE	0.3439	0.2105	0.0733	0.2990	0.1701	0.0687
	1600	bias	0.0047	0.0021	0.0075	0.0032	0.0045	0.0067
		std	0.2382	0.1471	0.0477	0.2200	0.1252	0.0449
		RMSE	0.2382	0.1471	0.0483	0.2200	0.1253	0.0454
x_{3n}	400	bias	0.0598	-0.0422	0.0141	0.1083	-0.0690	0.0118
		std	0.4852	0.2920	0.0999	0.3992	0.2316	0.1008
		RMSE	0.4889	0.2951	0.1009	0.4136	0.2416	0.1015
	900	bias	0.0459	-0.0193	0.0211	0.0493	-0.0232	0.0153
		std	0.3291	0.1969	0.0698	0.3104	0.1753	0.0696
		RMSE	0.3323	0.1979	0.0730	0.3143	0.1768	0.0713
	1600	bias	0.0111	0.0014	0.0079	0.0063	0.0037	0.0052
		std	0.2408	0.1449	0.0472	0.2248	0.1265	0.0473
		RMSE	0.2411	0.1449	0.0479	0.2249	0.1265	0.0476

Repetition=500, Simulation=100, $\beta_{10} = 0$, $\beta_{20} = 1$. $x_{kn} \sim N(0, k)$.

Table 2: Monte Carlo Simulation Results—-Misspecified Equilibrium

	n		$\lambda = 0.6$	$\beta_1 = 0$	$\beta_2 = 1$	$\lambda = 0.3$	$\beta_1 = 0$	$\beta_2 = 1$
x_{1n}	400	bias	0.0934	-0.0941	0.0007	0.1439	-0.1044	0.0020
		std	0.5027	0.3097	0.1109	0.4410	0.2527	0.1075
		RMSE	0.5113	0.3237	0.1109	0.4638	0.2734	0.1075
	900	bias	0.0070	-0.0294	0.0126	0.0234	-0.0218	0.0072
		std	0.3488	0.2161	0.0770	0.2929	0.1684	0.0717
		RMSE	0.3499	0.2181	0.0780	0.2938	0.1698	0.0720
	1600	bias	0.0182	-0.0308	0.0067	0.0057	-0.0035	0.0036
		std	0.2575	0.1615	0.0497	0.2318	0.1323	0.0491
		RMSE	0.2582	0.1644	0.0502	0.2319	0.1324	0.0492
x_{2n}	400	bias	0.0742	-0.0729	0.0144	0.1117	-0.0789	0.0093
		std	0.4878	0.2914	0.1030	0.4283	0.2430	0.1008
		RMSE	0.4934	0.3003	0.1040	0.4427	0.2555	0.1012
	900	bias	0.0328	-0.0384	0.0174	0.0391	-0.0255	0.0158
		std	0.3492	0.2101	0.0702	0.2997	0.1695	0.0666
		RMSE	0.3508	0.2136	0.0723	0.3022	0.1714	0.0684
	1600	bias	0.0203	-0.0277	0.0077	0.0062	-0.0027	0.0066
		std	0.2368	0.1457	0.0476	0.2200	0.1245	0.0451
		RMSE	0.2377	0.1483	0.0482	0.2201	0.1245	0.0456
x_{3n}	400	bias	0.0586	-0.0601	0.0131	0.1136	-0.0760	0.0118
		std	0.4824	0.2881	0.1007	0.4093	0.2364	0.1007
		RMSE	0.4860	0.2943	0.1016	0.4248	0.2483	0.1013
	900	bias	0.0570	-0.0432	0.0214	0.0527	-0.0298	0.0150
		std	0.3348	0.1981	0.0703	0.3089	0.1741	0.0697
		RMSE	0.3396	0.2028	0.0735	0.3134	0.1767	0.0713
	1600	bias	0.0227	-0.0243	0.0082	0.0057	-0.0005	0.0053
		std	0.2425	0.1439	0.0478	0.2272	0.1271	0.0470
		RMSE	0.2436	0.1460	0.0485	0.2272	0.1271	0.0473

Repetition=500, Simulation=100, $\beta_{10} = 0$, $\beta_{20} = 1$. $x_{kn} \sim N(0, k)$.

5. Conclusion

We consider a complete information binary choice game on an exogenous network in this paper. We assume players select the maximum NE, which produces Pareto optimal equilibrium of the game and is also an SCPE, to overcome the issue of multiple NE in a sample. We propose the estimation of the model by the MSM. After investigating the NED property of $y_{i,n}$ and developing some empirical process theory, we establish the consistency and asymptotically normality of the MSM estimator. Simulation results indicate that when the sample has a moderate or large size, the estimator has satisfactory finite sample properties.

Our studies can be extended in several directions. (1) It is possible that players do not choose the maximum NE, but reach other NE. In such cases, the problem might be partially identified. Because all NE's are between the maximum NE and the minimum one, we may consider moment inequality estimation. (2) It is clear that our study can be generalized to the ordered probit or logistic models. (3) Furthermore, relatively more efficient estimation methods than the MSM might be of interest to be considered. (4) In our NED analysis, we take links between individuals as exogenously given, so we have not taken values of elements in the weights matrix as endogenous. By expanding the model to take into account of network formation, one may then take care of possible selectivity of outcomes. (5) Because the GMM criterion function is a simple function, almost every $\theta \in \Theta$ is a local minimum point, it requires more investigation how to computationally improve the search of the global minimum point. (6) We have shown that under the condition, $\bar{\delta}\bar{l}_p < 1$, we have NED, on which our analysis is based on. However, if this condition fails, we do not know whether NED still holds or not. As a result, we do not know the asymptotic properties of our estimator. Thus, one of the possible future researches is on related properties when $\bar{\delta}\bar{l}_p < 1$ does not hold.

Appendices

A. Stochastic Equicontinuity for NED Random Fields

A.1. A Moment Inequality for NED Random Fields

In this section, let $Z_n = \{Z_{i,n}\}_{i=1}^n$ and $\varepsilon_n = \{\varepsilon_{i,n}\}_{i=1}^n$ be two generic random fields.

Assumption A.1. Let $2 < p_0 \leqslant q_0 \in \mathbb{R}$ and $2 \leqslant w_0 \in \mathbb{N}$ satisfy $\frac{1}{p_0} + \frac{w_0-1}{q_0} = \frac{1}{2}$. $M \equiv \max(1, \sup_{i,n} ||Z_{i,n}||_{L^{q_0}}) < \infty$; $\mathrm{E}\, Z_{i,n} = 0$ for all i and n.

Assumption A.1 implies that $q_0 \geqslant 2w_0$. Because $p_0 \leqslant q_0$, $\iota \equiv \sup_{i,n} ||Z_{i,n}||_{L^{p_0}} \leqslant M < \infty$. Next, we will establish a covariance inequality for products of $Z_{i,n}$ and a moment inequality for $\sum_{i=1}^n Z_{i,n}$.[9]

Lemma A.1. *Suppose Assumption 1 holds. Let $\varepsilon_n = \{\varepsilon_{i,n}\}_{i=1}^n$ be an α-mixing random field with α-mixing coefficient $\alpha(\cdot,\cdot,\cdot)$. Z_n satisfies Assumption A.1 and is an L_2-NED random field on*

[9]Similar results are Corollary A.2 and Theorem A.1 in Xu and Lee (2015b) with stronger regularity conditions such as $Z_{i,n}$'s are uniformly bounded. But they are not enough to establish SEC. Lemmas A.1 and A.2 relax some of those restrictions and do not require that $Z_{i,n}$ is uniformly bounded. In this paper, we introduce ι to bound the inequalities in Lemmas A.1 and A.2. ι is related to the construction of brackets when we show SEC.

ε_n *such that* $||Z_{i,n} - \mathrm{E}[Z_{i,n}|\mathcal{F}_{i,n}(s)]||_{L^2} \leqslant \psi(s)$, *where* $\mathcal{F}_{i,n}(s) \equiv \sigma(\{\varepsilon_{j,n} : d_{ij} \leqslant s\})$. *Denote* $r = \min\{d(\vec{i}_m, \vec{j}_l) : 1 \leqslant m \leqslant u, 1 \leqslant l \leqslant v\} > 0$. $w \equiv u + v \leqslant w_0$. *Then, for any* $0 < s < r/2$,

$$|\mathrm{cov}(Z_{i_1,n}\cdots Z_{i_u,n}, Z_{j_1,n}\cdots Z_{j_v,n})| \leqslant 4M^{w-1}[\alpha^{1/p_0}(u,v,r-2s) + w\psi(s)^{\frac{q_0-2w+2}{2q_0-2w+2}}]\iota.$$

Proof of Lemma A.1: Let $\mathcal{F}_{u,n}(s) \equiv \sigma(\cup_{m=1}^u \mathcal{F}_{i_m,n}(s))$, $U = \mathrm{E}[\prod_{m=1}^u Z_{i_m,n}|\mathcal{F}_{u,n}(s)]$, $\Delta U = \prod_{m=1}^u Z_{i_m,n} - U$. Similarly, define $\mathcal{F}_{v,n}(s)$, V and ΔV for $Z_{j_1,n}\cdots Z_{j_v,n}$.

Let $t \equiv \frac{u-1}{w_0-1}q_0 \leqslant q_0$. Under Assumption A.1, by the generalized Hölder inequality and Lyapunov's inequality,

$$||Z_{i_1,n}\cdots Z_{i_u,n}||_{L^2} \leqslant ||Z_{i_1,n}||_{L^{p_0}} \prod_{m=2}^u ||Z_{i_m,n}||_{L^t} \leqslant \iota \prod_{m=2}^u ||Z_{i_m,n}||_{L^{q_0}} \leqslant M^{u-1}\iota. \tag{A.1}$$

By Jensen's inequality,

$$||U||_{L^2} = \mathrm{E}^{1/2}\left\{\mathrm{E}^2\left[\prod_{p=1}^u Z_{i_p,n}\middle|\mathcal{F}_{u,n}(s)\right]\right\} \leqslant \mathrm{E}^{1/2}\left\{\mathrm{E}\left[\prod_{p=1}^u Z_{i_p,n}^2\middle|\mathcal{F}_{u,n}(s)\right]\right\} = \left\|\prod_{p=1}^u Z_{i_p,n}\right\|_{L^2} \leqslant \iota M^{u-1}.$$

Next, we will evaluate $||\Delta U||_{L^2}$. Let $A \equiv \frac{q_0}{w-1} \geqslant \frac{q_0}{w_0-1} > 2$. For the following derivations, we use the convention that $\prod_{m=1}^0 = \prod_{m=u+1}^u = 1$.

$$\begin{aligned}
||\Delta U||_{L^2} &= \left\|\prod_{p=1}^u Z_{i_p,n} - \mathrm{E}\left[\prod_{p=1}^u Z_{i_p,n}\middle|\mathcal{F}_{u,n}(s)\right]\right\|_{L^2} \leqslant \left\|\prod_{p=1}^u Z_{i_p,n} - \prod_{p=1}^u \mathrm{E}[Z_{i_p,n}|\mathcal{F}_{i_p,n}(s)]\right\|_{L^2} \\
&\leqslant \left\|\sum_{k=1}^u \left|\left\{\prod_{m=1}^{k-1}\mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\right\}\left(\prod_{m=k+1}^u Z_{i_m,n}\right)\cdot\{Z_{i_k,n} - \mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\}\right|\right\|_{L^2} \\
&\leqslant \sum_{k=1}^u \left\|\left\{\prod_{m=1}^{k-1}\mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\right\}\left(\prod_{m=k+1}^u Z_{i_m,n}\right)\cdot\{Z_{i_k,n} - \mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\}\right\|_{L^2} \qquad \text{(A.2)} \\
&\leqslant 2\sum_{k=1}^u \left\|\left\{\prod_{m=1}^{k-1}\mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\right\}\left(\prod_{m=k+1}^u Z_{i_m,n}\right)\right\|_{L^2}^{\frac{A-2}{2A-2}} \cdot \|Z_{i_k,n} - \mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\|_{L^2}^{\frac{A-2}{2A-2}} \cdot \\
&\quad \left\|\left\{\prod_{m=1}^{k-1}\mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\right\}\left(\prod_{m=k+1}^u Z_{i_m,n}\right)\cdot\{Z_{i_k,n} - \mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\}\right\|_{L^A}^{\frac{A}{2A-2}},
\end{aligned}$$

where the first inequality is by Theorem 10.12 in Davidson (1994), the second one is by Lemma C.2 in Xu and Lee (2015b), the third one is by Minkowski's inequality, and the fourth one is by $||B\rho||_{L^2} \leqslant 2(||\rho||_{L^2}^{A-2}||B||_{L^2}^{A-2}||B\rho||_{L^A}^{A})^{1/(2A-2)}$ when $A > 2$ (Lemma 17.15 in Davidson, 1994). Similarly to Eq. (A.1), $\left\|\{\prod_{m=1}^{k-1} \mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\} \cdot (\prod_{m=k+1}^{u} Z_{i_m,n})\right\|_{L^2} \leqslant M^{u-1}$. Notice that

$$\begin{aligned}
&\left\|\left\{\prod_{m=1}^{k-1} \mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\right\}\left(\prod_{m=k+1}^{u} Z_{i_m,n}\right) \cdot \{Z_{i_k,n} - \mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\}\right\|_{L^A} \\
\leqslant& \prod_{m=1}^{k-1} \|\mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\|_{L^{Au}} \cdot \prod_{m=k+1}^{u} \|Z_{i_m,n}\|_{L^{Au}} \cdot \|Z_{i_k,n} - \mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\|_{L^{Au}} \\
\leqslant& \prod_{m=1}^{k-1} \|\mathrm{E}[Z_{i_m,n}|\mathcal{F}_{i_m,n}(s)]\|_{L^{q_0}} \cdot \prod_{m=k+1}^{u} \|Z_{i_m,n}\|_{L^{q_0}} \cdot \left(\|Z_{i_k,n}\|_{L^{q_0}} + \|\mathrm{E}[Z_{i_k,n}|\mathcal{F}_{i_k,n}(s)]\|_{L^{q_0}}\right) \\
\leqslant& 2M^u,
\end{aligned}$$

where the first inequality is by the generalized Hölder inequality, the second one is by Lyapunov's inequality as $Au \leqslant q_0$ and Minkowski's inequality, and the last one is by Jensen's inequality. Hence, Eq. (A.2) implies that

$$||\Delta U||_{L^2} \leqslant 2u(M^{u-1})^{\frac{A-2}{2A-2}} \cdot \psi(s)^{\frac{A-2}{2A-2}} \cdot (2M^u)^{\frac{A}{2A-2}} = 2^{\frac{3A-2}{2A-2}} u M^{\frac{2uA-2u-A+2}{2A-2}} \psi(s)^{\frac{A-2}{2A-2}} \leqslant 4uM^u \psi(s)^{\frac{A-2}{2A-2}},$$

where the last inequality is based on $M \geqslant 1$, $\frac{2uA-2u-A+2}{2A-2} < u$ and $\frac{3A-2}{2A-2} < 2$ (implied by $A > 2$). Similar conclusions hold for V, ΔV, and $V + \Delta V$. Because $A \geqslant \frac{q_0}{w_0-1}$, by Lyapunov's inequality and the generalized Hölder inequality, $||V||_{L^{q_0/(w_0-1)}} \leqslant ||V||_{L^A} \leqslant \prod_{m=1}^{v} ||Z_{j_m,n}||_{L^{Av}} \leqslant \prod_{m=1}^{v} ||Z_{j_m,n}||_{L^{q_0}} \leqslant M^v$. Consequently, by Lemma A.2 in Jenish and Prucha (2012),

$$\begin{aligned}
&|\mathrm{cov}(U,\, V)| \leqslant 4\alpha^{1/p_0}(u,v,r-2s)||U||_{L^2}||V||_{L^{q_0/(w_0-1)}} \\
\leqslant& 4\alpha^{1/p_0}(u,v,r-2s) \cdot \iota M^{u-1} \cdot M^v = 4\iota\alpha^{1/p_0}(u,v,r-2s)M^{w-1}.
\end{aligned}$$

By these inequalities and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&|\mathrm{cov}(Z_{i_1,n} \cdots Z_{i_u,n}, Z_{j_1,n} \cdots Z_{j_v,n})| = |\mathrm{cov}(U + \Delta U,\, V + \Delta V)| \\
\leqslant& |\mathrm{cov}(U,\, V)| + |\mathrm{cov}(U,\, \Delta V)| + |\mathrm{cov}(\Delta U,\, V + \Delta V)| \\
\leqslant& 4\iota\alpha^{1/p_0}(u,v,r-2s)M^{w-1} + ||U||_{L^2}||\Delta V||_{L^2} + ||\Delta U||_{L^2}||V + \Delta V||_{L^2} \\
\leqslant& 4\iota\alpha^{1/p_0}(u,v,r-2s)M^{w-1} + \iota M^{u-1} \cdot 4vM^v\psi(s)^{\frac{A-2}{2A-2}} + 4uM^u\psi(s)^{\frac{A-2}{2A-2}} \cdot \iota M^{v-1} \\
=& 4M^{w-1}[\alpha^{1/p_0}(u,v,r-2s) + w\psi(s)^{\frac{q_0-2w+2}{2q_0-2w+2}}]\iota.
\end{aligned}$$

□

With the covariance inequality in Lemma A.1, we are ready to establish a moment inequality for $\mathrm{E}\left(\sum_{i=1}^{n} z_{i,n}\right)^q$, which can be regarded as a spatial NED version of Lemma 3.1 in Andrews and Pollard (1994). The proof for Lemma A.2 relies heavily on the techniques in the proof of Theorem A.1 in Xu and Lee (2015b). From Lemma A.1 in Jenish and Prucha (2009), under Assumption 1, there exists a constant $C_d > 0$ such that $|\{\vec{j} \in \mathbb{R}^d : d(\vec{i}, \vec{j}) \leqslant r\}| \leqslant C_d(\lfloor r \rfloor + 1)^d$ and $|\{\vec{j} \in \mathbb{R}^d : r \leqslant d(\vec{i}, \vec{j}) \leqslant r + 1\}| \leqslant C_d(\lfloor r \rfloor + 1)^{d-1}$, where C_d is related to the topological structure of the space individuals are located.

Lemma A.2. *Let Assumption 1 hold.* $\varepsilon_n = \{\varepsilon_{i,n}\}_{i=1}^n$ *is an* α*-mixing random field with* α*-mixing coefficient* $\alpha(u, v, r) = (u+v)^\tau \exp(-a_\varepsilon r)$ *for some constants* $\tau \geqslant 0$ *and* $a_\varepsilon > 0$*.* $Z_n = \{Z_{i,n}\}_{i=1}^n$ *satisfies Assumption A.1 and it is an* L_2*-NED random field on* ε_n *such that* $\sup_{i,n} ||Z_{i,n} - \mathrm{E}(Z_{i,n}|\mathcal{F}_{i,n}(s))||_{L^2} \leqslant C_Z \exp(-a_Z s)$ *for two positive constants* C_Z *and* a_Z*. Then there are two constants,* $C_{\varepsilon Z0Md}$ *and* $C_{\tau\varepsilon Z0Md}$*, where* $C_{\varepsilon Z0Md}$ *depending on* a_ϵ*,* C_Z*,* p_0*,* q_0*,* M *and* d*, and* $C_{\tau\varepsilon Z0Md}$ *depending on* a_ε*,* a_Z*,* C_Z*,* τ*,* p_0*,* q_0*,* M *and* d*, such that for any* $w \in \mathbb{N} \cap [2, w_0]$*,*

$$\left|\mathrm{E}\left(\sum_{i=1}^{n} Z_{i,n}\right)^w\right| \leqslant \frac{(2w-2)!}{(w-1)!} C_{\varepsilon Z0Md}^w [d(w-1)]! \max\left\{(n\iota C_{\tau\varepsilon Z0Md})^{w/2}, n\iota C_{\tau\varepsilon Z0Md}\right\}.$$

Proof of Lemma A.2: Let $P_w \equiv \{\{i_1, i_2, \cdots, i_w\} \in \mathbb{N}^w : 1 \leqslant i_1 \leqslant i_2 \leqslant \cdots \leqslant i_w \leqslant n$, but they are not all equal$\}$. P_w is a collection of w natural numbers between 1 and n. For any $p_w = \{i_1, i_2, \cdots, i_w\} \in P_w$, by Lemma A.1 in Xu and Lee (2015b), we can partition its elements into two non-empty mutually exclusive subsets $I_1(p_w)$ and $I_2(p_w)$, such that $I_1(p_w) \cup I_2(p_w) = p_w$, $d[I_1(p_w), I_2(p_w)] = r > 0$, and both $\cup_{i \in I_1(p_w)} \overline{B(\vec{i}, r/2)}$ and $\cup_{i \in I_2(p_w)} \overline{B(\vec{i}, r/2)}$ are path-connected. These imply that the partition has the largest distance among other possible partitions; and, within each partition, for any individual, there is another one such that their distance is $\leqslant r$. By Eq. (A.1),

$|\operatorname{E} Z_{i,n}^w| \leqslant ||Z_{i,n}^w||_{L^2} \leqslant \iota M^{w-1}$. Then,

$$
\begin{aligned}
&A_w(n) \equiv \sum_{1\leqslant i_1 \leqslant \cdots i_w \leqslant n} |\operatorname{E} Z_{i_1,n} \cdots Z_{i_w,n}| \\
&\leqslant \sum_{1\leqslant j\leqslant n} |\operatorname{E} Z_{i_j,n}^w| + \sum_{p_w \in P_w} \left| \operatorname{E} \prod_{j\in I_1(p_w)} Z_{j,n} \cdot \operatorname{E} \prod_{j\in I_2(p_w)} Z_{j,n} + \operatorname{cov}(\prod_{j\in I_1(p_w)} Z_{j,n}, \prod_{j\in I_2(p_w)} Z_{j,n}) \right| \\
&\leqslant n\iota M^{w-1} + \sum_{m=1}^{w-1} A_m(n) A_{w-m}(n) + \sum_{p_w\in P_w} \left| \operatorname{cov}(\prod_{j\in I_1(p_w)} Z_{j,n}, \prod_{j\in I_2(p_w)} Z_{j,n}) \right|.
\end{aligned}
\tag{A.3}
$$

For any natural number $1 \leqslant i \leqslant n$, define $P_w(i) \equiv \{\{i_1, i_2, \cdots, i_w\} \in P_w : i_1 = i\}$ and $P_w(i, \lfloor r \rfloor) \equiv \{p_w \in P_w(i) : \lfloor r \rfloor \leqslant d(I_1(p_w), I_2(p_w)) < \lfloor r \rfloor + 1\}$, where $\lfloor r \rfloor \equiv \max\{a \in \mathbb{Z} : a \leqslant r\}$. Then $P_w = \cup_{i=1}^{n-1} P_w(i) = \cup_{i=1}^{n-1} \cup_{r=0}^{\infty} P_w(i, \lfloor r \rfloor)$ and

$$
\sum_{p_w\in P_w} \left| \operatorname{cov}(\prod_{j\in I_1(p_w)} Z_{j,n}, \prod_{j\in I_2(p_w)} Z_{j,n}) \right| \leqslant \sum_{i=1}^{n-1} \sum_{\lfloor r \rfloor=0}^{\infty} \sum_{p_w\in P_w(i,\lfloor r \rfloor)} \left| \operatorname{cov}(\prod_{j\in I_1(p_w)} Z_{j,n}, \prod_{j\in I_2(p_w)} Z_{j,n}) \right|.
\tag{A.4}
$$

By Lemma A.1, where s is chosen such that $\frac{a_\varepsilon (r-2s)}{p_0} = a_Z s \frac{q_0 - 2w_0 + 2}{2q_0 - 2w_0 + 2}$,

$$
\begin{aligned}
&\left| \operatorname{cov}(\prod_{j\in I_1(p_w)} Z_{j,n}, \prod_{j\in I_2(p_w)} Z_{j,n}) \right| \leqslant 4M^{w-1} \left\{ w^{\frac{\tau}{p_0}} \exp\left[-\frac{a_\varepsilon (r-2s)}{p_0} \right] + w \left(C_Z e^{-a_Z s} \right)^{\frac{q_0-2w+2}{2q_0-2w+2}} \right\} \iota \\
&\leqslant 4M^{w-1} \left\{ w^{\frac{\tau}{p_0}} \exp\left[-\frac{a_\varepsilon (r-2s)}{p_0} \right] + w \left(C_Z e^{-a_Z s} \right)^{\frac{q_0-2w_0+2}{2q_0-2w_0+2}} \right\} \iota \\
&= 4\iota M^{w-1} (w^{\frac{\tau}{p_0}} + w C_Z^{\frac{q_0-2w_0+2}{2q_0-2w_0+2}}) \exp(-C_{\varepsilon Z0} r) \leqslant 4\iota M^{w-1} C_{\tau Z0} e^w \exp(-C_{\varepsilon Z0} r),
\end{aligned}
\tag{A.5}
$$

where the constant $C_{\varepsilon Z0} > 0$ depends on a_ε, a_Z, q_0 and p_0 and $C_{\tau Z0} > 0$ depends on τ, C_Z, p_0 and q_0. Fixing the position i and considering $P_w(i, \lfloor r \rfloor) \neq \emptyset$, we can establish a sequence of closed balls with radius r so that each ball contains at least one another point in $\{1 \leqslant i_1 \leqslant \cdots i_w \leqslant n\}$. So all points in $\{1 \leqslant i_1 \leqslant \cdots i_w \leqslant n\}$ can be covered sequentially by $(w-1)$ balls. Thus, when i is fixed, $\sum_{p_w \in P_w(i,\lfloor r \rfloor)} 1 \leqslant \{C_d(\lfloor r \rfloor + 1)^d\}^{w-1}$. By Eq. (A.4) and (A.5),

$$\begin{aligned}
&\sum_{p_w \in P_w} \left| \operatorname{cov}(\prod_{j \in I_1(p_w)} Z_{j,n}, \prod_{j \in I_2(p_w)} Z_{j,n}) \right| \\
\leqslant &\sum_{i=1}^{n-1} \sum_{\lfloor r \rfloor = 0}^{\infty} \sum_{p_w \in P_w(i,\lfloor r \rfloor)} 4\iota M^{w-1} C_{\tau Z0} e^w \exp(-C_{\varepsilon Z0} \lfloor r \rfloor) \\
\leqslant &4n\iota C_d^{w-1} M^{w-1} C_{\tau Z0} e^w \sum_{\lfloor r \rfloor = 0}^{\infty} (\lfloor r \rfloor + 1)^{d(w-1)} \exp(-C_{\varepsilon Z0} \lfloor r \rfloor) \\
\leqslant &4n\iota C_d^{w-1} M^{w-1} C_{\tau Z0} e^w \sum_{k=0}^{\infty} \int_{k+1}^{k+2} x^{d(w-1)} \exp\left[-C_{\varepsilon Z0}(x-2)\right] dx \\
\leqslant &4n\iota (C_d^{-1} M^{-1} C_{\tau Z0} e^{2C_{\varepsilon Z0}} C_{\varepsilon Z0}^{d-1})(C_d e C_{\varepsilon Z0}^{-d} M)^w [d(w-1)]!
\end{aligned}$$

Let $C_{\varepsilon Z0Md} \equiv C_d e C_{\varepsilon Z0}^{-d} M$ and $\bar{C}_{\varepsilon Z0Md} \equiv \left\{\min_{2 \leqslant w \leqslant w_0} (C_d e C_{\varepsilon Z0}^{-d} M)^w [d(w-1)]!\right\}^{-1} > 0$. Then,

$$\begin{aligned}
&n\iota M^{w-1} + 4n\iota (C_d^{-1} M^{-1} C_{\tau Z0} e^{2C_{\varepsilon Z0}} C_{\varepsilon Z0}^{d-1})(C_d e C_{\varepsilon Z0}^{-d} M)^w [d(w-1)]! \\
\leqslant &n\iota (\bar{C}_{\varepsilon Z0Md} M^{-1} + 4C_d^{-1} M^{-1} C_{\tau Z0} e^{2C_{\varepsilon X0}} C_{\varepsilon Z0}^{d-1}) C_{\varepsilon Z0Md}^w [d(w-1)]! \\
\equiv &n\iota C_{\tau\varepsilon Z0Md} C_{\varepsilon Z0Md}^w [d(w-1)]! \equiv V_w(n),
\end{aligned}$$

where $C_{\tau\varepsilon Z0Md} \equiv \bar{C}_{\varepsilon Z0Md} M^{-1} + 4C_d^{-1} M^{-1} C_{\tau Z0} e^{2C_{\varepsilon Z0}} C_{\varepsilon Z0}^{d-1}$. Combining this result with Eq. (A.3), we have $A_w(n) \leqslant \sum_{m=1}^{w-1} A_m(n) A_{w-m}(n) + V_w(n)$. By the Bohr–Mollerup theorem (Olver, 2010, p.138), $\ln V_w(n)$ is convex in w. Consequently, $V_p \leqslant V_q^{(p-2)/(q-2)} V_2^{(q-p)/(q-2)}$. By "A Technical Lemma"[10] in Doukhan and Louhichi (1999, p. 336), Lemma 12[11] in Doukhan and Louhichi (1999)

[10] If for every integers $2 \leqslant p \leqslant q-1$, $V_p \leqslant V_q^{(p-2)/(q-2)} V_2^{(q-p)/(q-2)}$, then for every integers m and q fulfilling $2 \leqslant m \leqslant q-1$,

$$\max(V_2^{m/2}, V_m) \max(V_2^{(q-m)/2}, V_{q-m}) \leqslant \max(V_2^{q/2}, V_q). \tag{A.6}$$

[11] Let $(U_q)_{q>0}$ and $(V_q)_{q>0}$ be two sequences of real numbers satisfying for some $\gamma > 0$, and for all $q \in \mathbb{N}$, $U_q \leqslant \sum_{m=1}^{q-1} U_m U_{q-m} + e^{q\gamma} V_q$, with $U_1 = 0 \leqslant V_1$. Suppose $\{V_q\}$ satisfies Eq. (A.6). Then for any $2 \leqslant q \in \mathbb{N}$,

$$U_q \leqslant \frac{e^{q\gamma}}{q} \begin{pmatrix} 2q-2 \\ q-1 \end{pmatrix} \max(V_2^{q/2}, V_q).$$

is applicable to give a bound for $A_w(n)$:

$$\begin{aligned}
&\left|\mathrm{E}\left(\sum_{i=1}^{n} Z_{i,n}\right)^{w}\right| \leqslant w! A_w(n) \\
\leqslant & w! \frac{(2w-2)!}{[(w-1)!]^2} \frac{C_{\varepsilon Z0Md}^{w}}{w} \max\left\{(n\iota C_{\tau\varepsilon X0Md} d!)^{w/2}, n\iota C_{\tau\varepsilon X0Md}[d(w-1)]!\right\} \\
\leqslant & \frac{(2w-2)!}{(w-1)!} C_{\varepsilon Z0Md}^{w} [d(w-1)]! \max\left\{(n\iota C_{\tau\varepsilon Z0Md})^{w/2}, n\iota C_{\tau\varepsilon Z0Md}\right\}.
\end{aligned}$$

□

A.2. Stochastic Equicontinuity

SEC is useful for obtaining an empirical CLT and asymptotic distributions of estimators (Pakes and Pollard, 1989, Andrews, 1994, Andrews and Pollard, 1994). Although there is a vast literature on the weak convergence of independent random variables or stationary time series, e.g., van der Vaart and Wellner (1996) and Dehling and Philipp (2002), there exist much fewer studies for triangular sequences with both heterogeneity and serial correlation. Among them, Andrews and Pollard (1994) examine the SEC of the empirical process of an α-mixing triangular sequence. We extend their research to investigate NED random fields.

Let (Θ, ρ) be a totally bounded pseudometric space. $\{Z_{i,n}(\varepsilon_n, \theta)\}_{i=1}^{n}$ is a triangular array. It is worth noticing that the problem we are studying is somewhat different from most studies in empirical process. Conventionally, given a sequence of independent or mixing $\{\varepsilon_i\}$, the weak convergence of the empirical process of $\{Z_i(\varepsilon_i, \theta)\}$ is investigated. In spatial econometrics and social network, usually dependent variables rely on all individuals' characteristics. That is to say, $Z_{i,n}$ depends on $\varepsilon_{1,n}, \cdots, \varepsilon_{n,n}$. For $Z = \{\{Z_{i,n}\}_{i=1}^{n}\}_{n=1}^{\infty}$ and $Y = \{\{Y_{i,n}\}_{i=1}^{n}\}_{n=1}^{\infty}$, denote $\rho(Z, Y) = \sup_{i,n} ||Z_{i,n} - Y_{i,n}||_{L^2}$. For simplicity of notations, let $Z_{i,n}(\theta) \equiv Z_{i,n}(\varepsilon_n, \theta)$, $\rho(\theta) \equiv \rho(Z(\theta)) \equiv \sup_{i,n} ||Z_{i,n}(\theta)||_{L^2}$, and $\rho(\theta_1, \theta_2) \equiv \rho(Z(\theta_1), Z(\theta_2))$. The empirical process for $Z(\theta)$ is denoted as $\nu_n(\theta) \equiv \nu_n(Z(\theta)) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Z_{i,n}(\theta) - \mathrm{E}\, Z_{i,n}(\theta)]$.

Definition 2. For any $\epsilon > 0$, the L^p-bracketing number $N(\epsilon, L^p) \equiv N(\epsilon, \Theta, L^p)$ of $Z_n(\theta) = \{Z_{i,n}(\theta)\}_{i=1}^{n}$ is the smallest value of $N \in \mathbb{N}$ for which there exist two sets of functions $\{Y_{i,n}(\theta_j^{(\epsilon)})\}_{j=1}^{N}$ and $\{\Delta Y_{i,n}(\theta_j^{(\epsilon)})\}_{j=1}^{N}$ evaluated at N points θ_j^{ϵ} for $j = 1, \cdots, N$, so that (1) for any $\theta \in \Theta$, there exists a natural number $j \in [1, N]$ such that $|Z_{i,n}(\theta) - Y_{i,n}(\theta_j^{(\epsilon)})| \leqslant \Delta Y_{i,n}(\theta_j^{(\epsilon)})$ and (2)

$\rho(\Delta Y) \equiv \sup_{i,j,n} ||\Delta Y_{i,n}(\theta_j^{(\epsilon)})||_{L^p} \leqslant \epsilon$. If there is no such an n, then $N(\epsilon, L^p) = \infty$.

Denote $N(\epsilon) \equiv N(\epsilon, L^2)$. One more convention will be useful to simplify arguments. If for some $\theta \in \Theta$, there are $j \neq j'$ such that $|Z_{i,n}(\theta) - Y_{i,n}(\theta_j^{(\epsilon)})| \leqslant \Delta Y_{i,n}(\theta_j)$ and $|Z_{i,n}(\theta) - Y_{i,n}(\theta_{j'}^{(\epsilon)})| \leqslant \Delta Y_{i,n}(\theta_{j'})$, then we only consider $\min\{j, j'\}$. Then given $\epsilon > 0$, $\theta \to \theta_j^{(\epsilon)}$ is a map. Thus, we can omit the j in $Y_{i,n}(\theta_j^{(\epsilon)})$ and $\Delta Y_{i,n}(\theta_j^{(\epsilon)})$, and the dependence of $\theta^{(\epsilon)}$ on θ will be reflected by θ directly. For instance, the approximating function for $Z_{i,n}(\bar{\theta})$ is denoted as $Y_{i,n}(\bar{\theta}^{(\epsilon)})$. If for any $\epsilon > 0$, $N(\epsilon, L^p) < \infty$, then $\{Z_n(\theta), L^p\}$ is totally bounded. Thus, the total boundedness is a necessary condition for finite bracketing numbers. Lemma A.3 will be used repeatedly in our subsequent proofs.

Lemma A.3. *(Andrews and Pollard, 1994) For any $w \geqslant 1$ and arbitrary random variables $Z_1, \cdots, Z_n$, $\|\max_{1 \leqslant i \leqslant n} |Z_i|\|_{L^w} \leqslant n^{1/w} \max_{1 \leqslant i \leqslant n} \|Z_i\|_{L^w}$.*

Theorem A.1 presents our main theorem for SEC.

Theorem A.1. *Let Assumption 1 hold. $\varepsilon_n = \{\varepsilon_{i,n}\}_{i=1}^n$ is an α-mixing random field with α-mixing coefficient $\alpha(u, v, r) = (u + v)^\tau \exp(-a_\varepsilon r)$ for some constants $\tau \geqslant 0$ and $a_\varepsilon > 0$. (Θ, ρ) is a pseudometric space. p_0, q_0 and w_0 are defined in Assumption A.1 and denote $r_0 = \frac{p_0^{-1} - q_0^{-1}}{2^{-1} - q_0^{-1}} \in (0, 1)$. Suppose for any $\epsilon > 0$, $N(\epsilon)$, the L^2-bracketing number of $Z_n(\theta) = \{Z_{i,n}(\theta)\}_{i=1}^n$, does not depend on n. For some even number $w \in [2, w_0]$, $\int_0^1 x^{-1/2} N(x^{1/r_0})^{1/w} dx < \infty$. Furthermore, $Z_n(\theta)$, $\{Y_{i,n}(\theta^{(\epsilon)})\}$ and $\{\Delta Y_{i,n}(\theta^{(\epsilon)})\}$ are all uniformly and geometrically L_2-NED random fields on ε_n: $||U_{i,n} - \mathrm{E}[U_{i,n}|\mathcal{F}_{i,n}(s)]||_{L^2} \leqslant C_Z e^{-a_Z s}$, where $U_{i,n} = Z_n(\theta)$, $Y_{i,n}(\theta^{(\epsilon)})$ or $\Delta Y_{i,n}(\theta^{(\epsilon)})$, and neither C_Z nor a_Z depends on θ or ϵ. In addition, their L^{q_0} norms are $\leqslant M$ uniformly in i, n, θ and ϵ. Then, for any $\epsilon > 0$, there exists a $\delta > 0$, such that*

$$\limsup_{n \to \infty} \left\| \sup_{\rho(\theta_1, \theta_2) < \delta} |\nu_n(\theta_1) - \nu_n(\theta_2)| \right\|_{L^w} < \epsilon.$$

Proof of Theorem A.1: We follow the idea in Andrews and Pollard (1994) to establish the result. For any $k \in \mathbb{N}$, construct $N(2^{-k/r_0})$ approximating functions $\{Y_{i,n}(\theta^{(k)})\}$, where $Y_{i,n}(\theta^{(k)}) \equiv Y_{i,n}(\theta^{(2^{-k/r_0})})$, and corresponding $\{\Delta Y_{i,n}(\theta^{(k)})\}$ such that $|Z_{i,n}(\theta) - Y_{i,n}(\theta^{(k)})| \leqslant \Delta Y_{i,n}(\theta^{(k)})$ for all i, n and $\theta \in \Theta$, and $||\Delta Y_{i,n}(\theta^{(k)})||_{L^2} \leqslant 2^{-k/r_0}$. By Proposition 6.10 in Folland (1999),

$$||\Delta Y_{i,n}(\theta^{(k)})||_{L^{p_0}} \leqslant ||\Delta Y_{i,n}(\theta^{(k)})||_{L^2}^{r_0} ||\Delta Y_{i,n}(\theta^{(k)})||_{L^{q_0}}^{1-r_0} \leqslant 2^{-k} M^{1-r_0}. \tag{A.7}$$

We will prove this theorem in three steps. Let $\gamma \in (\frac{1}{2}, 1)$ be a universal constant and $k_n \equiv \max\{k \in \mathbb{N} : 2^k \leqslant n^\gamma\}$.

Step 1: $\limsup_{n\to\infty} \left\|\sup_{\theta\in\Theta} |\nu_n(\theta) - \nu_n(Y(\theta^{(k_n)}))|\right\|_{L^w} = 0$.

The arguments for this step are as follows. By the definitions of $\Delta Y_{i,n}(\theta^{(k)})$ and k_n,

$$\sup_{i,n,\theta} ||\Delta Y_{i,n}(\theta^{(k_n)})||_{L^{p_0}} \leqslant M^{1-r_0}2^{-k_n} = O(n^{-\gamma}) = o(n^{-1/2}).$$

We have $\nu_n(\Delta Y(\theta^{(k_n)})) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\Delta Y_{i,n}(\theta^{(k_n)}) - \mathrm{E}\, \Delta Y_{i,n}(\theta^{(k)})]$. Because

$$\begin{aligned}
&|\nu_n(\theta) - \nu_n(Y(\theta^{(k_n)}))| \leqslant \frac{1}{\sqrt{n}} \sum_{i=1}^n [\Delta Y_{i,n}(\theta^{(k_n)}) + \mathrm{E}\, \Delta Y_{i,n}(\theta^{(k_n)})] \\
=&\nu_n(\Delta Y(\theta^{(k_n)})) + \frac{2}{\sqrt{n}} \sum_{i=1}^n \mathrm{E}\, \Delta Y_{i,n}(\theta^{(k_n)}) \leqslant \nu_n(\Delta Y(\theta^{(k_n)})) + 2\sqrt{n} \cdot M^{1-r_0}2^{-k_n} \\
=&\nu_n(\Delta Y(\theta^{(k_n)})) + o(1),
\end{aligned}$$

$\left\|\sup_{\theta\in\Theta} |\nu_n(\theta) - \nu_n(Y(\theta^{(k_n)}))|\right\|_{L^w} \leqslant \left\|\sup_{\theta\in\Theta} \nu_n(\Delta Y(\theta^{(k_n)}))\right\|_{L^w} + o(1)$. Here and in the following, we use max to emphasize the finiteness (although $|\Theta| = \infty$, there are only $N(2^{-k_n/r_0})$ various $\theta^{(k_n)}$'s). Since the NED of $\Delta Y_{i,n}(\theta^{(k_n)})$ is uniformly in θ, by Eq. (A.7) and Lemma A.2, $\sup_{\theta\in\Theta} || \sum_{i=1}^n [\Delta Y(\theta^{(k_n)}) - \mathrm{E}\, \Delta Y(\theta^{(k_n)})]||_{L^w} \leqslant C_1 \max\{(n2^{-k_n})^{1/2}, (n2^{-k_n})^{1/w}\} = C_1(n2^{-k_n})^{1/2}$ for some constant $C_1 > 0$, where the equality is built on $n2^{-k_n} \geqslant n^{1-\gamma} \geqslant 1$. Hence, by Lemma A.3 and $\int_0^1 x^{-1/2} N(x^{1/r_0})^{1/w} dx < \infty$, as $n \to \infty$,

$$\begin{aligned}
&\left\|\sup_{\theta\in\Theta} \nu_n(\Delta Y(\theta^{(k_n)}))\right\|_{L^w} \leqslant N(2^{-k_n/r_0})^{1/w} \max_{\theta\in\Theta} \left\|\nu_n(\Delta Y(\theta^{(k_n)}))\right\|_{L^w} \\
\leqslant&N(2^{-k_n/r_0})^{1/w} C_1 2^{-k_n/2} = C_1 \int_0^{2^{-k_n}} N(2^{-k_n/r_0})^{1/w} 2^{k_n/2} dx \\
\leqslant&C_1 \int_0^{2^{-k_n}} N(x^{1/r_0})^{1/w} x^{-1/2} dx \to 0.
\end{aligned}$$

Step 2: Given any $\epsilon > 0$, there exists an $m \in \mathbb{N}$ that depends merely on ϵ but does not depend on n, such that $\left\|\max_{\theta\in\Theta} \min_{\phi^{(m)}:\phi\in\Theta} \left|\nu_n(Y(\theta^{(k_n)})) - \nu_n(Y(\phi^{(m)}))\right|\right\|_{L^w} \leqslant \frac{\epsilon}{6}$ for large enough n.

The arguments for this step are as follows. Because m is fixed, whose value will be determined later, eventually, $k_n > m$ as n goes to ∞, so we will only consider $k_n > m$ in the following. We use the same chaining technique as in Andrews and Pollard (1994) to gap the difference between k_n and m. Since $Y(\theta^{(k)}) = Y(\bar{\theta}^{(k)}) \nRightarrow Y(\theta^{(k-1)}) = Y(\bar{\theta}^{(k-1)})$, $Y(\theta^{(k)}) - Y(\bar{\theta}^{(k-1)})$ might have $N(2^{-k})N(2^{1-k})$ different values. In order to reduce the number of such possible differences, we define $\theta^{[k]}$ inductively from k_n to m: $\theta^{[k_n]} \equiv \theta^{(k_n)}$; for $k \leqslant k_n$, $\theta^{[k-1]} = \arg\min_{\phi^{(k-1)}:\phi\in\Theta} \rho(Y(\theta^{[k]}), Y(\phi^{(k-1)}))$. Given $m+1 \leqslant k \leqslant k_n$, $\theta^{[k]} = \bar{\theta}^{(k)}$. Then

$$\begin{aligned}&||Y_{i,n}(\theta^{[k]}) - Y_{i,n}(\theta^{[k-1]})||_{L^2} \leqslant ||Y_{i,n}(\bar{\theta}^{(k)}) - Y_{i,n}(\bar{\theta}^{(k-1)})||_{L^2}\\ \leqslant&||Y_{i,n}(\bar{\theta}^{(k)}) - Z_{i,n}(\bar{\theta})||_{L^2} + ||Z_{i,n}(\bar{\theta}) - Y_{i,n}(\bar{\theta}^{(k-1)})||_{L^2} \leqslant 2^{-k/r_0} + 2^{-(k-1)/r_0} < 2\cdot 2^{-(k-1)/r_0}.\end{aligned}$$

As a result, by Proposition 6.10 in Folland (1999) and Minkowski's inequality,

$$\begin{aligned}&\sup_{i,n} ||Y_{i,n}(\theta^{[k]}) - Y_{i,n}(\theta^{[k-1]})||_{L^{p_0}} \leqslant \sup_{i,n} ||Y_{i,n}(\theta^{[k]}) - Y_{i,n}(\theta^{[k-1]})||_{L^2}^{r_0} \cdot ||Y_{i,n}(\theta^{[k]}) - Y_{i,n}(\theta^{[k-1]})||_{L^{q_0}}^{1-r_0}\\ \leqslant&\sup_{i,n}\left[2\cdot 2^{-(k-1)/r_0}\right]^{r_0}\left[||Y_{i,n}(\theta^{[k]})||_{L^{q_0}} + ||Y_{i,n}(\theta^{[k-1]})||_{L^{q_0}}\right]^{1-r_0} \leqslant 2M^{1-r_0}2^{-(k-1)}.\end{aligned}$$

Apply Lemma A.2 to $\{Y_{i,n}(\theta^{[k]}) - Y_{i,n}(\theta^{[k-1]}) - \mathrm{E}\,Y_{i,n}(\theta^{[k]}) + \mathrm{E}\,Y_{i,n}(\theta^{[k-1]})\}_{i=1}^n$, which is uniformly and geometrically NED on ε_n. When $k \leqslant k_n$, because $n2^{1-k} \geqslant n2^{-k_n} \geqslant 1$, by Lemma A.2, there are constants $C_{\varepsilon Z0Md}$, $C_{\tau\varepsilon 0Md}$, $\bar{C}_2$ and $C_2 > 0$ that do not depend on θ, k or n, such that

$$\begin{aligned}&\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n[Y_{i,n}(\theta^{[k]}) - Y_{i,n}(\theta^{[k-1]}) - \mathrm{E}\,Y_{i,n}(\theta^{[k]}) + \mathrm{E}\,Y_{i,n}(\theta^{[k-1]})]\right\|_{L^w}\\ \leqslant&\frac{1}{\sqrt{n}}\left[\frac{(2w-1)!}{(w-1)!}[d(w-1)]!\right]^{1/w} C_{\varepsilon Z0Md}\max(C_{\tau\varepsilon 0Md}^{1/2}, C_{\tau\varepsilon 0Md}^{1/w})\cdot\\ &\max((n4M^{1-r_0}2^{-(k-1)})^{1/2}, (n4M^{1-r_0}2^{-(k-1)})^{1/w})\\ \leqslant&\bar{C}_2\frac{1}{\sqrt{n}}\max((4M^{1-r_0})^{1/2}, (4M^{1-r_0})^{1/w})\max((n2^{-(k-1)})^{1/2}, (n2^{-(k-1)})^{1/w}) = C_2 2^{-k/2}.\end{aligned}$$

Then by Lemma A.3, $\left\|\max_\theta\left|\nu_n(Y(\theta^{[k]})) - \nu_n(Y(\theta^{[k-1]}))\right|\right\|_{L^w} \leqslant N^{1/w}(2^{-k/r_0})C_2 2^{-k/2}$. Hence,

$$\begin{aligned}&\left\|\max_\theta\left|\nu_n(Y(\theta^{[k_n]})) - \nu_n(Y(\theta^{[m]}))\right|\right\|_{L^w} \leqslant \sum_{k=m+1}^{k_n}\left\|\max_\theta\left|\nu_n(Y(\theta^{[k]})) - \nu_n(Y(\theta^{([k-1]}))\right|\right\|_{L^w}\\ \leqslant&C_2\sum_{k=m+1}^{\infty} N^{1/w}(2^{-k/r_0})2^{-k/2} = 2C_2\sum_{k=m+1}^{\infty}\int_{2^{-k-1}}^{2^{-k}} N^{1/w}(2^{-k/r_0})2^{k/2}dx\\ \leqslant&2C_2\sum_{k=m+1}^{\infty}\int_{2^{-k-1}}^{2^{-k}} N^{1/w}(x^{1/r_0})x^{-1/2}dx = 2C_2\int_0^{2^{-m-1}} N^{1/w}(x^{1/r_0})x^{-1/2}dx \leqslant \frac{\epsilon}{6},\end{aligned}$$

where the last step holds if we choose $m = m(\epsilon)$ large enough. Step 2 is established.

Step 3: Define the following equivalence relation: θ and $\bar{\theta}$ are equivalent, denoted $\theta \sim \bar{\theta}$, if and only if $\theta^{[m]} = \bar{\theta}^{[m]}$. This equivalence relation partitions Θ into $N(2^{-m/r_0})$ equivalence classes, $\mathcal{E}[1], \cdots, \mathcal{E}[N(2^{-m/r_0})]$. Steps 1 and 2 imply $\left\|\sup_{\theta\in\Theta}\left|\nu_n(\theta) - \nu_n(Y(\theta^{[m]}))\right|\right\|_{L^w} < \frac{\epsilon}{5}$ for large enough n. Consequently,

$$\left\|\sup_{\theta\sim\bar{\theta}}\left|\nu_n(\theta) - \nu_n(\bar{\theta})\right|\right\|_{L^w} \leqslant \left\|\sup_{\theta\sim\bar{\theta}}\left[|\nu_n(\theta) - \nu_n(\theta^{[m]})| + |\nu_n(\bar{\theta}^{[m]}) - \nu_n(\bar{\theta})|\right]\right\|_{L^w}$$
$$\leqslant \left\|\sup_{\theta}|\nu_n(\theta) - \nu_n(\theta^{[m]})|\right\|_{L^w} + \left\|\sup_{\bar{\theta}}|\nu_n(\bar{\theta}) - \nu_n(\bar{\theta}^{[m]})|\right\|_{L^w} < \frac{2\epsilon}{5}.$$

Define the distance of two classes by $d(\mathcal{E}[i], \mathcal{E}[j]) \equiv \inf\{\rho(\theta_1, \theta_2) : \theta_1 \in \mathcal{E}[i], \theta_2 \in \mathcal{E}[j]\}$. For any $\delta > 0$, there exists $\theta_{ij} \in \mathcal{E}[i]$ and $\theta_{ji} \in \mathcal{E}[j]$ such that $\rho(\theta_{ij}, \theta_{ji}) < d(\mathcal{E}[i], \mathcal{E}[j]) + \delta$. If $\theta \in \mathcal{E}[i]$, $\bar{\theta} \in \mathcal{E}[j]$ and $\rho(\theta, \bar{\theta}) < \delta$, then $\rho(\theta_{ij}, \theta_{ji}) < 2\delta$ and

$$|\nu_n(\theta) - \nu_n(\bar{\theta})| \leqslant |\nu_n(\theta) - \nu_n(\theta_{ij})| + |\nu_n(\theta_{ij}) - \nu_n(\theta_{ji})| + |\nu_n(\theta_{ji}) - \nu_n(\bar{\theta})|$$
$$\leqslant 2\sup_{\phi\sim\bar{\phi}}|\nu_n(\phi) - \nu_n(\bar{\phi})| + \max\{|\nu_n(\phi_{ij}) - \nu_n(\phi_{ji})| : \rho(\phi_{ij}, \phi_{ji}) < 2\delta\}.$$

$\{Z_{k,n}(\phi_{ij}) - Z_{k,n}(\phi_{ji}) - \mathrm{E}\, Z_{k,n}(\phi_{ij}) + \mathrm{E}\, Z_{k,n}(\phi_{ji})\}_{k=1}^{n}$ is geometrically L_2-NED (NED coefficient $e^{-a_Z s}$) on ε_n with NED scaling factor $2C_Z$. Its L^{q_0} norm is bounded by $4M$. By Proposition 6.10 in Folland (1999), $||Z_{k,n}(\phi_{ij}) - Z_{k,n}(\phi_{ji}) - \mathrm{E}\, Z_{k,n}(\phi_{ij}) + \mathrm{E}\, Z_{k,n}(\phi_{ji})||_{L^{p_0}} \leqslant (4M)^{1-r_0}(2\delta)^{r_0}$. Apply Lemma A.2 to $\nu_n(\phi_{ij}) - \nu_n(\phi_{ji}) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}[Z_{k,n}(\phi_{ij}) - Z_{k,n}(\phi_{ji}) - \mathrm{E}\, Z_{k,n}(\phi_{ij}) + \mathrm{E}\, Z_{k,n}(\phi_{ji})]$, with ι there equal to $(4M)^{1-r_0}(2\delta)^{r_0}$. There are two constants $\bar{C}_3 > 0$ and $C_3 > 0$, such that when $n \geqslant (4M)^{r_0-1}(2\delta)^{-r_0}$,

$$||\nu_n(\phi_{ij}) - \nu_n(\phi_{ji})||_{L^w} \leqslant \frac{1}{\sqrt{n}}\bar{C}_3 \max\{[n(4M)^{1-r_0}(2\delta)^{r_0}]^{1/2}, [n(4M)^{1-r_0}(2\delta)^{r_0}]^{1/w}\} = C_3\delta^{r_0/2}.$$

Because m does not depend on n and there are less than $N^2(2^{-m/r_0})$ pairs of $|\nu_n(\phi_{ij}) - \nu_n(\phi_{ji})|$, by Lemma A.3, when $n \geqslant (4M)^{r_0-1}(2\delta)^{-r_0}$,

$$\left\|\nu_n(\theta) - \nu_n(\bar{\theta})\right\|_{L^w} \leqslant \frac{4\epsilon}{5} + N^{2/w}(2^{-m/r_0})\max\{||\nu_n(\phi_{ij}) - \nu_n(\phi_{ji})||_{L^w} : \rho(\phi_{ij}, \phi_{ji}) < 2\delta\}$$
$$\leqslant \frac{4\epsilon}{5} + N^{2/w}(2^{-m/r_0})C_3\delta^{r_0/2} < \epsilon,$$

provided that $\delta = \delta(\epsilon)$ is chosen to be smaller enough. □

B. Asymptotic Distribution via SEC

As can be seen from Eq. (6), the criterion function for optimization is not continuous. Therefore, the usual Taylor expansion technique for obtaining asymptotic distributions of estimators does not work. Similar phenomena often appear in simulated estimation for discrete choice models. Theorem 3.3 in Pakes and Pollard (1989) gives a general approach for such cases. We generalize it to allow heterogeneity, so that it is applicable for our estimator in the main text. The nonstochastic criterion function $\bar{G}_n(\theta)$ in Pakes and Pollard (1989) does not depending on n, which holds for i.i.d. and stationary data. In Theorem B.1, heterogeneity is reflected by the n in $\bar{G}_n(\theta)$. In addition, to control heterogeneity, condition (vi) is also needed, which is trivial for stationary data. The proof for Theorem B.1 is essentially minor modifications of that for Theorem 3.3 in Pakes and Pollard (1989). This theorem is also applicable to non-differentiable functions. In this section, $||\cdot||$ denotes not only the Euclidean norm for a vector, but also the Frobenius matrix norm: $||(a_{ij})|| \equiv (\sum_{i,j} a_{ij}^2)^{1/2}$. $||Ax|| \leqslant ||A|| \cdot ||x||$ for any vector x and every conformable matrix A.

Theorem B.1. *Let $G_n(\theta)$ be a stochastic function and $\bar{G}_n(\theta)$ be a non-stochastic one. $\bar{G}_n(\theta_0) = 0$ for all n and $\liminf_{n\to\infty} |\bar{G}_n(\theta)| > 0$ for all $\theta \neq \theta_0$. Suppose $\hat{\theta}_n$ is a consistent estimate of θ_0, i.e., $\hat{\theta}_n \xrightarrow{p} \theta_0$. If (i) $||G_n(\hat{\theta}_n)|| \leqslant o_p(n^{-1/2}) + \inf_\theta ||G_n(\theta)||$, (ii) $\Gamma_n \equiv \partial\bar{G}_n(\theta_0)/\partial\theta' \to \Gamma$ and Γ has full column rank, (iii) for any nonstochastic positive sequence $\delta_n \to 0$, $\sup_{||\theta-\theta_0||<\delta_n} ||G_n(\theta) - \bar{G}_n(\theta) - G_n(\theta_0)||/[n^{-1/2} + ||G_n(\theta)|| + ||\bar{G}_n(\theta)||] = o_p(1)$, (iv) $\sqrt{n}G_n(\theta_0) \xrightarrow{d} N(0, V)$, (v) $\theta_0 \in \Theta^0$, where Θ^0 denotes the interior of the parameter space $\Theta \in \mathbb{R}^K$, and (vi) for each k, $\partial^2\bar{G}_{k,n}(\theta)/\partial\theta\partial\theta'$ is uniformly bounded in both n and θ, where $\bar{G}_{k,n}(\theta)$ is the k^{th} entry of $\bar{G}_n(\theta)$, then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (\Gamma'\Gamma)^{-1}\Gamma' V \Gamma(\Gamma'\Gamma)^{-1})$.*

Proof of Theorem B.1: (1) We first show the $\sqrt{n}$-consistency. $\hat{\theta}_n \xrightarrow{p} \theta_0$ implies that we can choose δ_n slow enough such that $\Pr\{||\hat{\theta}_n - \theta_0|| \geqslant \delta_n\} \to 0$. Then, by the triangle inequality and condition (iii), $||\bar{G}_n(\hat{\theta}_n)|| - ||G_n(\hat{\theta}_n)|| - ||G_n(\theta_0)|| \leqslant ||G_n(\hat{\theta}_n) - \bar{G}_n(\hat{\theta}_n) - G_n(\theta_0)|| \leqslant o_p(\frac{1}{\sqrt{n}}) + o_p(||G_n(\hat{\theta}_n)||) + o_p(||\bar{G}_n(\hat{\theta}_n)||)$. Thus, $||\bar{G}_n(\hat{\theta}_n)||[1 - o_p(1)] \leqslant o_p(\frac{1}{\sqrt{n}}) + ||G_n(\hat{\theta}_n)||[1 + o_p(1)] + ||G_n(\theta_0)||$. By conditions (i) and (iv), $||G_n(\theta_0)|| = O_p(\frac{1}{\sqrt{n}})$ and $||G_n(\hat{\theta}_n)|| \leqslant o_p(\frac{1}{\sqrt{n}}) + \inf_\theta ||G_n(\theta)|| \leqslant o_p(\frac{1}{\sqrt{n}}) + ||G_n(\theta_0)|| = O_p(\frac{1}{\sqrt{n}})$. It follows that $||\bar{G}_n(\hat{\theta}_n)|| = O_p(\frac{1}{\sqrt{n}})$. Conditions (ii) and (vi) imply that $||\bar{G}_n(\hat{\theta}_n)|| \geqslant C||\hat{\theta}_n - \theta_0||$ for some constant $C > 0$ with large probability as $\hat{\theta}_n$ is near enough to θ_0 when n is large enough. Hence, $\sqrt{n}||\hat{\theta}_n - \theta_0|| = O_p(1)$.

(2) Next, we establish the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Define $L_n(\theta) \equiv \Gamma_n(\theta - \theta_0) + G_n(\theta_0)$. By Taylor's theorem with the Lagrange form of the remainder and conditions (ii) and (vi), $|\bar{G}_{k,n}(\hat{\theta}_n) - \Gamma_{k,n}(\hat{\theta}_n - \theta_0)| = |(\hat{\theta}_n - \theta_0)' \frac{\partial^2 \bar{G}_{k,n}(\bar{\theta}_n)}{\partial\theta\partial\theta}(\hat{\theta}_n - \theta_0)| = O_p(||\hat{\theta}_n - \theta_0||^2)$, where $\bar{\theta}_n$ is a convex combination of θ_0 and $\hat{\theta}_n$. Then

$$\begin{aligned} &||G_n(\hat{\theta}_n) - L_n(\hat{\theta}_n)|| \leqslant ||G_n(\hat{\theta}_n) - \bar{G}_n(\hat{\theta}_n) - G_n(\theta_0)|| + ||\bar{G}_n(\hat{\theta}_n) - \Gamma_n(\hat{\theta}_n - \theta_0)|| \\ \leqslant & o_p(\frac{1}{\sqrt{n}}) + o_p(||G_n(\hat{\theta}_n)||) + o_p(||\bar{G}_n(\hat{\theta}_n)||) + o_p(||\hat{\theta}_n - \theta_0||) = o_p(\frac{1}{\sqrt{n}}). \end{aligned} \tag{B.1}$$

Define $\theta_n^* \equiv \arg\min_{\theta\in\Theta} ||L_n(\theta)|| = \theta_0 - (\Gamma_n'\Gamma_n)^{-1}\Gamma_n' G_n(\theta_0)$. Then by conditions (ii) and (iv), $\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, (\Gamma'\Gamma)^{-1}\Gamma' V \Gamma(\Gamma'\Gamma)^{-1})$. Thus, it suffices to show $\hat{\theta}_n = \theta_n^* + o_p(\frac{1}{\sqrt{n}})$.

Because $\theta_n^* = \theta_0 + O_p(n^{-1/2})$, condition (v) implies that $\Pr(\theta_n^* \in \Theta^0) \to 1$. By $\bar{G}_n(\theta_0) = 0$ and condition (vi), $||\bar{G}_n(\theta_n^*)|| \leqslant ||\Gamma_n(\theta_n^* - \theta_0)|| + o_p(||\theta_n^* - \theta_0||) = O_p(\frac{1}{\sqrt{n}})$. By condition (iii), $||G_n(\theta_n^*)|| - ||\bar{G}_n(\theta_n^*)|| - ||G_n(\theta_0)|| \leqslant o_p(\frac{1}{\sqrt{n}}) + o_p(||G_n(\theta_n^*)|| + ||\bar{G}_n(\theta_n^*)||)$. Thus, by condition (iv) and the previous two results, $||G_n(\theta_n^*)|| = O_p(\frac{1}{\sqrt{n}})$. Hence,

$$\begin{aligned} &||G_n(\theta_n^*) - L_n(\theta_n^*)|| \leqslant ||G_n(\theta_n^*) - \bar{G}_n(\theta_n^*) - G_n(\theta_0)|| + ||\bar{G}_n(\theta_n^*) - \Gamma_n(\theta_n^* - \theta_0)|| \\ \leqslant & o_p(\frac{1}{\sqrt{n}}) + o_p(||G_n(\theta_n^*)||) + o_p(||\bar{G}_n(\theta_n^*)||) + o_p(||\theta_n^* - \theta_0||) = o_p(\frac{1}{\sqrt{n}}). \end{aligned} \tag{B.2}$$

By Eq. (B.1) and (B.2), $||L_n(\hat{\theta}_n)|| \leqslant ||G_n(\hat{\theta}_n)|| + o_p(\frac{1}{\sqrt{n}}) \leqslant ||G_n(\theta_n^*)|| + o_p(\frac{1}{\sqrt{n}}) \leqslant ||L_n(\theta_n^*)|| + o_p(\frac{1}{\sqrt{n}})$. Squaring both sides, we have $||L_n(\hat{\theta}_n)||^2 = ||L_n(\theta_n^*)||^2 + o_p(n^{-1})$. Because $\theta_n^* \equiv \arg\min_{\theta\in\Theta} ||L_n(\theta)||$, by the Pythagorean Theorem, $||L_n(\hat{\theta}_n)||^2 = ||L_n(\theta_n^*)||^2 + ||L_n(\theta_n^*) - L_n(\hat{\theta}_n)||^2 = ||L_n(\theta_n^*)||^2 + ||\Gamma_n(\theta_n^* - \hat{\theta}_n)||^2$. Therefore, by condition (ii), $||\hat{\theta}_n - \theta_n^*|| = o_p(\frac{1}{\sqrt{n}})$. □

Lemma B.1. *Let $\{A_n(\theta) : \theta \in \Theta\}$ be a family of sequences of nonsingular, random matrices. There exists a nonsingular, nonrandom matrix A such that $\sup_{\theta:||\theta-\theta_0||<\delta_n} ||A_n(\theta) - A|| = o_p(1)$, whenever $\{\delta_n\}$ is a sequence of positive numbers that converges to zero. If conditions (ii), (iii), (iv) and (vi) of Theorem B.1 are satisfied by $G_n(\theta)$ and $\bar{G}_n(\theta)$, then they also hold if $G_n(\theta)$ is replaced by $A_n(\theta)G_n(\theta)$ and $\bar{G}_n(\theta)$ is replaced by $A\bar{G}_n(\theta)$.*

Proof of Lemma B.1: Conditions (ii), (iv) and (vi) are obvious. The proof for condition (iii) is almost the same as that for Lemma 3.5 in Pakes and Pollard (1989), thus it is omitted. □

The following theorem extends the preceding Theorem B.1 to the case where the true parameter vector is on the boundary of a parameter space, which has not been covered in Pakes and Pollard (1989).

Theorem B.2. *Let the parameter space be* $\Theta = [0, B_1] \times \Theta_{-1}$ *and* $\theta_0 = (\theta_{10}, \theta_{-1,0})$ *satisfies* $\theta_{10} = 0$ *and* $\theta_{-1,0} \in \Theta^0_{-1}$*, where* Θ^0_{-1} *denotes the interior of* $\Theta_{-1} \subseteq \mathbb{R}^{K-1}$*. Let* $G_n(\theta)$ *be a stochastic function and* $\bar{G}_n(\theta)$ *be a non-stochastic one.* $\bar{G}_n(\theta_0) = 0$ *for all* n *and* $\liminf_{n\to\infty} |\bar{G}_n(\theta)| > 0$ *for all* $\theta \neq \theta_0$*. Suppose* $\hat{\theta}_n$ *with* $\hat{\theta}_{1,n} \geqslant 0$ *is a consistent estimate of* θ_0*, i.e.,* $\hat{\theta}_n \xrightarrow{p} \theta_0$*.*

If (i) $||G_n(\hat{\theta}_n)|| \leqslant o_p(n^{-1/2}) + \inf_{\theta\in\Theta} ||G_n(\theta)||$*, (ii)* $\Gamma_n \equiv \partial\bar{G}_n(\theta_0)/\partial\theta' \to \Gamma$ *and* Γ *has full column rank, (iii) for any nonstochastic positive sequence* $\delta_n \to 0$*,* $\sup_{||\theta-\theta_0||<\delta_n} ||G_n(\theta) - \bar{G}_n(\theta) - G_n(\theta_0)||/[n^{-1/2} + ||G_n(\theta)|| + ||\bar{G}_n(\theta)||] = o_p(1)$*, (iv)* $\sqrt{n}G_n(\theta_0) \xrightarrow{d} N(0, V)$*, and (v) for each* k*,* $\partial^2\bar{G}_{k,n}(\theta)/\partial\theta\partial\theta'$ *is uniformly bounded in both* n *and* θ*, where* $\bar{G}_{k,n}(\theta)$ *is the* k^{th} *entry of* $\bar{G}_n(\theta)$*, then the limiting distribution of* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is as follows: with* $\frac{1}{2}$ *probability, it has a regular truncated normal density* $N(0, (\Gamma'\Gamma)^{-1}\Gamma'V\Gamma(\Gamma'\Gamma)^{-1})$ *with only the first coordinate truncated* $(\theta_1 > 0)$*; with another* $\frac{1}{2}$ *probability, its distribution is degenerated for the first component to be* $\theta_1 = 0$*, and the remaining component is a* $(K-1)$*-dimensional normal density* $N(0, (\Gamma'_{-1}\Gamma_{-1})^{-1}\Gamma'_{-1}V\Gamma_{-1}(\Gamma'_{-1}\Gamma_{-1})^{-1})$*, where* Γ_{-1} *is the submatrix of* Γ *without the first column.*

Proof of Theorem B.2: (1) The $\sqrt{n}$-consistency still holds because it does not rely on Condition (v) in Theorem B.1, as can be seen from the proof of Theorem B.1.

(2) Next, we establish the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. By the same argument as in Theorem B.1, $||G_n(\hat{\theta}_n) - L_n(\hat{\theta}_n)|| = o_p(n^{-1/2})$. Define $L_n(\theta) \equiv \Gamma_n(\theta - \theta_0) + G_n(\theta_0)$ and $\theta^*_n \equiv \arg\min_{\theta\in\mathbb{R}^K:\theta_1\geqslant 0} ||L_n(\theta)||$. If we first minimize $||L_n(\theta)||$ without the constraint $\theta_1 \geqslant 0$, then $\theta^*_n = \theta_0 - (\Gamma'_n\Gamma_n)^{-1}\Gamma'_nG_n(\theta_0)$. If $[\theta_0 - (\Gamma'_n\Gamma_n)^{-1}\Gamma'_nG_n(\theta_0)]_1 \leqslant 0$, then $\theta^*_{1,n} = 0$ with the the constraint $\theta_1 \geqslant 0$. Then, $\theta^*_{-1,n} = \theta_{-1,0} - (\Gamma'_{-1,n}\Gamma_{-1,n})^{-1}\Gamma'_{-1,n}G_{-1,n}(\theta_0)$. As as result, under conditions (ii) and (iv), the limiting distribution of $\sqrt{n}(\theta^*_n - \theta_0)$ is the one described in the theorem. Thus, it suffices to show $\hat{\theta}_n = \theta^*_n + o_p(n^{-1/2})$.

Because $\theta^*_n = \theta_0 + O_p(n^{-1/2})$, $\theta_{-1,0} \in \Theta^0_{-1}$ implies that $\Pr(\theta^*_{-1,n} \in \Theta^0_{-1}) \to 1$. By the same argument as in Theorem B.1, $||L_n(\hat{\theta}_n)||^2 = ||L_n(\theta^*_n)||^2 + o_p(n^{-1})$. Because $\theta^*_n \equiv \arg\min_{\theta\in\Theta} ||L_n(\theta)||$, when $\theta^*_{1,n} > 0$, by the same argument as in Theorem B.1, $||L_n(\hat{\theta}_n)||^2 = ||L_n(\theta^*_n)||^2 + ||\Gamma_n(\theta^*_n - \hat{\theta}_n)||^2$.

When $\theta^*_{1,n} = 0$,

$$||L_n(\hat{\theta}_n)||^2 = ||L_n(\theta^*_n)||^2 + ||L_n(\theta^*_n) - L_n(\hat{\theta}_n)||^2 + 2L_n(\theta^*_n)'[L_n(\hat{\theta}_n) - L_n(\theta^*_n)]$$
$$=||L_n(\theta^*_n)||^2 + ||\Gamma_n(\theta^*_n - \hat{\theta}_n)||^2 + 2L_n(\theta^*_n)'\Gamma_n(\hat{\theta}_n - \theta^*_n) \geqslant ||L_n(\theta^*_n)||^2 + ||\Gamma_n(\theta^*_n - \hat{\theta}_n)||^2,$$

where the last inequality is by $\theta^*_n \equiv \arg\min_{\theta\in\Theta} ||L_n(\theta)||$ subject to the non-negative constraint. In both cases, $||\Gamma_n(\theta^*_n - \hat{\theta}_n)||^2 \leqslant ||L_n(\hat{\theta}_n)||^2 - ||L_n(\theta^*_n)||^2 = o_p(n^{-1})$. Therefore, by condition (ii), $||\hat{\theta}_n - \theta^*_n|| = o_p(n^{-1/2})$. □

C. Probabilities of the Best NE and Its First and Second Order Derivatives

C.1. An Example with 3 Individuals

To motivate the materials in the following two subsections, we consider a game with 3 players first to obtain some intuitions: $y_1 = 1(\lambda_1 \frac{y_2+y_3}{2} + x_1 + \epsilon_1 > 0)$, $y_2 = 1(\lambda_2 \frac{y_1+y_3}{2} + x_2 + \epsilon_2 > 0)$ and $y_3 = 1(\lambda_3 \frac{y_1+y_2}{2} + x_3 + \epsilon_3 > 0)$. The equilibrium results according to the values of ϵ_2 are shown in Table 3. From Table 3, we have

$$\Pr(y_1 = 1|\epsilon_2 > -x_2) = [1 - F(-x_1 - \lambda_1)][1 - F(-x_3 - \lambda_3)] + [1 - F(-x_1 - \frac{\lambda_1}{2})]F(-x_3 - \lambda_3),$$

$$\Pr(y_1 = 1|-x_2-\frac{\lambda_2}{2} < \epsilon_2 < -x_2) = [1-F(-x_1-\lambda_1)][1-F(-x_3-\lambda_3)]+[1-F(-x_1-\frac{\lambda_1}{2})]F(-x_3-\lambda_3),$$

$$\Pr(y_1 = 1|-x_2-\lambda_2 < \epsilon_2 < -x_2-\frac{\lambda_2}{2}) = [1-F(-x_1-\lambda_1)][1-F(-x_3-\lambda_3)]+[1-F(-x_1)]F(-x_3-\lambda_3),$$

$$\Pr(y_1 = 1|-x_2-\lambda_2 < \epsilon_2 < -x_2-\frac{\lambda_2}{2}) = [1-F(-x_1)]+[F(-x_1)-F(-x_1-\frac{\lambda_1}{2})][1-F(-x_3-\frac{\lambda_3}{2})],$$

where $F(\cdot)$ is the CDF of ϵ_i, which is i.i.d. for $i = 1, 2, 3$. These probabilities do not depend on x_2 or λ_2, except its ranges, because when we calculate these probabilities on the plane of (ϵ_1, ϵ_3), the cutting lines, e.g., $\epsilon_1 = -x_1$ and $\epsilon_3 = -x_3 - \frac{\lambda_3}{2}$, do not depend on x_2 or λ_2.

Table 3: The Maximum NE (y_1, y_2, y_3)

	$\epsilon_2 > -x_2$			
	$\epsilon_1 < -x_1 - \lambda_1$	$-x_1 - \lambda_1 < \epsilon_1 < -x_1 - \frac{\lambda_1}{2}$	$-x_1 - \frac{\lambda_1}{2} < \epsilon_1 < -x_1$	$\epsilon_1 > -x_1$
$\epsilon_3 > -x_3$	(0,1,1)	(1,1,1)	(1,1,1)	(1,1,1)
$-x_3 - \frac{\lambda_3}{2} < \epsilon_3 < -x_3$	(0,1,1)	(1,1,1)	(1,1,1)	(1,1,1)
$-x_3 - \lambda_3 < \epsilon_3 < -x_3 - \frac{\lambda_3}{2}$	(0,1,0)	(1,1,1)	(1,1,1)	(1,1,1)
$\epsilon_3 < -x_3 - \lambda_3$	(0,1,0)	(0,1,0)	(1,1,0)	(1,1,0)
	$-x_2 - \frac{\lambda_2}{2} < \epsilon_2 < -x_2$			
	$\epsilon_1 < -x_1 - \lambda_1$	$-x_1 - \lambda_1 < \epsilon_1 < -x_1 - \frac{\lambda_1}{2}$	$-x_1 - \frac{\lambda_1}{2} < \epsilon_1 < -x_1$	$\epsilon_1 > -x_1$
$\epsilon_3 > -x_3$	(0,1,1)	(1,1,1)	(1,1,1)	(1,1,1)
$-x_3 - \frac{\lambda_3}{2} < \epsilon_3 < -x_3$	(0,1,1)	(1,1,1)	(1,1,1)	(1,1,1)
$-x_3 - \lambda_3 < \epsilon_3 < -x_3 - \frac{\lambda_3}{2}$	(0,0,0)	(1,1,1)	(1,1,1)	(1,1,1)
$\epsilon_3 < -x_3 - \lambda_3$	(0,0,0)	(0,0,0)	(1,1,0)	(1,1,0)
	$-x_2 - \lambda_2 < \epsilon_2 < -x_2 - \frac{\lambda_2}{2}$			
	$\epsilon_1 < -x_1 - \lambda_1$	$-x_1 - \lambda_1 < \epsilon_1 < -x_1 - \frac{\lambda_1}{2}$	$-x_1 - \frac{\lambda_1}{2} < \epsilon_1 < -x_1$	$\epsilon_1 > -x_1$
$\epsilon_3 > -x_3$	(0,0,1)	(1,1,1)	(1,1,1)	(1,1,1)
$-x_3 - \frac{\lambda_3}{2} < \epsilon_3 < -x_3$	(0,0,0)	(1,1,1)	(1,1,1)	(1,1,1)
$-x_3 - \lambda_3 < \epsilon_3 < -x_3 - \frac{\lambda_3}{2}$	(0,0,0)	(1,1,1)	(1,1,1)	(1,1,1)
$\epsilon_3 < -x_3 - \lambda_3$	(0,0,0)	(0,0,0)	(0,0,0)	(1,0,0)
	$\epsilon_2 < -x_2 - \lambda_2$			
	$\epsilon_1 < -x_1 - \lambda_1$	$-x_1 - \lambda_1 < \epsilon_1 < -x_1 - \frac{\lambda_1}{2}$	$-x_1 - \frac{\lambda_1}{2} < \epsilon_1 < -x_1$	$\epsilon_1 > -x_1$
$\epsilon_3 > -x_3$	(0,0,1)	(0,0,1)	(1,0,1)	(1,0,1)
$-x_3 - \frac{\lambda_3}{2} < \epsilon_3 < -x_3$	(0,0,0)	(0,0,0)	(1,0,1)	(1,0,1)
$-x_3 - \lambda_3 < \epsilon_3 < -x_3 - \frac{\lambda_3}{2}$	(0,0,0)	(0,0,0)	(0,0,0)	(1,0,0)
$\epsilon_3 < -x_3 - \lambda_3$	(0,0,0)	(0,0,0)	(0,0,0)	(1,0,0)

C.2. First Order Derivatives

Lemma C.1. *Suppose* $\sum_{i=1}^{K} a_i = 0$, $\sup_{1 \leqslant i \leqslant K} |a_i| \leqslant M$, *and* $\max_{1 \leqslant i < j \leqslant K} |b_i - b_j| < \epsilon$. *Then* $|\sum_{i=1}^{K} a_i b_i| \leqslant (K-1)M\epsilon$.

Proof of Lemma C.1: $\left|\sum_{i=1}^{K} a_i b_i\right| = \left|\sum_{i=1}^{K} a_i b_1 + \sum_{i=1}^{K} a_i(b_i - b_1)\right| = \left|\sum_{i=2}^{K} a_i(b_i - b_1)\right| \leqslant \sum_{i=2}^{K} |a_i| \cdot |b_i - b_1| \leqslant \epsilon \sum_{i=2}^{K} |a_i| = (K-1)M\epsilon$. □

Recall $B_f = \sup_x f(x)$, $B_W = \sup_n ||W_n||_\infty$, $B_X = \sup_{i,k,n} ||x_{ik,n}||_{L^2}$ and $B_Q = \sup_{i,k,n} ||q_{ik,n}||_{L^{q_0}}$. Notice that $x_{i,n}$ is part of $q_{i,n}$. Recall $|\{\vec{j} \in \mathbb{R}^d : d(\vec{i}, \vec{j}) \leqslant r\}| \leqslant C_d(\lfloor r \rfloor + 1)^d$ and $|\{\vec{j} \in \mathbb{R}^d : r \leqslant d(\vec{i}, \vec{j}) \leqslant r+1\}| \leqslant C_d(\lfloor r \rfloor + 1)^{d-1}$ for any $r \geqslant 0$ and for a constant $C_d > 0$. These relationships imply that $|\{\vec{j} \in \mathbb{R}^d : d(\vec{i}, \vec{j}) \leqslant m\bar{d}_0\}| \leqslant C_{d\bar{d}_0} m^{d-1}$ and $|\{\vec{j} \in \mathbb{R}^d : m\bar{d}_0 \leqslant d(\vec{i}, \vec{j}) \leqslant (m+1)\bar{d}_0\}| \leqslant C_{d\bar{d}_0} m^{d-1}$ for some constant $C_{d\bar{d}_0}$ and for all $m \in \mathbb{N}$.

Lemma C.2. $\frac{\partial \Pr(y_{i,n}=1|X_n,\theta)}{\partial \lambda} \leqslant B_f B_W C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}} < \infty$ *and* $\frac{\partial \Pr(y_{i,n}=1|\theta)}{\partial \lambda} \leqslant B_f B_W C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$ *for some constant* $C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$ depending only on d, $\bar{l}_p$, $\bar{\delta}$, $\bar{d}_0$ and m_0. $\left|\frac{\partial \operatorname{E}(y_{i,n} q_{il,n}|\theta)}{\partial \lambda}\right| \leqslant B_Q B_f B_W C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$.

$|\partial \Pr(y_{i,n}=1|\theta)/\partial \beta_j| \leqslant B_f B_X C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$ and $\left|\frac{\partial \operatorname{E}(y_{i,n} q_{il,n}|\theta)}{\partial \beta_j}\right| \leqslant B_f B_Q B_X C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$.

Proof of Lemma C.2: Consider the system $y_{i,n} = 1(\lambda_i \sum_{j=1}^n w_{ij,n} y_{j,n} + x_{i,n}\beta + \epsilon_{i,n} > 0)$ first, where $0 \leqslant \lambda_i \leqslant B_\lambda$. After establishing the properties of $\frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda_j}$, where the probability depends on parameter values, which are dropped for convenient and simplified notations, we will let $\lambda_1 = \cdots = \lambda_n = \lambda$. Given $m \in \mathbb{N}$ with $m \geqslant m_0$, if we change some values of $\{\epsilon_{j,n} : d^{ij} > m\}$ from $\tilde{\epsilon}_n^{(i,>m)}$ to $\bar{\epsilon}_n^{(i,>m)}$, by Proposition 1,

$$\Pr\left[\epsilon_n^{(i,\leqslant m)} : y_{i,n}(\epsilon_n^{(i,\leqslant m)}, \tilde{\epsilon}_n^{(i,>m)}, X_n) \neq y_{i,n}(\epsilon_n^{(i,\leqslant m)}, \bar{\epsilon}_n^{(i,>m)}, X_n) \middle| \tilde{\epsilon}_n^{(i,>m)}, \bar{\epsilon}_n^{(i,>m)}, X_n\right] \leqslant (\bar{l}_p \bar{\delta})^m.$$

As a result, given a k with $d_{ik} > m\bar{d}_0$, if we only change the value of $\epsilon_{k,n}$ from a to b (but keep all the other $\epsilon_{j,n}$'s), then

$$\Pr\left[\epsilon_n^{(i,\leqslant m)} : y_{i,n}(\epsilon_n^{(i,\leqslant m)}, a, \epsilon_n^{(i,>m)\backslash\{k\}}, X_n) \neq y_{i,n}(\epsilon_n^{(i,\leqslant m)}, b, \epsilon_n^{(i,>m)\backslash\{k\}}, X_n) \middle| \epsilon_n^{(i,>m)\backslash\{k\}}, X_n\right] \leqslant (\bar{l}_p \bar{\delta})^m.$$

Taking expectation with respect to $\epsilon_n^{(i,>m)\backslash\{k\}}$, we have that for any real numbers a and b,

$$|\Pr(y_{i,n}=1|\epsilon_{k,n}=a, X_n) - \Pr(y_{i,n}=1|\epsilon_{k,n}=b, X_n)| \leqslant (\bar{l}_p \bar{\delta})^m. \tag{C.1}$$

The critical values for $\epsilon_{k,n}$ are $-x_{k,n}\beta - \lambda C_0 \geqslant -x_{k,n}\beta - \lambda C_1 \geqslant \cdots \geqslant -x_{k,n}\beta - \lambda C_{p(k)}$, where $C_0 = 0$ and $\{C_0, \cdots, C_{p(k)}\} \in \{\sum w_{kj,n} y_{j,n} : y_{j,n} \in \{0,1\}\}$. Let $C_{-1} = -\infty$ and $C_{p(k)+1} = \infty$. Let $b_{jk} \equiv \Pr(y_{i,n}=1| - x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}, X_n)$. Then

$$\Pr(y_{i,n}=1|x_n) = \sum_{j=0}^{p(k)+1} b_{jk} \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n). \tag{C.2}$$

Because $y_{k,n} = 1(\lambda_k \sum_{j=1}^n w_{kj,n} y_{j,n} + x_{k,n}\beta + \epsilon_{k,n} > 0)$, it is the sign of $\lambda_k \sum_{j=1}^n w_{ij,n} y_{j,n} + x_{k,n}\beta + \epsilon_{k,n}$ that determines $y_{k,n}$. Given $\epsilon_{k,n} \in (-x_{k,n}\beta - \lambda C_j, -x_{k,n}\beta - \lambda C_{j-1})$ and any profile $y_{-k,n} \in \{0,1\}^{n-1}$, individual k is indifferent on the value of $\epsilon_{k,n}$, because the sign of $\lambda_k \sum_l w_{kl,n} y_{l,n} + x_{k,n}\beta + \epsilon_{k,n}$ is the same when $\epsilon_{k,n}$ varies over $(-x_{k,n}\beta - \lambda C_j, -x_{k,n}\beta - \lambda C_{j-1})$. Thus, $\Pr(y_{i,n} = 1|\epsilon_{k,n} = a, X_n)$ is identical for all $a \in (-x_{k,n}\beta - \lambda C_j, -x_{k,n}\beta - \lambda C_{j-1})$. This implies that for any

$a \in (-x_{k,n}\beta - \lambda C_j, -x_{k,n}\beta - \lambda C_{j-1})$,

$$b_{jk} = \int_{-x_k-\lambda C_j}^{-x_k-\lambda C_{j-1}} \Pr(y_{i,n}=1|\epsilon_{k,n}) dF(\epsilon_{k,n}| - x_{k,n}\beta - \lambda C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda C_{j-1}, X_n)$$

$$= \Pr(y_{i,n}=1|\epsilon_{k,n}=a, x_n) \int_{-x_k-\lambda C_j}^{-x_k-\lambda C_{j-1}} dF(\epsilon_{k,n}| - x_{k,n}\beta - \lambda C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda C_{j-1}, X_n)$$

$$= \Pr(y_{i,n}=1|\epsilon_{k,n}=a, X_n).$$

As a result, by Eq. (C.1), $\max_{0\leqslant i<j\leqslant 1+p} |b_{ik} - b_{jk}| \leqslant (\bar{l}_p\bar{\delta})^m$. In addition, once the interval $(-x_{k,n}\beta - \lambda C_j, -x_{k,n}\beta - \lambda C_{j-1})$ is given, we can analyze the maximum NE in $\mathbb{R}^{n-1}$, the space of $\epsilon_{-k,n}$. Using the critical values of individuals $\{1,\cdots,n\}\backslash\{k\}$, we can find zones where $y_{i,n}=1$ and calculate b_{jk}. Because the critical values of individuals $\{1,\cdots,n\}\backslash\{k\}$ do not depend on λ_k and x_k, $\partial b_{jk}/\partial\lambda_k = \partial b_{jk}/\partial(x_{k,n}\beta) = 0$. So,

$$\frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda_k} = \sum_{j=0}^{p(k)+1} \frac{\partial \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)}{\partial \lambda_k} b_{jk}. \tag{C.3}$$

Because $\left|\frac{\partial \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)}{\partial \lambda_k}\right| \leqslant B_f B_W$, and

$$\sum_{j=0}^{p(k)+1} \frac{\partial \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)}{\partial \lambda_k}$$

$$= \frac{\partial}{\partial \lambda_k} \sum_{j=0}^{p(k)+1} \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n) = \frac{\partial 1}{\partial \lambda_k} = 0,$$

it follows by Lemma C.1 that $0 \leqslant \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda_k} \leqslant [p(k)+1] B_f B_W (\bar{l}_p\bar{\delta})^m$ when $d_{ik} > m\bar{d}_0$ and $m \geqslant m_0$. Also, in general, because $\max_{0\leqslant i<j\leqslant 1+p} |b_{ik} - b_{jk}| < 1$, by Lemma C.1, $0 \leqslant \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda_k} \leqslant [p(k)+1] B_f B_W$. By Assumptions 1 and 7, $B_p \equiv \sup_{k,n} p(k) < \infty$. Hence, when $\lambda_1 = \cdots = \lambda_n = \lambda$,

$$\frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda} = \sum_{k=1}^{n} \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda_k} = \left(\sum_{k: d_{ik} \leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty} \sum_{k: m\bar{d}_0 < d_{ik} \leqslant (m+1)\bar{d}_0} \right) \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda_k}$$

$$\leqslant C_{d\bar{d}_0} m^d \cdot (B_p+1) B_f B_W + \sum_{m=m_0}^{\infty} C_{d\bar{d}_0} m^{d-1} \cdot (B_p+1) B_f B_W (\bar{l}_p\bar{\delta})^m$$

$$\leqslant (B_p+1) B_f B_W C_{d\bar{d}_0} \left[m^d + \sum_{m=m_0}^{\infty} m^{d-1} (\bar{l}_p\bar{\delta})^m \right] = B_f B_W C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}} < \infty,$$

where $C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$ depends only on d, $\bar{l}_p$, $\bar{\delta}$, $\bar{d}_0$ and m_0, since B_p relies merely on d and $\bar{d}_0$. Hence, $\frac{\partial \Pr(y_{i,n}=1)}{\partial \lambda} = \mathrm{E}\, \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial \lambda} \leqslant B_f B_W C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}$. As a result,

$$\left|\frac{\partial\, \mathrm{E}(y_{i,n} q_{il,n})}{\partial \lambda}\right| = \left|\frac{\partial\, \mathrm{E}[q_{il,n}\, \mathrm{E}(y_{i,n}|X_n)]}{\partial \lambda}\right| = \left|\mathrm{E}\left[q_{il,n} \frac{\partial\, \mathrm{E}(y_{i,n}|X_n)}{\partial \lambda}\right]\right|$$
$$\leqslant \mathrm{E}\left[|q_{il,n}| \left|\frac{\partial\, \mathrm{E}(y_{i,n}|X_n)}{\partial \lambda}\right|\right] \leqslant B_Q B_f B_W C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}}.$$

Next, $\frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} = \sum_{j=0}^{p+1} \frac{\partial \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)}{\partial (x_{k,n}\beta)} b_{jk}$. Because $|\partial \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)/\partial(x_{k,n}\beta)| \leqslant B_f$ and $\sum_{j=0}^{p+1} \partial \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)/\partial(x_{k,n}\beta) = \frac{\partial 1}{\partial (x_{k,n}\beta)} = 0$, by Lemma C.1, we have $0 \leqslant \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} \leqslant (p(k)+1)B_f$ and $0 \leqslant \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} \leqslant (p(k)+1)B_f(\bar{l}_p\bar{\delta})^m$ when $d_{ik} > m\bar{d}_0$ and $m \geqslant m_0$. For any $1 \leqslant j \leqslant K$,

$$\left|\frac{\partial \Pr(y_{i,n}=1)}{\partial \beta_j}\right| = \left|\mathrm{E} \sum_{k=1}^{n} \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} \frac{\partial (x_{k,n}\beta)}{\partial \beta_j}\right|$$
$$= \left|\left(\sum_{k: d_{ik} \leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty} \sum_{k: m\bar{d}_0 < d_{ik} \leqslant (m+1)\bar{d}_0}\right) \mathrm{E}\, \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} \frac{\partial (x_{k,n}\beta)}{\partial \beta_j}\right| \tag{C.4}$$
$$\leqslant B_f B_X (B_p+1) C_{d\bar{d}_0} \left[m^d + \sum_{m=m_0}^{\infty} m^{d-1} (\bar{l}_p\bar{\delta})^m\right] = B_f B_X C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}} < \infty.$$

As a result,

$$\left|\frac{\partial\, \mathrm{E}(y_{i,n} q_{il,n})}{\partial \beta_j}\right| = \left|\mathrm{E}\, q_{il,n} \sum_{k=1}^{n} \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} \frac{\partial (x_{k,n}\beta)}{\partial \beta_j}\right|$$
$$= \left|\left(\sum_{k: d_{ik} \leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty} \sum_{k: m\bar{d}_0 < d_{ik} \leqslant (m+1)\bar{d}_0}\right) \mathrm{E}\left[q_{il,n} \frac{\partial \Pr(y_{i,n}=1|X_n)}{\partial (x_{k,n}\beta)} \frac{\partial (x_{k,n}\beta)}{\partial \beta_j}\right]\right|$$
$$\leqslant B_f B_Q B_X (B_p+1) C_{d\bar{d}_0} \left[m^d + \sum_{m=m_0}^{\infty} m^{d-1} (\bar{l}_p\bar{\delta})^m\right] = B_f B_Q B_X C_{d\bar{d}_0 m_0 \bar{l}_p \bar{\delta}} < \infty.$$

□

C.3. Second Order Derivatives

For two second order continuously differentiable functions $g(x) = (g_1(x), \cdots, g_m(x))' : \mathbb{R}^n \to \mathbb{R}^m$ and $F(\cdot) : \mathbb{R}^m \to \mathbb{R}$, we have $\frac{\partial^2 F(g(x))}{\partial x_k \partial x_h} = \frac{\partial g(x)'}{\partial x_h} \frac{\partial^2 F(g(x))}{\partial g \partial g'} \frac{\partial g(x)}{\partial x_k} + \frac{\partial F(g(x))}{\partial g'} \frac{\partial^2 g(x)}{\partial x_k \partial x_h}$ for $h, k = 1, \cdots, n$, where $x = (x_1, \cdots, x_n)$ and $'$ represents transpose of matrices, not derivative.

Lemma C.3. $\sup_{i,n,\theta,X_n} \left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n,\theta)}{\partial \lambda^2}\right| \leqslant B_W^2 C_{dm_0 \bar{d}_0 f \bar{l}_p \bar{\delta}}$, $\sup_{i,n,l,\theta,X_n,j_1,j_2} \left|\frac{\partial^2 \mathrm{E}(y_{i,n} q_{il,n})}{\partial \beta_{j_1} \partial \beta_{j_2}}\right| \leqslant B_Q^3 C_{dm_0 \bar{d}_0 f \bar{l}_p \bar{\delta}}$, and $\sup_{i,j,l,n,\theta,X_n} \left|\frac{\partial^2 \mathrm{E}(y_{i,n} q_{il,n})}{\partial \lambda \partial \beta_j}\right| \leqslant B_X B_Q B_W C_{dm_0 \bar{d}_0 f \bar{l}_p \bar{\delta}}$.

Proof of Lemma C.3: First, we calculate $\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial \lambda_k^2}$. Differentiating Eq. (C.3) with respect to λ_k, we have

$$\frac{\partial^2 \Pr(y_{i,n} = 1|X_n)}{\partial \lambda_k^2} = \sum_{j=0}^{p(k)+1} \frac{\partial^2 \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1}|X_n)}{\partial \lambda_k^2} b_{jk}. \qquad \text{(C.5)}$$

Because $\left|\frac{\partial^2 \Pr(-x_{k,n}\beta - \lambda_k C_j < \epsilon_{k,n} < -x_{k,n}\beta - \lambda_k C_{j-1})}{\partial \lambda_k^2}\right| = |C_{j-1}^2 f'(-x_{k,n}\beta - \lambda_k C_{j-1}) - C_j^2 f'(-x_{k,n}\beta - \lambda_k C_j)| \leqslant 2B_{f'} B_W^2$, where $B_{f'} \equiv \max_x |f'(x)|$, and $0 \leqslant b_{jk} \leqslant 1$, we have $|\partial^2 \Pr(y_{i,n} = 1|X_n)/\partial \lambda_k^2| \leqslant 2B_{f'}(p(k) + 2)B_W^2$. When $d_{ik} > m\bar{d}_0$ and $m \geqslant m_0$, by Lemma C.1,

$$\left|\frac{\partial^2 \Pr(y_{i,n} = 1|X_n)}{\partial \lambda_k^2}\right| \leqslant 2B_{f'} B_W^2 [p(k) + 1](\bar{l}_p \bar{\delta})^m.$$

To calculate $\partial^2 \Pr(y_{i,n} = 1|X_n)/\partial \lambda_{k_1} \partial \lambda_{k_2}$ with $k_1 \neq k_2$, without loss of generality, assume that $d_{ik_1} \geqslant d_{ik_2}$. Denote the critical values for $\epsilon_{k_j,n}$ ($j = 1$ or 2) as $-x_{k_j,n}\beta - \lambda C_0^{(j)} \geqslant -x_{k_j,n}\beta - \lambda C_1^{(j)} \geqslant \cdots \geqslant -x_{k_j,n}\beta - \lambda C_{p(k)}^{(j)}$, where $C_0^{(j)} = 0$ and $\{C_0^{(j)}, \cdots, C_{p(k)}^{(j)}\} \in \{\sum w_{k_j l,n} y_{l,n} : y_{l,n} \in \{0, 1\}\}$. Let $C_{-1}^{(j)} = -\infty$ and $C_{p(k_{(j)})+1}^{(j)} = \infty$. Let $A_{j_1 k_1,n} = \Pr(-x_{k_1,n}\beta - \lambda_{k_1} C_{j_1}^{(1)} < \epsilon_{k_1,n} < -x_{k_1,n}\beta - \lambda_{k_1} C_{j_1-1}^{(1)}|X_n)$, $A_{j_2 k_2,n} = \Pr(-x_{k_2,n}\beta - \lambda_{k_2} C_{j_2}^{(2)} < \epsilon_{k_2,n} < -x_{k_2,n}\beta - \lambda_{k_2} C_{j_2-1}^{(2)}|X_n)$, and $b_{j_1 j_2,n} = \Pr(y_{i,n} = 1| - x_{k_1,n}\beta - \lambda_{k_1} C_{j_1}^{(1)} < \epsilon_{k_1,n} < -x_{k_1,n}\beta - \lambda_{k_1} C_{j_1-1}^{(1)}, -x_{k_2,n}\beta - \lambda_{k_2} C_{j_2}^{(2)} < \epsilon_{k_2,n} < -x_{k_2,n}\beta - \lambda_{k_2} C_{j_2-1}^{(2)}, X_n)$. Then $\Pr(y_{i,n} = 1) = \sum_{j_1=0}^{p(k_1)+1} \sum_{j_2=0}^{p(k_2)+1} A_{j_1 k_1,n} A_{j_2 k_2,n} b_{j_1 j_2,n}$. Given the intervals that $\epsilon_{k_1,n}$ and $\epsilon_{k_2,n}$ are located, and given other players' actions, $y_{k_1,n}$ and $y_{k_2,n}$ are determined, the values of $\epsilon_{k_1,n}$ and $\epsilon_{k_2,n}$ do not matter. In addition, given the intervals that $\epsilon_{k_1,n}$ and $\epsilon_{k_2,n}$ are located, we can divide $\mathbb{R}^{n-2}$, the space of $(\epsilon_{j,n})_{j \neq k_1,k_2}$, into some cuboids by the critical values of individuals in $\{1, \cdots, n\} \backslash \{k_1, k_2\}$. Then we can calculate the maximum NE in each cuboid and obtain $b_{j_1 j_2,n}$. Because the critical values of individuals in $\{1, \cdots, n\} \backslash \{k_1, k_2\}$ do not contain λ_{k_1}, λ_{k_2}, x_{k_1} and x_{k_2}, the partial derivatives of $b_{j_1 j_2,n}$ with respect to these variables equal 0. As a result,

$$\frac{\partial^2 \Pr(y_{i,n} = 1|X_n)}{\partial \lambda_{k_1} \partial \lambda_{k_2}} = \sum_{j_1=0}^{p(k_1)+1} \sum_{j_2=0}^{p(k_2)+1} \frac{\partial A_{j_1 k_1,n}}{\partial \lambda_{k_1}} \frac{\partial A_{j_2 k_2,n}}{\partial \lambda_{k_2}} b_{j_1 j_2,n}.$$

When $d_{ik_1} > m\bar{d}_0$ and $d^{ik_2} > m \geqslant m_0$, by Proposition 1(1), $\max_{1\leqslant j_1<j_1'\leqslant p(k_1)}|b_{j_1j_2,n} - b_{j_1'j_2,n}| \leqslant \bar{\delta}(\bar{l}_p\bar{\delta})^m$. On the other hand, when $d_{ik_1} > m\bar{d}_0$, $d^{ik_2} \leqslant m$ and $m \geqslant m_0$, we have $\max_{1\leqslant j_1<j_1'\leqslant p(k_1)}|b_{j_1j_2,n} - b_{j_1'j_2,n}| \leqslant (\bar{l}_p\bar{\delta})^m$ by Proposition 1(2). So, when $d_{ik_1} > m\bar{d}_0$ and $m \geqslant m_0$, $\max_{1\leqslant j_1<j_1'\leqslant p(k_1)}|b_{j_1j_2,n} - b_{j_1'j_2,n}| \leqslant (\bar{l}_p\bar{\delta})^m$. Notice that $\max_{l=1,2}\max_{1\leqslant j_l\leqslant p(k_l)}|\partial A_{j_lk_l,n}/\partial\lambda_{k_l}| \leqslant B_f||W_n||_\infty$. Consequently, when $d_{ik_1} > m\bar{d}_0$ and $m \geqslant m_0$, by Lemma C.1,

$$\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial\lambda_{k_2}}\right| \leqslant \sum_{j_2=0}^{p(k_2)+1}\left|\frac{\partial A_{j_2k_2,n}}{\partial\lambda_{k_2}}\right|\left|\sum_{j_1=0}^{p(k_1)+1}\frac{\partial A_{j_1k_1,n}}{\partial\lambda_{k_1}}b_{j_1j_2,n}\right|$$
$$\leqslant \sum_{j_2=0}^{p(k_2)+1}\left|\frac{\partial A_{j_2k_2,n}}{\partial\lambda_{k_2}}\right|(p(k_1)+1)B_fB_W(\bar{l}_p\bar{\delta})^m \leqslant (p(k_1)+1)(p(k_2)+1)B_f^2B_W^2(\bar{l}_p\bar{\delta})^m.$$

Because $b_{j_1j_2,n} \in [0,1]$, we always have

$$\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial\lambda_{k_2}}\right| \leqslant \left(\sum_{j_1=0}^{p(k_1)+1}\left|\frac{\partial A_{j_1k_1,n}}{\partial\lambda_{k_1}}\right|\right)\left(\sum_{j_2=0}^{p(k_2)+1}\left|\frac{\partial A_{j_2k_2,n}}{\partial\lambda_{k_2}}\right|\right) \leqslant (p(k_1)+2)(p(k_2)+2)B_f^2B_W^2.$$

By the bounds for $\partial^2\Pr(y_{i,n}=1|X_n)/\partial\lambda_k^2$ and $\partial^2\Pr(y_{i,n}=1|X_n)/\partial\lambda_{k_1}\partial\lambda_{k_2}$,

$$\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda^2}\right| = \left|\sum_{k_1,k_2}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial\lambda_{k_2}}\right|$$
$$\leqslant \left|\sum_k \frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_k^2}\right| + 2\left|\sum_{k_1}\sum_{k_2\neq k_1:d_{ik_2}\leqslant d_{ik_1}}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial\lambda_{k_2}}\right|$$
$$\leqslant \left(\sum_{k:d_{ik}\leqslant m_0\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k:m\bar{d}_0<d_{ik}\leqslant(m+1)\bar{d}_0}\right)\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_k^2}\right| +$$
$$2\left(\sum_{k_1:d_{ik_1}\leqslant m\bar{d}_0}\sum_{k_2\neq k_1:d_{ik_2}\leqslant d_{ik_1}} + \sum_{m=m_0}^{\infty}\sum_{k_1:m\bar{d}_0<d_{ik_1}\leqslant(m+1)\bar{d}_0}\sum_{k_2\neq k_1:d_{ik_2}\leqslant d_{ik_1}}\right)\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial\lambda_{k_2}}\right|$$
$$\leqslant B_W^2\left[C_{d\bar{d}_0}m_0^d\cdot 2B_{f'}(B_p+2) + \sum_{m=m_0}^{\infty}C_{d\bar{d}_0}m^{d-1}\cdot 2B_{f'}(B_p+1)(\bar{l}_p\bar{\delta})^m\right] +$$
$$2B_W^2\left[(C_{d\bar{d}_0}m_0^d)^2\cdot(B_p+2)^2B_f^2 + \sum_{m=m_0}^{\infty}C_{d\bar{d}_0}m^{d-1}\cdot C_{d\bar{d}_0}(m+1)^d\cdot(B_p+1)^2B_f^2(\bar{l}_p\bar{\delta})^m\right]$$
$$\leqslant B_W^2C_{dm_0\bar{d}_0f\bar{l}_p\bar{\delta}} < \infty,$$

for some constant $C_{dm_0\bar{d}_0 f\bar{l}_p\bar{\delta}}$ not depending on i, n, λ or x_n.

Similarly, $|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k,n}\beta)^2| \leqslant 2(p(k)+2)B_{f'}$ and $|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k,n}\beta)\partial\lambda_k| \leqslant 2B_{f'}(p(k)+2)B_W$. When $d_{ik} > m\bar{d}_0$ and $m \geqslant m_0$, $|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k,n}\beta)^2| \leqslant 2(p(k)+1)B_{f'}(\bar{l}_p\bar{\delta})^m$ and $|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k,n}\beta)\partial\lambda_k| \leqslant 2(p(k)+1)B_W B_{f'}(\bar{l}_p\bar{\delta})^m$. For $k_1 \neq k_2$, $|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial\lambda_{k_1}\partial(x_{k_2,n}\beta)| \leqslant (p(k_1)+2)(p(k_2)+2)B_f^2 B_W$ and $|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k_1,n}\beta)\partial(x_{k_2,n}\beta)| \leqslant (p(k_1)+2)(p(k_2)+2)B_f^2$. For $k_1 \neq k_2$, when $d_{ik_2} \leqslant d_{ik_1}$, $d_{ik_1} > m\bar{d}_0$ and $m \geqslant m_0$

$$|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k_1,n}\beta)\partial(x_{k_2,n}\beta)| \leqslant (p(k_1)+1)(p(k_2)+1)B_f^2(\bar{l}_p\bar{\delta})^m,$$

$$|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial\lambda_{k_1}\partial(x_{k_2,n}\beta)| \leqslant (p(k_1)+1)(p(k_2)+1)B_W B_f^2(\bar{l}_p\bar{\delta})^m,$$

$$|\partial^2 \Pr(y_{i,n}=1|X_n)/\partial(x_{k_1,n}\beta)\partial\lambda_{k_2}| \leqslant (p(k_1)+1)(p(k_2)+1)B_W B_f^2(\bar{l}_p\bar{\delta})^m.$$

With the above results, we have

$$\begin{aligned}
&\left|\frac{\partial^2 \mathrm{E}(y_{i,n}q_{ij,n})}{\partial\beta_{j_1}\partial\beta_{j_2}}\right| = \left|\mathrm{E}\sum_{k_1,k_2}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial(x_{k_1,n}\beta)\partial(x_{k_2,n}\beta)}x_{j_1k_1,n}x_{j_2k_2,n}q_{ij,n}\right| \\
\leqslant&\left|\mathrm{E}\sum_{k}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial(x_{k,n}\beta)^2}x_{j_1k,n}x_{j_2k,n}q_{ij,n}\right| + 2\left|\mathrm{E}\sum_{k_1}\sum_{k_2\neq k_1:d_{ik_2}\leqslant d_{ik_1}}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial(x_{k_1,n}\beta)\partial(x_{k_2,n}\beta)}x_{j_1k_1,n}x_{j_2k_2,n}q_{ij,n}\right| \\
\leqslant&\left(\sum_{k:d_{ik}\leqslant m_0\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k:\, m\bar{d}_0<d_{ik}\leqslant(m+1)\bar{d}_0}\right)\mathrm{E}\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial(x_{k,n}\beta)^2}x_{j_1k,n}x_{j_2k,n}q_{ij,n}\right| + \\
&2\left(\sum_{k_1:d_{ik_1}\leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k_1:m\bar{d}_0<d_{ik_1}\leqslant(m+1)\bar{d}_0}\right)\sum_{k_2\neq k_1:d_{ik_2}\leqslant d_{ik_1}}\mathrm{E}\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial(x_{k_1,n}\beta)\partial(x_{k_2,n}\beta)}x_{j_1k_1,n}x_{j_2k_2,n}q_{ij,n}\right| \\
\leqslant&B_Q^3\left[C_{d\bar{d}_0}m_0^d\cdot 2B_{f'}(B_p+2) + \sum_{m=m_0}^{\infty}C_{d\bar{d}_0}m^{d-1}\cdot 2B_{f'}(B_p+1)(\bar{l}_p\bar{\delta})^m\right] + \\
&2B_Q^3\left[(C_{d\bar{d}_0}m_0^d)^2\cdot(B_p+2)^2B_f^2 + \sum_{m=m_0}^{\infty}C_{d\bar{d}_0}m^{d-1}\cdot C_d(m+1)^d\cdot(B_p+1)^2B_f^2(\bar{l}_p\bar{\delta})^m\right] \\
\leqslant&B_Q^3 C_{dm_0\bar{d}_0 f\bar{l}_p\bar{\delta}} < \infty,
\end{aligned}$$

and

$$
\begin{aligned}
&\left|\frac{\partial^2 \mathrm{E}(y_{i,n}q_{ij,n})}{\partial\lambda\partial\beta_j}\right| = \left|\mathrm{E}\sum_{k_1,k_2}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial(x_{k_2,n}\beta)}x_{jk_2,n}q_{ij,n}\right| \\
\leqslant&\left|\mathrm{E}\sum_{k}\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k}\partial(x_{k,n}\beta)}x_{jk,n}q_{ij,n}\right| + \left|\mathrm{E}\left(\sum_{k_1}\sum_{k_2:d_{ik_2}\leqslant d_{ik_1}} + \sum_{k_2}\sum_{k_1:d_{ik_1}\leqslant d_{ik_2}}\right)\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial(x_{k_2,n}\beta)}x_{jk_2,n}q_{ij,n}\right| \\
\leqslant&\left(\sum_{k:d_{ik}\leqslant m_0\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k:\, m\bar{d}_0<d_{ik}\leqslant(m+1)\bar{d}_0}\right)\mathrm{E}\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k}\partial(x_{k,n}\beta)}x_{jk,n}q_{ij,n}\right| + \\
&\left(\sum_{k_1:d_{ik_1}\leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k_1:m\bar{d}_0<d_{ik_1}\leqslant(m+1)\bar{d}_0}\right)\sum_{k_2\neq k_1:d_{ik_2}\leqslant d_{ik_1}}\mathrm{E}\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial(x_{k_2,n}\beta)}x_{jk_2,n}q_{ij,n}\right| + \\
&\left(\sum_{k_2:d_{ik_2}\leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k_2:m\bar{d}_0<d_{ik_2}\leqslant(m+1)\bar{d}_0}\right)\sum_{k_1\neq k_2:d_{ik_1}\leqslant d_{ik_2}}\mathrm{E}\left|\frac{\partial^2 \Pr(y_{i,n}=1|X_n)}{\partial\lambda_{k_1}\partial(x_{k_2,n}\beta)}x_{jk_2,n}q_{ij,n}\right| \\
\leqslant& B_X B_Q B_W\left[C_{d\bar{d}_0}m_0^d\cdot 2B_{f'}(B_p+2) + \sum_{m=m_0}^{\infty}C_{d\bar{d}_0}m^{d-1}\cdot 2B_{f'}(B_p+1)(\bar{l}_p\bar{\delta})^m\right] + \\
&2B_X B_Q B_W\left[(C_{d\bar{d}_0}m_0^d)^2\cdot(B_p+2)^2B_f^2 + \sum_{m=m_0}^{\infty}C_{d\bar{d}_0}m^{d-1}\cdot C_{d\bar{d}_0}(m+1)^d\cdot(B_p+1)^2B_f^2(\bar{l}_p\bar{\delta})^m\right] \\
\leqslant& B_X B_Q B_W C_{dm_0\bar{d}_0 f\bar{l}_p\bar{\delta}} < \infty.
\end{aligned}
$$

□

D. Proofs for the Main Text

Lemma D.1. *For all real numbers a_1, a_2, b_1, b_2, $|\max(a_1,a_2)-\max(b_1,b_2)| \leqslant \max(|a_1-b_1|,|a_2-b_2|) \leqslant |a_1-b_1|+|a_2-b_2|$.*

Proof of Lemma D.1: If a_1-a_2 and b_1-b_2 have the same sign, then $(a_1 \geqslant a_2)\wedge(b_1 \geqslant b_2)$ or $(a_1 \leqslant a_2)\wedge(b_1\leqslant b_2)$. Clearly, the conclusion holds.

If a_1-a_2 and b_1-b_2 have different signs, without loss of generality, we assume $a_1 \geqslant a_2$ and $b_1 \leqslant b_2$. The following are all possible cases: $b_1\leqslant b_2\leqslant a_2\leqslant a_1$, $b_1\leqslant a_2\leqslant b_2\leqslant a_1$, $b_1\leqslant a_2\leqslant a_1\leqslant b_2$, $a_2\leqslant b_1\leqslant b_2\leqslant a_1$, $a_2\leqslant b_1\leqslant a_1\leqslant b_2$, and $a_2\leqslant a_1\leqslant b_1\leqslant b_2$. One can check that the conclusion holds under each case. □

Lemma D.2. *$\{X_{i,n}^{(1)}\}_{i=1}^n$ and $\{X_{i,n}^{(2)}\}_{i=1}^n$ are two random fields on $\{\epsilon_{i,n}\}_{i=1}^n$. If $||X_{i,n}^{(j)}-\mathrm{E}[X_{i,n}^{(j)}|\mathcal{F}_{i,n}(s)]||_{L^2}\leqslant$*

$d_{i,n}^{(j)}\psi^{(j)}(s)$ *for* $j = 1, 2$, *where* $\mathcal{F}_{i,n}(s) = \sigma(\{\epsilon_{j,n} : d_{ij} \leqslant s\})$, *then*

$$|| \max_{j=1,2} X_{i,n}^{(j)} - \mathrm{E}[\max_{j=1,2} X_{i,n}^{(j)}|\mathcal{F}_{i,n}(s)]||_{L^2} \leqslant (d_{i,n}^{(1)} + d_{i,n}^{(2)}) \max_{j=1,2} \psi^{(j)}(s).$$

Proof of Lemma D.2: By Lemma D.1, the conclusion holds because NED is kept under Lipschitz transformations. □

Proof of Proposition 1: If $\{j : \infty > d^{ij} > m\} = \emptyset$, $\bar{y}_{i,n} = y_{i,n}(\epsilon_n^{(i,\leqslant m)}, x_n^{(i,\leqslant m)}) = \tilde{y}_{i,n}$. The conclusions hold trivially.

Next, we consider the case $\{j : d^{ij} = m+1\} \neq \emptyset$. Denote $\chi_{i,n}(\theta_i, x_{i,n}) \equiv 1(-\lambda_i||W_n||_\infty - x_{i,n}\beta \leqslant \epsilon_{i,n} < -x_{i,n}\beta)$. When $\chi_{i,n}(\theta_i, x_{i,n}) = 0$, $y_{i,n}(\epsilon_n, X_n)$ will not be affected by other players' decisions. When $\chi_{i,n}(\theta_i, x_{i,n}) = 1$, it is possible (but not necessary) that other players' actions will affect player i's decision. Call an individual i is *susceptible* at $(\theta_i, x_{i,n})$ iff $\chi_{i,n}(\theta_i, x_{i,n}) = 1$; otherwise, i is *unaffected* at $(\theta_i, x_{i,n})$ iff $\chi_{i,n}(\theta_i, x_{i,n}) = 0$. Because of the i.i.d. of $\epsilon_{i,n}$, conditional on X_n, $\chi_{i,n}(\theta_i, x_{i,n})$ and $\chi_{j,n}(\theta_j, x_{j,n})$ are independent for $i \neq j$.

First, we shall argue that, if $\bar{y}_{i,n} \neq \tilde{y}_{i,n}$, then there exists a susceptible path (i.e., a path of susceptible individuals at (θ, X_n)) $j_m \to j_{m-1} \to \cdots \to j_1 \to i$ with $d^{ij_m} = m$. Suppose the conclusion does not hold. Denote $A \subseteq \{j : d^{ij} < m\}$ the set of all individuals of whom each will be on at least one susceptible path ending in i. Then $A \cap \{j : d^{ij} = m\} = \emptyset$ and $B \equiv \{j \in \{j : d^{ij} \leqslant m\} \backslash A : d^{kj} = 1 \,\text{for some}\, k \in A\} \neq \emptyset$. From the definitions of A and B, B is unaffected. Thus, when we consider the behaviors of individuals in $A \cup B$, we can ignore the behavior of the rest players. Now consider the subgame formed by individuals in $A \cup B$. By the unique best NE for the subgame, $\bar{y}_n|_{A\cup B} = \tilde{y}_n|_{A\cup B}$. This contradicts $\bar{y}_{i,n} \neq \tilde{y}_{i,n}$. Hence, the conclusion follows.

Next, we will show conclusions (1) and (2). Recall $\sup_{i,n,\theta_i,x_{i,n}} \mathrm{E}(\chi_{i,n}(\theta_i, x_{i,n})|x_{i,n}) = \bar{\delta}$ for some $\bar{\delta} \in (0,1)$. Denote $j_0 = i$.

(1)

$$\begin{aligned}
&\Pr\left(\bar{y}_{i,n} \neq \tilde{y}_{i,n} | \, x_n^{(i,\leqslant m)}, \bar{x}_n^{(i,>m)}, \tilde{x}_n^{(i,>m)}, \bar{\epsilon}_n^{(i,>m)}, \tilde{\epsilon}_n^{(i,>m)}\right)\\
\leqslant &\Pr(\exists \,\text{a susceptible path}\, j_m \to j_{m-1} \to \cdots \to j_1 \to i \,\text{with}\, d^{ij_m} = m | x_n^{(i,\leqslant m)})\\
\leqslant &\sum_{j_m \to j_{m-1} \to \cdots \to j_1 \to i : d^{ij_m} = m} \Pr(i, j_1, \cdots, j_m \,\text{are all susceptible}| x_n^{(i,\leqslant m)})\\
\leqslant &\sum_{j_m \to j_{m-1} \to \cdots \to j_1 \to i : d^{ij_m} = m} \prod_{p=0}^{m} \Pr(\chi_{j_p} = 1 | x_n^{(i,\leqslant m)})\\
\leqslant &\sum_{j_m \to j_{m-1} \to \cdots \to j_1 \to i : d^{ij_m} = m} \prod_{p=0}^{m} \bar{\delta} \leqslant \bar{\delta}(\bar{l}_p \bar{\delta})^m.
\end{aligned}$$

(2) Similarly,

$$\begin{aligned}
&\Pr\left(\bar{y}_{i,n} \neq \tilde{y}_{i,n} | \, \epsilon_{k,n}, x_n^{(i,\leqslant m)}, \bar{x}_n^{(i,>m)}, \tilde{x}_n^{(i,>m)}, \bar{\epsilon}_n^{(i,>m)}, \tilde{\epsilon}_n^{(i,>m)}\right)\\
\leqslant &\sum_{j_m \to j_{m-1} \to \cdots \to j_1 \to i : d^{ij_m} = m} \Pr(i, j_1, \cdots, j_m \text{are all susceptible}| \epsilon_{k,n}, x_n^{(i,\leqslant m)})\\
\leqslant &\sum_{j_m \to j_{m-1} \to \cdots \to j_1 \to i : d^{ij_m} = m} \prod_{p=0}^{m} \Pr(\chi_{j_p} = 1 | \epsilon_{k,n}, x_n^{(i,\leqslant m)}).
\end{aligned} \tag{D.1}$$

Conditional on $\{\epsilon_{k,n}, x_n^{(i,\leqslant m)}\}$, $\Pr(\chi_{j_p} = 1 | \epsilon_{k,n}, x_n^{(i,\leqslant m)}) = \Pr(\chi_{j_p} = 1 | x_n^{(i,\leqslant m)}) \leqslant \bar{\delta}$ for $j_p \neq k$, and $\Pr(\chi_k = 1 | \epsilon_{k,n}, x_n^{(i,\leqslant m)}) \leqslant 1$ no matter whether k is susceptible or not. So $\prod_{p=0}^{m} \Pr(\chi_{j_p} = 1 | \epsilon_{k,n}, x_n^{(i,\leqslant m)}) \leqslant \bar{\delta}^m$ and $\Pr\left(\bar{y}_{i,n} \neq \tilde{y}_{i,n} | \, \epsilon_{k,n}, x_n^{(i,\leqslant m)}, \bar{x}_n^{(i,>m)}, \tilde{x}_n^{(i,>m)}, \bar{\epsilon}_n^{(i,>m)}, \tilde{\epsilon}_n^{(i,>m)}\right) \leqslant (\bar{l}_p \bar{\delta})^m$. □

Proof of Corollary 1: Given a pair of nonstochastic $(\tilde{x}_n^{(i,>m)}, \tilde{\epsilon}_n^{(i,>m)})$, denote $\tilde{y}_{i,n}(\theta) = y_{i,n}(\epsilon_n^{(i,\leqslant m)}, \tilde{\epsilon}_n^{(i,>m)}, x_n^{(i,\leqslant m)}, \tilde{x}_n^{(i,>m)} | \theta)$ as a function of $\epsilon_n^{(i,\leqslant m)}$ and $x_n^{(i,\leqslant m)}$. By Proposition 1, $\mathrm{E}\, |y_{i,n}(\theta) - \tilde{y}_{i,n}(\theta)| \leqslant (\bar{l}_p \bar{\delta})^m$. Hence, by Theorem 10.12 in Davidson (1994, p.151),

$$||y_{i,n}(\theta) - \mathrm{E}[y_{i,n}(\theta) | x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m]||_{L^2}^2 \leqslant ||y_{i,n}(\theta) - \tilde{y}_{i,n}(\theta)||_{L^2}^2 \leqslant (\bar{l}_p \bar{\delta})^m.$$

Then, the first conclusion holds. Because $\{j : d^{ij} \leqslant m\} \subseteq \{j : d_{ij} \leqslant m\bar{d}_0\}$ under Assumption 1, by Theorem 10.12 in Davidson (1994), $||y_{i,n}(\theta) - \mathrm{E}[y_{i,n}(\theta) | \mathcal{F}_{i,n}(m\bar{d}_0)]||_{L^2} \leqslant ||y_{i,n}(\theta) - \mathrm{E}[y_{i,n}(\theta) | x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m]||_{L^2} \leqslant (\bar{\delta}\bar{l}_p)^{m/2}$. □

Proof of Corollary 2: The first conclusion holds because

$$
\begin{aligned}
&||y_{i,n}(\theta)x_{ik,n} - \mathrm{E}[y_{i,n}(\theta)x_{ik,n}|x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m]||_{L^2}^2 \\
\leqslant &\mathrm{E}\left\{x_{ik,n}^2\, \mathrm{E}\left\{\left[y_{i,n}(\theta) - \mathrm{E}\left(y_{i,n}(\theta)|x_{j,n}, \epsilon_{j,n}, d^{ij} \leqslant m\right)\right]^2 \middle| X_n\right\}\right\} \leqslant \mathrm{E}\left[x_{ik,n}^2(\bar{\delta}\bar{l}_p)^m\right] \leqslant B_X^2(\bar{\delta}\bar{l}_p)^m,
\end{aligned}
$$

where the second inequality follows by similar arguments as in Corollary 1. The rest conclusions hold similarly. □

Proof of Proposition 2: Because $W_n \neq 0$ from Assumption 3, there exists an i with $\sum_{j=1}^n w_{ij,n} > 0$. By presumption, $\mathrm{E}[\Pr(y_{i,n} = 1|X_n, \theta) - y_{i,n}|X_n] = \Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta) = 1|X_n) - \Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta_0) = 1|X_n) = 0$. Denote $F(\cdot)$ as the CDF of the $\epsilon_{i,n}$. Given $x_{j1,n}$ (for all $j \neq i$), we consider the case that $x_{j2,n}\beta_{20} \to -\infty$ for all j with $d^{ij} = 1$. Thus,

$$
1 - F(-x_{i,n}\beta_0) = \Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta) = 1|X_n). \tag{D.2}
$$

We discuss the identification by three mutually exclusive but exhaustive situations. (1) If $sign(\beta_{20}) = sign(\beta_2)$, then $x_{j2,n}\beta_{20} \to -\infty$ implies $x_{j,n}\beta \to -\infty$. Thus, $1 - F(-x_{i,n}\beta_0) = 1 - F(-x_{i,n}\beta)$. Because $F(\cdot)$ is strictly increasing, $x_{i,n}\beta = x_{i,n}\beta_0$. So, $\mathrm{E}(x_{i,n}'x_{i,n})\beta = \mathrm{E}(x_{i,n}'x_{i,n})\beta_0$. Because $\mathrm{E}(x_{i,n}'x_{i,n})$ has full rank, $\beta = \beta_0$. Next, given $x_{j1,n}$, consider $x_{j2,n}\beta_{20} \to +\infty$ for all j with $d^{ij} = 1$. Then, $\Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta) = 1|X_n) = \Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta_0) = 1|X_n)$ implies $F(-\lambda\sum_{j=1}^n w_{ij,n} - x_{i,n}\beta_0) = F(-\lambda_0\sum_{j=1}^n w_{ij,n} - x_{i,n}\beta_0)$. Hence, $\lambda_0 = \lambda$. (2) If $sign(\beta_{20}) = -sign(\beta_2)$, $x_{j2,n}\beta_{20} \to -\infty$ implies $x_{j2,n}\beta_2 \to +\infty$. Then $\Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta_0) = 1|X_n) = \Pr(\epsilon_n : y_{i,n}(\epsilon_n, X_n, \theta) = 1|X_n)$ implies $F(-x_{i,n}\beta_0) = F(-\lambda\sum_{j=1}^n w_{ij,n} - x_{i,n}\beta)$. So,

$$
x_{i,n}\beta_0 = \lambda\sum_{j=1}^n w_{ij,n} + x_{i,n}\beta. \tag{D.3}
$$

Next, let $x_{j2,n}\beta_{20} \to +\infty$. Then $x_{j2,n}\beta_2 \to -\infty$ and $F(-\lambda_0\sum_{j=1}^n w_{ij,n} - x_{i,n}\beta_0) = F(-x_{i,n}\beta)$. Thus,

$$
\lambda_0\sum_{j=1}^n w_{ij,n} + x_{i,n}\beta_0 = x_{i,n}\beta. \tag{D.4}
$$

By Eq. (D.3) and (D.4), $\lambda_0\sum_{j=1}^n w_{ij,n} = -\lambda\sum_{j=1}^n w_{ij,n}$. Because $\lambda_0 \geqslant 0$, $\lambda \geqslant 0$ and $\sum_{j=1}^n w_{ij,n} > 0$, we have $\lambda = \lambda_0 = 0$. In consequence, Eq. (D.3) implies that $\beta = \beta_0$. (3) When $\beta_2 = 0$, the LHS of Eq. (D.2) depends on $x_{i2,n}$ but the RHS does not depend on $x_{i2,n}$. This is impossible. □

Proof of Theorem 1: Under Assumptions 5 and 9 on compact parameter spaces and identification, to establish the conclusion, it is sufficient to show that $\sup_{\theta\in\Theta}|Q_n(\theta)-\bar{Q}_n(\theta)| \xrightarrow{p} 0$ and that $\{\bar{Q}_n(\theta)\}_{n=K+1}^{\infty}$ is equicontinuous. Consider $\sup_{\theta\in\Theta}|Q_n(\theta)-\bar{Q}_n(\theta)| \xrightarrow{p} 0$ first. Notice

$$
\begin{aligned}
&\sup_{\theta\in\Theta}|Q_n(\theta)-\bar{Q}_n(\theta)| \leqslant \sup_{\theta\in\Theta}\left|\left\|\Omega_n^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^n \hat{g}_{i,n}(\theta)\right\|^2 - \left\|\Omega^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^n \hat{g}_{i,n}(\theta)\right\|^2\right| \\
&+\sup_{\theta\in\Theta}\left|\left\|\Omega^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^n \hat{g}_{i,n}(\theta)\right\|^2 - \left\|\Omega^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^n \mathrm{E}\, g_{i,n}(\theta)\right\|^2\right| \\
\leqslant&\sup_{\theta\in\Theta}\left\|(\Omega_n(\theta)-\Omega(\theta))^{1/2}\frac{1}{n}\sum_{i=1}^n \hat{g}_{i,n}(\theta)\right\|^2 + \\
&\sup_{\theta\in\Theta}\left|\left[\frac{1}{n}\sum_{i=1}^n[\hat{g}_{i,n}(\theta)+\mathrm{E}\, g_{i,n}(\theta)]\right]'\Omega(\theta)\left[\frac{1}{n}\sum_{i=1}^n[\hat{g}_{i,n}(\theta)-\mathrm{E}\, g_{i,n}(\theta)]\right]\right| \\
\leqslant& o_p(1)+\sup_{\theta\in\Theta}\left\|\Omega^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^n[\hat{g}_{i,n}(\theta)+\mathrm{E}\, g_{i,n}(\theta)]\right\|\cdot\left\|\Omega^{1/2}(\theta)\frac{1}{n}\sum_{i=1}^n[\hat{g}_{i,n}(\theta)-\mathrm{E}\, g_{i,n}(\theta)]\right\|,
\end{aligned} \tag{D.5}
$$

where the last inequality is based on Assumptions 8 and 9(1) and the Cauchy-Schwartz inequality. Because $\sup_{i,k,n,\theta}||\hat{g}_{ik,n}(\theta)||_{L^2}<\infty$, it suffices to show $\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{i=1}^n[\hat{g}_{ik,n}(\theta)-\mathrm{E}\, g_{ik,n}(\theta)]| \xrightarrow{p} 0$. By the WLLN for NED (Jenish and Prucha, 2012), $\frac{1}{n}\sum_{i=1}^n(y_{i,n}-\mathrm{E}\, y_{i,n})q_{ik,n}=o_p(1)$. Since

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n[\hat{g}_{ik,n}(\theta)-\mathrm{E}\, g_{ik,n}(\theta)]\right| \\
=&\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{R}\sum_{r=1}^R y_{i,n}(\epsilon_n^{(r)},X_n,\theta)-y_{i,n}\right]q_{ik,n}-[\mathrm{E}\, y_{i,n}(\epsilon_n,X_n,\theta)-\mathrm{E}\, y_{i,n}]q_{ik,n}\right| \\
=&\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{r=1}^R\frac{1}{n}\sum_{i=1}^n\left[y_{i,n}(\epsilon_n^{(r)},X_n,\theta)-\mathrm{E}\, y_{i,n}(\epsilon_n,X_n,\theta)\right]q_{ik,n}\right|+\left|\frac{1}{n}\sum_{i=1}^n(y_{i,n}-\mathrm{E}\, y_{i,n})q_{ik,n}\right| \\
\leqslant&\frac{1}{R}\sum_{r=1}^R\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n\left[y_{i,n}(\epsilon_n^{(r)},X_n,\theta)-\mathrm{E}\, y_{i,n}(\epsilon_n,X_n,\theta)\right]q_{ik,n}\right|+o_p(1),
\end{aligned}
$$

it is sufficient to show $\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n[y_{i,n}(\epsilon_n,X_n,\theta)-\mathrm{E}\, y_{i,n}(\epsilon_n,X_n,\theta)]q_{ik,n}\right| = o_p(1)$, because $\mathrm{E}\, y_{i,n}(\epsilon_n^{(r)},X_n,\theta)=\mathrm{E}\, y_{i,n}(\epsilon_n,X_n,\theta)$. We apply the bracketing method in empirical process in three steps to show this. Empirical process method is needed because $y_{i,n}(\epsilon_n,X_n,\theta)$ is neither smooth nor even continuous in θ. Let $y_{i,n}(\epsilon_n,X_n\beta,\lambda)\equiv y_{i,n}(\epsilon_n,X_n,\theta)$ based on the explicit index form $X_n\beta$ of the regressors in the model.

First, we construct the brackets. Notice that for any $\delta > 0$ and $\theta_1 = (\lambda_1, \beta_1')' \in \Theta$ with $\overline{B(\theta_1, \delta)} \equiv \{\theta \in \mathbb{R}^{K+1} : ||\theta - \theta_1||_\infty \leqslant \delta\} \subseteq \Theta$, if $\bar{\theta} \in \overline{B(\theta_1, \delta)}$, then $|x_{i,n}\bar{\beta} - x_{i,n}\beta_1| \leqslant K||x_{i,n}||\delta$. Denote $||X_n|| \equiv (||x_{1,n}||, \cdots, ||x_{n,n}||)'$. Because $Y_n(\epsilon_n, X_n\beta, \lambda)$ is non-decreasing in λ and $X_n\beta$, $Y_n(\epsilon_n, X_n\beta_1 - K||X_n||\delta, \lambda_1 - \delta) \leqslant Y_n(\epsilon_n, X_n\bar{\beta}, \bar{\lambda}) \leqslant Y_n(\epsilon_n, X_n\beta_1 + K||X_n||\delta, \lambda_1 + \delta)$. Because $a \leqslant b \leqslant c$ implies that $\min(ad, cd) \leqslant bd \leqslant \max(ad, cd)$ for any $d \in \mathbb{R}$,

$$
\begin{aligned}
&\underline{yq}_{ik,n}(\theta_1) \equiv \min[y_{i,n}(\epsilon_n, X_n\beta_1 - K||X_n||\delta, \lambda_1 - \delta)q_{ik,n}, y_{i,n}(\epsilon_n, X_n\beta_1 + K||X_n||\delta, \lambda_1 + \delta)q_{ik,n}] \\
\leqslant &y_{i,n}(\epsilon_n, X_n\bar{\beta}, \bar{\lambda})q_{ik,n} \\
\leqslant &\max[y_{i,n}(\epsilon_n, X_n\beta_1 - K||X_n||\delta, \lambda_1 - \delta)q_{ik,n}, y_{i,n}(\epsilon_n, X_n\beta_1 + K||X_n||\delta, \lambda_1 + \delta)q_{ik,n}] \equiv \overline{yq}_{ik,n}(\theta_1).
\end{aligned}
$$

Because Θ is a cuboid and $\overline{B(\theta_1, \delta)}$ is a cube in $\mathbb{R}^{K+1}$, we can find $N(\delta) < \infty$ cubes $\{\overline{B(\theta_j, \delta)}\}_{j=1}^{N(\delta)}$ such that $\cup_{j=1}^{N(\delta)}\overline{B(\theta_j, \delta)} = \Theta$. We want to emphasize that $N(\delta)$ and $\{\theta_j\}_{j=1}^{N(\delta)}$ depend only on δ and Θ, but not on n. Thus, $\{[\underline{yq}_{ik,n}(\theta_j), \overline{yq}_{ik,n}(\theta_j)]\}_{j=1}^{N(\delta)}$ is a bracketing set for $\{y_{i,n}(\epsilon_n, X_n\beta, \lambda)q_{ik,n} : (\lambda, \beta')' \in \Theta\}$.

Second, we show that it is an $L^2(\epsilon)$-bracket by choosing a suitable δ. Notice that

$$
\begin{aligned}
&||\overline{yq}_{ik,n}(\theta_j) - \underline{yq}_{ik,n}(\theta_j)||_{L^2} \\
=&||[y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j + \delta) - y_{i,n}(\epsilon_n, X_n\beta_j - K||X_n||\delta, \lambda_j - \delta)]q_{ik,n}||_{L^2} \\
\leqslant&B_Q||y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j + \delta) - y_{i,n}(\epsilon_n, X_n\beta_j - K||X_n||\delta, \lambda_j - \delta)||_{L^{p_0}} \\
=&B_Q \,\mathrm{E}^{1/p_0}[y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j + \delta) - y_{i,n}(\epsilon_n, X_n\beta_j - K||X_n||\delta, \lambda_j - \delta)],
\end{aligned}
\tag{D.6}
$$

where a power of p_0 is the same as power one because $y_{i,n}$'s take values 0 or 1. By similar argument as that for Eq. (C.4), we have

$$\begin{aligned}
&\left|\frac{\partial \operatorname{E} y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j)}{\partial \delta}\right| = \left|\operatorname{E}\frac{\partial \operatorname{E}[y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j)|X_n]}{\partial \delta}\right| \\
=&\left|\operatorname{E}\sum_{k=1}^{n}\frac{\partial \operatorname{E}[y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j)|X_n]}{\partial (x_{k,n}\beta)}K||x_{k,n}||\right| \\
=&\left|\left(\sum_{k:d_{ik}\leqslant m\bar{d}_0} + \sum_{m=m_0}^{\infty}\sum_{k:m\bar{d}_0<d_{ik}\leqslant(m+1)\bar{d}_0}\right)\operatorname{E}\frac{\partial \operatorname{E}[y_{i,n}(\epsilon_n, X_n\beta_j + K||X_n||\delta, \lambda_j)|X_n]}{\partial (x_{k,n}\beta)}K||x_{k,n}||\right| \\
\leqslant& K^2 B_X B_f (B_p+2) C_{d\bar{d}_0}\left[m^d + \sum_{m=m_0+1}^{\infty} m^{d-1}\cdot(\bar{l}_p\bar{\delta})^m\right] = K^2 B_X B_f C_{d\bar{d}_0 m_0 \bar{l}_p\bar{\delta}} < \infty.
\end{aligned} \tag{D.7}$$

By Eq. (D.7) and Lemma C.2, we have

$$\operatorname{E}[y_{i,n}(\epsilon_n, X_n\beta_j+K||X_n||\delta, \lambda_j+\delta)-y_{i,n}(\epsilon_n, X_n\beta_j-K||X_n||\delta, \lambda_j-\delta)] \leqslant 2\delta(K^2B_X+B_W)B_fC_{d\bar{d}_0m_0\bar{l}_p\bar{\delta}}.$$

Thus, when $\delta \leqslant (\epsilon/B_Q)^{p_0}/[2(K^2B_X + B_W)B_fC_{d\bar{d}_0m_0\bar{l}_p\bar{\delta}}]$, we have $||\overline{yq}_{ik,n}(\theta_j) - \underline{yq}_{ik,n}(\theta_j)||_{L^2} \leqslant \epsilon$.

Finally, the uniform convergence $\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n[y_{ik,n}(\epsilon_n, X_n, \theta) - \operatorname{E} y_{ik,n}(\epsilon_n, X_n, \theta)]q_{ik,n}\right| = o_p(1)$ holds by the following argument as that for Theorem 2.4.1 in van der Vaart and Wellner (1996). By Corollary 2 and Lemma D.2, both $\{\underline{yq}_{ik,n}(\theta_j)\}_{i=1}^n$ and $\{\overline{yq}_{ik,n}(\theta_j)\}_{i=1}^n$ are geometrically L_2-NED uniformly in i, j and n. For each $\theta \in \Theta$, there exists a $1 \leqslant j \leqslant N(\delta)$ such that $||\theta - \theta_j||_\infty < \delta$. Then

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n[y_{i,n}(\epsilon_n, X_n\beta, \lambda) - \operatorname{E} y_{i,n}(\epsilon_n, X_n\beta, \lambda)]q_{ik,n} \leqslant \frac{1}{n}\sum_{i=1}^n[\overline{yq}_{ik,n}(\theta_j) - \operatorname{E} y_{i,n}(\epsilon_n, X_n\beta, \lambda)q_{ik,n}] \\
=&\frac{1}{n}\sum_{i=1}^n[\overline{yq}_{ik,n}(\theta_j) - \operatorname{E}\overline{yq}_{ik,n}(\theta_j)] + \frac{1}{n}\sum_{i=1}^n \operatorname{E}\{[y_{i,n}(\epsilon_n, X_n\beta_j + KB_X\delta\iota_n, \lambda_j+\delta) - y_{i,n}(\epsilon_n, X_n\beta, \lambda)]q_{ik,n}\} \\
\leqslant& \max_{1\leqslant j'\leqslant N(\delta)}\frac{1}{n}\sum_{i=1}^n[\overline{yq}_{ik,n}(\theta_{j'}) - \operatorname{E}\overline{yq}_{ik,n}(\theta_{j'})] + \epsilon.
\end{aligned}$$

Notice that the first term on the RHS of the above equation does not depend on θ and equals $o_p(1)$ by the WLLN for NED. Thus, $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n[y_{ik,n}(\epsilon_n, X_n, \theta) - \operatorname{E} y_{ik,n}(\epsilon_n, X_n, \theta)]q_{ik,n} \leqslant o_p(1) + \epsilon$. Similarly, $\inf_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n[y_{ik,n}(\epsilon_n, X_n, \theta) - \operatorname{E} y_{ik,n}(\epsilon_n, X_n, \theta)]q_{ik,n} \geqslant o_p(1) - \epsilon$. Because $\epsilon > 0$ is arbitrary, the uniform convergence in probability holds.

It remains to show that $\{\bar{Q}_n(\theta)\}_{n=K+1}^{\infty}$ is equicontinuous. It suffices to show $\frac{1}{n}\sum_{i=1}^n \mathrm{E}\, g_{ik,n}(\theta) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}\{[y_{i,n}(\epsilon_n, X_n\beta, \lambda) - y_{i,n}(\epsilon_n, X_n\beta_0, \lambda_0)]q_{ik,n}\}$ is equicontinuous, which is implied by Lemma C.2. □

Proof of Proposition 3: $g_{i,n}(\theta) = [\Pr(y_{i,n} = 1|X_n, \theta) - y_{i,n}]q'_{i,n}$. To apply Theorem A.1, we need to calculate the bracketing number. The required brackets have been constructed in the proof of Theorem 1 and we have shown that when $\delta \leqslant (\epsilon/B_Q)^{p_0}/[2(K^2 B_X + B_W)B_f C_{d\bar{d}_0 m_0 n_0 \bar{\delta}}]$, $||\overline{yq}_{ik,n}(\theta_j) - \underline{yq}_{ik,n}(\theta_j)||_{L^2} \leqslant \epsilon$. To cover the parameter space $[0, B_\lambda] \times \prod_{k=1}^K [-B_{\beta_k}, B_{\beta_k}]$, we need $\lceil \frac{B_\lambda}{\delta} \rceil \times \prod_{k=1}^K \left\lceil \frac{2B_{\beta_k}}{\delta} \right\rceil = O(\epsilon^{-p_0(1+K)})$ cubes in $\mathbb{R}^{1+K}$. With $w_0 \geqslant w > 2p_0(K+1)r_0^{-1}$, we have $\frac{p_0(K+1)}{wr_0} + \frac{1}{2} < 1$. As a result, for some constant $C > 0$,

$$\int_0^1 N(x^{1/r_0})^{1/w} x^{-1/2} dx \leqslant C \int_0^1 x^{-[p_0(K+1)/wr_0 + 1/2]} dx < \infty.$$

Then the condition $\int_0^1 N(x^{1/r_0})^{1/w} x^{-1/2} dx < \infty$ in Theorem A.1 holds. Consequently, Theorem A.1 is applicable and this proposition holds. □

Proof of Corollary 3: Denote $y_{i,n}^{(r)}(\theta) \equiv y_{i,n}(\epsilon_n^{(r)}, X\beta, \lambda)$.

$$\begin{aligned}
&\limsup_{n\to\infty} \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \left| n^{-1/2} \sum_{i=1}^n [\hat{g}_{ik,n}(\theta_1) - \hat{g}_{ik,n}(\theta_2)] \right| \right\|_{L^q} \\
\leqslant &\limsup_{n\to\infty} \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \left| n^{-1/2} \sum_{i=1}^n \frac{1}{R} \sum_{r=1}^R [y_{i,n}^{(r)}(\theta_1) - y_{i,n}^{(r)}(\theta_2)] q_{ik,n} \right| \right\|_{L^q} \\
\leqslant &\limsup_{n\to\infty} \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \frac{1}{R} \sum_{r=1}^R \left| n^{-1/2} \sum_{i=1}^n [y_{i,n}^{(r)}(\theta_1) - y_{i,n}^{(r)}(\theta_2)] q_{ik,n} \right| \right\|_{L^q} \\
\leqslant &\limsup_{n\to\infty} \frac{1}{R} \sum_{r=1}^R \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \left| n^{-1/2} \sum_{i=1}^n [y_{i,n}^{(r)}(\theta_1) - y_{i,n}^{(r)}(\theta_2)] q_{ik,n} \right| \right\|_{L^q} \\
\leqslant &\limsup_{n\to\infty} \left\| \sup_{||\theta_1-\theta_2||_\infty<\eta} \left| n^{-1/2} \sum_{i=1}^n [g_{ik,n}(\theta_1) - g_{ik,n}(\theta_2)] \right| \right\|_{L^q} \leqslant \epsilon,
\end{aligned}$$

where the last inequality is from Proposition 3. □

Proof of Theorem 2: Let $G_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^n g_{i,n}(\theta)$ and $\bar{G}_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \mathrm{E}\, g_{i,n}(\theta)$. When $\lambda_0 > 0$, we apply Theorem B.1 and Lemma B.1 to prove the conclusion.

Apparently, Corollary 3 is sufficient for condition (iii) in Theorem B.1 and Lemma C.3 implies condition (vi) there. It remains to check condition (iv) in Theorem B.1, whether the CLT holds.

Consider $m \geqslant m_0$. Denote $\Delta y_{i,n}^{(r,m)} = y_{i,n}(\epsilon_n^{(r)}, X_n, \theta) - \mathrm{E}[y_{i,n}(\epsilon_n^{(r)}, X_n, \theta)|x_{j,n}, \epsilon_{j,n}^{(r)},\, d_{ij} \leqslant m\bar{d}_0]$ and $\Delta y_{i,n}^{(m)} = y_{i,n} - \mathrm{E}(y_{i,n}|x_{j,n}, \epsilon_{j,n},\, d_{ij} \leqslant m\bar{d}_0)$. By Theorem 10.12 in Davidson (1994) and Minkowski's inequality

$$
\begin{aligned}
&||\hat{g}_{ik,n}(\theta) - \mathrm{E}[\hat{g}_{ik,n}(\theta)|x_{j,n}, \epsilon_{j,n}, \epsilon_{j,n}^{(r)}, r = 1, \cdots, R,\, d_{ij} \leqslant m\bar{d}_0]||_{L^2} \\
\leqslant& ||\frac{1}{R}\sum_{r=1}^{R} \Delta y_{i,n}^{(r,m)} q_{ik,n} - \Delta y_{i,n}^{(m)} q_{ik,n}||_{L^2} \leqslant \frac{1}{R}\sum_{r=1}^{R} ||\Delta y_{i,n}^{(r,m)} q_{ik,n}||_{L^2} + ||\Delta y_{i,n}^{(m)} q_{ik,n}||_{L^2} \leqslant 2B_Q(\bar{l}_p\bar{\delta})^{m/2}.
\end{aligned}
$$

So, $\{\hat{g}_{ik,n}(\theta)\}_{i=1}^n$ is uniformly and geometrically L_2-NED on $\{x_{i,n}, \epsilon_{i,n}, \epsilon_{i,n}^{(r)}, r = 1, \cdots, R\}_{i=1}^n$. Next, we calculate its variance. Denote $\nu(g_n(\theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [y_{i,n} - \mathrm{E}(y_{i,n}|X_n, \theta_0)]q_{i,n}$ and $\nu(g_n^{(r)}(\theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [y_{i,n}(\epsilon_n^{(r)}, X_n, \theta_0) - \mathrm{E}(y_{i,n}|X_n, \theta_0)]q_{i,n}$. Notice that $\nu(g_n^{(r)}(\theta_0))$ and $\nu(g_n(\theta_0))$ are independent conditional on X_n, and $\mathrm{E}[\nu(g_n(\theta_0))|X_n] = \mathrm{E}[\nu(g_n^{(r)}(\theta_0))|X_n] = 0$. By $\mathrm{var}(Y) = \mathrm{E}\,\mathrm{var}(Y|X) + \mathrm{var}\,\mathrm{E}(Y|X)$,

$$
\begin{aligned}
&\mathrm{var}\, \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ik,n}(\theta) = \mathrm{var}\left[\frac{1}{R}\sum_{r=1}^R \nu(g_n^{(r)}(\theta_0)) - \nu(g_n(\theta_0))\right] \\
=& \mathrm{E}\left\{\mathrm{var}\left[\frac{1}{R}\sum_{r=1}^R \nu(g_n^{(r)}(\theta_0)) - \nu(g_n(\theta_0))\middle| X_n\right]\right\} \\
=& \mathrm{E}\left\{\frac{1}{R^2}\sum_{r=1}^R \mathrm{var}\left[\nu(g_n^{(r)}(\theta_0))\middle| X_n\right] + \mathrm{var}\left[\nu(g_n(\theta_0))\right| X_n]\right\} \\
=& (1 + \frac{1}{R})\,\mathrm{E}\left\{\mathrm{var}\left[\nu(g_n(\theta_0))\right| X_n]\right\} = (1 + \frac{1}{R})\,\mathrm{var}\,\nu(g_n(\theta_0)) \to (1 + \frac{1}{R})V.
\end{aligned}
$$

Hence, the CLT of NED in Jenish and Prucha (2012) is applicable under Assumptions 1, 4, 12 and 14, and we have $\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{i,n}(\theta_0) \xrightarrow{d} N(0, (1 + \frac{1}{R})V)$. Then this theorem holds by Theorem B.1 and Lemma B.1.

When $\lambda_0 = 0$, the conclusion is by Theorem B.2 and Lemma B.1. □

Proof of Proposition 4: (1) Similarly to the proof of Theorem 1, it is sufficient to show $\sup_{\theta\in\Theta} |Q_n^{(b)}(\theta) - \bar{Q}_n(\theta)| \xrightarrow{p} 0$. By the same argument as in Eq. (D.5), we only need to show

$\sup_{\theta\in\Theta} |\frac{1}{n}\sum_{i=1}^n [\hat{g}_{ik,n}^{(b)}(\theta) - \mathrm{E}\, g_{ik,n}(\theta)]| \xrightarrow{p} 0$.

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n [\hat{g}_{ik,n}^{(b)}(\theta) - \mathrm{E}\, g_{ik,n}(\theta)]\right| \leqslant \left|\frac{1}{n}\sum_{i=1}^n [y_n(\epsilon_n^{(b)}, X_n, \theta_0) - y_n(\epsilon_n^{(b)}, X_n, \hat{\theta}_n)] q_{ik,n}\right| + \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n \left\{[\hat{\Pr}(y_{i,n} = 1|X_n, \theta) - y_n(\epsilon_n^{(b)}, X_n, \theta_0)] q_{ik,n} - \mathrm{E}\, g_{ik,n}(\theta)\right\}\right|.$$

Because the first term on the RHS is $o_p(1)$ by Proposition 3 and $\hat{\theta}_n = \theta_0 + o_p(1)$, and the second term on the RHS is $o_p(1)$ from the proof of Theorem 1, the uniform convergence holds.

(2) It suffices to check conditions (iii) and (iv) of Theorem B.1 when $\lambda_0 \in (0, B_\lambda)$. A sufficient condition for condition (iii) is the SEC of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ik,n}^{(b)}(\theta)$. Notice that the term $y_n(\epsilon_n^{(b)}, X_n, \hat{\theta}_n)$ in $\hat{g}_{ik,n}^{(b)}(\theta)$ does not depend on θ, then the SEC of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ik,n}^{(b)}(\theta)$ can be established by the same argument as in Proposition 3 and Corollary 3. Next, we verify condition (iv).

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ik,n}^{(b)}(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [\hat{\Pr}(y_{i,n} = 1|X_n, \theta) - y_n(\epsilon_n^{(b)}, X_n, \theta_0)] q'_{i,n} + \frac{1}{\sqrt{n}}\sum_{i=1}^n [y_n(\epsilon_n^{(b)}, X_n, \theta_0) - y_n(\epsilon_n^{(b)}, X_n, \hat{\theta}_n)] q'_{i,n}.$$

Because the second term on the RHS is $o_p(1)$ by the SEC of $\frac{1}{\sqrt{n}}\sum_{i=1}^n y_n(\epsilon_n^{(b)}, X_n, \theta) q'_{i,n}$, and the first term $\xrightarrow{d} N(0, (1 + \frac{1}{R})V)$ from the proof of Theorem 2, condition (iv) holds.

Because conditions (iii) and (iv) of Theorem B.2 are the same as those of Theorem B.1, the conclusion also holds when $\lambda_0 = 0$. □

References

[1] ANDREWS, D. W. (1994): "Empirical Process Methods in Econometrics," Handbook of Econometrics, 4, 2247-2294.

[2] ANDREWS, D. W. AND POLLARD, D. (1994): "An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes," International Statistical Review, 62(1), 119-132.

[3] BAJARI, P., H. HONG, AND S. P. RYAN (2010): "Identification and Estimation of a Discrete Game of Complete Information," Econometrica, 78, 1529-1568.

[4] BERNHEIM, B. D., B. PELEG, AND M. D. WHINSTON (1987): "Coalition-Proof Nash Equilibria I. Concepts," Journal of Economic Theory, 42, 1-12.

[5] BESAG, J. (1974): "Spatial Interaction and the Statistical Analysis of Lattice Systems," Journal of the Royal Statistical Society (Series B), 192-236.

[6] CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," Econometrica, 75, 1243-1284.

[7] DAVIDSON, J. (1994): Stochastic Limit Theory: An Introduction for Econometricians. Oxford university press.

[8] DE JONG, R. M., AND T. WOUTERSEN (2011): "Dynamic Time Series Binary Choice," Econometric Theory, 27, 673-702.

[9] DEHLING, H., AND W. PHILIPP (2002): Empirical Process Techniques for Dependent Data. Springer.

[10] DOUKHAN, P., AND S. LOUHICHI (1999): "A New Weak Dependence Condition and Applications to Moment Inequalities," Stochastic Processes and their Applications, 84, 313-342.

[11] FOLLAND, G. B. (1999): Real Analysis: Modern Techniques and Their Applications. John Wiley & Sons.

[12] GAETAN, C., X. GUYON, AND K. BLEAKLEY (2010): Spatial Statistics and Modeling. Springer.

[13] HOSHINO, T. (2009): "GMM Estimation of Spatial Autoregressive probit Models-An analysis of the implementation of the planning system in Japan," Working Paper.

[14] HOLLOWAYA, G., SHANKARA, B. AND RAHMAN S. (2002): "Bayesian spatial probit estimation: a primer and an application to HYV rice adoption," Agricultural Economics, 27(3), 2002.

[15] HSIEH, C.-S. AND LEE, L.-F. (2014): “Specification and Estimation of Network Formation and Network Interaction Models with the Exponential Probability Distribution”, Working Paper.

[16] JENISH, N. AND PRUCHA, I. R. (2009): “Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields,” Journal of Econometrics, 150, 86-98.

[17] JENISH, N. AND PRUCHA, I. R. (2012): “On Spatial Processes and Asymptotic Inference Under Near-Epoch Dependence,” Journal of Econometrics, 170, 178-190.

[18] JIA, P. (2008): “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” Econometrica, 76, 1263-1316.

[19] KRAUTH, B. V. (2006): “Simulation-Based Estimation of Peer Effects,” Journal of Econometrics, 133, 243-271.

[20] LACOMBE AND LESAGE (2013): “Use and Interpretation of Spatial Autoregressive probit,” Working Paper.

[21] LESAGE, J., AND R. K. PACE (2009): Introduction to Spatial Econometrics. CRC press.

[22] LIU, X., PATACCHINI, E. AND ZENOU, Y. (2014): “Endogenous Peer Effects: Local Aggregate or Local Average?”, Working Paper.

[23] MENZEL, K. (2014): “Large Matching Markets as Two-Sided Demand Systems,” Working paper, New York University.

[24] — (2015): “Inference for Games with Many Players,” Working paper, New York University.

[25] MILGROM, P., AND J. ROBERTS (1990): “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities,” Econometrica, 58, 1255-1277.

[26] — (1995): “Strongly Coalition-Proof Equilibria in Games with Strategic Complementarities.”

[27] MIYAUCHI, Y. (2014): “Structural Estimation of a Pairwise Stable Network with Nonnegative Externality,” Working paper, MIT.

[28] MOLINARI, F., AND A. ROSEN (2008): “The Identification Power of Equilibrium in Games: The Supermodular Case,” Journal of Business & Economic Statistics, 26, 297-302.

[29] OLVER, F. W. (2010): NIST Handbook of Mathematical Functions. Cambridge University Press.

[30] PAKES, A., AND D. POLLARD (1989): “Simulation and the Asymptotics of Optimization Estimators,” Econometrica, 57 (5), 1027-1057.

[31] TRAIN, K. E. (2009): Discrete Choice Methods with Simulation. Cambridge university press.

[32] UETAKE, K., AND Y. WATANABE (2013): “Estimating Supermodular Games Using Rationalizable Strategies,” Advances in Econometrics, 31, 233-247

[33] VAN DER VAART, A. W., AND J. A. WELLNER (1996): Weak Convergence. Springer.

[34] XU, X. AND Lee, L. (2015a): “Maximum Likelihood Estimation of a Spatial Autoregressive Tobit Model,” Journal of Econometrics, 188, 264-288.

[35] — (2015b): “Sieve Maximum Likelihood Estimation of the Spatial Autoregressive Tobit model,” Working Paper, The Ohio State University.