Nonparametric Identification and Estimation of Additive Social Interaction Models with Homophily

Xirong Chen∗  Darong Dai†  Wenzheng Gao‡

October 26, 2017

Abstract

We study strategic social interaction among economic agents that are connected through the phenomena of homophily. In particular, we measure homophily effects by the differences between players’ socioeconomic characteristics. Under the symmetric equilibrium selection mechanism, we establish a nonparametric approach to identify the structural model and propose a computationally feasible two-step estimation procedure. The asymptotic properties of the two-step estimator are derived under context of large games, i.e., the number of players going to infinity. Finally, we apply the identification and estimation methods to study the peer effects on youth smoking behaviors using data of adolescents in the United States, our empirical findings show positive and statistically significant peer effects and demonstrate the empirical importance of including homophily effect in our model.

Keywords: Social Network, Homophily Effect, Global Interaction, Bayesian Nash Equilibrium

JEL classification: C14; C31; C72 and Z13

∗School of Economics and Management, Beihang University. Email: xirong.chenbeihang@163.com
†Department of Economics, Texas A&M University, College Station 77843, USA. Email: dai496@tamu.edu
‡Department of Economics, Texas A&M University, College Station 77843, USA. Email : wgao@tamu.edu

We are grateful to Yonghong An and Qi Li for their continuous guidance and advice. Helpful comments and suggestions from Vikram Manjunath, Zheng Fang, Jeremy Fox, Geert Ridder, Xun Tang, Haiqing Xu and seminar participants at Texas A&M University, 2016 Texas Econometrics Camp and 2017 Asian meeting of the Econometric Society are also gratefully acknowledged. Any remaining errors are our own responsibility.
1 Introduction

Social interaction models study how economic agents interact with each other through their decision making processes with respect to the socioeconomic activity. Recent empirical studies have found evidence of interaction effects on crime (Glaeser et al., 1996), employment (Calvó-Armengol & Jackson, 2004), in-school achievements (Calvó et al., 2009), adolescent behavior (Gaviria & Raphael, 2001; Nakajima, 2007; Badev, 2013), among others. In the previous literature, social interaction can be modeled as either a Manski type linear-in-mean regression model or a strategic game played in a social network. These two approaches are widely used in studying social interaction effects with continuous and discrete outcomes respectively, e.g., see Brock & Durlauf (2001), Bramoullé et al. (2009), Bisin et al. (2011) and Xu (2015), however one potential problem associated with these studies is that they treat other agents in a network equally important for a given economic agent and ignore a pervasive phenomenon in social network: homophily, which is the principle that “similarity breeds connection” (McPherson et al., 2001).

In this paper we construct a social interaction model under the framework of simultaneous move game with incomplete information and adopt the solution concept of Bayesian Nash Equilibrium (BNE). In the game each player chooses an action from a finite set and the payoff function consists of three parts: direct utility from the chosen action, strategic effect from other players’ actions and a stochastic component representing player’s private information. The three components are assumed to be additively separable, similar payoff structure has been studied in De Paula & Tang (2012).

One innovation in this paper is to make use of the homophily principle when measuring the strategic effects of other player’s actions. In sociology, homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people, therefore intuitively we would expect that for a particular player, the strategic effect from another player’s action will be strong if they are similar to each other in terms of socioeconomic attribute. The similarity between two players is represented by a social distance function, which measures the difference between two players’ socioeconomic characteristics, and we restrict the strategic effect to be decreasing as the social distance between two players increases. Our specification of homophily effect is motivated by the previous work of Akerlof (1997) and Conley & Topa (2002), who argue that agents close to each other in terms of socioeconomic characteristics interact strongly while those who are socially distant have little interactions. This specification makes our model different from previous literature, our method can demonstrate how a social network connects each agent to the other and reflects the impact of homophily network structure on agents’ social actions.

Motivated by the commonly adopted data structure in the social interaction litera-
ture, the identification and estimation strategies in this paper are developed under “a large game” setting, meaning that the number of players in a network is fairly large. Identification and estimation in a large game are difficult because of two reasons. First, such games will usually generate multiple equilibria, which leads to the incompleteness of econometric models (Tamer, 2003). Second, players’ actions are interdependent in a large social network, resulting in problems for identifying and estimating player’s equilibrium probability of actions. We solve the first problem by employing a symmetric equilibrium selection mechanism proposed in Leung (2015), which allows for the existence of multiple equilibria but requires those equilibria to be symmetric. The second problem is addressed by imposing a conditional independence assumption, which requires players’ private information to be independently and identically distributed conditional on all the public information and is commonly used in the literature of incomplete information games.

The identification proceeds in two steps. The first step is to identify the equilibrium conditional choice probabilities (CCPs), which is guaranteed by the symmetric equilibrium selection mechanism and conditional independence assumption. The second step is to identify payoff primitives. Specifically, we extend the method proposed in Matzkin (1993) to the context of game theoretical models in order to identify the deterministic part of the payoff function as a whole. The key is to establish a rank ordering property regarding CCPs, which means that actions with higher deterministic payoffs are more likely to be chosen by players. Then by exploring the variation of CCPs and homophily effects, direct utility and strategic effect can be identified separately.

Based upon the identification methodology, we propose a computationally feasible two-step method to nonparametrically estimate the model primitives and establish its consistency. As a result of the symmetric equilibrium selection mechanism, players with the same characteristics can be treated as repeated observations of the same player. Therefore in the first step we can estimate the CCPs using a conventional kernel-type estimator. In the second step, we nonparametrically estimate the parameters of interest by a smoothed version of the pairwise maximum score method proposed in Fox (2007). Furthermore, under a semiparametric setting, we show that the first-stage nonparametric estimation has no impact on the asymptotic behavior of second-stage estimation under mild conditions and derive the asymptotic distribution of smoothed pairwise maximum score estimator.

In the empirical application, we apply our methods to study the peer effects on youth smoking behavior using the data from the National Longitudinal Study of Adolescent Health (Add Health). The Add Health is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year. It contains student’s social network data, as well as their socioeconomic characteristics, which are indispensable for our analysis. We treat each school in the dataset as an observation of
a social network and apply the proposed two-step method to estimate the peer effects on students’ smoking behavior using data from 7 schools, each of which has more than 800 observations. We find positive and statistically significant peer effects for all schools, which is similar to other empirical findings of peer effects on youth smoking behavior using different datasets. See e.g., Nakajima (2007) and Soetevent & Kooreman (2007). Our empirical finding indicates that smoking behavior from a student’s schoolmates will make that student more likely to consume cigarette. We also compare the empirical results with and without imposing the homophily effects, the comparison indicates that without considering the homophily effects, most of the estimated peer effects become insignificant, which demonstrates the empirical importance of including homophily effects in our model.

One of our main contributions in this paper is to employ a novel way to incorporate homophily effect into a social interaction model. Liu & Xu (2015) study the social interaction model with homophily and use the dependence of private information between players to represent the homophily effect. Here we adopt a different approach by using the difference between players’ observed socioeconomic characteristics to explicitly model the homophily effect of the social network, which we believe is more appropriate under the context of incomplete information game because the private information is unobserved between players and hence they can not use it to measure the “closeness” between each other. Under our setting, the homophily effect can be easily calculated using data.

This paper also adds to the growing literature of identification and inference of discrete games with incomplete information. Most of the previous discussions focus on “small-game” settings and assume the observability of a large number of repetitions for the same game in order to identify and estimate the models, see e.g., Aradillas-Lopez 2010; Bajari et al. 2010; De Paula & Tang, 2012; Wan & Xu, 2014). Instead, our identification and estimation methods are based on one observation of a large game and thus are more suitable for the commonly used data structure like the Add Health data in the social interaction literature. A similar paper that considers the large game setting with incomplete information game is Leung (2015), but the objectives of our paper and Leung (2015) are different since his work studies on social network formation while we focus on social interactions in a given network. Our approach treats the network formation process as exogenously given, hence we can use the variation of homophily effect to help identify model primitives. To the best of our knowledge, Badev (2013) is the only paper that considers both network formation and social interactions in networks, but he imposes a Gumbel distribution assumption and uses a MCMC algorithm to identify and estimate the model, which departs from the nonparametric method proposed in this paper.

It is worth mentioning that our identification method is fully nonparametric while most of the previous works in social interaction and incomplete information game adopt paramet-
ric or semiparametric method for identification. For example, Manski (1993) and Bramouillé et al. (2009) assume that the utility function is linear, Brock & Durlauf (2001), Bajari et al. 2010 and Xu (2015) impose a parametric distributional assumption in order to identify the model. Therefore our results are more general and robust to misspecification of those parametric assumptions and provide new insights into the identification methodology of this literature. Lewbel & Tang (2015) also consider nonparametric identification of incomplete information games by using a “special regressor” that is independent of private information. In contrast we allow for the endogeneity of all covariates and achieve point identification by imposing some mild assumptions on the payoff function that can be supported by economic theory.

Last but not the least, our work contributes to the literature of nonparametric estimation by providing a two-step estimator and establishing its uniform consistency using the empirical process methods developed by Pollard (1984) and Van der Vaart & Wellner (1996). If, additionally the payoff function is of a parametric form, we show that the first-step nonparametric estimator is asymptotically orthogonal to the second-step smoothed pairwise maximum score estimation under mild restrictions and hence establish the asymptotic normality for the pairwise smoothed maximum score estimator. Therefore by providing a sufficient condition for asymptotic orthogonality under the context of smoothed maximum score estimation, our work is also related to the literature of semiparametric M-estimation, see e.g., Andrews (1994), Newey (1994) and Ichimura & Lee (2010), but the difference between our work and previous literature is that because of the distribution-free setting and nonsmooth population objective function, our two-step semiparametric estimation will converges at a rate slower than the usual $\sqrt{n}$ rate, which makes it more difficult to derive the rate of convergence and obtain the asymptotic distribution.

The rest of the paper is organized as follows. Section 2 presents the setting and basic assumptions of our model. Section 3 provides the identification method. Section 4 discusses the estimation method and establishes the asymptotic behavior of our proposed estimator. Section 5 contains empirical analysis of peer effects on youth smoking behaviors. Section 6 concludes. All proofs are provided in Appendix.

2 The model

2.1 Setting

We consider a incomplete information game played in a social network. There are $n$ players indexed by $i \in N \equiv \{1, 2, \ldots, n\}$. In this game each player simultaneously choose a discrete action $Y_i \in A \equiv \{0, 1, 2, \ldots, K\}$. Let $X_i \in \mathcal{X} \subseteq \mathbb{R}^d$ and $Z_{ik} \in \mathcal{Z} \subseteq \mathbb{R}^q$ be the vectors of $i$’s payoff relevant state variables. Here $X_i$ represents player $i$’s socioeconomic characteristics.
and $Z_{ik}$ is a vector of observable attributes related to player $i$’s action $k \in A$, which may be different for each player. For example, consider the example of college choice decision, $A$ is the set of colleges available for the student and $X_i$ can be student $i$’s family income, age and so on, while $Z_{ik}$ will be college $k$’s tuition fee and distance to his home, which in general varies across different students. Moreover, player $i$ also observes a vector of choice-specific payoff shocks $\epsilon_i \equiv \{\epsilon_{i0}, \epsilon_{i1}, \ldots, \epsilon_{ik}\} \in \mathbb{R}^{K+1}$, which is private information.

Player $i$’s payoff from choosing an action $k \in A$ is specified as

$$U_{ik}(Y_{-i}, X_i, X_{-i}, Z_{ik}, \epsilon_i) = \alpha(X_i, Z_{ik}) + \sum_{j \neq i} \beta(Y_j, X_i, Z_{ik}) \cdot \gamma(H_{ij}) + \epsilon_{ik}, \quad (1)$$

where $Y_{-i}$ and $X_{-i}$ denotes the action profile and socioeconomic characteristics of all the players except $i$, $\alpha(\cdot)$ is a choice-specific function, and $\beta(\cdot)$ represents the strategic effects of the actions of other player on his payoff. Because only the differences of choice-specific payoffs matter to players, without loss of generality we normalize the payoff of action 0 to be 0. $H_{ij}$ is the distance between $X_i$ and $X_j$, i.e.

$$H_{ij} \equiv d(X_i, X_j) \quad (2)$$

for a standard distance function $d(\cdot)$. We use $H_{ij}$ to measure the socioeconomic difference between player $i$ and player $j$. Based on the theory of homophily in social network, people are more likely to associate and bond with similar others, so in our model $\gamma(H_{ij})$ represents the homophily effect of the social network, Formally, we impose the following assumption:

**Assumption 1.** (Homophily) For all $i, j \in N$, $\gamma(\cdot) : \mathbb{R} \mapsto [0, 1]$ is monotonically decreasing in $H_{ij}$ and $\sum_{j \neq i} \gamma(H_{ij}) = 1$.

One example of such function is $\gamma(H_{ij}) = H_{ij}^{-1} / \sum_{l \in N} H_{il}^{-1}$, which is also the functional form we adopted in the empirical studies. Under Assumption 1, the second part of player $i$’s payoff function can be viewed as a weighted average of the strategic effects of all other players in the same game, where the weights correspond to the homophily effects between player $i$ and other players. This specification makes our model different from the commonly used “linear-in-mean” approach in the literature, which assumes that each player’s action will be affected by the average behavior of all other players (see e.g. Manski,1993; Bramoullé et al., 2009). Under our setting, each player’s action will be affected by a weighted average of other player’s actions, where the weight corresponds to the socioeconomic difference between different players. Therefore $\gamma(\cdot)$ can demonstrate how a social network connects each agent and reflect the impact of homophily network structure on agents’ actions. In the previous example, it is not difficult to see that our specification of the interaction structure include the “linear-in-mean” approach as a special case by setting $H_{ij}$ to be a constant for
all \( i \) and \( j \).

### 2.2 Equilibrium

In this static incomplete information game, each player’s strategy is based on her prior beliefs about the probability distribution of other player’s actions. Let \( Z_i = (Z_{i0}^T, Z_{i1}^T, \ldots, Z_{iK}^T)^T \), \( S_i = (X_i^T, Z_{i0}^T)^T \) and \( S = (S_1, S_2, \ldots, S_n) \) be all the public information associated with player \( i \). Also let \( \theta = (\alpha, \beta(0, \cdot, \cdot), \beta(1, \cdot, \cdot), \ldots, \beta(K, \cdot, \cdot))^T \) be the structural parameters of the game, following the Bayesian Nash Equilibrium (BNE) solution concept, player \( i \)'s equilibrium strategy, denoted as \( Y^*_i \), can be written as

\[
Y^*_i(S, \epsilon_i; \theta) = \arg\max_{k \in A} E[U_{ik}(Y_{-i}, X_i, X_{-i}, Z_{ik}, \epsilon_i) | S, \epsilon_i]
= \arg\max_{k \in A} \left\{ \alpha(X_i, Z_{ik}) + \sum_{l=0}^{K} \beta(l, X_i, Z_{ik}) \sum_{j \neq i} \Pr(Y^*_j(S, \epsilon_j; \theta) = l | S, \epsilon_i) \gamma(H_{ij}) \right\} + \epsilon_{ik}.
\]

(3)

In order to characterize the BNE solution we impose the following assumption

**Assumption 2.** *(Conditional Independence)* Conditional on \( S \), \( \{\epsilon_{ik}\}_{i \in N, k \in A} \) is identically and independently distributed with a continuously differentiable and strictly increasing distribution function \( F_{\epsilon_{ik}|S}(\cdot) \).

Assumption 2 is commonly imposed in the literature on identification and estimation of static games with incomplete information and social interaction models (see, e.g., De Paula & Tang (2012), Bajari et al. (2010) and Xu (2015)). Under this conditional independence assumption, \( \Pr(Y^*_j(S, \epsilon_j; \theta) = l | S, \epsilon_i) = \Pr(Y^*_j(S, \epsilon_j; \theta) = l | S) \) for \( j \neq i \). Following the literature in incomplete information game, we let \( \sigma_{ik}(S; \theta) = \Pr(Y^*_i(S, \epsilon_i; \theta) = k | S) \) be the equilibrium conditional choice probability of player \( i \) choosing action \( k \). To simplify notation, let

\[
V_i(X_i, Z_{ik}, S) \equiv \alpha(X_i, Z_{ik}) + \sum_{l=0}^{K} \left[ \beta(l, X_i, Z_{ik}) \sum_{j \neq i} \sigma_{jl}(S; \theta) \gamma(H_{ij}) \right],
\]

(4)
then a BNE solution (given state \( S \)) can be characterized by

\[
\sigma_{ik}(S; \theta) = Pr[(V_i(X_i, Z_{ik}, S) + \epsilon_{ik} > V_i(X_i, Z_{ih}, S) + \epsilon_{ih})|S], \quad \forall h \in A \setminus \{k\}
\]

\[
= Pr[\epsilon_{ih} < (V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S) + \epsilon_{ik})|S], \quad \forall h \in A \setminus \{k\}
\]

\[
= \int_{\epsilon \in \mathbb{R}} \prod_{h \neq k} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S)) f_{\epsilon_{ik}|S}(\epsilon) d\epsilon,
\]

(5)

where \( f_{\epsilon_{ik}|S}(\cdot) \) denotes the (conditional) density function of \( \epsilon_{ik} \).

For any given \((S; \theta)\) and based on (5), we can define a mapping \( \Gamma^{(S;\theta)} : \Delta \to \Delta \) such that

\[
\Gamma^{(S;\theta)}(\{\sigma_{ik}(S; \theta)\}_{i \in N, k \in A}) \equiv \left( \Gamma_1^{(S;\theta)}(\{\sigma_{ik}(S; \theta)\}_{i \neq 1, k \in A}), \ldots, \Gamma_N^{(S;\theta)}(\{\sigma_{ik}(S; \theta)\}_{i \neq N, k \in A}) \right)^T
\]

with

\[
\Gamma_j^{(S;\theta)}(\{\sigma_{ik}(S; \theta)\}_{i \neq j, k \in A}) = (\sigma_{j0}, \ldots, \sigma_{jk})^T
\]

\[
\equiv \left( \Gamma_0^{(S;\theta)}(\{\sigma_{ik}(S; \theta)\}_{i \neq j, k \in A}), \ldots, \Gamma_K^{(S;\theta)}(\{\sigma_{ik}(S; \theta)\}_{i \neq j, k \in A}) \right)^T
\]

where \( \Delta \) denotes a simplex of dimension \( n \cdot (K + 1) \).

In general, this mapping may have multiple fixed points and hence multiple equilibria, among which we just focus on those symmetric equilibria in this paper. To this end, we first define some permutation functions. Define \( \pi_{ij} : N \to N \) as a permutation of the indices \( i \) and \( j \) of players. Specifically, \( \pi_{ij} \) maps the index \( i \) to the index \( j \), \( j \) to \( i \), and \( i' \) to itself for all \( i' \neq i, j \). Similarly, define \( \pi_{ij}^X \) as a function that permutes the \( i \)-th and \( j \)-th elements of any \( X \equiv (X_1, \ldots, X_n)^T \in X^n \); and \( \pi_{ij}^Z \) as a function that permutes the \( i \)-th and \( j \)-th elements of any \( Z \equiv (Z_1, \ldots, Z_n)^T \in \mathbb{Z}^{n(K+1)} \). We thus have the set of permutations \( \Pi \equiv \{ (\pi_{ij}, \pi_{ij}^X, \pi_{ij}^Z) | i, j \in N \} \) with the generic element written as \( \pi(\cdot) \).

**Definition 1.** An equilibrium belief \( \sigma \in \Delta \) is symmetric if for any \( \theta \in \Theta \), \( i \in N \), \( k \in A \) and \( \pi \in \Pi \), we have \( \sigma_{ik}(S; \theta) = \sigma_{\pi(i)k}(\pi(S); \theta) \).

Here, symmetry means that, for any action \( k \in A \), pairs of agents with the same attributes choose this action with the same conditional probability. Even if such a symmetric equilibrium exists, there might still be multiple equilibria for any given draw \((S, \epsilon)\). We hence, as in Leung (2015), need to define a selection mechanism. First, we introduce a sequence of auxiliary random vectors \( \{\xi^n|n \in N\} \) with an arbitrary finite dimension such that \((S^n, \xi^n) \perp \epsilon^n \) for all \( n \in N \), in which \( S^n \) and \( \epsilon^n \) represent the sequentialization of \( S \) and
\(\epsilon\) using the number of players. In particular, we can make sense of \(\xi^n\) as a public signal that players may use to coordinate on a particular equilibrium \(^1\). Most importantly, we assume that \(\xi^n\) is payoff irrelevant and accordingly, define the equilibrium selection mechanism as a measurable function \(\rho_n : (S^n, \xi^n; \theta) \to \sigma^n \in \Delta^{SE}(S^n; \theta) \subseteq \Delta\), where \(\sigma^n\) denotes the sequentialization of \(\sigma\) using the number of players and \(\Delta^{SE}(S^n; \theta)\) is the set of symmetric equilibria (SE). This mapping thus formalizes the way in which players coordinate on a symmetric equilibrium, and also it does not rely on the privately informed vector \(\epsilon_i\) for all \(i \in N\).

**Assumption 3. (Equilibrium Selection)** There exist sequences of equilibrium selection mechanisms \(\{\rho_n | n \in N\}\) and public signals \(\{\xi^n | n \in N\}\) such that for \(n\) sufficiently large, \(\Delta^{SE}(S^n; \theta)\) is nonempty, and also for any \(Y \equiv (Y_1, Y_2, ..., Y_n)^T\) with \(Y_i \in A\),

\[
Pr(Y^n = \overline{Y}|S^n) = \sum_{\sigma^n \in \Delta^{SE}(S^n; \theta)} Pr(\rho_n (S^n, \xi^n; \theta) = \sigma^n|S^n) \prod_{i=1}^n \sigma^n_i (\overline{Y}_i|S^n),
\]

where \(Y^n\) represents the sequentialization of \(Y \equiv (Y_1, Y_2, ..., Y_n)^T\) using the number of players.

Intuitively Assumption 3 means that given one observation of the game, only one symmetric equilibrium is realized in the data. But we allow the symmetric equilibrium to be different across different observations of the same game. To guarantee that \(\Delta^{SE}(S^n; \theta)\) is nonempty, we need to impose following assumptions about the exchangeability of players and the continuity of payoff functions so that a symmetric BNE always exists,

**Assumption 4. (Anonymity)** For all \(\theta \in \Theta, i \in N, k \in A\) and any realization \(\epsilon_{ik} \in \mathbb{R}\), payoffs \(U_{ik}(\cdot)\) are anonymous in the sense that, for any permutation \(\pi \in \Pi\), we have \(U_{ik}(\sigma_{-i}, S, \epsilon_{ik}) = U_{\pi(i)k}(\sigma_{-\pi(i)}, \pi(S), \epsilon_{\pi(i)k})\).

In a word, under anonymity, payoffs do not depend on the particular labels assigned to players but only on their attributes and equilibrium beliefs, which is a natural assumption under the context of large number of players in the game (See, e.g., Leung (2015) and Menzel (2016)), therefore player labels in the data set have no economic relevance. It also ensures that the equilibria are extensively robust in the sense of Kalai (2004) even if the simultaneous-play assumption is relaxed.

**Assumption 5. (Continuity)** For all \(\theta \in \Theta, i \in N, k \in A\) and any realization \(\epsilon_{ik} \in \mathbb{R}\), payoffs \(U_{ik}(\sigma_{-i}, S, \epsilon_{ik})\) are continuously differentiable in \(S\).

\(^1\)The inclusion of \(\xi^n\) is to ensure that the selection mechanism is nondegenerate, see Leung (2015) for details.
Assumption 5 is a regularity condition to ensure that the mapping $\Gamma(S;\theta)$ has a fixed point. Consequently the existence of a symmetric BNE can be guaranteed by the following theorem:

**Theorem 1.** Suppose Assumptions 2-5 hold, then there exists a symmetric Bayesian-Nash equilibrium.

*Proof.* See Appendix A. □

### 3 Identification

In this section we provide a nonparametric method to explore the identifiability of the structural parameter $\theta$ in the sense similar to Matzkin (1992, 1993), i.e., different values of $\theta$ will result in different choice probabilities. Matzkin (1992, 1993) discusses the nonparametric identification in the discrete choice model with single agent, we modify Matzkin’s definition of identification and apply to the game theoretic model in this paper. To be specific, the identification is implemented in two steps: The first step is to identify players’ conditional choice probability of equilibrium actions and the second step is to identify structural parameters of the payoff function.

As mentioned in the previous section, we focus on one market and the equilibrium selection mechanism ensures that we only have one equilibrium give the one observation of that market, the CCPs are therefore implicitly identified, hence $\theta$ will be identified as well if different values of $\theta$ lead to different CCPs.

In the second step, we achieve identification by restricting the functional form of the payoff function in the social network and proceed as follows: first we identify the composite function $V_i(X_i, Z_{ik}, S)$ for all $i \in N$ and $k \in A$ using a modification of the approach in Matzkin (1993), specifically we impose the following restrictions on the payoff function of the game, which includes some monotonicity and continuity assumptions. Next we identify the structural parameter $\theta$ by imposing a rank condition similar to Bajari et al. (2010) and Xu (2015). First, we introduce the following definition of identification:

**Definition 2.** For all $i \in N$ and $k \in A$, the function $V_i(X_i, Z_{ik}, S)$ is identified in the set $\mathcal{V}$ if for all $V_i'(X_i, Z_{ik}, S) \in \mathcal{V}$ such that $V_i'(X_i, Z_{ik}, S) \neq V_i(X_i, Z_{ik}, S)$, there exist a set $\tilde{S} \in \mathcal{S}$ with positive Lebesgue measure and for all $S \in \tilde{S}$ we have $\sigma_{ik}(S; V) \neq \sigma_{ik}(S; V')$.

In this definition, $\sigma_{ik}(S; V)$ denotes the CCP of player $i$ choosing action $k$ with the emphasis of dependence on $V$, where $V = (V_1, V_2, \cdots, V_n)$ for $i \in N$. Definition 2 simply means that identification can be achieved if different values of $V_i(X_i, Z_{ik}, S)$ lead to different CCPs. In order to obtain the identification result, we impose the following assumptions:
**Assumption 6.** (Monotonic Transformation) For all $V_i$ and $V_i' \in \mathcal{V}$ such that $V_i \neq V_i'$, there does not exist a strict increasing function $m : V_i(\cdot) \to \mathbb{R}$ such that $V_i'(X_i, Z_{ik}, S) = m \circ V_i(X_i, Z_{ik}, S)$ for all $Z_{ik} \in \mathcal{Z}$.

By Assumption 6, no two functions in $\mathcal{V}$ are monotone transformations of each other, this assumption is similar to Assumption 1.3 in Matzkin (1993) and guarantees that for an arbitrary player $i$, no two payoff functions induce the same preorder on $\{Z_{i0}, Z_{i1}, \ldots, Z_{iK}\}$. Matzkin (1993) provides several sufficient conditions for Assumption 6, which includes concavity or homogeneity of $V_i(X_i, Z_{ik}, S)$, see Matzkin (1993) for details.

**Assumption 7.** (Monotonicity) There exists $l \in A$ such that $V_i$ is strictly increasing with respect to $Z_{il}$ for all $V_i \in \mathcal{V}$ and $Z_{il}$ has a everywhere positive Lebesgue density conditional on $S \setminus \{Z_{il}\}$.

Assumption 7 is the key assumption for identification, it means that at least one element of $Z_i$ has a continuous support and that $v_i(\cdot)$ is strictly monotonic on that regressor conditional on all the public information in the game. Note that this assumption is different from the “special regressor” literature initiated by Lewbel (2000), which requires regressor be independent with the private information. We believe that the requirement of monotonicity is less restrictive than independence because it can be motivated by economic theory whereas the independence assumption is hard to justify. The advantage of using the special regressor is that one can also identify the distribution of the private information (see, e.g., Lewbel & Tang (2015)), however this is not the goal of this paper because we are interested in identifying the value of structural parameters. The key step of identification is to establish the so-called rank ordering property, which is defined as follows.

**Definition 3.** The rank ordering property is satisfied if for a given player $i$ and for actions $k, l \in A$,

$$V_i(X_i, Z_{ik}, S) > V_i(X_i, Z_{il}, S)$$

if and only if

$$\sigma_{ik}(S, V_i) > \sigma_{il}(S, V_i).$$

Definition 3 states that the equilibrium belief of player $i$’s action will be rank ordered by the deterministic part of her payoff function. Actions with higher deterministic payoffs are more likely to be chosen. This is a property that was first introduced in Manski (1975), we modify it under the setup of our model. Then we have the following identification theorem:

**Proposition 1.** Under Assumptions 1-7, $V_i(X_i, Z_{ik}, S)$ is identified in the set $\mathcal{V}$ for all $i \in N$ and $k \in A$.

*Proof.* See Appendix A. \(\square\)
We briefly summarize the intuition of our identification strategy: the function $V_i(\cdot)$ is identified by exploring the variation of choice specific characteristics $Z_{ik}$ for $k \in A$, specifically suppose we have two payoff function candidates $V_i(\cdot)$ and $V'_i(\cdot)$ such that $V_i(\cdot) \neq V'_i(\cdot)$. By Assumption 6 and 7, there exists a choice $l \in A$ and an nonempty set $\tilde{S} \subset S$ such that $V_i(\cdot)$ and $V'_i(\cdot)$ will impose opposite preference ordering on options $k$ and $l$ for all $S \in \tilde{S}$, i.e., under payoff function $V_i(\cdot)$ agent $i$ may prefer $k$ to $l$ but under $V'_i(\cdot)$ she will prefer $l$ to $k$ and vice versa. Hence by Assumption 2 and equilibrium condition (5), the rank ordering property holds, then it must be that either $\sigma_{ik}(S;V) \neq \sigma_{ik}(S;V')$ or $\sigma_{il}(S;V) \neq \sigma_{il}(S;vV')$. Therefore by Definition 2 $V_i(\cdot)$ is identified.

Once $V_i(\cdot)$ is identified, the structural parameter $\theta$ can be identified accordingly by exploring the variation of equilibrium beliefs. Specifically we need the variation of the product of equilibrium belief and homophily effect to be sufficiently large, note that since the equilibrium beliefs will add up to one, to avoid the multicollinearity problem we normalize $\beta_k(0,\cdot,\cdot) = 0$ for all $k \in A$, i.e., the choice of action 0 by other players will have no impact on player $i$’s action. Then let $\phi_{il}(S) = \sum_{j \neq i} \sigma_{ji}(S;\theta) \cdot \gamma(H_{ij})$ and $\phi_i(S) = (1,\phi_{i1}(S),\cdots,\phi_{iK}(S))^T$, we introduce the following rank condition.

**Assumption 8.** (Rank Condition) For sufficiently large game size $n$, the matrix $E[\phi_i(S) \cdot \phi_i(S)^T|X_i,Z_{ik}]$ is invertible, i.e.,

$$\liminf_{n \to \infty} \det(E[\phi_i(S) \cdot \phi_i(S)^T|X_i,Z_{ik}]) > 0. \quad (6)$$

Assumption 8 is testable and similar to conditions imposed in Bajari et al. (2010) and Xu (2015), then by simple algebra, we have

$$\theta_k = \{E[\phi_i(S) \cdot \phi_i(S)^T|X_i,Z_{ik}]\}^{-1}E[\phi_i(S) \cdot V_i(X_i,Z_{ik},S_{-ik})|X_i,Z_{ik}]. \quad (7)$$

Consequently we can identify $\theta_k$ for all $k \in A$.

**Theorem 2.** Under Assumptions 1-7, the structural functions $\theta$ are nonparametrically identified in a large game.

**Proof.** The proof follows directly from the discussion above and is hence omitted. \qed

To summarize, our identification strategy is to first identify the deterministic payoff function for any given player using any given action over a positive Lebesgue measure set in the space of public information. Then, by imposing some properties on the support of payoff functions as well as equilibrium beliefs, a closed form expression for $\theta$ can be derived.
4 Estimation

In this section, we discuss the estimation of the structural parameters of our model in a nonparametric setting. The proposed estimation method is a two-step method and the estimator is shown to be uniformly consistent. Moreover, under a semiparametric setting, we prove that although the first stage estimator converges at a speed lower than the parametric root-n rate, the convergence speed of the second stage estimator will not be affected and is asymptotically normal. We believe that the semiparametric inference can help the applied researchers to get a better understanding of the estimation procedure and perform empirical analysis.

4.1 Nonparametric estimation

The estimation method consists of two steps: the first step is to nonparametrically estimate the equilibrium beliefs \( \{\sigma_{ik}(S;\theta)\}_{i,N}, k,A \), which can be done using standard nonparametric technique. Since the identification of \( \theta \) requires at least one element of \( S \) to be continuously distributed, we use the kernel smoothing method and focus on the case that all components of \( S \) are continuously distributed for the purpose of notational simplicity. As in Leung (2015), the symmetric equilibrium selection mechanism alleviates the curse of dimensionality problem caused by the large dimension of \( S \) and enable us to obtain the estimates with only a single network observation. The intuition is that players with same characteristics can be treated as repeated observations of a single player.

Because of the symmetric equilibrium selection mechanism, we can write \( \sigma_{ik}(S;\theta) \) as \( \rho_{nk}(S_i,S_{-i}) \), where \( S_{-i} = S \setminus S_i \) and \( \rho_{nk}(S_i,S_{-i}) \) is a function that is invariant to permutations of the component \( S_j \) of \( S_{-i} \). In order to facilitate derivation of asymptotic result we consider the following well-known class of smooth function\(^2\): for \( 0 < \alpha < \infty \), let \( C^\alpha_M(X) \) denote the class of functions \( f : X \rightarrow \mathbb{R} \) with \( \|f\|_\alpha \leq M \), where for any \( m \)-dimensional vector of non-negative integers \( k = (k_1,k_2,\cdots,k_m) \):

\[
\|f\|_\alpha \equiv \max \sup_{|k| \leq \alpha} |D^k f(x)| + \max \sup_{|k|=\alpha} \frac{|D^k f(x) - D^k f(y)|}{\|x-y\|^\alpha},
\]

where \( |k| \equiv \sum_{i=1}^m k_i \), \( \alpha \) denotes the greatest integer smaller than \( \alpha \) and \( D^k \) is the differential operator

\[
D^k \equiv \frac{\partial^{\left|k\right|}}{\partial x_1^{k_1} \cdots \partial x_m^{k_m}}.
\]

We use \( \{\hat{\sigma}_{ik}(S)\}_{i,N}, k,A \) to denote the nonparametric estimator for \( \{\sigma_{ik}(S;\theta)\}_{i,N}, k,A \) and let \( \hat{\phi}_i(S) = \sum_{j \in N \setminus \{i\}} \hat{\sigma}_{ij}(S;\theta) \cdot \gamma(H_{ij}) \) and \( \hat{\phi}_i(S) = (1,\hat{\phi}_{i1}(S),\cdots,\hat{\phi}_{iK}(S))^T \). The

---

\(^2\)See, e.g., Van der Vaart & Wellner (1996) Section 2.7.1.
nonparametric estimator will have the following form:

\[
\hat{\phi}_{ik}(S) = \sum_{j \neq i} \left[ \frac{\sum_{j=1}^{n} 1(Y_j = k)K(S_j - S_i)}{\sum_{j=1}^{n} K(S_j - S_i)} \right] \gamma(H_{ij}),
\]

(8)

where \(K(\cdot)\) is a high order product kernel function and \(h_1 = \prod_{r=1}^{d+q(K+1)} h_{1r}\). The first stage estimator can be viewed as a weighted U-statistics and under the following conditions, we show that this first stage estimator is consistent.

**Theorem 3.** Under the following conditions, \(\hat{\phi}_{ik}(S) - \phi_{ik}(S; \theta) = o_p(1)\) for all \(i \in N\) and \(k \in A\).

(a) Assumptions 2 and 3 hold,
(b) \(\rho_{nk}(S_i, S_j) \in C^M_s(S)\),
(c) \(\{S_i : i \in N\}\) is independent and identically distributed with a \(\nu\)-times differentiable density \(f(\cdot)\) bounded away from zero,
(d) \(K(\cdot) : \mathbb{R}^{d+q(K+1)} \rightarrow [0, 1]\) is a \(\nu\)th order product kernel function,
(e) As \(n \rightarrow \infty\), \(\max_{1 \leq r \leq d+q(K+1)} h_{1r} \rightarrow 0\) and \(nh_1 \rightarrow \infty\).

**Proof.** See Appendix B.

Since under current assumptions, we can not identify the distribution of the private information \(\epsilon\), traditional estimation method like maximum likelihood estimation cannot be used. Instead in the second step we can proceed to use a smoothed version of the pairwise maximum score method in Fox (2007) to estimate the model, which does not require one to know the distribution of \(\epsilon\). Specifically

\[
\hat{\theta} \in \arg\max_{\theta \in \Theta} Q_n(\theta, \hat{\phi}, h_2) \equiv \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k} G \left( \frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2} \right),
\]

(9)

where \(G(\cdot)\) is a differentiable function on \(\mathbb{R}\) satisfying the following conditions:

\(G1\). \(|G(v)| < M\) for some finite \(M\) and all \(v \in \mathbb{R}\),

\(G2\). \(\lim_{v \rightarrow -\infty} G(v) = 0\) and \(\lim_{v \rightarrow \infty} G(v) = 1\),

\(G3\). \(G(\cdot)\) is Lipschitz continuous, i.e., \(|G(v) - G(w)| \leq c \cdot |v - w|\) for all \(v, w \in \mathbb{R}\) and some \(c \geq 0\),

\(G4\). \(G'(\cdot)\) is a \(\nu\)th order kernel function (\(\nu \geq 2\)).
As pointed out in Horowitz (1992), here $G(\cdot)$ is analogous to a cumulative distribution function. Note that $h_2$ is the smoothing parameter satisfying $\lim_{n \to \infty} h_2 = 0$ and $\lim_{n \to \infty} n \cdot h_2 = \infty$. To ensure consistency of the estimator, we need to impose the following assumptions:

**Assumption 9.** The collection of the subgraphs of all $\theta \in \Theta$ forms a Vapnik-Chervonenkis (VC) class.

Assumption 9 is a fairly weak technical condition on the space of $\theta$, intuitively it requires that the number of distinct subsets of the space of $\theta$ does not grow "too fast". For a formal definition and examples of VC class, see Van der Vaart & Wellner (1996). Note that this assumption will be automatically satisfied if $\Theta$ is finite dimensional, i.e., under the parametric setting.

**Assumption 10.** There exists a metric $\| \cdot \|_\Theta$ such that (i) $\Theta$ is compact with respect to $\| \cdot \|_\Theta$; (ii) $\theta_n \in \Theta$ converges to $\theta$ uniformly if $\| \theta_n - \theta \|_\Theta \to 0$.

Assumption 10 is commonly assumed in the nonparametric and semiparametric econometrics literature (see, e.g., Gallant & Nychka (1987), Matzkin (1993) and Ai & Chen (2003)). It restricts the space of structural parameters as well as the choice of the norm $\| \cdot \|_\Theta$. As pointed out in Ai & Chen (2003), it will be satisfied if the infinite dimensional space $\Theta$ consists of bounded and smooth functions. Therefore without loss of generality we also impose the following assumption:

**Assumption 11.** There exists some $C < \infty$ such that $\| \theta \|_\Theta < C$ for all $\theta \in \Theta$.

Let

$$Q(\theta, \phi) \equiv E \left[ \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k} 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right]$$

be the probability limit of $Q_n(\theta, \hat{\phi}, h_2)$, in order to establish consistency we need to first introduce several auxiliary lemmas.

**Lemma 1.** $Q_n(\theta, \hat{\phi}, h_2)$ converges to $Q(\theta, \phi)$ uniformly with probability approaching 1.

*Proof.** See Appendix B. 

**Lemma 2.** $Q(\theta, \phi)$ is continuous in $\theta \in \Theta$.

*Proof.** See Appendix B.

**Lemma 3.** $Q(\theta, \phi)$ is uniquely maximized at $\theta^* \in \Theta$, which is the true value of the parameters.
Proof. See Appendix B.

By using Lemma 1-3, the next theorem establishes the uniform consistency of our proposed estimator:

**Theorem 4.** Given Assumption 1-8 and 9-11, \( \hat{\theta} \) is uniformly consistent for \( \theta^* \), i.e., \( \| \hat{\theta} - \theta^* \|_{\Theta} = o_p(1) \).

**Proof.** By Lemma 1-3 and Assumption 10, conditions (i)-(iv) of Theorem 2.1 in Newey & McFadden (1994) are satisfied, then immediately we can get \( \| \hat{\theta} - \theta^* \|_{\Theta} = o_p(1) \).

### 4.2 Semiparametric estimation and inference

In this subsection we restrict the space of the structural parameters to be finite dimensional space \( \Theta \subseteq \mathbb{R}^{d+q} \) and discuss about the semiparametric estimation and inference of our model. Specifically we focus on the case where \( K = 2 \) and let \((X_i, Z_{ik}) \equiv S_{ik} \) and \( \alpha_k(X_i, Z_{ik}) = S_{ik}^T \alpha_k, \beta_k(l, X_i, Z_{ik}) = \beta_{kl} \) for all \( i \in N \) and \( k \in A \). Without loss of generality we normalize the payoff of action 0 to be 0, i.e., \( U_{i0} = 0 \) for all \( i \in N \). Note that identification also requires \( \beta_{k0} = 0 \), then the objective function becomes

\[
Q_n(\theta, \hat{\phi}_1, h_2) = \frac{1}{n} \sum_{i=1}^n [2 \cdot 1(Y_i = 1) - 1] G \left( \frac{S_{i1}^T \alpha_1 + \beta_1 \sum_{j \neq i} \hat{\phi}_j(S; \theta)\gamma(H_{ij})}{h_2} \right) = \frac{1}{n} \sum_{i=1}^n [2 \cdot 1(Y_i = 1) - 1] G \left( \frac{S_{i1}^T \alpha_1 + \beta_1 \hat{\phi}_1}{h_2} \right) = \frac{1}{n} \sum_{i=1}^n [2 \cdot 1(Y_i = 1) - 1] G \left( \frac{w_{i1}^T \theta}{h_2} \right),
\]

where \( w_1 = (S_{i1}^T, \hat{\phi}_1)^T \) and we can see that the objective function has a similar form as in Horowitz (1992). In order to characterize the asymptotic distribution of \( \hat{\theta} \) we first introduce some additional notations: write \( S_1 = (S_{11}, S_{1i})^T, \tilde{w}_1 = (S_{i1}^T, \phi_1)^T, \alpha_1 = (\alpha_{11}, \hat{\alpha}_1^T)^T, \tilde{\theta} = (\tilde{\alpha}_1^T, \beta_1^T)^T \) and define

\[
B_n(\theta, \hat{\phi}_1, h_2) = \frac{\partial Q_n(\theta, \hat{\phi}_1, h_2)}{\partial \theta}
\]

and

\[
H_n(\theta, \hat{\phi}_1, h_2) = \frac{\partial^2 Q_n(\theta, \hat{\phi}_1, h_2)}{\partial \theta \partial \theta^T}.
\]

Let \( p(w_1^T \theta | S) \) denote the conditional density of \( w_1^T \theta \) on \( \theta \), which is positive everywhere with respect to the Lebesgue measure by Assumption 12 (a) and (c) imposed below. For each positive integer \( t \), define

\[
p^{(t)}(w_1^T \theta | S) = \frac{\partial^t p(w_1^T \theta | S)}{\partial (w_1^T \theta)^t}.
\]
whenever the derivative exists, and define \( p^{(0)}(w_T^T \theta | S) = p(w_T^T \theta | S) \). Let \( F(\cdot | w_T^T \theta, S) \) denote the cumulative distribution function of \( \epsilon \) on \( w_T^T \theta \) and \( S \). For each positive integer \( t \), define

\[
F^{(t)}(-w_T^T \theta | w_T^T \theta, S) = \frac{\partial^t F(-w_T^T \theta | w_T^T \theta, S)}{\partial (w_T^T \theta)^t}.
\] (11)

For each \( \nu \geq 2 \), define the \( (d+q) \times 1 \) vector \( B \) and the \( (d+q) \times (d+q) \) matrices \( D \) and \( H \) by

\[
B = -2 \int_{-\infty}^{\infty} u^r G'(u) du \sum_{t=1}^{\nu} \left\{ \left[ t!(\nu-t)! \right]^{-1} E \left[ F^{(t)}(0|0,S)p^{(\nu-t)}(0|S)|\tilde{w} \right] \right\},
\]

\[
D = \int_{-\infty}^{\infty} [G'(u)]^2 du E \left[ \tilde{w}_1 \tilde{w}_1^T p(0|S) \right],
\]

\[
H = 2E \left[ \tilde{w}_1 \tilde{w}_1^T F^{(1)}(0|0,S)p(0|S) \right].
\]

It is worth mentioning that when deriving the asymptotic distribution of our semiparametric maximum score estimator, \( D \) and \( H \) have roles that are analogous to the outer product and Hessian forms off the information matrix in maximum likelihood estimation. The regularity conditions imposed for the asymptotic distribution result are stated as follows.

**Assumption 12.** (a) \( |\alpha_{11}| = 1 \) and \( \tilde{\theta} \) is contained in a compact subset \( \hat{\Theta} \) of \( \mathbb{R}^{d+q} \); (b) \( \text{Median}(\epsilon|S) = 0 \); (c) the support of the distribution of \( w \) is contained in any proper linear subspace of \( \mathbb{R}^{d+q+1} \); (d) \( \Pr(Y = 1|S) \in (0,1) \) for almost every \( S \); (e) the distribution of \( S_{11} \) conditional on \( S \) has everywhere positive density with respect to the Lebesgue measure; (f) \( \lim_{n \to \infty} \log n/(nh_2^2) = 0 \); (g) the component of \( \tilde{w}_1 \), \( \tilde{w}_1 \tilde{w}_1^T \) and \( \tilde{w}_1 \tilde{w}_1^T \tilde{w}_1 \tilde{w}_1^T \) have finite third absolute moments; (h) There exists some \( M < \infty \) such that for all \( t \leq \nu \), all \( w_T^T \theta \) in a neighborhood of 0 and almost every \( S \), \( p^{(t)}(w_T^T \theta | S) \) and \( F^{(t)}(-w_T^T \theta | w_T^T \theta, S) \) exist and are continuous functions of \( w_T^T \theta \) satisfying \( |p^{(t)}(w_T^T \theta | S)| < M \) and \( |F^{(t)}(-w_T^T \theta | w_T^T \theta, S)| < M \). In addition, \( |p(w_T^T \theta | S)| < M \) for all \( w_T^T \theta \) and almost every \( S \). (i) The support of \( \tilde{S}_1 \) is bounded; (j) \( H \) is negative definite; (k) \( \tilde{\theta} \) is an interior point of \( \hat{\Theta} \).

Assumption 12 (a)-(e) are used to establish the rank ordering property under a semi parametric setting and are standard in the maximum score estimation literature; see e.g., Manski (1985), Horowitz (1992) and Chen et al. (2014). Assumption 12 (f) is analogous to an under-smooth condition assumptions made in kernel density estimation. Assumption 12 (g) and (h) ensure the existence of \( B, D \) and \( H \) as well as the convergence of certain sequences of integrals when deriving the asymptotic normality, see Horowitz (1992) for details. Assumption 12 (i)-(k) are standard in asymptotic distribution theory.

**Assumption 13.** If \( \hat{\phi}_1 - \phi_1 = O_p(r_n) \), where \( r_n \) is a nonstochastic positive real sequence, then \( r_n = o(1/\sqrt{n\tilde{h}_2}) \).
Assumption 13 requires that the first-stage nonparametric estimator $\hat{\phi}_1$ converges to $\phi_1$ faster than $1/\sqrt{nh_2}$. Note that $r_n$ will be determined by the dimension of the continuous part of $S_i$ and under the current semi-parametric setting, only one element of $S_i$ is required to be continuous, therefore this assumption is not restrictive. Then by applying Taylor expansion and modify the results in Horowitz (1992) we have the following theorem:

**Theorem 5.** Suppose Assumption 12 and 13 hold, let $\lambda < \infty$ be the limit of $nh_2^{2\nu+1}$ as $n \to \infty$, then

$$\sqrt{nh_2}(\hat{\theta} - \theta) \xrightarrow{d} N(-\sqrt{\lambda}B, H^{-1}DH^{-1}).$$

(12)

**Proof.** See Appendix C.

Note that the proof does not trivially follow from Horowitz (1992) because under our setting $\{Y_i\}_{i \in N}$ is not an independent random sequence. But since $\{Y_i\}_{i \in N}$ is independent conditional on $S$, our strategy is to first derive the conditional asymptotic distribution of $\hat{\theta}$ and then prove that unconditionally it will converge to the same distribution.

To apply the result in Theorem 5, it is necessary to consistently estimate $A$, $D$ and $H$. Let $\hat{\theta}$ be a consistent estimator of $\theta$ based on $h_2 \propto n^{-1/(2\nu+1)}$ and by using Theorem 3 in Horowitz (1992), we are able to get consistent estimators for $A$, $D$ and $H$ as

$$\hat{B} = h_2^{\ast - \nu}B_n(\hat{\theta}, \hat{\phi}_1, h_2^\ast),$$

(13)

$$\hat{D} = \frac{h_2}{n} \sum_{i=1}^{n} b_n(\hat{\theta}, \hat{\phi}_1, h_2) b_n(\hat{\theta}, \hat{\phi}_1, h_2)^T,$$

(14)

$$\hat{H} = H_n(\hat{\theta}, \hat{\phi}_1, h_2),$$

(15)

where $h_2^\ast \propto n^{-\delta/(2\nu+1)}$ for some $\delta \in (0, 1)$ and

$$b_n(\hat{\theta}, \hat{\phi}_1, h_2) = [2 \cdot 1(Y_i = 1) - 1] \left( \frac{\hat{w}_{i1}}{h_2} \right) G' \left( \frac{w_{i1}^T\theta}{h_2} \right).$$

(16)

It is generally acknowledged that $\hat{\theta}$ can be quite sensitive to the choice of the bandwidth $h_2$. In practice, the optimal bandwidth is chosen to minimize the mean square error of $\hat{\theta}$ and is selected by a plug-in method proposed in Horowitz (1992): Given $\nu$, choose any $h_2 \propto n^{-1/(2\nu+1)}$ and any $h_2^{\ast} \propto n^{-\delta/(2\nu+1)}$ for $0 < \delta < 1$. Obtain the smoothed maximum score estimator $\hat{\theta}$ based on $h_2$, and use $\hat{\theta}$ and $h_2^{\ast}$ to compute $\hat{B}$, $\hat{D}$ and $\hat{H}$. Then compute

---

3In the kernel estimation, the kernel function $K(\cdot)$ will be replaced by a indicator function for discrete variable and the rate of convergence for the mixed variables is the same as the case involving only the subset of continuous variables, see Li & Racine (2007) for details.
optimal $h_2$ by the following formula:

$$h_2 = \left[ \frac{\text{Tr}(\hat{H}^{-1}\hat{H}^{-1}\hat{D})}{2n\nu\hat{B}^T\hat{H}^{-1}\hat{H}^{-1}\hat{B}} \right]^{\frac{1}{2\nu+1}},$$

(17)
in which case

$$n^{\frac{\nu}{2\nu+1}}(\hat{\theta} - \theta) \overset{d}{\rightarrow} N\left(-\left(\frac{\text{Tr}(H^{-1}H^{-1}D)}{2\nu\hat{B}^T\hat{H}^{-1}\hat{H}^{-1}\hat{B}}\right)^{\frac{\nu}{2\nu+1}}H^{-1}\hat{B}, \left(\frac{\text{Tr}(H^{-1}H^{-1}D)}{2\nu\hat{B}^T\hat{H}^{-1}\hat{H}^{-1}\hat{B}}\right)^{\frac{1}{2\nu+1}}H^{-1}DH^{-1}\right).$$

5 An empirical application

In this section we use our proposed method to analyze the peer effects on youth smoking behavior. Recently, there is a growing body of empirical literature on studying the peer effects on adolescents smoking behavior, see e.g., Gaviria & Raphael (2001), Nakajima (2007), Soetevent & Kooreman (2007), Krauth (2007) and reference therein. The data we use is obtained from the National Longitudinal Study of Adolescent Health (Add Health), which is a database designed to study the relationship between the social environment and adolescents’ behavior. It contains a nationally representative sample of students in grades 7-12 from 80 high schools and 52 middles schools in the United States during the 1994-1995 school year. In the data every student is asked to complete a questionnaire to provide information about his or her socioeconomic characteristics as well as school-related behavior and friendship. The sample contains information on 90,118 students.4

Our empirical strategy is to treat each school in Add Health dataset as a unique social network, since different schools may achieve different (symmetric) equilibria, we estimate peer effects on a school-by-school case. All the respondents in our empirical analysis are selected from 7 largest schools with more than 800 observations each and the total number of observations $n = 6,342$. Following the literature, the covariates we choose include age, GPA, race information, gender and family background (whether mother has gone to college and father has a job). The missing observation in mother’s education has been treated as 0. We also include a dummy variable indicating whether the student has participated in any clubs, organizations or teams at school. The summary statistics for variables used in our empirical analysis are presented in Table 1.

It is well known that nonparametric kernel method suffers from the “curse of dimensionality”, i.e., its convergence rate is inversely related to the dimension of covariates involved and this problem will be even worse if covariates are discrete. Therefore in order to allevi-

---

4See the Add Health website (http://www.cpc.unc.edu/projects/addhealth) for a detailed description of surveys and data.
Table 1: Descriptive Statistics of Key Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>15.629</td>
<td>1.267</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>Female</td>
<td>0.487</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>GPA</td>
<td>2.960</td>
<td>0.500</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>White</td>
<td>0.753</td>
<td>0.432</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.116</td>
<td>0.320</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black</td>
<td>0.113</td>
<td>0.316</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>0.043</td>
<td>0.203</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mother college</td>
<td>0.470</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Father work</td>
<td>0.974</td>
<td>0.160</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>No club</td>
<td>0.150</td>
<td>0.357</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.382</td>
<td>0.486</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

To address the dimensionality problem, in the first stage estimation we use the smoothing method proposed in Hall et al. (2004), the first stage bandwidth $h$ is selected by the cross-validation method. In the second stage estimation, the objective function we use is similar to (10) and the second stage bandwidth $h_2$ is selected by the plug-in method proposed in section 4.2. The smoothing function $G(\cdot)$ is chosen to be the integral of a fourth-order kernel for nonparametric density estimation (Horowitz, 2002). The homophily effects are calculated by introducing a social distance function $\gamma(\cdot)$. Specifically let $H_{ij} = \|X_i - X_j\|$ be the Euclidean norm and use

$$\gamma(H_{ij}) = \frac{H_{ij}^{-1}}{\sum_{l \in N} H_{il}^{-1}}.$$  \hspace{1cm} (18)

It can be easily verified that (18) satisfies Assumption 1.

Our empirical results are presented in Table 2. The standard errors are computed using Theorem 5. Because the consistency of our estimator requires normalizing the coefficient of one continuous covariate to be 1 or -1, we normalize the coefficient of GPA to be equal to -1.\(^5\) For all 7 schools in the data, we find positive and statistically significant (at 1%) peer effects on smoking, means that smoking behavior from a student’s schoolmates will make that student more likely to consume cigarette. Gaviria & Raphael (2001), Nakajima (2007) and Soetevent & Kooreman (2007) use different datasets and find similar results. From our results, it is clear that age has positive effect on student’s smoking behavior, which is consistent with previous literature; see, e.g., Nakajima (2007) and Soetevent & Kooreman (2007). Father working for pay is negatively correlated with smoking, we believe this indicates (to some extent) that the student’s family income is negatively correlated with smoking.

\(^5\) The negative effect of GPA on smoking has been confirmed by many previous literature, see, e.g., Bisin
Table 2: Estimation Results (with homophily effect)

<table>
<thead>
<tr>
<th>Variable</th>
<th>School 1</th>
<th>School 2</th>
<th>School 3</th>
<th>School 4</th>
<th>School 5</th>
<th>School 6</th>
<th>School 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>2.588***</td>
<td>0.050**</td>
<td>0.753***</td>
<td>0.132*</td>
<td>2.352***</td>
<td>0.149**</td>
<td>0.060**</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.024)</td>
<td>(0.076)</td>
<td>(0.071)</td>
<td>(0.293)</td>
<td>(0.067)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.866***</td>
<td>0.068</td>
<td>-9.962***</td>
<td>-3.162*</td>
<td>-6.354***</td>
<td>2.034</td>
<td>-5.047***</td>
</tr>
<tr>
<td></td>
<td>(1.447)</td>
<td>(1.377)</td>
<td>(1.696)</td>
<td>(0.680)</td>
<td>(0.563)</td>
<td>(1.060)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>-32.393***</td>
<td>0.338</td>
<td>-7.383***</td>
<td>8.738***</td>
<td>-37.598***</td>
<td>-2.326***</td>
<td>1.801***</td>
</tr>
<tr>
<td></td>
<td>(1.756)</td>
<td>(1.268)</td>
<td>(1.696)</td>
<td>(0.623)</td>
<td>(4.916)</td>
<td>(0.396)</td>
<td>(0.363)</td>
</tr>
<tr>
<td>Black</td>
<td>-14.921***</td>
<td>0.341</td>
<td>-8.294***</td>
<td>-3.799***</td>
<td>-36.904***</td>
<td>12.274***</td>
<td>1.171***</td>
</tr>
<tr>
<td></td>
<td>(0.927)</td>
<td>(1.269)</td>
<td>(1.635)</td>
<td>(0.813)</td>
<td>(2.170)</td>
<td>(0.333)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>-23.795***</td>
<td>-4.850***</td>
<td>7.057***</td>
<td>-1.652</td>
<td>-10.406***</td>
<td>-2.185***</td>
<td>-80.839***</td>
</tr>
<tr>
<td></td>
<td>(1.815)</td>
<td>(1.417)</td>
<td>(1.801)</td>
<td>(1.238)</td>
<td>(0.545)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.411)</td>
<td>(1.417)</td>
<td>(0.396)</td>
<td>(5.165)</td>
<td>(0.416)</td>
<td>(0.935)</td>
</tr>
<tr>
<td>No club</td>
<td>2.528***</td>
<td>-0.039</td>
<td>4.797***</td>
<td>-4.142***</td>
<td>-0.214</td>
<td>-8.312***</td>
<td>-0.161</td>
</tr>
<tr>
<td></td>
<td>(0.679)</td>
<td>(0.226)</td>
<td>(0.000)</td>
<td>(0.283)</td>
<td>(2.217)</td>
<td>(0.099)</td>
<td></td>
</tr>
<tr>
<td>Mother college</td>
<td>-7.638***</td>
<td>-2.681***</td>
<td>-0.142</td>
<td>-0.897***</td>
<td>-5.961***</td>
<td>15.000***</td>
<td>2.220***</td>
</tr>
<tr>
<td></td>
<td>(1.130)</td>
<td>(0.128)</td>
<td>(0.364)</td>
<td>(0.796)</td>
<td>(2.432)</td>
<td>(0.287)</td>
<td></td>
</tr>
<tr>
<td>Father work</td>
<td>-1.052***</td>
<td>-5.268***</td>
<td>-10.296***</td>
<td>-10.914***</td>
<td>-36.661***</td>
<td>-19.213***</td>
<td>-4.116***</td>
</tr>
<tr>
<td></td>
<td>(0.458)</td>
<td>(0.437)</td>
<td>(1.413)</td>
<td>(0.623)</td>
<td>(4.815)</td>
<td>(3.837)</td>
<td>(0.946)</td>
</tr>
<tr>
<td>Peer effects</td>
<td>6.794***</td>
<td>5.191***</td>
<td>12.686***</td>
<td>3.086***</td>
<td>7.865***</td>
<td>9.152***</td>
<td>2.346***</td>
</tr>
<tr>
<td></td>
<td>(2.299)</td>
<td>(0.649)</td>
<td>(2.639)</td>
<td>(0.609)</td>
<td>(0.856)</td>
<td>(2.231)</td>
<td>(0.770)</td>
</tr>
<tr>
<td>Observations</td>
<td>805</td>
<td>818</td>
<td>846</td>
<td>973</td>
<td>855</td>
<td>1205</td>
<td>840</td>
</tr>
<tr>
<td>Standard errors</td>
<td>in parentheses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| * 10% significant, ** 5% significant, *** 1% significant. The coefficient of GPA is normalized to -1.

For the purpose of comparison, we also estimate the model without imposing homophily effects, i.e.,

$$\gamma(H_{ij}) = \frac{1}{n-1}. \quad (19)$$

Under this setup, our model incorporates a similar setting as a Manski-type linear-in-mean model. The results are listed in Table 3, we can see that the estimated peer effects become statistically insignificant among 6 of all 7 schools included. The only exception is school 5, from which we obtain a negatively significant peer effects. This comparison demonstrates the empirical importance of including homophily effects in our model.

6 Conclusion

This paper develops a structural model of strategic social interactions that emphasizes the impact of homophily effects on agents’ socioeconomic decisions. Our model assumes that individuals are affected by all players within the same social network (global interaction), but the strength of interactions decays as the social distance between players increases. Therefore our specification reflects the homophily principle in sociology: similarity breeds connection. By imposing a symmetric equilibrium selection mechanism, we allow the exis-
Table 3: Estimation Results (without homophily effect)

<table>
<thead>
<tr>
<th>Variable</th>
<th>School 1</th>
<th>School 2</th>
<th>School 3</th>
<th>School 4</th>
<th>School 5</th>
<th>School 6</th>
<th>School 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.096***</td>
<td>-0.556***</td>
<td>0.537***</td>
<td>0.516***</td>
<td>2.645***</td>
<td>0.007</td>
<td>0.063**</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.131)</td>
<td>(0.245)</td>
<td>(0.064)</td>
<td>(0.113)</td>
<td>(0.052)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.535</td>
<td>10.672</td>
<td>-10.690</td>
<td>-9.900***</td>
<td>-0.060</td>
<td>7.910</td>
<td>-6.611***</td>
</tr>
<tr>
<td></td>
<td>(1.200)</td>
<td>(116.879)</td>
<td>(31.353)</td>
<td>(2.691)</td>
<td>(2.407)</td>
<td>(252.989)</td>
<td>(0.520)</td>
</tr>
<tr>
<td>White</td>
<td>-2.445***</td>
<td>1.829***</td>
<td>-1.981***</td>
<td>-1.709***</td>
<td>-0.117</td>
<td>-26.113</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>(0.403)</td>
<td>(0.242)</td>
<td>(0.607)</td>
<td>(0.712)</td>
<td>(2.640)</td>
<td>(84.410)</td>
<td>(0.256)</td>
</tr>
<tr>
<td>Black</td>
<td>-5.058***</td>
<td>12.359</td>
<td>-2.905***</td>
<td>-9.762***</td>
<td>-0.363</td>
<td>8.670</td>
<td>-0.453*</td>
</tr>
<tr>
<td></td>
<td>(0.636)</td>
<td>(117.076)</td>
<td>(0.568)</td>
<td>(2.691)</td>
<td>(2.694)</td>
<td>(253.076)</td>
<td>(0.260)</td>
</tr>
<tr>
<td></td>
<td>(0.451)</td>
<td>(100.156)</td>
<td>(0.000)</td>
<td>(0.742)</td>
<td>(2.048)</td>
<td>(84.194)</td>
<td>(0.301)</td>
</tr>
<tr>
<td>Female</td>
<td>1.497***</td>
<td>8.646</td>
<td>-6.240***</td>
<td>3.351***</td>
<td>-5.865***</td>
<td>-0.054</td>
<td>-6.090***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(100.301)</td>
<td>(1.145)</td>
<td>(0.200)</td>
<td>(2.065)</td>
<td>(0.169)</td>
<td>(0.317)</td>
</tr>
<tr>
<td>No club</td>
<td>0.128</td>
<td>11.356</td>
<td>3.940***</td>
<td>-4.936***</td>
<td>-0.112</td>
<td>-35.546</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.336)</td>
<td>(117.114)</td>
<td>(0.351)</td>
<td>(0.308)</td>
<td>(0.467)</td>
<td>(168.603)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>Mother</td>
<td>-1.433***</td>
<td>0.182</td>
<td>-0.183</td>
<td>-47.370***</td>
<td>0.355</td>
<td>1.865***</td>
<td>5.066***</td>
</tr>
<tr>
<td>college</td>
<td></td>
<td>(1.708)</td>
<td>(16.615)</td>
<td>(0.160)</td>
<td>(0.000)</td>
<td>(0.429)</td>
<td>(0.214)</td>
</tr>
<tr>
<td>Father</td>
<td>-0.516***</td>
<td>-1.230***</td>
<td>-2.666**</td>
<td>-6.605***</td>
<td>-5.848***</td>
<td>-0.778***</td>
<td>-4.879***</td>
</tr>
<tr>
<td>work</td>
<td>(0.217)</td>
<td>(0.115)</td>
<td>(1.061)</td>
<td>(2.663)</td>
<td>(1.934)</td>
<td>(0.202)</td>
<td>(0.196)</td>
</tr>
<tr>
<td>Peer effects</td>
<td>0.832</td>
<td>-33.619</td>
<td>-3.694</td>
<td>-0.510</td>
<td>-62.560***</td>
<td>-10.762</td>
<td>1.055</td>
</tr>
<tr>
<td></td>
<td>(3.757)</td>
<td>(578.515)</td>
<td>(6.041)</td>
<td>(5.883)</td>
<td>(8.815)</td>
<td>(441.097)</td>
<td>(0.871)</td>
</tr>
<tr>
<td>Observations</td>
<td>805</td>
<td>818</td>
<td>846</td>
<td>973</td>
<td>855</td>
<td>1205</td>
<td>840</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* 10% significant, ** 5% significant, *** 1% significant.

The coefficient of GPA is normalized to -1.

tence of multiple equilibria across different networks and establish nonparametric identification of the model and propose a computationally feasible two-step estimation procedure that is robust to misspecification of distribution assumption and the presence of multiple equilibria. In the empirical application we use our method to analyze the peer effects on youth smoking using Add Health data and find strong empirical evidence of peer effects among adolescents within the same school. Furthermore by comparing the empirical results with and without specifying the homophily effect, our findings demonstrate the empirical importance of including homophily effect in our model.

The work presented in this paper indicates various possible extensions for future research. An example is to use different equilibrium solution concept that allows both local (i.e., the agent’s neighbors or friends) and global interactions between players. Horst & Scheinkman (2006) provides some examples of such equilibria. Another, perhaps more interesting issue, is to identify the social distance function $\gamma(\cdot)$. Here we assume that $\gamma(\cdot)$ is known, which can be viewed as a normalization assumption. Developing methods to identify and estimate $\gamma(\cdot)$ in our framework will be of both theoretical and empirical importance and calls for future work.
Appendix A
Equilibrium and Identification

Proof of Theorem 1:
This proof mainly relies on an application of Schauder Fixed Point Theorem and Arzelà-Ascoli Theorem. We use the same approach developed by Leung (2015) in proving Theorem 1 and we shall complete it in 5 steps.

Step 1. Let \( \Delta \) be the space of continuously differentiable functions that are matrix-valued with codomain being the set of \( n \times (K+1) \) matrices whose entries lie in \([0, 1]\). Endow \( \Delta \) with the norm \( \| \cdot \| \equiv \sup_{S \subseteq \mathbb{R}^n} |f(S)| \), in which \( S \subseteq \mathbb{R}^{d+q(K+1)} \) is a compact Hausdorff space, so that \( \Delta \) is a subset of the Banach space \( C(S, \| \cdot \|) \). Pick \( \Sigma \) as the subset consisting of such functions satisfying properties: (1) They are symmetric everywhere on \( S \); (2) They are everywhere continuously differentiable; (3) They are equicontinuous; (4) They have columns that have a sum of one.

Step 2. It's obvious that for any \( \sigma, \sigma' \in \Sigma \) and \( \chi \in [0, 1] \), \( \chi \sigma(S) + (1-\chi) \sigma'(S) \) still meet the above (1)-(4) properties. Hence, we have \( \chi \sigma(S) + (1-\chi) \sigma'(S) \in \Sigma \), which confirms the convexity of \( \Sigma \).

Step 3. For any \( k \in A \), let \( U_{ik}(\sigma-i, S, \epsilon_i) \) be the payoff function at the true parameter \( \theta \). To emphasize the dependence of \( \Gamma \) on \( S \), we w.o.l.g. rewrite \( \Gamma(\theta; S) \) defined above as \( \Gamma(\theta; \sigma, S) \), for any \( \sigma \in \Sigma \). By making use of Assumption 4, we are led to

\[
\Gamma(\theta; \sigma, S) = \int_{\mathbb{R}} \left[ \prod_{h \neq k} F_{\epsilon_h}(\sigma-h(S)) - \prod_{h \neq k} F_{\epsilon_h}(\sigma-h(S)) \right] f_{\epsilon_i}(\sigma, S) \, d\epsilon_i.
\]

Hence, we can claim that \( \Gamma(\theta; \cdot, S) \) maps symmetric functions to symmetric functions. Also, notice from (5) that \( \Gamma(\theta; \cdot, S) \) is continuous in \( S \) and \( \sigma \) and also preserves equicontinuity. Also, by using the definition of \( \Gamma(\theta; \cdot, S) \), it is straightforward to show that the columns have a sum of one.

Step 4. We show that \( \Gamma(\theta; S) \) is a subset of a compact space. And we just need to show that \( \Sigma \) is compact. Noting that \( S \) is compact, \( \Sigma \) is uniformly bounded, and \( \Sigma \) is equicontinuous by our construction, we apply the Arzelà-Ascoli Theorem to obtain that \( \Sigma \) is also compact. Hence, it suffices to show that \( \Sigma \) is relatively compact. It is relatively compact. Hence, it suffices to show that \( \Sigma \) is also closed, and we prove it by using Assumption 2 and 5 assures that \( \Gamma(\theta; \cdot, S) \) is continuous in \( S \) and \( \sigma \) and also preserves equicontinuity. Also, by using the definition of \( \Gamma(\theta; \cdot, S) \), it is straightforward to show that the columns have a sum of one.

Step 5. Step 2. For any \( k \in A \), let \( U_{ik}(\sigma-i, S, \epsilon_i) \) be the payoff function at the true parameter \( \theta \). To emphasize the dependence of \( \Gamma \) on \( S \), we w.o.l.g. rewrite \( \Gamma(\theta; \sigma, S) \), for any \( \sigma \in \Sigma \). By making use of Assumption 4, we are led to

\[
\Gamma(\theta; \sigma, S) = \int_{\mathbb{R}} \left[ \prod_{h \neq k} F_{\epsilon_h}(\sigma-h(S)) - \prod_{h \neq k} F_{\epsilon_h}(\sigma-h(S)) \right] f_{\epsilon_i}(\sigma, S) \, d\epsilon_i,
\]

where \( \epsilon \) is a random variable with distribution \( F_{\epsilon} \), and \( \sigma \) is a parameter. Hence, we have \( \sigma(S) + (1-\sigma(S)) \in \Sigma \), which confirms the convexity of \( \Sigma \).
Hence, the desired contradiction. Letting \( \{\sigma^m\}_m \) be a sequence in \( \Sigma \) that converges to a limit written as \( \sigma^* \). Suppose that \( \sigma^* \notin \Sigma \). Then, for any \( i \in \mathbb{N} \) and \( k \in \mathcal{A} \), we should have \( \sigma^*_{ik}(S;\theta) \neq \sigma_{\pi(i)k}^{m}(\pi(S);\theta) \). However, \( \sigma^m \in \Sigma \) for any \( m \) implies that \( \sigma^m_{ik}(S;\theta) = \sigma_{\pi(i)k}^{m}(\pi(S);\theta) \) for any \( m \), hence leading to \( \sigma^*_{ik}(S;\theta) = \sigma_{\pi(i)k}^*(\pi(S);\theta) \) by continuity. This, as a result, establishes the desired contradiction.

Step 5. A canonical application of the Schauder Fixed Point Theorem completes the proof.

Proof of Theorem 1: The proof is a modification of the argument in Matzkin (1993), let \( V_i(X_i, Z_{ik}, S) \) and \( V'_i(X_i, Z_{ik}, S) \) be that \( V_i(X_i, Z_{ik}, S) \neq V'_i(X_i, Z_{ik}, S) \), by Assumption 7, \( \exists l \in \mathcal{A} \) and \( Z_{il} \) process an everywhere positive Lebesgue density conditional on \( S \setminus \{Z_{il}\} \) and \( V_i \) and \( V'_i \) are strictly increasing with respect to \( Z_{il} \). Then by Assumption 6 and the argument in Matzkin (1993), there exist a set \( \mathcal{S} \subset \mathcal{S} \) with positive Lebesgue measure such that \( \forall S \in \mathcal{S} \), either

\[
V_i(X_i, Z_{ik}, S) > V_i(X_i, Z_{il}, S) \text{ and } V'_i(X_i, Z_{ik}, S) < V'_i(X_i, Z_{il}, S)
\]

or

\[
V_i(X_i, Z_{ik}, S) < V_i(X_i, Z_{il}, S) \text{ and } V'_i(X_i, Z_{ik}, S) > V'_i(X_i, Z_{il}, S).
\]

Suppose without loss of generality that the first case holds, by Assumption 2, \( F_{\epsilon_{ih}|S}(\cdot) \) is strictly increasing, then we can get

\[
F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{il}, S)) > F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{il}, S) - V_i(X_i, Z_{ik}, S))
\]

and

\[
F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S)) > F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ih}, S) - V_i(X_i, Z_{ik}, S)).
\]

Hence,

\[
\sigma_{ik}(S;\theta) - \sigma_{il}(S;\theta) = \int_{\epsilon \in \mathbb{R}} \left[ \prod_{h \neq k} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S)) \right] f_{\epsilon_{ih}|S}(\epsilon) d\epsilon
\]

\[
- \int_{\epsilon \in \mathbb{R}} \left[ \prod_{h \neq l} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{il}, S) - V_i(X_i, Z_{ih}, S)) \right] f_{\epsilon_{ih}|S}(\epsilon) d\epsilon
\]

> 0.
Therefore

\[ V_i(X_i, Z_{ik}, S) > V_i(X_i, Z_{il}, S) \implies \sigma_{ik}(S; \theta) > \sigma_{il}(S; \theta) \]

Similarly we can prove that

\[ V'_i(X_i, Z_{ik}, S) < V'_i(X_i, Z_{il}, S) \implies \sigma_{ik}(S; \theta') < \sigma_{il}(S; \theta') \]

So for all \( S \in \mathcal{S} \) either

\[ \sigma_{ik}(S; \theta) \neq \sigma_{ik}(S; \theta') \]

or

\[ \sigma_{il}(S; \theta) \neq \sigma_{il}(S; \theta') \]

Thus we have identified \( V_{ik}(X_i, Z_{ik}, S) \) for all \( k \in A \) and \( i \in N \). \( \square \)

**Appendix B  Nonparametric estimation**

**Proof of Theorem 3:**

\[
\hat{\phi}_{ik}(S) - \phi_{ik}(S) = \sum_{j \neq i}^{n} \hat{\sigma}_{jk}(S)\gamma(H_{ij}) - \sum_{j \neq i}^{n} \sigma_{jk}(S)\gamma(H_{ij}) = \sum_{j \neq i}^{n} [\hat{\sigma}_{jk}(S) - \sigma_{jk}(S)]\gamma(H_{ij})
\]

\[
= \sum_{j \neq i}^{n} \left\{ \frac{\sum_{l=1}^{n} 1(Y_l = k)K\left( \frac{S_l - S_{ij}}{h_1} \right)}{\sum_{l=1}^{n} 1(\gamma(H_{ij}))} - E[1(Y_j = k)|S] \right\} \gamma(H_{ij})
\]

\[
= \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \left\{ \frac{1(Y_l = k) - E[1(Y_l = k)|S]}{f(S_l)} \right\} K\left( \frac{S_l - S_{ij}}{h_1} \right) \gamma(H_{ij})
\]

\[
= \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \left\{ \frac{1(Y_l = k) - E[1(Y_l = k)|S]}{f(S_l)} \right\} K\left( \frac{S_l - S_{ij}}{h_1} \right) \gamma(H_{ij}) +
\]

\[
\sum_{j \neq i}^{n} \left\{ \frac{1}{nh_1} \sum_{l=1}^{n} E[1(Y_l = k)|S]K\left( \frac{S_l - S_{ij}}{h_1} \right)}{f(S_l)} - E[1(Y_j = k)|S] \right\} \gamma(H_{ij}) + (s.o.)
\]

\[
\equiv A_{n1} + A_{n2} + (s.o.),
\]
where \((s.o.)\) denotes the terms of smaller order. We first show that
\[ A_{n1} = O_p(1/\sqrt{n}) \], note that by Law of Iterated Expectation
\[
\mathbb{E}(A_{n1}) = \mathbb{E}(\mathbb{E}(A_{n1}|S))
\]
\[
= \mathbb{E} \left[ \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \left\{ \frac{\mathbb{E}[1(Y_l = k)|S] - \mathbb{E}[1(Y_l = k)|S]}{f(S_l)} \right\} K \left( \frac{S_i - S_j}{h_1} \right) \gamma(H_{ij}) \right]
\]
\[ = 0. \]

since \(\sum_{j \neq i}^{n} \gamma(H_{ij}) = 1\). To simplify notation define
\[
K_{ij} = K \left( \frac{S_i - S_j}{h_1} \right)
\]  
(20)
and
\[
D_{lji} = \left\{ \frac{1(Y_l = k) - \mathbb{E}[1(Y_l = k)|S]}{f(S_l)} \right\} K_{ij} \gamma(H_{ij}).
\]  
(21)

Then we have
\[
\text{Var}(A_{n1}|S) = \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \sum_{m=1}^{n} \sum_{o \neq p}^{n} \text{Cov}[D_{lji}, D_{mop}|S]
\]
\[
= \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \text{Var}[D_{lji}|S] + \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \sum_{o \neq p}^{n} \text{Cov}[D_{lji}, D_{top}|S] +
\]
\[
\frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{m=1}^{n} \sum_{j \neq i, p}^{n} \text{Cov}[D_{lji}, D_{mjp}|S] + (s.o.)
\]
\[ \equiv B_{n1} + B_{n2} + B_{n3} + (s.o.). \]

Since \(\mathbb{E}(D_{lji}|S) = 0\),
\[
B_{n1} = \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \mathbb{E}(D_{lji}^2|S)
\]
\[ = \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \frac{\mathbb{E}[1(Y_l = k)|S] \cdot \{1 - \mathbb{E}[1(Y_l = k)|S]\}}{f^2(S_l)} K_{ij}^2 \gamma^2(H_{ij})
\]
\[ \leq \frac{1}{4n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \frac{K_{ij}^2 \gamma^2(H_{ij})}{f^2(S_l)}
\]
Since \( f(\cdot) \) is bounded away from zero, we know for some \( C < \infty \),
\[
\mathbb{E}(B_{n1}) = \mathbb{E}[\mathbb{E}(B_{n1}|S_i, S_j)]
\leq \frac{C}{4n^2h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \int \int \gamma^2(H_{ij}) \left( \int K^2_l f(S_l|S_i, S_j) dS_l \right) f(S_i)f(S_j) dS_idS_j
\]
\[
= \frac{C}{4n^2h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \int \int \gamma^2(H_{ij}) \left( \int K^2(v)f(S_j + h_1v) dv \right) f(S_i)f(S_j) dS_idS_j
\]
\[
= O \left( \frac{1}{nh_1} \right).
\]

Similarly,
\[
B_{n2} = \frac{1}{n^2h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \sum_{o \neq p}^{n} \mathbb{E}(D_{lj}D_{lo}|S)
\leq \frac{C}{4n^2h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \sum_{o \neq p}^{n} K_{lj}K_{lo}\gamma(H_{ij})\gamma(H_{po})
\]
\[
f(S_j)f(S_o)
\]
and then
\[
\mathbb{E}(B_{n2}) \leq \frac{C}{4n^2h_1^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{o \neq p}^{n} \int \int \int \int K_{lj}K_{lo}\gamma(H_{ij})\gamma(H_{po})f(S_l)f(S_j)f(S_o)f(S_p) dS_l dS_j dS_o dS_p
\]
\[
= O \left( \frac{1}{n} \right).
\]

Following a similar argument we can show that \( \mathbb{E}(B_{n3}) = O_p(1/n) \), hence by the law of total variance,
\[
Var(A_{n1}) = \mathbb{E}[Var(A_{n1})] + Var[\mathbb{E}(A_{n1})]
\]
\[
= \mathbb{E}(B_{n1}) + \mathbb{E}(B_{n2}) + \mathbb{E}(B_{n3})
\]
\[
= O \left( \frac{1}{nh_1} \right) + O \left( \frac{1}{n} \right) + O \left( \frac{1}{n} \right)
\]
\[
= O \left( \frac{1}{nh_1} \right).
\]

Hence \( A_{n1} = O_p(1/\sqrt{nh_1}) \).

By condition (b), we know that \( \rho_{nk}(S_i, S_{-i}) \) is \( s \)-times differentiable and the derivatives are uniformly bounded. Then in order to do multivariate Taylor expansion, we first introduce some multi-index notations: for \( \alpha \in \mathbb{N}^{d+q(K+1)} \) and \( S_i \in \mathbb{R}^{d+q(K+1)} \), define the \( \alpha \)-th
order derivative of \( \rho_{nk}(S_i, S_{-i}) \) at \( S_i \) as
\[
\rho_{nk}^{(\alpha)}(S_i) = \frac{\partial^{|\alpha|} \rho_{nk}}{\partial S_{i_1}^{\alpha_1} \partial S_{i_2}^{\alpha_2} \cdots \partial S_{i_{d+q(K+1)}}^{\alpha_{d+q(K+1)}}}, \quad |\alpha| \leq s
\]
where \( |\alpha| = \sum_{i=1}^{d+q(K+1)} \alpha_i \) and \( S_i \) denotes the \( j \)th component of \( S_i \). Also let \( \alpha! = \prod_{i=1}^{d+q(K+1)} \alpha_i! \) and \( S^\alpha = \prod_{i=1}^{d+q(K+1)} S_i^{\alpha_i} \), then by Taylor expansion of \( \sigma_{ik}(S; \theta) \) at \( S_j \),
\[
\sigma_{ik}(S; \theta) = \rho_{nk}(S_i, S_{-i}) = \rho_{nk}(S_j, S_{-i}) + \sum_{1 \leq |b| < s} \frac{\rho_{nk}^b(S_j)}{b!} (S_l - S_j)^b + \sum_{|b| = s} C_b(S_m)(S_l - S_j)^b
\]
where \( \lim_{S_m \to S_j} C_b(S_i) = 0 \). By symmetry, \( \rho_{nk}(S_j, S_{-i}) = \sigma_{jk}(S; \theta) \), hence
\[
A_{n2} = \sum_{j \neq i} \frac{1}{nh_1} \sum_{l=1}^{n} \frac{K \left( \frac{S_l - S_i}{h_1} \right) \gamma(H_{ij})}{\bar{f}(S_j)} \mathbb{E}[1(Y_j = k)|S] - \mathbb{E}[1(Y_j = k)|S] + \sum_{1 \leq |b| < s} \frac{B_b}{b!} |\mathbb{E}((S_l - S_j)^b)|
\]
where the term \( B_b \) is the upper bound for \( \rho_{nk}^b(S_i) \) and \( \max_{l \in \mathcal{N}} C_b(S_i) < B_b \). By standard argument of Taylor expansion and change of variables, we have
\[
A_{n2} = O_p \left( \frac{1}{\sqrt{nh_1}} + \sum_{r=1}^{d+q(K+1)} h_{1r}^v \right).
\]
Hence
\[
\hat{\phi}_{ik}(S) - \phi_{ik}(S) = O_p \left( \frac{1}{\sqrt{nh_1}} + \sum_{r=1}^{d+q(K+1)} h_{1r}^v \right) = o_p(1). \quad (22)
\]
Proof of Lemma 1: Let

$$Q_n^*(\theta, \phi) \equiv \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} 1(\phi_i^T \theta_k > \phi_i^T \theta_h).$$

Then by Triangle inequality,

$$\sup_{\theta \in \Theta} |Q_n(\theta, \hat{\phi}, h_2) - Q(\theta, \phi)|$$

$$\leq \sup_{\theta \in \Theta} |Q_n(\theta, \hat{\phi}, h_2) - Q_n(\theta, \phi, h_2)| + |Q_n(\theta, \phi, h_2) - Q_n^*(\theta, \phi)| + |Q_n^*(\theta, \phi) - Q(\theta, \phi)|$$

$$\leq \sup_{\theta \in \Theta} |Q_n(\theta, \hat{\phi}, h_2) - Q_n(\theta, \phi, h_2)| + \sup_{\theta \in \Theta} |Q_n(\theta, \phi, h_2) - Q_n^*(\theta, \phi)| + \sup_{\theta \in \Theta} |Q_n^*(\theta, \phi) - Q(\theta, \phi)|$$

$$\equiv A_{n1} + A_{n2} + A_{n3}.$$

Next we need to prove that $A_{ni} = o_p(1)$ for $i = 1, 2, 3$.

$$A_{n1} = \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} G \left( \frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2} \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) \right|$$

$$\leq \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k}^{K} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} 1(Y_i = k) \left[ G \left( \frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2} \right) - G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) \right] \right|$$

$$\leq \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k}^{K} \sup_{\theta \in \Theta} \left[ G \left( \frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2} \right) - G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) \right] \cdot \frac{1}{n}.$$

By condition $G3$,

$$\left| \frac{1}{n} \left[ G \left( \frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2} \right) - G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) \right] \right| \leq c \cdot \left| \frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{n h_2} - \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{n h_2} \right|$$

$$= c \cdot \left| \frac{(\hat{\phi}_i^T - \phi_i^T) \cdot (\theta_k - \theta_h)}{n h_2} \right| = o_p(1)$$

by Theorem 3, Assumption 11 and Cauchy-Schwarz Inequality. Thus we know $A_{n1} = o_p(1)$.
by Slutsky’s theorem.

\[ A_{n2} = \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k} 1(Y_i = k) \sum_{h \neq k} G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k} 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right| \]

\[ \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k} \sup_{\theta \in \Theta} \left| G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) - 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right| \]

\[ \equiv B_{n1}(a) + B_{n2}(a), \]

where

\[ B_{n1}(a) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k} \sup_{\theta \in \Theta} \left| G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) - 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right| \cdot 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| > a) \]

and

\[ B_{n2}(a) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k} \sup_{\theta \in \Theta} \left| G \left( \frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2} \right) - 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right| \cdot 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a). \]

Since \( \lim_{n \to \infty} h_2 = 0 \), conditions G1 and G2 imply that \( B_{n1}(a) \to 0 \) for each \( a > 0 \) as \( n \to \infty \). As for \( B_{n2}(a) \), since by condition G1, \( G(\cdot) \) is bounded by \( M \), we know

\[ B_{n2}(a) \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k} M \sup_{\theta \in \Theta} 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) = M \sum_{k=1}^{K} \sum_{h \neq k} \frac{1}{n} \sum_{i=1}^{n} \sup_{\theta \in \Theta} 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a). \]

By Lemma 2.6.17 and 2.6.18 in Van der Vaart and Wellner (1996), we know \( \{1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) : \theta \in \Theta \} \) is VC-subgraph given Assumption 9. Thus Glivenko-Cantelli Theorem (see, e.g., Theorem 2.4.3 in Van Der Vaart and Wellner (1996)) implies that

\[ \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) - \mathbb{E} \left[ 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) \right] \right| = o_p(1). \]  

(23)

Let \( r(\cdot) : \Theta \to \Theta \) be such that \( r(Z_d) = v_i(Z_d, S \setminus \{Z_d\}) \), then \( r^{-1}(\cdot) \) exists by Assumption 7, thus by Triangle Inequality and Law of Iterated Expectation,

\[ \mathbb{E} \left[ 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) \right] \leq \int_{s \in S} \left[ \int_{r^{-1}(a)} f_{Z_d | S}(z)dz \right] f_S(s)ds, \]  

(24)

where \( f_{Z_d | S}(\cdot) \) denotes the conditional density function of \( Z_d \) given \( S \) and \( f_S(\cdot) \) is the density function of \( S \). By Assumption 7, the integral in brackets of (24) is continuous, hence by
making \( a \) arbitrarily close to 0, it will converge to 0 uniformly over \( \theta \in \Theta \). Since it is also bounded by 1, using Lebesgue Dominated Convergence Theorem, we can immediately get 

\[
E \left[ 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) \right]
\]

converges to 0 uniformly over \( \theta \in \Theta \). Thus by (23),

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} 1(|\phi_i^T \theta_k - \phi_i^T \theta_h| \leq a) \right| = o_p(1).
\]

Again by Slutsky’s Theorem, \( B_n(\alpha) = o_p(1) \), so \( A_n = o_p(1) \) as well.

\[
A_{n3} = \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} 1(\phi_i^T \theta_k > \phi_i^T \theta_h) - \mathbb{E} \left[ \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right] \right|
\]

\[
\leq \sum_{k=1}^{K} \sum_{h \neq k}^{K} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} 1(Y_i = k) \cdot 1(\phi_i^T \theta_k > \phi_i^T \theta_h) - \mathbb{E} \left[ 1(Y_i = k) \cdot 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right] \right|
\]

\[
= o_p(1)
\]

by Glivenko-Cantelli Theorem. Consequently

\[
\sup_{\theta \in \Theta} |Q_n(\theta, \hat{\phi}, h_2) - Q(\theta, \phi)| = A_{n1} + A_{n2} + A_{n3} = o_p(1).
\]

\[\square\]

**Proof of Lemma 2:** By Law of Iterated Expectation and Assumption 7,

\[
Q(\theta, \phi) = \mathbb{E} \left[ \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right]
\]

\[
= \sum_{k=1}^{K} \sum_{h \neq k}^{K} \int_{s \in \mathcal{S}} \int_{r^{-1}(\phi_i^T \theta_h)}^{\infty} 1(Y_i = k) f_{Z_i|S}(z) dz \cdot f_S(s) ds.
\]

Since \( r(\cdot) : \mathcal{Z} \mapsto \Theta \) is continuous and strictly increasing by Assumption 5, \( r^{-1}(\cdot) : \Theta \mapsto \mathcal{Z} \) is also continuous. By Assumption 10, \( \Theta \) is compact with respect to \( \| \cdot \|_{\Theta} \), thus \( r^{-1}(\cdot) \) is uniformly continuous. Suppose there exists a sequence of functions \( \{\theta_{nh}\}_{n \in \mathbb{N}} \) in \( \Theta \) and \( \|\theta_{nh} - \theta_h\|_{\Theta} \to 0 \) as \( n \to \infty \), then by Assumption 10,

\[
\sup_{\theta \in \Theta} \|\theta_{nh} - \theta_h\|_{\Theta} = o(1).
\]

(25)
By definition of uniform continuity, ∃ δ > 0 such that if ∥θ_nh − θ_h∥Θ < δ, then

\[ |r^{-1}(\phi_i^T \theta_{nh}) - r^{-1}(\phi_i^T \theta_h)| < \epsilon \]  \hspace{1cm} (26)

for all ε > 0. By (25), ∃ n_0 ∈ ℤ such that for all n ≥ n_0,

\[ \sup_{\theta \in \Theta} ∥\theta_nh − \theta_h∥Θ < \delta, \]

then (26) will hold. Define r_n(Z_d) = φ_i^T θ_{nh}, then by Triangle Inequality and (26),

\[
\begin{align*}
\sup_{\theta \in \Theta} |r_n^{-1}(\phi_i^T \theta_{nh}) - r_n^{-1}(\phi_i^T \theta_h)| & \leq \sup_{\theta \in \Theta} |r_n^{-1}(\phi_i^T \theta_{nh}) - r_n^{-1}(\phi_i^T \theta_{nh})| + \sup_{\theta \in \Theta} |r_n^{-1}(\phi_i^T \theta_{nh}) - r_n^{-1}(\phi_i^T \theta_h)| \\
& \leq \sup_{\theta \in \Theta} |r_n^{-1}(\phi_i^T \theta_{nh}) - r_n^{-1}(\phi_i^T \theta_h)| \\
& < \epsilon.
\end{align*}
\]

Since ε > 0 is arbitrary, we conclude that r_n^{-1}(φ_i^T θ_{nh}) → r_n^{-1}(φ_i^T θ_h) uniformly over θ ∈ Θ. Hence by Lebesgue Dominated Convergence Theorem,

\[
\int_{r_n^{-1}(\phi_i^T \theta_{nh})}^{\infty} 1(Y_i = k)f_{Z_d|S(z)}dz - \int_{r_n^{-1}(\phi_i^T \theta_h)}^{\infty} 1(Y_i = k)f_{Z_d|S(z)}dz = o(1).
\]

Thus the integral in brackets of (9) is continuous in θ. Since this integral is also bounded by 1, again by Dominated Convergence Theorem, we conclude that Q(θ, φ) is continuous in θ.

Proof of Lemma 3: Using a similar argument as in the proof of Lemma 2, we know

\[
Pr(\phi_i^T \theta_k = \phi_i^T \theta_h) = \int_{S \in S} \left[ \int_{r_n^{-1}(\phi_i^T \theta_h)}^{\infty} f_{Z_d|S(z)}dz \right] f_S(s)ds = 0. \hspace{1cm} (27)
\]

Using Law of Iterated Expectation, rewrite Q(θ, φ) as

\[
Q(\theta, \phi) = \sum_{k=1}^{K} \sum_{h \neq k}^{K} E \left[ \sigma_{ik}(S; \theta)1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right]. \hspace{1cm} (28)
\]

By Theorem 1, we know θ* is identified, thus by the proof of Theorem 1, we can get

\[
\phi_i^T \theta_k^* > \phi_i^T \theta_h^* \iff \sigma_{ik}(S; \theta^*) > \sigma_{ih}(S; \theta^*)
\]

32
for any \((k, h) \in A \times A\) such that \(k \neq h\). Therefore \(Q(\theta, \phi)\) will be globally maximized by \(\theta^*\) since by (12) there is no tie in choice probability. Now we need to prove that \(Q(\theta, \phi)\) is uniquely maximized by \(\theta^*\), suppose by contradiction there exists another \(\theta' \in \Theta\) such that \(\theta' \neq \theta^*\) and \(\theta'\) maximizes \(Q(\theta, \phi)\), then by the proof of Theorem 1, we know there will exist some \((k, l) \in A \times A\) such that

\[
\phi_i^T \theta^*_k > \phi_i^T \theta^*_l \quad \text{and} \quad \phi_i^T \theta^*_k < \phi_i^T \theta^*_l
\]

or

\[
\phi_i^T \theta^*_l < \phi_i^T \theta^*_k \quad \text{and} \quad \phi_i^T \theta^*_k > \phi_i^T \theta^*_l.
\]

Therefore it is not possible for \(\theta'\) to maximize \(Q(\theta, \phi)\) as well, this is a contradiction, so we can conclude that \(Q(\theta, \phi)\) will be uniquely maximized by \(\theta^*\).

\[\Box\]

### Appendix C  Semiparametric estimation

**Proof of Theorem 5:** Under our setting, \(\{Y_i\}_{i \in N}\) is not an independent random sequence, therefore the results in Horowitz (1992) cannot be directly used. However since \(Y_i \perp Y_j\) conditional on \(S\) for all \(i \neq j\), we can instead derive the conditional asymptotic distribution of \(\hat{\theta}_1\) and show that asymptotically the conditional and unconditional distribution \(\hat{\theta}\) are equivalent.

Let \(\hat{\theta}\) be a smoothed maximum score estimator, then we know with probability approaching 1, \(B_n(\hat{\theta}, \hat{\phi}_1, h_2) = 0\), hence by Taylor expansion,

\[
B_n(\theta, \hat{\phi}_1, h_2) + H_n(\hat{\theta}, \hat{\phi}_1, h_2)(\hat{\theta} - \theta) = 0,
\]

where \(\hat{\theta}\) lies between \(\theta\) and \(\hat{\theta}\). Therefore

\[
\sqrt{nh_2} B_n(\theta, \hat{\phi}_1, h_2) + H_n(\theta, \hat{\phi}_1, h_2)\sqrt{nh_2}(\theta - \hat{\theta}) = 0.
\]

Then we have

\[
\sqrt{nh_2}(\theta - \hat{\theta}) = -H_n(\hat{\theta}, \hat{\phi}_1, h_2)^{-1} \sqrt{nh_2} B_n(\theta, \hat{\phi}_1, h_2)
\]

\[
= -H_n(\hat{\theta}, \hat{\phi}_1, h_2)^{-1} \sqrt{nh_2}\{B_n(\theta, \hat{\phi}_1, h_2) - \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S] + \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S]\}
\]

\[
\equiv -H_n(\hat{\theta}, \hat{\phi}_1, h_2)^{-1} \sqrt{nh_2}\{C_n + \mathbb{E}[B_n(\theta, \phi_1, h_2)|S]\}.
\]
We first show that $\sqrt{n} h_2 C_n = o_p(1)$, Let

$$b_i(\theta, \hat{\phi}_1, h_2) = [2 \cdot 1(Y_i = 1) - 1] \left( \frac{\tilde{w}_{i1}}{h_2} \right) C' \left( \frac{w_{i1}^T \theta}{h_2} \right),$$

(32)

then $\sqrt{n} h_2 B_n(\theta, \hat{\phi}_1, h_2) = \left( \sqrt{h_2/n} \right) \sum_{i=1}^n b_i(\theta, \hat{\phi}_1, h_2)$. By Law of Iterated Expectation

$$\mathbb{E}(\sqrt{n} h_2 C_n) = 0.$$  

(33)

$$\mathbb{E}(\sqrt{n} h_2 C_n)^2 = \mathbb{E}\{\mathbb{E}[\sqrt{n} h_2 C_n]^2 | S]\}
= \frac{h_2}{n} \mathbb{E}\left\{ \mathbb{E}\left[ \left( \sum_{i=1}^n b_i(\theta, \hat{\phi}_1, h_2) - \mathbb{E}(b_i(\theta, \hat{\phi}_1, h_2) | S) \right) \right] | S \right\}
= \frac{h_2}{n} \sum_{i=1}^n \mathbb{E}\left\{ \mathbb{E}\left[ \left( b_i(\theta, \hat{\phi}_1, h_2) - \mathbb{E}(b_i(\theta, \hat{\phi}_1, h_2) | S) \right) \right] | S \right\}
= o(1),$$

where the third equality is by conditional independence and the fact that $\mathbb{E}[b_i(\theta, \hat{\phi}_1, h_2) - \mathbb{E}(b_i(\theta, \hat{\phi}_1, h_2) | S)] | S = 0$ and the last equality is because $\text{Var}(b_i(\theta, \hat{\phi}_1, h_2))$ is bounded by Assumption 12. Hence

$$\sqrt{n} h_2 C_n = o_p(1).$$

(34)

By Law of Iterated Expectation and Mean Value Theorem,

$$\sqrt{n} h_2 \mathbb{E}\{\mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2) | S]\}
= \sqrt{n} h_2 \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)]
= \sqrt{n} h_2 \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)] + \sqrt{n} h_2 \frac{\partial \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)]}{\partial \hat{\phi}_1} (\hat{\phi}_1 - \phi_1),$$

where $\hat{\phi}_1$ is between $\hat{\phi}_1$ and $\phi_1$. By Assumption 13, $\sqrt{n} h_2 (\hat{\phi}_1 - \phi_1) = o_p(1)$ and $\frac{\partial \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)]}{\partial \hat{\phi}_1}$ is bounded by Condition G4 and Assumption 12, hence we know

$$\sqrt{n} h_2 \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)] = \sqrt{n} h_2 \mathbb{E}[B_n(\theta, \phi_1, h_2)] + o_p(1).$$

(35)

By Lemma 5 in Horowitz (1992),

$$\lim_{n \to \infty} \sqrt{n} h_2 \mathbb{E}[B_n(\theta, \phi_1, h_2)] = \sqrt{\lambda B}. 34$$
Therefore
\[
\lim_{n \to \infty} \sqrt{nh_2}E[B_n(\theta, \hat{\phi}_1, h_2)] = \sqrt{\lambda}B + o_p(1). \tag{36}
\]
Therefore by Lebesgue Dominated Convergence Theorem and Lemma 5 in Horowitz (1992),
\[
\lim_{n \to \infty} \text{Var}\{\sqrt{nh_2}E[B_n(\theta, \hat{\phi}_1, h_2)|S]\} = D + o_p(1). \tag{37}
\]
By (36) and (37) and apply Lindeberg-Feller’s Central Limit Theorem, we have
\[
D^{-\frac{1}{2}} \sqrt{nh_2}[E[B_n(\theta, \hat{\phi}_1, h_2)|S] - E(B_n(\theta, \hat{\phi}_1, h_2))] \overset{d}{\to} \mathcal{N}(0, I_{d+q}), \tag{38}
\]
where \(I_{d+q}\) is an identity matrix with dimension \(d + q\). Furthermore it is easy to verify that if \(\hat{\phi}_1 - \phi_1 = o_p(1)\), then
\[
H_n(\hat{\theta}, \hat{\phi}_1, h_2) - H_n(\hat{\theta}, \phi_1, h_2) = o_p(1)
\]
by condition G4. Hence by Law of Iterated Expectation and Lemma 8 and Lemma 9 in Horowitz (1992), the stochastic equicontinuity of \(H_n(\theta, \hat{\phi}_1, h_2)\) holds at \(\theta\) and then we know
\[
H_n(\hat{\theta}, \hat{\phi}_1, h_2) = H + o_p(1) \tag{39}
\]
since \(\hat{\theta}\) is a consistent estimator for \(\theta\). By Slutsky’s Theorem, (31), (38), (39) together imply that
\[
\sqrt{nh_2}(\hat{\theta} - \theta) \overset{d}{\to} \mathcal{N}(-\sqrt{\lambda}H^{-1}B, H^{-1}DH^{-1}).
\]

References


