ESSAYS ON MICROECONOMETRICS WITH APPLICATIONS TO SOCIAL NETWORK

A Dissertation

by

WENZHENG GAO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Co-Chair of Committee,	Yonghong An
Committee Members,	Zheng Fang
	Silvana Krasteva
	Ximing Wu
Head of Department,	Timothy Gronberg

May 2018

Major Subject: Economics

ABSTRACT

This dissertation includes three chapters on microeconometrics with applications to social network. In the first chapter, we study identification and estimation of peer effects in a game theoretical social interaction model with incomplete information. We show that players' equilibrium choice probabilities and peer effects can be identified in the presence of measurement errors in network connections by exploiting the nonparametric methodology developed for nonclassical measurement error models. Based on the identification methodology, a semiparametric estimation method is established and applied to study the peer effects on youth alcohol drinking behaviors using data of adolescents in the United States, our empirical findings show that peer effects will be significantly underestimated if measurement errors are ignored.

In the second chapter, 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.

In the third chapter, we study bandwidth selection method for the smoothed maximum score estimator. The smoothed maximum score estimator is a semiparametric estimator for binary response model, which is very useful for many economics and statistics applications. The method for selecting the smoothing parameter (bandwidth) in smoothed maximum score estimator is analogous to the plug-in method in kernel density estimation. It requires initial “pilot" values of the bandwidth to obtain the optimal bandwidth. The method has the disadvantage of not being fully data-driven. In this paper, we propose a data-driven bandwidth selection method by minimizing a cross-validated criterion function. Simulation results show that our proposed method performs better than existing methods.

DEDICATION

To Mom and Dad.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude as my doctoral study at TAMU is drawing to a close. First of all, I am deeply indebted to my advisors Qi Li and Yonghong An for their continuous guidance and advice that develop this dissertation. I can never finish my dissertation without the help and encouragement from them. I have also benefited a lot from the way they think and conduct research. I would say that they are definitely the best advisors anyone can ask for.

I would also like to thank my dissertation committee members Zheng Fang, Silvana Krasteva and Ximing Wu for extremely helpful comments and suggestions regarding my research. I am especially grateful for their help and advice when I was on the job market.

Gratitude also goes to all my friends and classmates at TAMU. Thank you all for the friendship and my life has been blessed for knowing all of you. I am especially grateful to Darong Dai and Naibao Zhao, for all the generous help and insightful discussions we have had together over the last five years.

Finally, I am eternally thankful to my parents, Mingtang Gao and Qing Liu. I can never be the person I am today without the unconditional love and support from them.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Qi Li (chair), Professor Yonghong An (co-chair), Professor Zheng Fang and Professor Silvana Krasteva of the Department of Economics, and Professor Ximing Wu of the Department of Agricultural Economics.

The data analyzed for Chapter 2 and 3 was conducted with restricted access to the National Longitudinal Study of Adolescent to Adult Health data. The analyses depicted in Chapter 3 and 4 were conducted in part by Darong Dai, Xirong Chen and Zheng Li. Portions of this dissertation were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by Department of Economics at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

This dissertation develops microeconometric methods to nonparametrically identify and estimate peer effects in social networks. In particular, the peer effect is modeled as strategic effect in an incomplete information game with large number of players.

In the first chapter, we study identification and estimation of peer effects in a game theoretical social interaction model with incomplete information. This paper is motivated by the fact that econometric analysis of networks has long suffered the issue of measurement errors, usually in the form of missing or spurious data in network connections. The presence of measurement errors mainly results from the sources of network data, which predominantly are surveys and questionnaires soliciting self-reports ([1]). Applied researchers typically construct network from data and naively treat this network as the true network of interest, ignoring the problem of measurement errors. The main objective of this essay is to develop an econometric framework to identify and estimate peer effects in the presence of measurement errors in network data. The identification proceeds in two steps. In the first step, we show that under semi-anonymously symmetric equilibrium, the CCPs can be nonparametrically identified when the number of players is fairly large. Specifically, we prove that the game theoretical model can be fitted into the measurement error models as proposed in [2] and [3] and then provide two different methods to identify the CCPs. The first (point) identification method is implemented by incorporating an instrumental variable and applying a spectral decomposition technique to the observed distributions in data and is also the method used in the empirical application. The second (partial) identification method is developed in the case when a valid instrumental variable is not available in the data and follows the direct misclassification proposed in [3]. After the identification of the CCPs, the second step is to identify payoff primitives, which are shown to be point-identified using standard techniques as in [4] and [5]. The estimation method developed in this paper is similar to [2] and is a directly application of the semiparametric sieve maximum likelihood estimator (MLE) framework developed in [6]. Applying the methodology developed in this paper to study adolescent behaviors in United

States, we estimate the peer effects of teenagers' alcohol drinking behaviors using the data from the National Longitudinal Study of Adolescent an Adult Health (Add Health). In the empirical application, we estimate peer effects on adolescent alcohol drinking behaviors using two methods: the first one is the usual MLE ignoring measurement error and the second one is the proposed sieve MLE. Our results indicates that when the measurement errors in network data are ignored, the peer effects estimated using MLE are statistically significant and qualitatively similar to those empirical results in [7] and [8], who use Add Health data and the National Education Longitudinal Study (NELS) to study peer effects on youth behaviors. However, the estimate for peer effects is biased in the presence of measurement errors. Using the proposed sieve estimator, we find a significant and much larger (nearly 100%) estimate of peer effects on youth alcohol drinking behaviors. Therefore, our work also contributes to the empirical literature studying peer effects by providing empirical evidence of the existence of measurement error in Add Health data and illustrating the consequences when ignoring it.

In the second chapter, 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. 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. Motivated by the commonly adopted data structure in the social interaction literature, 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. 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 [9] 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. In the empirical application, we apply our methods to study the peer effects on youth smoking behavior using the Add Health data. 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., [10] and [11]. 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.

In the third chapter, we study bandwidth selection method for the smoothed maximum score estimator. The smoothed maximum score estimator is proposed by [12] and is a semiparametric estimator for binary response model. Binary response models are very useful for many economics and statistics applications. See [13] for a review of econometric applications of binary response models. In this model, we do not impose parametric assumptions on the distribution of the error term. Therefore, the parameter of interest cannot be estimated by maximum likelihood method that has been widely used for probit and logit models. If the error term and covariates are independent of each other, various semiparametric methods (e.g., [14], [15], [16] and [17]) can be used to obtain a consistent estimator of β. The maximum score estimator (MS) of [18, 19] allows for the dependence of the distribution of u on x in an unknown and general way (heteroskedasticity of an unknown form). However, since the objective function is discontinuous, the convergence rate of the maximum score estimator is $n^{-1/3}$, and its limiting distribution is non-standard ([20]). [12] develops a smoothed version of Manski's maximum score estimator, which is asymptotically normal and has a faster convergence rate. The convergence rate could approach $n^{-1/2}$, depending on the strength of certain smoothness conditions. The idea of Horowitz's smoothed maximum score estimator (SMS) is analogous to the nonparametric estimation of cumulative distribution function (CDF), and involves replacing the indicator function by a continuously differentiable function in the objective function of the maximum score estimation. The continuously differentiable function retains the essential features of an indicator function. It is generally acknowledged that kernel smoothing method can be very sensitive to the selection of bandwidth. Different bandwidths can lead to completely different results. In terms of bandwidth selection, [12] proposes a method that is analogous to the plug-in method in kernel density estimation.The method requires initial "pilot" values of the bandwidth to compute the SMS estimator, and then uses this estimator to obtain the optimal bandwidth. This method has the disadvantage of not being fully data-driven, since the estimated optimal bandwidth depends on the initial selection of bandwidth. In this paper, we propose an alternative method to obtain the bandwidth. Unlike the conventional plug-in method, we choose the bandwidth by minimizing a cross-validated criterion function. It is completely data-driven and does not require the selection of the initial bandwidth. We use Monte Carlo simulations to examine the finite sample performance of our proposed method. The results show that our proposed method performs better than existing methods.

2. IDENTIFICATION AND ESTIMATION OF PEER EFFECTS IN MIS-MEASURED SOCIAL NETWORKS

2.1 Introduction

In recent years, a growing body of literature studies social networks and their implications for economic outcomes, see e.g., [21] for an extensive review of the literature. A social network is represented by network graph, which contains a set of connections (edges) among a collection of economic agents (nodes). For instance in a school-based friendship network, nodes will be students and edges may represent friendship connections among them. This paper is motivated by the fact that econometric analysis of networks has long suffered the issue of measurement errors, usually in the form of missing or spurious data in network connections. The presence of measurement errors mainly results from the sources of network data, which predominantly are surveys and questionnaires soliciting self-reports ([1]). Applied researchers typically construct network from data and naively treat this network as the true network of interest, ignoring the problem of measurement errors. The main objective of this article is to develop an econometric framework to identify and estimate peer effects in the presence of measurement errors in network data.

Ever since the seminar work of [22], network-based peer effects have been studied extensively in econometrics, see, e.g., [23], [24], [25], [26], [27] and so forth. However, most of the previous work assumes that the observed network in data represents the true network structure. A few exceptions include [28] and [29]. Nevertheless, the difference between their work and ours is that their methods can only handle the missing data problem by restricting the measurement errors to be "one-sided", meaning that if two nodes are connected according to the data, then the econometrician knows there are no measurement errors. However, as is mentioned in [30], in surveys respondents sometimes reports relations that are not actually present and hence lead to spurious network connections in the data. Our work, on the other hand, provides a unified approach to solve both the missing and spurious problem in network data. To the best of our knowledge, our paper is the first to address these two problems simultaneously.

The model studied in this paper is an incomplete information game theoretical model with binary choice. Each player's payoff function consists of three components: direct utility from the chosen action, peer effects from socially connected players' actions and a stochastic component representing payoff shocks. The payoff shocks are players' private information with commonly known distribution. The three components are assumed to be additively separable, similar payoff structure has been studied in [31]. The game studied in this paper belongs to the semi-anonymous graphical game discussed in [21] and [32], in the sense that player's choice is influenced mainly by the relative population of a given action among his or her neighbors and does not depend on the specific identities of the neighbors who take the action.

It is well known that identification and estimation of empirical game with incomplete information can be difficult when multiple equilibria exist. This is because the usual identification and estimation methods are developed under the many-game paradigm, which requires observing many repetitions of the same n-player game in order to identify and estimate the equilibrium conditional choice probabilities (CCPs) and payoff primitives. Therefore, multiple equilibria will cause problem under many-game setting since different equilibria may exist among the repetitions of the same game and simply pooling those repetitions together can only allow us to identify a mixture of CCPs under different equilibria, making it hard to recover payoff primitives from CCPs ([33]). The identification and estimation methods in this paper are developed under the large-game setting, i.e., the number of players going to infinity. Under the setup of large game, multiple equilibria will no longer be problematic because we do not need to pool information cross-sectionally. At the first glance, large-game setting makes identification and estimation of CCPs impossible since we can only observe a single action of each player and hence do not have enough variation to identify CCPs. The solution in this paper is to focus on the equilibria that are semi-anonymously symmetric, meaning that players with same characteristics and relative proportion of a given action among her neighbors will have same (ex-ante) probabilities for choosing an action in equilibrium.

Consequently, players with same characteristics and relative proportion of a given action among his or her neighbors can be viewed as same-type players and equilibrium actions of those players will generate variations that can help identify the CCPs.

The identification proceeds in two steps. In the first step, we show that under semi-anonymously symmetric equilibrium, the CCPs can be nonparametrically identified when the number of players is fairly large. Specifically, we prove that the game theoretical model can be fitted into the measurement error models as proposed in [2] and [3] and then provide two different methods to identify the CCPs. The first (point) identification method is implemented by incorporating an instrumental variable and applying a spectral decomposition technique to the observed distributions in data and is also the method used in the empirical application. The second (partial) identification method is developed in the case when a valid instrumental variable is not available in the data and follows the direct misclassification proposed in [3]. After the identification of the CCPs, the second step is to identify payoff primitives, which are shown to be point-identified using standard techniques as in [4] and [5].

It is worth mentioning that the identification of CCPs does not trivially follows [2] and [3] because in our model, the CCPs are conditional on all player's characteristics and the network structure. The network structure is represented by an $n \times n$ random matrix, where n is the sample size. Therefore, the dimension of measurement errors will also be $n \times n$, resulting in a high dimensionality problem and the results in [2] and [3] cannot be directly applied since their methods requires the dimension of measurement errors to be fixed. Recent development in high dimensional measurement errors models all focus on linear models , see, e.g., [34] and [35]. Hence neither can their methods be applied to the nonlinear model in this paper. In this paper, we solve the high dimensionality problem by requiring all equilibria to be semi-anonymously symmetric and prove that as long as the number of players is fairly large, the CCPs are asymptotically equivalent to the ones that are only conditional on each player's own characteristics and a scalar valued function summarizing actions of her neighbors. Similar idea has been proposed in [36] and [37] in studying network formation models. We extend their ideas to the context of our model by allowing the

CCPs to be conditional not only players' own characteristics, but also a scalar valued function for the actions of their neighbors in the network.

The estimation method developed in this paper is similar to [2] and is a directly application of the semiparametric sieve maximum likelihood estimator (MLE) framework developed in [6]. Under the setup of our model, the observed data is weakly dependent conditional on all public information in the game. We show that the sieve MLE framework developed in [2] under independent and identical distributed (i.i.d) data context can be extended to the (conditional) weakly dependent case and establish its consistency and asymptotic distribution. The Monte Carlo experiments demonstrate that our proposed estimator performs well in finite samples.

Applying the methodology developed in this paper to study adolescent behaviors in United States, we estimate the peer effects of teenagers' alcohol drinking behaviors using the data from the National Longitudinal Study of Adolescent an Adult 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 (Wave I). The instrument variable we use are obtained from second wave survey, which was conducted one year after Wave I. Given that the Wave II data was surveyed after one year of Wave I, it is convincing that the measurement errors in the two waves are independent with each other, conditioning on the latent true value of network. Hence the exclusion restrictions for identification are satisfied. In the empirical application, we estimate peer effects on adolescent alcohol drinking behaviors using two methods: the first one is the usual MLE ignoring measurement error and the second one is the sieve MLE. Our results indicates that when the measurement errors in network data are ignored, the peer effects estimated using MLE are statistically significant and qualitatively similar to those empirical results in [7] and [8], who use Add Health data and the National Education Longitudinal Study (NELS) to study peer effects on youth behaviors. However, the estimate for peer effects is biased in the presence of measurement errors. Using the proposed sieve estimator, we find a significant and much larger (nearly 100%) estimate of peer effects on youth alcohol drinking behaviors. Therefore, our work also contributes to the empirical literature studying peer effects by providing empirical evidence of the existence of measurement error in Add Health data and illustrating the consequences when ignoring it.

Recently there has been some studies in measurement error issues of social network. However, little work has been done in developing methods to identify and estimate network-based peer effects in presence of measurement errors. [30] characterize different forms of measurement errors in network data and studies the sensitivity of network statistics to those measurement errors. [38] study a Manski-type linear-in-means model and applies their method to investigate the peer effects using Add Health data, which is one of the most commonly used dataset in network econometrics. They find that even though the structure of the friendship network tends to change substantially between two waves of the survey, the estimated peer effects are qualitatively similar. Based on this finding, they cast doubt about relying on self-reported friendship links to study peer effects. Nevertheless, their method can not solve the problem caused by measurement errors. [28] proposes a method to point-identify the peer effects in a complete information game that also allows the existence of measurement errors in network connections. The difference between his work and ours is that he restricts the measurement errors to be “one-sided", meaning that if two players are connected according to the data, then no measurement errors will exist. Consequently, his method can not deal with the case of spurious data. [29] provide analytical and numerical examples to illustrate the severity of the biases in estimated peer effects caused by measurement errors and propose a two-step graphical reconstruction procedure to correct the biases. Nevertheless, in order for the graphic reconstruction procedure to be valid, the measurement errors still need to be one-sided. Our method, on the other hand, allows the measurement errors to be two-sided, making it unique and novel in the literature of network econometrics.

The rest of the essay is organized as follows. Section 2.2 presents the setting and basic assumptions of our model. Section 2.3 provides the (point) identification method with the help of instrumental variable. Section 2.4 discusses the (partial) identification method without instrumental variable. Section 2.5 discusses the estimation method and establishes the asymptotic behavior of our proposed estimator. Section 2.6 illustrates the finite sample performance of the proposed estimator by conducting several Month Carlo experiments. Section 2.7 contains empirical analysis of peer effects on youth alcohol drinking behaviors and Section 2.8 concludes. All proofs are provided in Appendix A.1, and Appendix A.2 contains additional simulation results.

2.2 Model Setup

We consider a simultaneous-move incomplete information game played in a social network[1]. There are n players indexed by $i \in N$, where $N \equiv \{1, 2, \ldots, n\}$ is the set of all players. In this game, each player simultaneously chooses a discrete action $Y_i \in A \equiv \{0, 1\}$. Let $X_i \in \mathcal{X} \subset \mathbb{R}^d$ be the vector of player i's socioeconomic characteristics. Also let $G^*_{ij} = 1$ if player i nominates j as her friend and $G^*_{ij} = 0$ otherwise. Note that the edges in the network graph are directed, which implies that friendship need not be symmetric, i.e., $G^*_{ij} \neq G^*_{ji}$ is allowed. We use $G^*_i = (G^*_{i1}, G^*_{i2}, \cdots, G^*_{i,i-1}, G^*_{i,i+1}, \cdots, G^*_{in})^T \in \{0, 1\}^n$ and $G^* = (G^{*T}_1, G^{*T}_2, \cdots, G^{*T}_n)^T \in \mathcal{G}$ to denote the network connections for player i and the network structure in this game, respectively. Note that we use the superscript " $*$ " to emphasize that G^* represents the true network structure without measurement errors, which is observed by all players but remains unknown to econometricians because of measurement errors. Then player i's payoff function for choosing action 1 is specified as

$$U_{i1} = \alpha(X_i) + W(Y_{N^*_i}, G^*_i)\beta(X_i) - \epsilon_i, \tag{2.1}$$

where $N^*_i = \{j | G^*_{ij} = 1\}$ is the set of friends of i and $Y_{N^*_i}$ denotes the vector of actions taken by friends of i. Following the literature in empirical game with binary actions, we normalize the payoff for action 0 to be zero, i.e., $U_{i0} = 0$. Note that the cardinality of N^*_i, denoted as $|N^*_i|$, is called the *degree* of player i. In this payoff function, $W(Y_{N^*_i}, N^*_i) = \sum_{j \in N^*_i} Y_j / |N^*_i| \in [0, 1]$ is a continuous and bounded function summarizing the average actions of player i's friends and is assumed to be known by both players and econometricians. To simplify notation we use W^*_i to denote $W(Y_{N^*_i}, N^*_i)$. Note that this game is a semi-anonymous graphical game in the sense of [39] and [21], i.e., the player's choice is influenced mainly by the relative population of a given action among his or her neighbors and is not dependent on the specific identities of the neighbors who

[1] Based on the empirical data that will be used, the social network we considered here is friendship network.

take the action. ϵ_i is the payoff shock, which is assumed to be private information with commonly known distribution. Let $X^c \equiv (X_1^T, X_2^T, \cdots, X_n^T)^T \in \mathcal{X}^n$ be the matrix collecting all players' characteristics. In order to characterize the equilibrium we first impose some assumptions

ASSUMPTION 2.2.1. *(i) $\{\epsilon_i\}_{i\in N}$ is i.i.d with an absolutely continuous cumulative distribution function (CDF) $F_\epsilon(\cdot)$ and bounded probability density function (PDF) $f_\epsilon(\cdot)$. (ii) $\{X_i\}_{i\in N}$ is i.i.d with compact support $\mathcal{X}$. (iii) The support of W_i^* is compact.*

Assumption 2.2.1 is commonly imposed in the literature on identification and estimation of static games with incomplete information and social interaction models (see, e.g., [31], [40] and [27]). Condition (i) ensures the continuity of player's equilibrium choice probabilities, which is a necessary condition for the existence of equilibrium. Conditions (ii) and (iii) are used to establish the uniform convergence of the large game, which will be discussed in Section 2.3.

Since the payoff of action 0 has been normalized to be zero, in this incomplete information game player i will choose action 1 if the expected utility of action 1 is positive, where the expectation is conditional on all the public information (X^c, G^*) and her private information ϵ_i. Therefore, any Bayesian Nash Equilibrium (BNE) can be characterized by a profile of strategies $\{Y_i(X^c, G^*, \cdot)\}_{i\in N}$ such that for all i and ϵ_i,

$$Y_i(X^c, G^*, \epsilon_i) = \mathbb{1}\left\{\epsilon_i \le \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*, \epsilon_i]\right\}, \tag{2.2}$$

where $\mathbb{1}(\cdot)$ is the indicator function. By Assumption 2.2.1, Y_i and Y_j will be conditionally independent with each other for all $i \neq j$, and then

$$\mathbb{E}[W_i^*|X^c, G^*, \epsilon_i] = \mathbb{E}[W_i^*|X^c, G^*].$$

As a result, in any BNE the joint distribution of $Y_{N_i^*}$ conditional on i's information (X^c, G^*, ϵ_i)

takes the form

$$Pr(Y_{N_i^*} = y_{N_i^*}|X^c, G^*) = \prod_{j \in N_i^*} p_j(X^c, G^*)^{y_j^*}(1 - p_j(X^c, G^*))^{1-y_j^*} \tag{2.3}$$

and $p_i(X^c, G^*) \equiv Pr\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]|X^c, G^*\}$ is i's equilibrium probability of choosing 1 given X^c and G^*. Following the literature of empirical game with incomplete information, instead of pure strategies we use players (ex-ante) conditional choice probabilities (CCPs) to characterize the Bayesian Nash Equilibrium:

Definition 2.2.1. *Given X^c and G^*, the Bayesian Nash Equilibrium (BNE) is a collection of CCPs $\{p_i(X^c, G^*)\}_{i \in N}$ satisfying the following conditions:*

$$p_i(X^c, G^*) = Pr\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]|X^c, G^*\}$$

for all $i = 1, 2, \cdots, n$.

The CCPs can be computed as the fixed point of the correspondence with each coordinate-function component given by Definition 2.2.1. The fixed point is guaranteed to exist by Kakutani Fixed-point Theorem. In general, the fixed point may not be unique so we may have multiple fixed points and hence multiple equilibria. In this paper, the identification and estimation strategies are developed under the “large game" setting as in Leung (2015), meaning that the number of players in the game is approaching infinity and we do not need to pull different observations of the same game for the purpose of identification and estimation. Therefore, multiple equilibria will no longer be problematic and the CCPs can be identified directly from the equilibrium actions in the data if no measurement errors are present, and payoff primitives can be identified accordingly, using standard argument as in [40] and [27]. However, with measurement errors in the network structure, point identification of CCPs requires additional information, which will be discussed in Section 2.3.

In this paper, we focus on equilibria that are symmetric in player's characteristics X and W^*, which summarizes the actions of her friends. Specifically we introduce the definition of *semi-*

anonymously symmetric equilibrium:

Definition 2.2.2. *The conditional choice probabilities* $\{p_i(X^c, G^*)\}_{i\in N}$ *are semi-anonymously symmetric if* $X_i = X_j$ *and* $W_i^* = W_j^*$ *implies that*

$$p_i(X^c, G^*) = p_j(X^c, G^*) \text{ for all } i, j \in N.$$

This definition implies that two players will have same (ex-ante) conditional choice probabilities of choosing action 1 if they have similar characteristics and the proportion of a given action among their neighbors are the same. Under the setup of large game, it makes sense to assume symmetric equilibria because otherwise we will not have sufficient variation in the data to help identify CCPs. In semi-anonymously symmetric equilibria, players with same X and W^* can be viewed as same type players and the CCP for each player can be identified by exploring the variations generated by the equilibrium actions of players sharing the same type. Therefore, the assumption of semi-anonymously symmetric equilibria will be crucial for identification and estimation as we assume only one large game is observed. In this paper, we will focus on studying semi-anonymously symmetric BNE[2].

To ensure that the semi-anonymously symmetric equilibria exist under the large game setting, we impose the following assumptions

ASSUMPTION 2.2.2. *(i)* $|N_i^*| \to \infty$ *as* $n \to \infty$ *for all* $i \in N$*; (ii) There exists a* $C_\beta < \infty$ *such that* $|\beta(X_i)| \le C_\beta$ *for all* $X_i \in \mathcal{X}$ *and all* $i \in N$.

Under Assumption 2.2.2 (i), the degree of each player is unbounded, meaning that the number of neighbors for each player goes to infinity as $n \to \infty$. Condition (ii) requires the peer effect $\beta(\cdot)$ to be bounded. Then the following lemma shows that W_i^* and its conditional expectation are asymptotically equivalent.

Lemma 2.2.1. *Under Assumptions 2.2.1-2.2.2,* $W_i^* - \mathbb{E}(W_i^* | X^c, G^*) = o_p(1)$.

[2] Formally as in [36], an equilibrium selection mechanism should be imposed in order for us to focus on semi-anonymously symmetric equilibria, we omit this part for notational simplicity.

Lemma 2.2.1 indicates that under the large game setting, the endogenous variable W_i^* ([22]) becomes exogenous since X^c and G^* are independent with ϵ_i. Then the following proposition shows that there will always exist a semi-anonymously symmetric equilibrium in the large game.

Proposition 2.2.1. *Under Assumptions 2.2.1-2.2.2, there always exists a semi-anonymously symmetric BNE for any $X^c \in \mathcal{X}^n$ and $G^* \in \mathcal{G}$ under the large setting.*

2.3 Identification with Instrumental Variable

In this section, we identify model elements $\{\{p_i(X^c, G^*)\}_{i \in N}, \alpha(\cdot), \beta(\cdot)\}$. In the data the econometrician will observe players' equilibrium actions $\{Y_i\}_{i \in N}$[3], their socioeconomic characteristics X^c and a contaminated measurement G for G^*. As is discussed in Section 2.1, the measurement errors in G are caused by the missing or spurious data in network connections. Figure 2.1 below provides an graphical illustration of the missing data problem in network analysis. In this figure, each blue node represents a player and the line (edge) connecting two nodes represents their friendship connections. Because of missing data, the econometrician may analyze the network structure in (a) even though the true network is the much more complicated one in (b), and obviously econometric analysis based on these two structures will give us very different results. Similarly the spurious data case can be illustrated by reversing (a) and (b) in Figure 2.1. The identification method proceeds in two steps: first we identify CCPs by exploring additional information provided by an instrumental variable, and then in the second step we identify payoff primitives $\alpha(\cdot)$ and $\beta(\cdot)$.

2.3.1 Conditional Choice Probabilities

Since G_{ij}^* and its counterpart G_{ij} in the contaminated measurement G are binary, the measurement errors will be nonclassical in general. Therefore, the identification of CCPs requires the availability of an instrumental variable G' (maybe a repeated measurement of G^*). The technical challenge we will encounter when identifying the CCPs is the high dimensionality problem: the dimensions of X^c and G^* are $d \times n$ and $n \times n$ respectively, where n is the sample size. So the

[3]We assume that coordination failure of the equilibrium never happens in the data.

Figure 2.1: Missing Data

(a) Network with measurement errors

(b) True network

dimensions of X^c and G^* will increase as the sample size increases. Consequently, the first step of identification cannot directly follows the technique introduced in Hu & Schennach (2008) for nonclassical measurement errors with fixed dimensions. Furthermore, our model is nonlinear, so neither will recent development in the literature of high dimensional measurement error (see e.g., [34] and [35]) be used since they all focus on linear models. In this paper, we propose a difference method to solve the high dimensionality problem,

The idea of our identification method is to explore the asymptotic behavior of the n-player game when n approaches infinity. Thanks to the semi-anonymously symmetric equilibria, we can show that the CCP of an individual converges to some limiting CCP that conditional on her own characteristics and a scalar valued function of her friends' behaviors. This asymptotic feature is crucial for applying the results in [2].

From Definition 2.2.1 the probability that player i select action 1 conditional on characteristic profile X^c and true network structure G^* is

$$p_i(X^c, G^*) = Pr\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]|X^c, G^*\} \tag{2.4}$$

where the probability operation is calculated with respect to ϵ_i. Under the semi-anonymously symmetric BNE, we expect $p_i(X^c, G^*)$ converges to a limit given by

$$p_i(X_i, W_i^*) = Pr\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^* | X_i, W_i^*\}. \tag{2.5}$$

We refer $p_i(X_i, W_i^*)$ as the CCP derived from a "large game" with infinite number of players. In this game, conditional on her own characteristics X_i and a function W_i^* summarizing her friends' actions, player i will choose actions myopically, without considering all other public information in this game. The limiting CCPs can be computed as the fixed point of the correspondence with each coordinate-function component given by (2.5). The following proposition establishes the convergence result:

Proposition 2.3.1. *Under Assumptions 2.2.1-2.2.2,* $\sup_{i \in N} |p_i(X^c, G^*) - p_i(X_i, W_i^*)| = o_p(1)$.

The intuition behind Proposition 2.3.1 is that under the semi-anonymously symmetric BNE, players with same characteristics X and relative proportion of a given action among her friends W^* can be viewed as same-type players. Consequently, for player i, her CCP $p_i(X^c, G^*)$ can be identified by exploring the variations of equilibrium actions from players sharing the same type with her. Under the large game setting, the information provided those variations turns out to be the limiting CCP $p_i(X_i, W_i^*)$. Therefore, under the large game setting instead of $p_i(X^c, G^*)$ we can identify $p_i(X_i, W_i^*)$, which can be done by using the technique in [2]. Let $W_i = W_i(Y_{N_i}, G_i)$ and $W_i' = W_i(Y_{N_i'}, G_i')$, where N_i and N_i' are defined analogously to N_i^*, as the set of friends for i under G and G', respectively. Let $\mathcal{Y}$, $\mathcal{W}$, $\mathcal{W}^*$ and $\mathcal{W}'$ denote the supports of the distributions of the random variables Y, W^*, W and W' respectively. We consider W^*, W and W' to be jointly continuously distributed and impose the following assumptions:

Assumption 2.3.1. *The joint density of Y and X, W^*, W, W' admits a bounded density with respect to the product measure of some dominating measure μ (defined on $\mathcal{Y}$) and the Lebesgue measure on $\mathcal{X} \times \mathcal{W} \times \mathcal{W}^* \times \mathcal{W}'$. All marginal and conditional densities are also bounded.*

We use the notation $f_{A|B}(\cdot)$ to denote the density of random variable A conditional on random variable B and use lowercase letter a and b to denote the realized value for A and B. Besides, we use $\mathcal{A}$ and $\mathcal{B}$ to denote the support of A and B. To state the identification result, we first make some assumptions about the conditional densities.

ASSUMPTION 2.3.2. *(i)* $f_{Y|X,W^*,W,W'}(\cdot) = f_{Y|X,W^*}(\cdot)$ *and*
(ii)$f_{W|X,W^*,W'}(\cdot) = f_{W|W^*}(\cdot)$ *and* $f_{W^*|X,W'}(\cdot) = f_{W^*|W'}(\cdot)$ *for all* $(Y,X,W^*,W,W') \in \mathcal{Y} \times \mathcal{X} \times \mathcal{W} \times \mathcal{W}^* \times \mathcal{W}'$.

Assumption 2.3.2 (i) indicates that W and W' do not provide any additional information about Y than W^* already provides while Assumption 2.3.2 (ii) specifies that W' and X does not provide any more information about W than W^* already provides and X does not provide any more information about W^* than W' already provides. These assumptions can be interpreted as standard exclusion restrictions. Note that Assumption 2.3.2 is general enough to include both the classical and nonclassical measurement error cases. If W' is a repeated measurement of W^*, this assumption can be implied by that the two measurements W and W' be mutually independent conditional on W^*.

To state the next assumption, it is useful to define an integral operator $L_{A|B}$, which maps $\mathcal{G}(\mathcal{A})$ to $L_{A|B}g \in \mathcal{G}(\mathcal{B})$ defined by

$$[L_{A|B}g](a) \equiv \int f_{A|B}(a|b)g(b)db,$$

where $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(\mathcal{B})$ are spaces of $g(\cdot)$ with domains $\mathcal{A}$ and $\mathcal{B}$, respectively.

ASSUMPTION 2.3.3. *The integral operators* $L_{W|W^*}$ *and* $L_{W'|W}$ *are injective.*

The operator $L_{A|B}$ is said to be injective if its inverse $L^{-1}_{A|B}$ is defined over the range of the operator $L_{A|B}$. Intuitively it means that there is enough variation in the density of A for different values of B. A simple example where $L_{A|B}$ is not injective is when $f_{A|B}(\cdot)$ is fixed on $\mathcal{A}$ for any $B \in \mathcal{B}$. Injectivity assumption is weak and commonly imposed in the literature of nonparametric IV methods. As pointed out in [2], under Assumption 2.3.1, a sufficient condition for the injectivity of $L_{b|a}$ is the bounded completeness of $f_{A|B}(\cdot)$. Formally $f_{A|B}(\cdot)$ is *bounded complete* if the only solution $\delta(a)$ to

$$\int_{\mathcal{A}} \delta(a) f_{A|B}(a|b) da = 0 \text{ for all } b \in \mathcal{B}$$

is $\delta(a) = 0$ for all bounded $\delta(a) \in \mathcal{L}^1(\mathcal{A})$. Primitive conditions for bounded completeness can be found in [41] and are fairly weak.

ASSUMPTION 2.3.4. *For all $X \in \mathcal{X}$, the set $\{Y : f_{Y|X,W^*}(\cdot) \neq f_{Y|X,\tilde{W}^*}(\cdot)\}$ has positive probability for any W^* and $\tilde{W}^* \in \mathcal{W}^*$ such that $W^* \neq \tilde{W}^*$.*

Assumption 2.3.4 will be violated if the distribution of Y conditional on X and W^* is identical at two different values of W^*. Since Y is binary, this assumption is equivalent to a monotonicity assumption on $p(X, W^*)$. For example if $p(X, W^*)$ is strictly monotone in W^*, this condition will be satisfied.

ASSUMPTION 2.3.5. *There exists a known functional M such that $M[f_{W|W^*}(\cdot)] = W^*$ for all $W^* \in \mathcal{W}^*$.*

M is a very general functional that maps a density to a real number and that defines some measures of location. As is mentioned in [2], examples of M include, but are not limited to, the mean, the mode, and the τth quantile. Then following the insight in [2], we have the following result of identification.

Proposition 2.3.2. *Under Assumptions 2.3.1-2.3.5, given the true observed density $f_{YW|X,W'}$, the equation*

$$f_{YW|X,W'}(y, w|x, w') = \int_{\mathcal{W}^*} f_{Y|X,W^*}(y|x, w^*) f_{W|W^*}(w|w^*) f_{W^*|W'}(w^*|w') dw^* \tag{2.6}$$

admits a unique solution $(f_{Y|X,W^*}, f_{W|W^*}, f_{W^*|W'})$ for all $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $w \in \mathcal{W}$ and $w' \in \mathcal{W}'$.

We provide a heuristic argument of the proof here, detailed explanation can be found in [2]. Equation (2.6) can be established by Assumption 2.3.2 and then shown to define the operator equivalence relationship

$$L_{Y;W|X,W'} = L_{W|W^*} \Delta_{Y;X,W^*} L_{W^*|W'}, \tag{2.7}$$

where $L_{Y;W|XW'}$ is defined similarly to $L_{W|W'}$ and $\Delta_{Y;X,W^*}$ is the operator mapping the function $g(W^*)$ to the function $f_{Y|X,W^*}(\cdot)g(W^*)$ for given $Y \in \mathcal{Y}$ and $X \in \mathcal{X}$. Note that by Assumption 2.3.2 and integrating (2.7) over all $Y \in \mathcal{Y}$, we can obtain another equivalence relationship

$$L_{W|W'} = L_{W|W^*}L_{W^*|W'}. \tag{2.8}$$

Then by Assumption 2.3.3, equation (2.8) and rearranging terms in (2.7), we can obtain

$$L_{Y;W|X,W'}L_{W|W'}^{-1} = L_{W|W^*}\Delta_{Y;X,W^*}L_{W|W^*}^{-1},$$

which means that $L_{Y;W|X,W'}L_{W|W'}^{-1}$ admits an eigenvalue-eigenfunction decomposition with eigenvalues $f_{Y|X,W^*}(\cdot)$ and eigenfunctions $f_{W|W^*}(\cdot)$. The uniqueness of this decomposition follows from combining technique in spectral analysis with Assumptions 2.3.4 and 2.3.5.

By Proposition 2.3.2 we have identified $f_{Y|X,W^*}(\cdot)$. Since for each $i \in N$,

$$f_{Y|X,W^*}(Y_i|X_i, W_i^*) = p_i(X_i, W_i^*)^{Y_i}(1 - p_i(X_i, W_i^*))^{1-Y_i}, \tag{2.9}$$

$p_i(X_i, W_i^*)$ is identified. Hence by Proposition 2.3.1, the conditional choice probability $p_i(X^c, G^*)$ is identified under the large game setting.

2.3.2 Payoff Primitives

In this subsection, we identify payoff primitives $\alpha(\cdot)$ and $\beta(\cdot)$. The idea of identification is to first identify $v(X_i, W_i^*) = \alpha(X_i) + \beta(X_i)W_i^*$, and then $\alpha(X_i)$ and $\beta(X_i)$ can be identified separately by exploring the variation of W_i^*. Note that

$$\begin{aligned} p_i(X_i, W_i^*) &= Pr[\alpha(X_i) + \beta(X_i)W_i^* - \epsilon_i \geq 0|X_i, W_i^*] \\ &= F_{\epsilon|X,W^*}[\alpha(X_i) + \beta(X_i)W_i^*], \end{aligned} \tag{2.10}$$

where and $F_{\epsilon|X,W^*}(\cdot)$ is the conditional CDF of ϵ_i and we assume that it is strictly increasing.

Let $\mathcal{V}$ denote the set to which the function $v(\cdot)$ belongs, and let $\mathcal{F}$ denote the set to which $F_{\epsilon|X,W^*}(\cdot)$ belongs. The identification of $v(\cdot)$ does not directly follows the identification of CCPs because different values of $v(\cdot)$ may still lead to the same CCP, providing the existence of such distribution functions for ϵ. Consequently, if for any two functions v and v' in $\mathcal{V}$, we can find distributions $F_{\epsilon|X,W^*}$ and $F'_{\epsilon|X,W^*}$ in $\mathcal{F}$ such that the pairs $(v, F_{\epsilon|X,W^*})$ and $(v', F'_{\epsilon|X,W^*})$ generate the same conditional choice probability $p_i(X^c, G^*)$, v and v' are said to observationally equivalent.

Definition 2.3.1. *Any two functions $v(\cdot)$ and $v'(\cdot)$ in $\mathcal{V}$ are said to be observationally equivalent if there exist $F_{\epsilon|X,W^*}(\cdot)$ and $F'_{\epsilon|X,W^*}(\cdot)$ in $\mathcal{F}$ such that for all $X \in \mathcal{X}$ and $W^* \in \mathcal{W}^*$, $F_{\epsilon|X,W^*}[v(X, W^*)] = F'_{\epsilon|X,W^*}[v'(X, W^*)]$*

Definition 2.3.1 implies that the set of functions indistinguishable from $v(\cdot)$ is

$$\mathcal{V}_{o.e} \equiv \{v' \in \mathcal{V} : \exists\, F_{\epsilon|X,W^*}, F'_{\epsilon|X,W^*} \in \mathcal{F} \text{ s.t. } F'_{\epsilon|X,W^*}[v'(X, W^*)] = F_{\epsilon|X,W^*}[v(X, W^*)]\}. \quad (2.11)$$

Following the insight in [4], we provide a lemma that shows what properties $\mathcal{V}$ has to satisfy to guarantee the identification of $v(\cdot) \in \mathcal{V}$.

Lemma 2.3.1. $\mathcal{V}_{o.e} = \{v(X, W^*)\}$ *if and only if there does not exist a strictly increasing function $g : v(\mathcal{X}, \mathcal{W}^*) \mapsto \mathbb{R}$ such that $v' = g \circ v$ on $\mathcal{X} \times \mathcal{W}^*$.*

Lemma 2.3.1 implies that the function v is identified up to a monotone transformation, g. For example the ratios of derivatives of v are identified. In order to identify v we can either restrict the function g for the purpose of normalization or the set of functions v in such a way that no two different functions in this set can be strictly increasing transformation of each other. As is mentioned in [4], one of the normalizations is that for some given value $\overline{X} \in \mathcal{X}$,

$$g(v(\overline{X}, W_i^*)) = W_i^*. \quad (2.12)$$

This type of normalization can be viewed as the generalization of the identification method provided in [19]. Examples of restrictions on the set of function v include homogeneity and additive separability. Specifically suppose for some $\overline{X} \in \mathcal{X}$, some $\overline{W}^* \in \mathcal{W}^*$ and all $\lambda > 0$,

$$v(\lambda\overline{X}, \lambda\overline{W}^*) = \lambda\delta \tag{2.13}$$

where $v(\overline{X}, \overline{W}^*) = \delta$ for some $\delta \in \mathbb{R}$. Then following the arguments in [9], we can show that it is impossible to write two different functions in the set of v as strictly increasing transformation of each other. Besides, if economic theory indicates that

$$v(X_i, W_i^*) = r(X_i) + W_i^* \tag{2.14}$$

where $r(\overline{X}) = \delta$ for some $\overline{X} \in \mathcal{X}$ and $\delta \in \mathbb{R}$, then again one can prove that no two different functions v can be written as strictly increasing transformation of each other.

After the identification of $v(X_i, W_i^*)$, since $v(X_i, W_i^*) = \alpha(X_i) + \beta(X_i)W_i^*$ and $\alpha(\cdot)$, $\beta(\cdot)$ only depend on X_i, we can separately identify the structural functions $\alpha(\cdot)$ and $\beta(\cdot)$ by relying on the information provided by two individuals with same characteristics X but different W^*.

Theorem 2.3.1. *Under Assumptions 2.2.1-2.3.3, the CCPs $\{p_i(X^c, G^*)\}_{i\in N}$ and the payoff primitives $\alpha(\cdot)$ and $\beta(\cdot)$ are nonparametrically identified.*

2.4 Partial Identification without Instrumental Variable

The identification argument in the previous section depends on the availability of an instrumental variable G'. Sometimes we may not be able to find such a variable, and therefore in this section we discuss the identification without using the instrumental variable and show that in this case the CCPs will be partially identified.

2.4.1 Conditional Choice Probabilities

Similar to the previous section, we focus on semi-anonymously symmetric equilibrium and the first step is to identify the limiting CCP $p(X, W^*)$. The identification method in this section is closely related to the *direct misclassification approach* in [3], which requires the support of W^* to be discrete. Since $W^* \in [0, \overline{w}]$ is continuous, following An (2017) we can discretize it by the following method of discretization:

$$W_d^* = \begin{cases} 1 \text{ if } W^* \in [0, w(1)], \\ 2 \text{ if } W^* \in (w(1), w(2)], \\ \cdots \\ M \text{ if } W^* \in (w(M-1), \overline{w}], \end{cases}$$

where the support of W^* is divided into M $(M \geq 2)$ intervals by the $M-1$ cutoff points $w(1), w(2), \cdots, w(M-1)$ and satisfy $0 < w(1) < w(2) < \cdots < w(M-1) < \overline{w}$. Similarly we can also discretize W into W_d. Both W_d^* and W_d take values from $\mathcal{M} \equiv \{1, 2, \cdots, M\}$ but the cutoff points for discretizing W^* and W can be different. By Proposition 2.3.1 we need to identify the discretized limiting CCP $p(X, W_d^* = j),\ j \in \mathcal{M}$.

Specifically we characterize the relationship between the observable conditional distribution of W_d and the unobservable conditional distribution of W_d^* as

$$\begin{bmatrix} Pr(W_d = 1|X) \\ \vdots \\ Pr(W_d = M|X) \end{bmatrix} = \begin{bmatrix} Pr(W_d = 1|X, W_d^* = 1) & \cdots & Pr(W_d = 1|X, W_d^* = M) \\ \vdots & \ddots & \vdots \\ Pr(W_d = M|X, W_d^* = 1) & \cdots & Pr(W_d = M|X, W_d^* = M) \end{bmatrix} \begin{bmatrix} Pr(W_d^* = 1|X) \\ \vdots \\ Pr(W_d^* = M|X) \end{bmatrix}. \quad (2.15)$$

Let $\mathcal{P}^W$ denote the column vector $[Pr(W_d = j|X), j \in \mathcal{M}]$, $\mathcal{P}^{W^*}$ denote the column vector $[Pr(W_d^* = j|X), j \in \mathcal{M}]$ and $\varXi^*$ be the matrix of elements $\{Pr(W_d = i|X, W_d^* = j), \}_{i,j \in \mathcal{M}}$. Then (2.15) can be written compactly as

$$\mathcal{P}^W = \varXi^* \mathcal{P}^{W^*} \quad (2.16)$$

Let $H[\varXi^*]$ be the identification region of $\varXi^*$, which is of central importance in our identification method and will be characterized by probabilistic constraints and constraints coming from validation studies and theories developed in economics. Examples of those constrains include restricting $Pr(W_d = j|X, W_d^* = j)$ to be constant and known, requiring $Pr(W_d = j|X, W_d^* = j)$ to be monotonic in j and imposing a lower bound on $Pr(W_d = j|X, W_d^* = j)$.

The intuition of identification can be summarized as follows: first notice that the discretized CCP $p(X, W_d^* = j)$ is related to $Pr(W_d = i|X, W_d^* = j)$ through the law of total probability:

$$p(X, W_d^* = j) = \sum_{i \in \mathcal{M}} Pr(Y = 1|X, W_d = i, W_d^* = j) \cdot Pr(W_d = i|X, W_d^* = j),\ j \in \mathcal{M}. \quad (2.17)$$

In order to characterize the identification region of $Pr(Y = 1|X, W_d^* = j)$, we need to first identify $Pr(Y = 1|X, W_d = i, W_d^* = j)$, which is related to the observed CCP $p(X, W_d = i)$ and $Pr(W_d^* = j|X, W_d = i)$ through the following relationship:

$$p(X, W_d = i) = \sum_{j \in \mathcal{M}} Pr(Y = 1|X, W_d = i, W_d^* = j) \cdot Pr(W_d^* = j|X, W_d = i), i \in \mathcal{M}. \quad (2.18)$$

Therefore, if we can identify $Pr(W_d^* = j|X, W_d = i)$, then by (2.18) we can identify $Pr(Y = 1|X, W_d = i, W_d^* = j)$. Consequently, the identification of $p(X, W_d^* = j)$ follows directly from (2.17) and information contained in $H[\Xi^*]$.

Following the intuition above we need to first identify $Pr(W_d^* = j|X, W_d = i)$. If we can solve the system of equations in (2.16) and uniquely recover $\mathcal{P}^{W^*}$, then $Pr(W_d^* = j|X, W_d = i)$ can be identified by combining the information in $H[\Xi^*]$ with $\mathcal{P}^{W^*}$ and the observed probabilities $\mathcal{P}^W$. Solving (2.16) requires the matrix Ξ^* to be full rank. Therefore, without loss of generality we impose the following assumption

ASSUMPTION 2.4.1. *For all $j \in \mathcal{M}$, $\mathcal{P}(W_d = j|X, W_d^* = j) > \frac{1}{2}$.*

Assumption 2.4.1 requires the probability of “correct reporting" to be greater than $\frac{1}{2}$ for each of the values that W_d^* can take. Validation studies indicate that this requirement is often satisfied in practice. This assumption implies that for any $\Xi \in H(\Xi^*)$, it is strictly diagonally dominant and hence non-singular (Theorem 6.1.10 in [42]). Consequently, we have

$$\mathcal{P}^{W^*}(\Xi^*) = (\Xi^*)^{-1}\mathcal{P}^W, \quad (2.19)$$

where the parenthesis is to emphasize that $\mathcal{P}^{W^*}$ depend on the information from Ξ^*. Let $\mathcal{P}_j^W(\Xi^*)$ denote the jth element of $\mathcal{P}^{W^*}(\Xi^*)$ and similarly define $\mathcal{P}_i^{W^*}$ and Ξ_{ij}^*, and then the identification results are given in the following proposition:

Proposition 2.4.1. *Given the set $H(\Xi^*)$, the sharp lower and upper bounds for the discretized CCP $p(X, W_d^* = j),\ j \in \mathcal{M}$ are given respectively by*

$$L_j = \inf_{\Xi \in H(\Xi^*)} \sum_{i=1}^{M} \frac{Pr(Y=1|X, W_d = i) - [1 - \varsigma_{ji}(\Xi)]}{\varsigma_{ji}(\Xi)} \cdot \Xi_{ij}$$

and

$$U_j = \sup_{\Xi \in H(\Xi^*)} \sum_{i=1}^{M} \frac{Pr(Y=1|X, W_d = i)}{\varsigma_{ji}(\Xi)} \cdot \Xi_{ij},$$

where

$$\varsigma_{ji}(\Xi) = \frac{\Xi_{ij}\mathcal{P}_j^W(\Xi)}{\mathcal{P}_i^{W^*}},\ i, j \in \mathcal{M}.$$

2.4.2 Payoff Primitives

After the partial identification of the discretized CCPs, we need to identify the function $v(\cdot)$. By Proposition 2.4.1, $L_j \le F_{\epsilon|X,W^*}[v(X, W_d^* = j)] = p(X, W_d^* = j) \le U_j$. Therefore, another function $v'(\cdot) \in \mathcal{V}$ will be observationally equivalent to $v(\cdot)$ if the induced CCP by $v'(\cdot)$ will also fall between these lower and upper bounds. Formally the set of functions indistinguishable from $v(\cdot)$ is

$$\mathcal{V}_{o.e} \equiv \{v' \in \mathcal{V} : \exists F'_{\epsilon|X,W^*} \in \mathcal{F} \text{ s.t. } F'_{\epsilon|X,W^*}[v'(X, W_d^* = j)] < L_j \text{ or } F'_{\epsilon|X,W^*}[v'(X, W_d^* = j)] > U_j\}. \tag{2.20}$$

We characterize the sufficient condition for identification of the function $v(\cdot)$ in the following lemma:

Lemma 2.4.1. *If for any $v'(\cdot) \in \mathcal{V}$ with $v'(\cdot) \neq v(\cdot)$ and any distribution function $F'_{\epsilon|X,W^*}(\cdot) \in \mathcal{F}$, $Pr[L_j \le F'_{\epsilon|X,W^*}[v'(X, W_d^* = j)] \le U_j] = 0$, then $\mathcal{V}_{o.e} = \{v(X, W_d^* = j)\}$.*

If the sufficient condition in Lemma 2.4.1 is satisfied, $v(X, W_d^* = j)$ will be point-identified, then the payoff primitives $\alpha(\cdot)$ and $\beta(\cdot)$ can be separately identified by using a similar argument as in Section 2.3.2.

2.5 Estimation

Based on the identification equation (2.6), we propose a semiparametric sieve maximum likelihood estimator for the unknown parameters. The density function $f_{Y|X,W^*}(y|x,w^*)$ will be parametrized as $f_{Y|X,W^*}(y|x,w^*;\theta_0)$, where $\theta_0 \in \Theta \subset \mathbb{R}^{d+1}$ is a finite-dimensional parameter vector and the subscript “0" means the true value of the parameter. We assume that θ is identified if $f_{Y|X,W^*}(\cdot)$ is identified so the parametrization does not include redundant degrees of freedom. The unknown density functions $f_{W|W^*}$ and $f_{W^*|W'}$ will be estimated by nonparametric method. Specifically we will approximate $f_{W|W^*}$ and $f_{W^*|W'}$ by truncated series and estimate all parameters within a semiparametric maximum likelihood hood framework.

As in [43], we impose standard smoothness restrictions and assume that the unknown functions $f_{W|W^*}$ and $f_{W^*|W'}$ belongs to a Hölder space. For any $d \times 1$ vector $a = (a_1, a_2, \cdots, a_d)^T$ of nonnegative integers, let $|a| = \sum_{k=1}^{d} a_k$ and for any $u \in \mathcal{U} \subset \mathbb{R}^d$, we denote the $|a|$-th derivative of a function $h : \mathcal{U} \mapsto \mathbb{R}$ as

$$\nabla^a h(u) = \frac{\partial^{|a|}}{\partial u_1^{a_1} \cdots \partial u_d^{a_d}} h(u).$$

For some $\xi > 0$, let $\underline{\xi}$ be the largest integer smaller than ξ and let $\|\cdot\|_E$ denote and Euclidean norm. The Hölder space $\Lambda^\xi(\mathcal{U})$ of order ξ is a space of functions $h : \mathcal{U} \mapsto \mathbb{R}$ such that the first $\underline{\xi}$ derivative is bounded and the $\underline{\xi}$=th derivative is Hölder continuous with exponent $\xi - \underline{\xi} \in (0, 1]$, i.e.,

$$\max_{|a|=\underline{\xi}} |\nabla^a h(u) - \nabla^a h(u')| \leq \text{const.}(\|u - u'\|_E)^{\xi - \underline{\xi}}$$

for all $u, u' \in \mathcal{U}$ and some constant. The Hölder space becomes a Banach space when endowed with the Hölder norm as follows:

$$\|h\|_{\Lambda^\xi} = \sup_{u \in \mathcal{U}} |g(u)| + \max_{|a|=\underline{\xi}} \sup_{u \neq u'} \frac{|\nabla^a h(u) - \nabla^a h(u')|}{(\|u - u'\|_E)^{\xi - \underline{\xi}}}.$$

The Hölder ball (with radius c) is defined as $\Lambda_c^\xi(\mathcal{U}) \equiv \{h \in \Lambda^\xi(\mathcal{U}) : \|h\|_{\Lambda^\xi} \leq c \leq \infty\}$. It is well known that power series, splines, Fourier series and wavelets all can approximate functions in $\Lambda_c^\xi(\mathcal{U})$ well.[4]

ASSUMPTION 2.5.1. *(i) $f_{W|W^*} \in \Lambda_c^\xi(\mathcal{W} \times \mathcal{W}^*)$ with some $\xi > 1$ and $\int_{\mathcal{W}} f_{W|W^*}(w|w^*)dw = 1$ for all $w^* \in \mathcal{W}^*$; (ii) $f_{W^*|W'} \in \Lambda_c^\xi(\mathcal{W}^* \times \mathcal{W}')$ with some $\xi > 1$ and $\int_{\mathcal{W}^*} f_{W^*|W'}(w^*|w')dw^* = 1$ for all $w' \in \mathcal{W}'$.*

To simplify notations we use f_1 and f_2 to denote $f_{W|W^*}$ and $f_{W^*|W'}$, respectively. We assume that the unknown functions f_1 and f_2 belongs to the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ defined below:

$$\mathcal{F}_1 = \{f_1(\cdot) : \text{ Assumptions 2.3.3, 2.3.5 and 2.5.1 (i) hold}\},$$

$$\mathcal{F}_2 = \{f_2(\cdot) : \text{ Assumptions 2.3.3 and 2.5.1 (ii) hold}\}.$$

By Proposition 2.3.2 and Kullback-Leibler information criterion, the true value of parameters $\gamma_0 = (\theta_0^T, f_1, f_2)^T$ can be solved by

$$\gamma_0 = \underset{\gamma=(\theta^T, f_1, f_2)^T \in \Gamma}{\operatorname{argmax}} \mathbb{E}\left[\ln \int_{\mathcal{W}^*} f_{Y|X,W^*}(y|x, w^*; \theta) f_1(w|w^*) f_2(w^*|w') dw^*\right],$$

where $\Gamma = \Theta \times \mathcal{F}_1 \times \mathcal{F}_2$. Let $\{p_j^{k_n}(\cdot), j = 1, 2, \cdots\}$ be a sequence of known basis functions (power series, splines, Fourier series, etc.). To approximate f_1 and f_2, we use a tensor-product linear sieve basis, denoted as $p^{k_n}(\cdot,\cdot) = (p_1^{k_n}(\cdot,\cdot), p_2^{k_n}(\cdot,\cdot), \cdots, p_{k_n}^{k_n}(\cdot,\cdot))^T$. The integer k_n is the smoothing parameter, which is required to grow with n so that the approximation error decreases to zero. To conduct sieve approximation we replace $\mathcal{F}_1$ and $\mathcal{F}_2$ by sieve spaces $\mathcal{F}_{1n}$ and $\mathcal{F}_{2n}$ where

$$\mathcal{F}_{1n} = \left\{f_1 : f_1(w|w^*) = p^{k_n}(w, w^*)^T \mu, \ \forall\, \mu \text{ s.t. Assumptions 2.3.3, 2.3.5 and 2.5.1 (i) hold}\right\}$$

[4] Note that we assume the support of W, W^* and W' are bounded, in case of unbounded support we need to use a weighted Hölder ball $\Lambda_c^{\xi,\omega}(\mathcal{U})$ with some weighting function $\omega(\cdot)$ to facilitate the treatment of functions defined on unbounded domains, see [44] for details.

and

$$\mathcal{F}_{2n} = \left\{ f_2 : f_2(w^*|w') = p^{k_n}(w^*, w')^T \zeta, \; \forall\, \zeta \text{ s.t. Assumptions 2.3.3 and 2.5.1 (ii) hold} \right\}.$$

Then we estimate γ_0 by $\hat{\gamma} = (\hat{\theta}^T, \hat{f}_1, \hat{f}_2)^T$ as

$$\hat{\gamma} = \underset{\gamma \in \Gamma_n}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \ln \int_{\mathcal{W}^*} f_{Y|X,W^*}(Y_i|X_i, w^*; \theta) f_1(W_i|w^*) f_2(w^*|W_i') dw^*,$$

where $\Gamma_n = \Theta \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$. In practice, this integration can be conveniently implemented by different numerical techniques including Simpson's rules, Gaussian quadrature and so forth.

We apply the result in [45] to establish consistency of the sieve estimator $\hat{\gamma}$ for γ_0 under a norm $\|\cdot\|_s$, which is defined as below:

$$\|\gamma\|_s = \|\theta\|_E + \|f_1\|_\infty + \|f_2\|_\infty$$

for Euclidean norm $\|\cdot\|_E$ and sup norm $\|\cdot\|_\infty$.

In general, conditional on X^c and G^*, $W_i \perp\!\!\!\perp W_j$ if $N_i \cap N_j = \emptyset$, i.e., players i and j do not share common friends according to the observed data[5]. To ensure the consistency of the sieve estimator, the interdependence of $\{W_i\}_{i \in N}$ should disappear as $n \to \infty$. If the social network has a "circle" structure, as is illustrated in Figure 2.2 (b), $\{W_i\}_{i \in N}$ will be independent, conditioning on X^c and G^*. However, assuming the observed network is a circle is too restrictive and will ignore many interesting network structures. Let $C_i \equiv \{j \in N \backslash \{i\} : N_i \cap N_j \neq \emptyset\}$ denote the set of players that share common friends with i. If we can bound the number of players who share comm friends in the large network, then it is easy to verify that $\{W_i\}_{i \in N}$ will have finite dependence. To formalize this intuition, define $C = \cup_{i \in N} C_i$ to be the set of players sharing common friends in the network. We need to bound the cardinality of C: if there exists a $m < \infty$ such that $C <\leq m$, then it is

[5]Since W_i contains information about player i's friends' actions, and their actions are independent with each other when conditioning on X^c and G^*, therefore if i and j share common friends, then those friends' actions will enter both W_i and W_j, making W_i and W_j dependent with each other.

straightforward to verify that conditional on X^c and G^*, $\{W_i\}_{i\in N}$ will be m-dependent. Similar logic also applies to $\{W_i'\}_{i\in N}$. Therefore we impose the following assumption:

Figure 2.2: Illustration of Observed Network Structures

(a) Common friends

(b) Circle

Remark. *Figure 2.2 provides a graphic illustration of our notations. In Figure 2.2 (a), i and k are friends with each other and they share a common friend j, therefore $N_i \cap N_k = \{j\}$, $C_i = \{k\}$, $C_k = \{i\}$, $C_j = \emptyset$ and $C = \{i, k\}$ in this 3-players network. In Figure 2.2 (b) i, j and k do not share a common friend with each other and hence $C_i = C_j = C_k = C = \emptyset$.*

ASSUMPTION 2.5.2. *$\{W_i\}_{i=1}^n$ and $\{W_i'\}_{i=1}^n$ are m-dependent and identically distributed, conditional on X^c and G^*.*

By Assumption 2.5.2, the observed data $\{(Y_i, X_i, W_i, W_i')_{i=1}^n\}$ is weakly dependent (m-dependent) and identically distributed, conditional on X^c and G^*.

As in [2], we also define the projection of $\gamma \in \mathit{\Gamma}$ onto the space $\mathit{\Gamma}_n$ as

$$\mathit{\Pi}_n \gamma \equiv \underset{\gamma=(\theta^T, f_1, f_2)^T \in \mathit{\Gamma}_n}{\operatorname{argmax}} \mathbb{E}\left[\ln \int_{\mathcal{W}^*} f_{Y|X,W^*}(y|x, w^*; \theta) f_1(w|w^*) f_2(w^*|w') dw^*\right]$$

and impose the following assumption:

ASSUMPTION 2.5.3. *$\|\mathit{\Pi}_n \gamma_0 - \gamma_0\| = o(n^{-1/4})$ as $k_n \to \infty$ and $k_n/n \to 0$.*

Assumption 2.5.3 implies that the sieve can approximate the true value γ_0 arbitrarily well and guarantee that the number of terms in the sieve grows slower than than the sample size so that the bias and variance of sieve approximation can be controlled. It can be satisfied by using many commonly used sieve functions such as power series, splines and so forth.

ASSUMPTION 2.5.4. *(i)* $\Theta \subset \mathbb{R}^{d+1}$ *is compact; (ii)* $\theta_0 \in \mathrm{int}(\Theta)$.

Assumption 2.5.4 (i) ensures that the parameter space Γ is compact under the pseudo norm $\|\cdot\|_s$ and is commonly imposed in the nonparametric and semiparametric econometrics literature, see [46] for detailed discussion about this condition. Assumption 2.5.4 (ii) is standard and requires θ_0 to be an “interior" solution.

Define $D = (y, x, w, w')$ for $y \in \{0, 1\}$, $x \in \mathcal{X}$, $w \in \mathcal{W}$ and $w' \in \mathcal{W}'$ and follow [2], we also impose the following restrictions on the log-likelihood function:

ASSUMPTION 2.5.5. *(i)* $\mathbb{E}[(\ln f_{YW|W',X}(D,\gamma))^2]$ *is bounded; (ii)* $\ln f_{YW|W',X}(D,\gamma)$ *is Hölder continuous in* γ.

Assumption 2.5.5 guarantees a Hölder continuity property for the log-likelihood function, specifically it imposes an envelope condition on the derivative of the log-likelihood function and will be used to characterize a stochastic equicontinuity condition in [47]. With previous assumptions, we establish the consistency of $\hat{\gamma}$ in the following theorem:

Theorem 2.5.1. *Under Assumptions 2.2.1-2.3.5 and 2.5.1-2.5.5, we have* $\|\hat{\gamma} - \gamma_0\|_s = o_p(1)$.

Theorem 2.5.1 provides the consistency result under the norm $\|\cdot\|_s$. Nevertheless, it is relatively difficult to derive the asymptotic normality and $\sqrt{n}$ consistency result under $\|\cdot\|_s$ since it is too strong to obtain a convergence rate faster than $n^{-1/4}$. Following [43], we employ a weaker norm $\|\cdot\|$ to establish the asymptotic normality of $\hat{\theta}$.

Before introducing the norm $\|\cdot\|$, we first review the concept of pathwise derivative, consider γ_1 and $\gamma_2 \in \Gamma$, and assume the existence of a continuous path $\{\gamma(\tau) : \tau \in [0, 1]\}$ in Γ such that $\gamma(0) = \gamma_1$ and $\gamma(1) = \gamma_2$. Also assume that Γ is convex at the true value γ_0 in the sense that for any $\gamma \in \mathit{\Gamma}$ and $\tau \in (0,1)$, $(1-\tau)\gamma_0 + \tau\gamma \in \mathit{\Gamma}$. If $\ln f_{YW|W'X}(D, (1-\tau)\gamma_0 + \tau\gamma)$ is continuously differentiable at $\tau = 0$ for almost all D and any $\gamma \in \mathit{\Gamma}$, the first pathwise derivative of $\ln f_{YW|W'X}$ at γ_0 evaluated at the direction $[\gamma - \gamma_0]$ can be defined as

$$\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[\gamma - \gamma_0] \equiv \frac{d \ln f_{YW|W'X}((1-\tau)\gamma_0 + \tau\gamma)}{d\tau}\bigg|_{\tau=0} \text{ a.s. } D.$$

We define the inner product $\langle \gamma, \gamma \rangle$ as

$$\langle \gamma, \gamma \rangle = \mathbb{E}\left[\left(\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[\gamma]\right)^2\right]$$

with the induced Fisher norm $\|\gamma\|$ defined as $\sqrt{\langle \gamma, \gamma \rangle}$. Now we derive the asymptotic distribution of $\hat{\theta}$. As before we must first introduce some notations. Let $\overline{\mathit{\Gamma}} = \mathbb{R}^{d+1} \times \overline{\mathcal{F}}$ with $\overline{\mathcal{F}} = \overline{\mathcal{F}_1 \times \mathcal{F}_2} - \left\{(f_{W|W^*}, f_{W^*|W'})^T\right\}$ denote the closure of the linear span of $\mathit{\Gamma} - \{\gamma_0\}$ under the norm $\|\cdot\|$ and define the Hilbert space $(\overline{\mathit{\Gamma}}, \|\cdot\|)$. Then we can write

$$\begin{aligned}\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[\gamma - \gamma_0] =& \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{d\theta}[\theta - \theta_0] + \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df_1}[f_1 - f_{W|W^*}] \\ &+ \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df_2}[f_2 - f_{W^*|W'}].\end{aligned}$$

For each component θ_j of θ, $j = 1, 2, \cdots, d+1$, let $m_j^* \equiv (f_{1j}^*, f_{2j}^*)^T \in \overline{\mathcal{F}}$ denote the solution to

$$\min_{m_j \in \overline{\mathcal{F}}} \mathbb{E}\left\{\left(\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{d\theta_j} - \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df_1}[f_{1j}] - \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df_2}[f_{2j}]\right)^2\right\}.$$

Define $m^* = (m_1^*, m_2^*, \cdots, m_{d+1}^*)$,

$$\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df}[m_j^*] = \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df_1}[f_{1j}^*] + \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df_2}[f_{2j}^*],$$

$$\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df}[m^*] = \left(\frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df}[m_1^*], \cdots, \frac{d \ln f_{YW|W'X}(D,\gamma_0)}{df}[m_{d+1}^*]\right)^T$$

and the column vector

$$G_{m^*}(D) = \frac{d \ln f_{YW|W'X}(D, \gamma_0)}{d\theta} - \frac{d \ln f_{YW|W'X}(D, \gamma_0)}{df}[m^*]. \tag{2.21}$$

In order to show that $\hat{\theta}$ has a multivariate normal distribution asymptotically, we instead can show that $\lambda^T \hat{\theta}$ has a normal distribution for all $\lambda \in \mathbb{R}^{d+1}$. Following Ai & Chen (2003) for $s(\gamma) = \lambda^T \theta$ with $\lambda \neq 0$, which is a linear functional on $\overline{\Gamma}$, we have

$$\sup_{0 \neq \gamma - \gamma_0 \in \overline{\Gamma}} \frac{|s(\gamma) - s(\gamma_0)|^2}{\|\gamma - \gamma_0\|^2} = \lambda^T \left(\mathbb{E}\{G_{m^*}(D) G_{m^*}(D)^T\}\right)^{-1} \lambda.$$

Therefore, in order for the functional $s(\gamma)$ to be bounded, $\mathbb{E}\{G_{m^*}(D) G_{m^*}(D)^T\}$ has to be positive definite. Then by Riesz representation theorem, there exists a representer γ^* such that

$$s(\gamma) - s(\gamma_0) = \langle \gamma^*, \gamma - \gamma_0 \rangle \text{ for all } \gamma \in \Gamma, \tag{2.22}$$

where $\gamma^* = (\gamma^*_\theta, \gamma^*_f) \in \overline{\Gamma}$ with $\gamma^*_\theta = (\mathbb{E}\{G_{m^*}(D) G_{m^*}(D)^T\})^{-1} \lambda$ and $\gamma^*_f = -m^* \gamma^*_\theta$. (2.22) implies that under suitable conditions, it is sufficient to find the asymptotic distribution of $\langle \gamma^*, \hat{\gamma} - \gamma_0 \rangle$ to obtain that of $s(\hat{\gamma}) - s(\gamma_0) = \lambda^T (\hat{\theta} - \theta_0)$.

Define

$$\mathcal{N}_{0n} = \{\gamma \in \Gamma_n : \|\gamma - \gamma_0\|_s = o(1), \|\gamma - \gamma_0\| = o(n^{-1/4})\}$$

and define $\mathcal{N}_0$ the same way with Γ_n replaced by Γ. For $\gamma \in \Gamma_n$, let $\gamma^*(\gamma, \varepsilon_n) = (1 - \varepsilon_n)\gamma + \varepsilon_n(\gamma^* + \gamma_0)$ with $\varepsilon_n = o(n^{-1/2})$ be a local alternative value and denote $P_n \gamma^*(\gamma, \varepsilon_n)$ by a projection of $\gamma^*(\gamma, \varepsilon_n)$ onto Γ_n. The following conditions are sufficient for the $\sqrt{n}$-normality of $\hat{\theta}$:

ASSUMPTION 2.5.6. *(i) There exists a measurable function $c(D)$ with $\mathbb{E}[c(D)^4] < \infty$ such that $|\ln f_{YW|W'X}(D, \gamma)| \leq c(D)$ for all D and $\gamma \in \Gamma_n$; (ii) $\ln f_{YW|W'X}(D, \gamma) \in \Lambda^\xi_c(\mathcal{Y} \times \mathcal{W} \times \mathcal{W}' \times \mathcal{X})$ for some constant $c > 0$ with $\xi > d_D/2$, for all $\gamma \in \Gamma_n$, where d_D is the dimension of D. (iii) Γ is convex in γ_0 and $f_{Y|W^*,X;\theta}$ is pathwise differentiable at θ_0; (iv) There exists c_1 and $c_2 > 0$ such* *that* $c_1 KL(\gamma,\gamma_0) \leq \|\gamma-\gamma_0\|_s^2 \leq c_2 KL(\gamma,\gamma_0)$ *holds for all* $\gamma \in \Gamma_n$ *with* $\|\gamma-\gamma_0\|_s = o(1)$*, where* $KL(\gamma,\gamma_0) \equiv \mathbb{E}\left[\ln \frac{f_{YW|W'X}(D,\gamma_0)}{f_{YW|W'X}(D,\gamma)}\right]$ *is the Kullback-Leibler information.*

Assumption 2.5.6 (i) and (ii) impose an envelope condition and a smoothness condition on the log likelihood function. Condition (iii) ensures that the Fisher norm $\|\cdot\|$ is well defined. Condition (iv) guarantees that the population criterion function can be approximated locally by $\|\gamma-\gamma_0\|^2$. Assumption 2.5.6 together with Assumption 2.5.3 ensure that under the weaker norm $\|\cdot\|$, the sieve estimator will converge at the rate $n^{-1/4}$, which is a prerequisite to derive the asymptotic distribution of $\hat{\theta}$.

ASSUMPTION 2.5.7. *(i)* $\mathbb{E}\{G_{m^*}(D)G_{m^*}(D)^T\}$ *is bounded and positive-definite; (ii) There exists a* $\gamma_n^* = P_n\gamma^*(\gamma,\varepsilon_n) \in \Gamma_n - \{\gamma_0\}$ *such that* $\|\gamma_n^*-\gamma^*\| = o(n^{-1/4})$*; (iii) For all* $\gamma \in \mathcal{N}_{0n}$*,* $KL(\gamma,\gamma_0) = \frac{1}{2}\|\gamma-\gamma_0\|^2(1+o(1))$.

Assumption 2.5.7 (i) implies that θ_0 is locally identified. Condition (ii) requires that the Riesz representer v^* can be approximated well by the sieve space, which is necessary to ensure that the bias of the sieve estimator is asymptotically negligible. Condition (iii) indicates that $KL(\cdot,\cdot)$ is locally equivalent to $\|\cdot\|^2$, which characterizes the local quadratic behavior of the criterion difference, i.e., Condition B.2 in [47]. By checking conditions in Theorem 2 of [47], we show that the estimator for the structural parameter θ_0 is $\sqrt{n}$ consistent and follows an asymptotic normal distribution.

Theorem 2.5.2. *Under Assumptions 2.2.1-2.3.5 and 2.5.1-2.5.7,* $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,V)$*, where*

$$V = \left[\mathbb{E}\{G_{m^*}(D_1)G_{m^*}(D_1)^T\}\right]^{-1} + 2\sum_{j=1}^{m}\mathbb{E}\left\{\left[\mathbb{E}\{G_{m^*}(D_1)G_{m^*}(D_1)^T\}\right]^{-1} G_{m^*}(D_1)G_{m^*}(D_j)^T \left[\mathbb{E}\{G_{m^*}(D_j)G_{m^*}(D_j)^T\}\right]^{-1}\right\}.$$

2.6 Monte Carlo Simulations

This section illustrates the finite sample performance of the proposed estimator using simulated data. Specifically we consider a simple binary game with linear payoff:

$$U_{i1} = \alpha_0 + \alpha_1 X_i + \beta W_i(Y_{N_i^*}, G_i^*) - \epsilon_i$$

with

$$W_i(Y_{N_i^*}, G_i^*) = \frac{\sum_{j \in N_i^*} Y_j}{\sum_{j \neq i} G_{ij}^*}$$

representing the proportion of player i's friends that will choose action 1. We assume that $\epsilon | X, W^* \sim N(0,1)$ so that the density $f(Y_i^* | X_i, W_i^*)$ will have the form

$$f(Y_i^*|X_i, W_i^*) = \Phi(\alpha_0 + \alpha_1 X_i + \beta W_i^*)^{Y_i^*}[1 - \Phi(\alpha_0 + \alpha_1 X_i + \beta W_i^*)]^{1-Y_i^*},$$

where $\Phi(\cdot)$ is the CDF of standard normal random variable.

In the simulations, the payoff covariate X_i is randomly drawn from a standard normal distribution. Furthermore, we generate the latent random social network as follows: the whole sample is divided into 20 equally sized subnetworks with each having $n/20$ players and those subnetworks are placed on a line, indexed as $1, 2, \cdots$. For any two players i and j, $i \neq j$ within the same subnetwork, the true network connections $G_{ij}^* \in \{0,1\}$ is drawn independently from the probability mass distribution $(1 - \frac{20}{n}, \frac{20}{n})$, specifically $G_{ij}^* = \mathbb{1}(\eta_{g^*} > 1 - \frac{20}{n})$ for $\eta_{g^*} \sim U(0,1)$. Moreover, $G_{ii}^* = 0$ for all $i \in N$. To ensure the weak dependence of network data, we require that for players in different subnetworks l and m, if $|l - m| = 1$, $G_{ij}^* = \mathbb{1}(\eta_{g^*} > 1 - \frac{20}{n} \cdot \frac{1}{10})$; If $|l - m| > 1$, $G_{ij}^* = 0$. The instrumental variable G' are generated as $G'_{ij} = \mathbb{1}(0.6\eta_{g^*} + 0.4\eta_z > 0.2)$, where $\eta_z \sim U(0,1)$ for players within the same subnetwork and $G'_{ij} = G_{ij}^*$ for players in different subnetworks. The correlation between G^* and G' is captured by η_{g^*}.

The distribution of the measurement error is specified in the matrix $\mathbf{P}_{G_{ij}|G_{ij}^*}$ with $Pr(G_{ij} = k | G_{ij}^* = k')$ for k and $k' \in \{0,1\}$ in each entry. The specification uses the constant misclassification probabilities for players in the same subnetwork as follows:

$$\mathbf{P}_{G_{ij}|G^*_{ij}} = \begin{bmatrix} Pr(G_{ij}=0|G^*_{ij}=0) & Pr(G_{ij}=0|G^*_{ij}=1) \\ Pr(G_{ij}=1|G^*_{ij}=0) & Pr(G_{ij}=1|G^*_{ij}=1) \end{bmatrix}.$$

Note that the elements on the diagonal of $\mathbf{P}_{G_{ij}|G^*_{ij}}$ represent the probability of correct reporting. For a given value of G^*_{ij}, the value of G_{ij} is generated according to $\mathbf{P}_{G_{ij}|G^*_{ij}}$ and another independent random variable $\eta_g \sim U(0,1)$ as $G_{ij} = \mathbb{1}(\eta_g > 1 - Pr(G_{ij} = 1|G^*_{ij}))$ for players in the same subnetwork and $G_{ij} = G^*_{ij}$ for players in different subnetworks. The values of $\mathbf{P}_{G_{ij}|G^*_{ij}}$ will be specified for the experiment. In the current experiment we choose

$$\mathbf{P}_{G_{ij}|G^*_{ij}} = \begin{bmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{bmatrix}.$$

Simulation results under different specifications of $\mathbf{P}_{G_{ij}|G^*_{ij}}$ are provided in Appendix B. We have performed several experiments with the number of players $n = 1000$. In each iteration of the experiment, we first compute the semi-anonymously symmetric BNE by solving the fixed point of the equilibrium mapping given the underlying parameter value $\alpha_0 = 1$, $\alpha_1 = 1$ and $\beta = 1$. With the numerical solution in hand, we can simulate the equilibrium actions Y for each player.

Regrading estimation, we consider three maximum likelihood estimators: (i) the inconsistent estimator obtained when we ignore measurement errors and treat G as the true network, (ii) the infeasible estimator obtained using the latent true network graph G^* and (iii) the proposed sieve estimator using the IV G'. According to Assumption 2.3.5 the identification restriction imposed for the sieve MLE is the zero mode assumption, i.e., $M[f] = \operatorname{argmax}_{x\in\mathcal{X}} f(x)$. The sieves of unknown functions f_1 and f_2 are constructed through tensor product bases of truncated univariate trigonometric series. Since W_i, W_i^* and $W_i' \in [0,1]$ for all $i \in N$, we have

$$\begin{aligned} f_{W|W^*}(w|w^*) &= \sum_{i=0}^{i_n}\sum_{j=0}^{j_n} \mu_{ij} q_i(w-w^*) q_j(w^*) \\ &= \sum_{i=0}^{i_n}\sum_{j=0}^{j_n} [\mu_{1ij}\cos(i\pi)(w-w^*) + \mu_{2ij}\sin(i\pi)(w-w^*)]\cos(j\pi)w^* \end{aligned}$$

and

$$\begin{aligned} f_{W^*|W'}(w^*|w') &= \sum_{i=0}^{i_n}\sum_{j=0}^{j_n} \zeta_{ij} q_i(w^*-w') q_j(w') \\ &= \sum_{i=0}^{i_n}\sum_{j=0}^{j_n} [\zeta_{1ij}\cos(i\pi)(w^*-w') + \zeta_{2ij}\sin(i\pi)(w^*-w')]\cos(j\pi)w', \end{aligned}$$

where

$$q_i(w-w^*) = (\cos(i\pi)(w-w^*), \sin(i\pi)(w-w^*))^T \text{ and } q_j(w^*) = \cos(j\pi)w^*,$$

$$q_i(w^*-w') = (\cos(i\pi)(w^*-w'), \sin(i\pi)(w^*-w'))^T \text{ and } q_j(w') = \cos(j\pi)w',$$

and $\mu_{ij} = (\mu_{1ij}, \mu_{2ij})$, $\zeta_{ij} = (\zeta_{1ij}, \zeta_{2ij})$. As in [2], it is fairly straightforward to show that the restriction $\int_{\mathcal{W}} f_1(w|w^*)dw = 1$ implies that $\mu_{100} = \frac{1}{2}$ and $\mu_{10j} = 0$ for all $j = 0, 1, \cdots, j_n$ and similarly $\int_{\mathcal{W}^*} f_1(w^*|w')dw^* = 1$ implies $\zeta_{100} = \frac{1}{2}$ and $\zeta_{10j} = 0$ for all $j = 0, 1, \cdots, j_n$. Furthermore, the zero mode restriction implies that $\sum_{i=1}^{i_n} \frac{(-1)^i}{i}\mu_{2ij} = 0$ and $\sum_{i=1}^{i_n} \frac{(-1)^i}{i}\zeta_{2ij} = 0$ for all $j = 0, 1, \cdots, j_n$. We will incorporate these restrictions when maximizing the sieve MLE objective function.

The implementation of the sieve method requires appropriate selection of the smoothing parameter $k_n = (i_n + 1)(j_n + 1)$. A formal selection method for k_n and proof its asymptotic validity is beyond the scope of this paper. However, since it is well known that the asymptotic distribution semiparametric sieve estimators is identical in a wide range of smoothing parameter sequences, following [2] and [48] we choose the smoothing parameter by locating a range of values where the estimates are not very sensitive to small variations in i_n and j_n in simulations.

The simulation results are provided in Table 2.1. For each estimator, we report the mean, the standard deviation (Std.dev), and the mean squared error (MSE) of the estimators averaged over all 500 replications. The simulation results indicate that if we ignore the measurement errors and naively conduct MLE, the estimated peer effects will be severely biased. On the other hand our proposed Sieve MLE performs well in reducing the bias and MSE caused by the presence of measurement errors for the parameter of interest β, which represents the peer effects. Furthermore, our method can also reduce the bias and MSE in estimates of α_0. We change the values of smoothing parameters and the sieve estimates are not very sensitive those changes and hence suggests that the selected smoothing parameters are valid. In Appendix B we provide additional simulation results.

Table 2.1: Simulation Results

	Parameter(=True Value)								
	$\alpha_0 = 1$			$\alpha_1 = 1$			$\beta = 1$ (Peer Effects)		
	Mean	Std.dev	MSE	Mean	Std.dev	MSE	Mean	Std.dev	MSE
Ignoring meas. error	1.4023	1.1326	1.4420	0.9114	0.0744	0.0134	0.0560	1.3149	2.6166
Accurate data	1.0096	0.0890	0.0080	1.0144	0.0818	0.0069	1.0093	0.1363	0.0186
Sieve MLE	0.9824	0.2682	0.0721	0.9137	0.1097	0.0195	1.0862	0.5406	0.2991
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 2, j_n = 3$ in f_2.									
Ignoring meas. error	1.4099	1.1298	1.4418	0.9092	0.0745	0.0138	0.0398	1.3119	2.6396
Accurate data	1.0078	0.0871	0.0076	1.0095	0.0857	0.0074	0.9983	0.1404	0.0197
Sieve MLE	1.0055	0.2581	0.0665	0.9063	0.0927	0.0174	1.0077	0.5464	0.2980
Smoothing parameters: $i_n = 3, j_n = 4$ in f_1; $i_n = 3, j_n = 4$ in f_2.									
Ignoring meas. error	1.3307	1.1168	1.3541	0.9056	0.0772	0.0149	0.1300	1.2997	2.4428
Accurate data	0.9991	0.0916	0.0084	1.0079	0.0837	0.0071	1.0112	0.1316	0.0174
Sieve MLE	0.9879	0.2552	0.0651	0.9104	0.1030	0.0186	1.0554	0.5576	0.3134
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 6, j_n = 4$ in f_2.									
n=1000, replication=500									

2.7 Empirical Application

In this section, we use our proposed method to analyze the peer effects on youth alcohol drinking behaviors. Recently, there is a growing body of empirical literature on studying the peer effects on adolescents behaviors, see e.g., [8], [10], [7], [38] and references therein. The data we used is obtained from the National Longitudinal Study of Adolescent to Adult Health (Add Health), which is a database designed to investigate the relationship between the social environment and adolescents' behaviors. The Wave I data contains a nationally representative sample of more than 90,000 students in grades 7-12 from 80 high schools and 52 middles schools in the United States during the 1994-1995 school year and the second wave surveyed almost 15,000 of the same students one year after Wave I. In the data every student was asked to complete a questionnaire to provide information about his or her socioeconomic characteristics as well as school-related behaviors and friendship network.[6]

A unique feature of the Add Health data is that it contains information about respondents' social network information. In the survey questionnaire, each student is asked to provide his or her friendship information by nominating at most five male and female best friends. However, the restriction on the number of friends to be nominated can plausibly lead to measurement error because students with more than 10 friends will not be able to provide information on all his or her friendship network. [38] compare peer effects by using Wave I and Wave II network data and find qualitatively similar peer effects, despite the fact that in the data friendship network tend to change substantially between two waves. Based on this empirical finding, they cast doubt about the accuracy of self-reported friendship links.

Our empirical strategy is to use the Wave II network data as the instrumental variable for the Wave I network G. Given the fact that the Wave II data was surveyed after one year of Wave I, it is convincing that the exclusion restrictions in Assumption 2.3.2 are satisfied. Following the literature, the covariates we used include age, GPA, race information, gender and family income with sample size $n = 1,528$. The summary statistics for variables used in our empirical analysis are presented in Table 2.2.

We use a similar specification of the payoff function and sieve basis functions as in the simulation and the smoothing parameter k_n is also selected by finding a range of values where the

[6]See the Add Health website (http://www.cpc.unc.edu/projects/addhealth) for a detailed description of surveys and data.

Table 2.2: Descriptive Statistics (1,528 observations)

Variable	Mean	Std. Dev.	Min	Max
Female	0.4980	0.5002	0	1
Age	15.5157	1.5762	12	20
White	0.7251	0.4466	0	1
GPA	3.1291	0.5731	1.75	4
Income	53.1198	45.2244	0	900
Alcohol	0.4928	0.5001	0	1

The unit for income is in thousand dollars.

estimates are not very sensitive to small variations in i_n and j_n. Our empirical results are presented in Table 2.3. Note that we also report the results of MLE when ignoring measurement errors for the purpose of comparison. The standard errors in parentheses are calculated using 400 bootstrap samples.

Table 2.3: Empirical Results

Variable	MLE ignoring meas. error	Sieve MLE
Female	0.0401	0.2662
	(0.0675)	(0.7418)
Age	0.1866***	0.1004***
	(0.0221)	(0.0241)
White	0.2875***	0.1093***
	(0.0776)	(0.0340)
GPA	-0.4282***	-0.4920**
	(0.0605)	(0.2102)
Income	0.0013	0.0008
	(0.0008)	(0.0040)
Constant	-1.9648***	-1.3716*
	(0.4092)	(0.7162)
Peer Effects	0.5688***	1.1357***
	(0.1257)	(0.2327)

Standard errors in parentheses are calculated using 400 bootstrap samples
Smoothing parameters: $i_n = 5, j_n = 3$ in f_1; $i_n = 4, j_n = 5$ in f_2.
* 10% significant, ** 5% significant, *** 1% significant.

In Table 2.3, most of the estimated coefficients are significant at the 10% significance level. By comparing the results in Columns 2 and 3, we find that when the measurement errors in network data are ignored, the peer effects estimated using MLE are positive and statistically significant (i.e., 0.5688 with a standard error 0.1257). The estimated peer effects ignoring measurement errors is qualitatively similar to those empirical results in [7] and [8], who use Add Health data and the NELS data to study school-based peer effects on youth behaviors. However, the estimated peer effects are biased in the presence of measurement errors. Using the proposed sieve estimator, we find consistent and much larger estimated peer effects (1.1357 with a standard error of 0.2304) on youth alcohol drinking behaviors. This is equivalent to a 39.15% difference in the average partial effect of actions from peer group. Hence our empirical results demonstrate that if the measurement errors in network data are ignored, and then peer effects will be significantly underestimated.

The accuracy of the estimated peer effects can be crucial for policymakers who wish to establish some methods to control for teenagers' smoking and alcohol drinking behaviors. For example policymakers can impose additional sales taxes on cigarette and alcoholic beverage. Nevertheless, in order to determine the optimal tax rates they need to know how sensitive will adolescent react to the increases in tax, i.e., tax elasticities. As is demonstrated in [10], the smoking and alcohol actions from peer group can have significant social multiplier effects on the tax elasticities. Consequently, the effectiveness of tax policies depends upon whether we can estimate the peer effects accurately.

2.8 Conclusion

We have developed an econometric framework to nonparametrically identify CCPs and peer effects of a network game with incomplete information, allowing for the presence of measurement error in network connections. In particular, we show that under the large game setting, the CCPs are asymptotically equivalent to the ones that are conditional on players' own characteristics and a scalar valued function of their network structure. Hence the CCPs can be nonparametrically identified by applying the method in [2]. Then the payoff primitives are proved to be identified up to a monotone transformation. We also propose a semiparametric method to consistently estimate the peer effects. As an application of the proposed methods, we study the peer effects of adolescent alcohol drinking behaviors and find that the peer effects will be significantly underestimated when measurement are ignored.

It is interesting to see what we can do for the inference of peer effects without the availability of an instrumental variable. From the analysis in Section 2.4 we know that the CCPs will no longer be point-identified, instead we will obtain a sharp identification region for the CCPs. With the partially identified CCPs, the peer effects may or may not be point-identified, depending on the specific identification assumptions imposed in the model. The estimation method proposed in [3] can be used to estimate the sharp bounds. Then peer effects can be estimated by exploring the literature of moment inequality models. We leave this for future research.

Another important extension of the current methodology would be to apply it to the new but fast growing area of network formation econometrics. Under the setup of network formation model, we will have measurement error in outcome variables and a method dealing with the identification of CCPs in this case will be required.

3. NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF ADDITIVE SOCIAL INTERACTION MODELS WITH HOMOPHILY

3.1 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 ([49], employment ([50]), in-school achievements ([7]), adolescent behavior ([8]; [10]; [51]), 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 [23], [26], [52] and [27]. 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" ([53]).

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 [31].

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 [54] and [55], 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 literature, 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 ([33]). 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 [36], 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 [9] 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 [56]. 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., [10] and [11]. 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. [57] 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., [58]; [40]; [40]; [59]). 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 [57], but the objectives of our paper and [57] 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, [51] 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 work in social interaction and incomplete information game adopts parametric or semiparametric method for identification. For example, [22] and [26] assume that the utility function is linear, [23], [40] and [27] 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. [60] 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 [61] and [62]. 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., [63], [64] and [65], 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 essay is organized as follows. Section 3.2 presents the setting and basic assumptions of our model. Section 3.2 provides the identification method. Section 3.4 discusses the estimation method and establishes the asymptotic behavior of our proposed estimator. Section 3.5 contains empirical analysis of peer effects on youth smoking behaviors. Section 3.6 concludes. All proofs are provided in Appendix B.

3.2 The model

3.2.1 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

$$U_{ik}(Y_{-i}, X_i, X_{-i}, Z_{ik}, \epsilon_i) = \alpha(X_i, Z_{ik}) + \sum_{j \neq i} \beta(Y_j, X_i, Z_{ik}) \cdot \gamma(H_{ij}) + \epsilon_{ik}, \tag{3.1}$$

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.

$$H_{ij} \equiv d(X_i, X_j) \tag{3.2}$$

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 3.2.1. *(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 3.2.1, 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. [22]; [26]). 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.

3.2.2 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 $Z_i = (Z_{i0}^T, Z_{i1}^T, \cdots, Z_{iK}^T)^T$, $S_i = (X_i^T, Z_i^T)^T$ and $S = (S_1, S_2, \cdots, S_n)$ 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

$$\begin{aligned} Y_i^*(S, \epsilon_i; \theta) &= \underset{k \in A}{\operatorname{argmax}}\ \mathbb{E}[U_{ik}(Y_{-i}, X_i, X_{-i}, Z_{ik}, \epsilon_i)|S, \epsilon_i] \\ &= \underset{k \in A}{\operatorname{argmax}} \left\{ \alpha(X_i, Z_{ik}) + \sum_{l=0}^{K} \left[\beta(l, X_i, Z_{ik}) \sum_{j \neq i} Pr(Y_j^*(S, \epsilon_j; \theta) = l|S, \epsilon_i) \gamma(H_{ij}) \right] + \epsilon_{ik} \right\}. \end{aligned} \tag{3.3}$$

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

ASSUMPTION 3.2.2. *(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 3.2.2 is commonly imposed in the literature on identification and estimation of static games with incomplete information and social interaction models (see, e.g., [31], [40] and

[27]). 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 $\sigma_{ik}(S;\theta) \equiv Pr(Y_i^*(S,\epsilon_i;\theta) = k|S)$ be the equilibrium conditional choice probability of player i choosing action k. To simplify notation, let

$$V_i(X_i, Z_{ik}, S) \equiv \alpha(X_i, Z_{ik}) + \sum_{l=0}^{K}\left[\beta(l, X_i, Z_{ik})\sum_{j\neq i}\sigma_{jl}(S;\theta)\gamma(H_{ij})\right]. \tag{3.4}$$

Then a BNE solution (given state S) can be characterized by

$$\begin{aligned}
\sigma_{ik}(S;\theta) &= Pr[(V_i(X_i, Z_{ik}, S) + \epsilon_{ik} > V_i(X_i, Z_{ih}, S) + \epsilon_{ih})|S], \ \ \forall h \in A\backslash\{k\} \\
&= Pr[\epsilon_{ih} < (V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S) + \epsilon_{ik})|S], \ \ \forall h \in A\backslash\{k\} \\
&= \int_{\epsilon\in\mathbb{R}}\left[\prod_{h\neq k} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S))\right] f_{\epsilon_{ik}|S}(\epsilon)d\epsilon,
\end{aligned} \tag{3.5}$$

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

For any given $(S;\theta)$ and based on (3.5), we can define a mapping $\Gamma^{(S;\theta)} : \Delta \to \Delta$ such that

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

with

$$\begin{aligned}
\Gamma_j^{(S;\theta)}(\{\sigma_{ik}(S;\theta)\}_{i\neq j,k\in A}) &= (\sigma_{j0}, ..., \sigma_{jK})^T \\
&\equiv \left(\Gamma_{j0}^{(S;\theta)}(\{\sigma_{ik}(S;\theta)\}_{i\neq j,k\in A}), ..., \Gamma_{jK}^{(S;\theta)}(\{\sigma_{ik}(S;\theta)\}_{i\neq j,k\in A})\right)^T
\end{aligned}$$

,

where Δ 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, π_{ij} maps the index i to the index j, j to i, and i' to itself for all $i' \neq i, j$. Similarly, define π_{ij}^X as a function that permutes the ith and jth elements of any $X \equiv (X_1, .., X_n)^T \in \mathcal{X}^n$; and π_{ij}^Z as a function that permutes the ith and jth elements of any $Z \equiv (Z_1, .., Z_n)^T \in \mathcal{Z}^{n(K+1)}$. We thus have the set of permutations $\Pi \equiv \left\{ \left(\pi_{ij}, \pi_{ij}^X, \pi_{ij}^Z\right) | i, j \in N \right\}$ with the generic element written as $\pi(\cdot)$.

Definition 3.2.1. *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, ϵ). We hence, as in [36], 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 ϵ^n represent the sequentialization of S and ϵ using the number of players. In particular, we can make sense of ξ^n as a public signal that players may use to coordinate on a particular equilibrium [1]. Most importantly, we assume that ξ^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 σ^n denotes the sequentialization of σ 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 ϵ_i for all $i \in N$.

ASSUMPTION 3.2.3. *(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 $\overline{Y} \equiv (\overline{Y}_1, ..., \overline{Y}_i, ..., \overline{Y}_n)^T$ with $\overline{Y}_i \in A$,*

$$Pr(Y^n = \overline{Y} | S^n) = \sum_{\sigma^n \in \Delta^{SE}(S^n; \theta)} Pr(\rho_n (S^n, \xi^n; \theta) = \sigma^n | S^n) \prod_{i=1}^{n} \sigma_i^n \left(\overline{Y}_i | S^n\right),$$

[1]The inclusion of ξ^n is to ensure that the selection mechanism is nondegenerate, see Leung (2015) for details.

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

Intuitively Assumption 3.2.3 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 3.2.4. *(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 $U_{ik}\left(\sigma_{-i}, S, \epsilon_{ik}\right) = U_{\pi(i)k}\left(\sigma_{-\pi(i)}, \pi(S), \epsilon_{\pi(i)k}\right)$.*

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., [36] and [66]). Therefore, player labels in the data set have no economic relevance. It also ensures that the equilibria are extensively robust in the sense of [39] even if the simultaneous-play assumption is relaxed.

ASSUMPTION 3.2.5. *(Continuity) For all $\theta \in \Theta$, $i \in N$, $k \in A$ and any realization $\epsilon_{ik} \in \mathbb{R}$, payoffs $U_{ik}\left(\sigma_{-i}, S, \epsilon_{ik}\right)$ are continuously differentiable in S.*

Assumption 3.2.5 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 3.2.1. *Suppose Assumptions 3.2.2-3.2.5 hold. Then there exists a symmetric Bayesian-Nash equilibrium.*

3.3 Identification

In this section, we provide a nonparametric method to explore the identifiability of the structural parameter θ in the sense similar to [67] and [9], i.e., different values of θ will result in different choice probabilities. [67] and [9] discuss 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 θ will be identified as well if different values of θ 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 [9], 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 θ by imposing a rank condition similar to [40] and [27]. First, we introduce the following definition of identification:

Definition 3.3.1. *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, $\sigma_{ik}(S; V_i)$ 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 3.3.1 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 3.3.1. *(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 3.3.1, no two functions in $\mathcal{V}$ are monotone transformations of each other, this assumption is similar to Assumption 1.3 in [9] and guarantees that for an arbitrary player i, no two payoff functions induce the same preorder on $\{Z_{i0}, Z_{i1}, \cdots, Z_{iK}\}$. [9] provides several sufficient conditions for Assumption 3.3.1, which includes concavity or homogeneity of $V_i(X_i, Z_{ik}, S)$, see [9] for details.

ASSUMPTION 3.3.2. *(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 $S \backslash \{Z_{il}\}$.*

Assumption 3.3.2 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 [68], 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., [60]). 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 3.3.2. *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 3.3.2 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 [18], we modify it under the setup of our model. Then we have the following identification theorem:

Proposition 3.3.1. *Under Assumptions 3.2.1-3.3.2, $V_i(X_i, Z_{ik}, S)$ is identified in the set $\mathcal{V}$ for all $i \in N$ and $k \in A$.*

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 3.3.1 and 3.3.2, there exists a choice $l \in A$ and an nonempty set $\tilde{\mathcal{S}} \subset \mathcal{S}$ such that $V_i(\cdot)$ and $V_i'(\cdot)$ will impose opposite preference ordering on options k and l for all $S \in \tilde{\mathcal{S}}$, 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 3.2.2 and equilibrium condition (3.5), 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 3.3.1 $V_i(\cdot)$ is identified.

Once $V_i(\cdot)$ is identified, the structural parameter θ 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 $\phi_{il}(S) = \sum_{j \neq i} \sigma_{jl}(S; \theta) \cdot \gamma(H_{ij})$ and $\phi_i(S) = (1, \phi_{i1}(S), \cdots, \phi_{iK}(S))^T$, we introduce the following rank condition.

Assumption 3.3.3. *(Rank Condition) For sufficiently large game size n, the matrix $\mathbb{E}[\phi_i(S) \cdot \phi_i(S)^T | X_i, Z_{ik}]$ is invertible, i.e.,*

$$\liminf_{n \to \infty} \det(\mathbb{E}[\phi_i(S) \cdot \phi_i(S)^T | X_i, Z_{ik}]) > 0. \tag{3.6}$$

Assumption 3.3.3 is testable and similar to conditions imposed in [40] and [27]. Then by simple algebra, we have

$$\theta_k = \{\mathbb{E}[\phi_i(S) \cdot \phi_i(S)^T | X_i, Z_{ik}]\}^{-1} \mathbb{E}[\phi_i(S) \cdot V_i(X_i, Z_{ik}, S_{-ik}) | X_i, Z_{ik}]. \tag{3.7}$$

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

Theorem 3.3.1. *Under Assumptions 3.2.1-3.3.2, the structural functions θ are nonparametrically identified in a large game.*

Proof. The proof follows directly from the discussion above and is hence omitted. □

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 θ can be derived.

3.4 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.

3.4.1 Nonparametric estimation

The estimation method consists of two steps: the first step is to nonparametrically estimate the equilibrium beliefs $\{\sigma_{ik}(S;\theta)\}_{i\in N,\ k\in A}$, which can be done using standard nonparametric technique. Since the identification of θ 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 [36], 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 $S_{-i} = S\backslash S_i$ 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 $C_M^\alpha(\mathcal{X})$ denote the class of functions $f : \mathcal{X} \mapsto \mathbb{R}$ with $\|f\|_\alpha \leq M$, where for any m-dimensional vector of non-negative integers $k = (k_1, k_2, \cdots, k_m)$:

$$\|f\|_\alpha \equiv \max_{|k|\leq \underline{\alpha}} \sup_x |D^k f(x)| + \max_{|k|=\underline{\alpha}} \sup_{x,y} \frac{|D^k f(x) - D^k f(y)|}{\|x-y\|^{\alpha-\underline{\alpha}}},$$

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

$$D^k \equiv \frac{\partial^{|k|}}{\partial x_1^{k_1} \cdots \partial x_m^{k_m}}.$$

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

$$\hat{\phi}_{ik}(S) = \sum_{j\neq i}\left[\frac{\sum_{j=1}^n 1(Y_j = k) K(\frac{S_j - S_i}{h_1})}{\sum_{j=1}^n K(\frac{S_j - S_i}{h_1})}\right]\gamma(H_{ij}), \tag{3.8}$$

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 3.4.1. *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 3.2.2 and 3.2.3 hold,

(b) $\rho_{nk}(S_i, S_{-i}) \in C_M^\alpha(\mathcal{S})$,

(c)$\{S_i\ :\ i \in N\}$ is independent and identically distributed with a ν-times differentiable density

[2]See, e.g., Van der Vaart & Wellner (1996) Section 2.7.1.

$f(\cdot)$ bounded away from zero,

(d)$K(\cdot) : \mathbb{R}^{d+q(K+1)} \mapsto [0,1]$ is a νth order product kernel function,

(e)As $n \to \infty$, $\max_{1 \leq r \leq d+q(K+1)} h_{1r} \to 0$ and $nh_1 \to \infty$.

Since under current assumptions, we can not identify the distribution of the private information ϵ, 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 [56] to estimate the model, which does not require one to know the distribution of ϵ. Specifically

$$\hat{\theta} \in \underset{\theta \in \Theta}{\operatorname{argmax}}\ Q_n(\theta, \hat{\phi}, h_2) \equiv \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} G\left(\frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2} \right), \tag{3.9}$$

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

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

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

$G3$. $G(\cdot)$ 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 νth order kernel function ($\nu \geq 2$).

As pointed out in [12], 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 3.4.1. *The collection of the subgraphs of all $\theta \in \Theta$ forms a Vapnik-Chervonenkis (VC) class.*

Assumption 3.4.1 is a fairly weak technical condition on the space of θ, intuitively it requires that the number of distinct subsets of the space of θ does not grow "too fast". For a formal definition and examples of VC class, see [62]. Note that this assumption will be automatically satisfied if Θ is finite dimensional, i.e., under the parametric setting.

ASSUMPTION 3.4.2. *There exists a metric $\|\cdot\|_\Theta$ such that (i) Θ is compact with respect to $\|\cdot\|_\Theta$; (ii) $\theta_n \in \Theta$ converges to θ uniformly if $\|\theta_n - \theta\|_\Theta \to 0$.*

Assumption 3.4.2 is commonly assumed in the nonparametric and semiparametric econometrics literature (see, e.g.,[46], [22] and [43]). It restricts the space of structural parameters as well as the choice of the norm $\|\cdot\|_\Theta$. As pointed out in [43], it will be satisfied if the infinite dimensional space Θ consists of bounded and smooth functions. Therefore, without loss of generality we also impose the following assumption:

ASSUMPTION 3.4.3. *There exists some $C < \infty$ such that $\|\theta\|_\Theta < C$ for all $\theta \in \Theta$.*

Let

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

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

Lemma 3.4.2. *$Q(\theta, \phi)$ is continuous in $\theta \in \Theta$.*

Lemma 3.4.3. *$Q(\theta, \phi)$ is uniquely maximized at $\theta^* \in \Theta$, which is the true value of the parameters.*

By using Lemma 3.4.1-3.4.3, the next theorem establishes the uniform consistency of our proposed estimator:

Theorem 3.4.2. *Given Assumption 3.2.1-3.3.3 and 3.4.1-3.4.3, $\hat{\theta}$ is uniformly consistent for θ^*, i.e., $\|\hat{\theta} - \theta^*\|_\Theta = o_p(1)$.*

Proof. By Lemma 3.4.1-3.4.3 and Assumption 3.4.2, conditions (i)-(iv) of Theorem 2.1 in [69] are satisfied. Then immediately we can get $\|\hat{\theta} - \theta^*\|_\Theta = o_p(1)$. □

3.4.2 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

$$
\begin{aligned}
Q_n(\theta, \hat{\phi}_1, h_2) &= \frac{1}{n}\sum_{i=1}^{n}[2 \cdot 1(Y_i = 1) - 1]G\left(\frac{S_{i1}^T\alpha_1 + \beta_1\sum_{j\neq i}\hat{\sigma}_{j1}(S;\theta)\gamma(H_{ij})}{h_2}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}[2 \cdot 1(Y_i = 1) - 1]G\left(\frac{S_{i1}^T\alpha_1 + \beta_1\hat{\phi}_{i1}}{h_2}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}[2 \cdot 1(Y_i = 1) - 1]G\left(\frac{w_{i1}^T\theta}{h_2}\right), \qquad (3.10)
\end{aligned}
$$

where $w_1 = (S_1^T, \hat{\phi}_1)^T$ and we can see that the objective function has a similar form as in [12]. In order to characterize the asymptotic distribution of $\hat{\theta}$ we first introduce some additional notations: write $S_1 = (S_{11}, \tilde{S}_1^T)^T$, $\tilde{w}_1 = (\tilde{S}_1^T, \phi_1)^T$, $\alpha_1 = (\alpha_{11}, \tilde{\alpha}_1^T)^T$, $\tilde{\theta} = (\tilde{\alpha}_1^T, \beta_1^T)^T$ and define

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

and

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

Let $p(w_1^T\theta|S)$ denote the conditional density of $w_1^T\theta$ on θ, which is positive everywhere with respect to the Lebesgue measure by Assumption 3.4.4 (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 $p^{(0)}(w_1^T\theta|S) = p(w_1^T\theta|S)$. Let $F(\cdot|w_1^T\theta, S)$ denote the cumulative distribution function of ϵ on $w_1^T\theta$ and S. For each positive integer t, define

$$F^{(t)}(-w_1^T\theta|w_1^T\theta, S) = \frac{\partial^t F(-w_1^T\theta|w_1^T\theta, S)}{\partial(w_1^T\theta)^t}. \tag{3.11}$$

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

$$\begin{aligned}
B &= -2\int_{-\infty}^{\infty} u^\nu G'(u)du \sum_{t=1}^{\nu}\left\{[t!(\nu-t)!]^{-1}\mathbb{E}\left[F^{(t)}(0|0,S)p^{(\nu-t)}(0|S)\tilde{w}\right]\right\},\\
D &= \int_{-\infty}^{\infty}[G'(u)]^2 du\mathbb{E}\left[\tilde{w}_1\tilde{w}_1^T p(0|S)\right],\\
H &= 2\mathbb{E}\left[\tilde{w}_1\tilde{w}_1^T F^{(1)}(0|0,S)p(0|S)\right].
\end{aligned}$$

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 3.4.4. *(a) $|\alpha_{11}| = 1$ and $\tilde{\theta}$ is contained in a compact subset $\tilde{\Theta}$ of $\mathbb{R}^{d+q}$; (b) $Median(\epsilon|S) = 0$; (c) the support of the distribution of w is not 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) $\lim_{n\to\infty} \log n/(nh_2^4) = 0$; (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 $w_1^T\theta$ in a neighborhood of 0 and almost every S, $p^{(t)}(w_1^T\theta|S)$ and $F^{(t)}(-w_1^T\theta|w_1^T\theta, S)$ exist and are continuous functions of $w_1^T\theta$ satisfying $\left|p^{(t)}(w_1^T\theta|S)\right| < M$ and $\left|F^{(t)}(-w_1^T\theta|w_1^T\theta, S)\right| < M$. In addition, $|p(w_1^T\theta|S)| < M$ for all $w_1^T\theta$ 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 $\tilde{\Theta}$.*

Assumption 3.4.4 (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., [19], [12] and

[70]. Assumption 3.4.4 (f) is analogous to an under-smooth condition assumptions made in kernel density estimation. Assumption 3.4.4 (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 [12] for details. Assumption 3.4.4 (i)-(k) are standard in asymptotic distribution theory.

ASSUMPTION 3.4.5. *If $\hat{\phi}_1 - \phi_1 = O_p(r_n)$, where r_n is a nonstochastic positive real sequence, then $r_n = o(1/\sqrt{nh_2})$.*

Assumption 3.4.5 requires that the first-stage nonparametric estimator $\hat{\phi}_1$ converges to ϕ_1 faster than $1/\sqrt{nh_2}$. Note that r_n will be determined by the dimension of the continuous part of S_i[3] 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 [12] we have the following theorem:

Theorem 3.4.3. *Suppose Assumption 3.4.4 and 3.4.5 hold, let $\lambda < \infty$ be the limit of $nh_2^{2\nu+1}$ as $n \to \infty$. Then*

$$\sqrt{nh_2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(-\sqrt{\lambda}H^{-1}B, H^{-1}DH^{-1}). \tag{3.12}$$

Note that the proof does not trivially follow from [12] 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 3.4.3, it is necessary to consistently estimate A, D and H. Let $\hat{\theta}$ be a consistent estimator of θ based on $h_2 \propto n^{-1/(2\nu+1)}$ and by using Theorem 3 in Horowitz

[3] In 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 [71] for details.

(1992), we are able to get consistent estimators for A, D and H as

$$\hat{B} = h_2^{*-\nu} B_n(\hat{\theta}, \hat{\phi}_1, h_2^*), \tag{3.13}$$

$$\hat{D} = \frac{h_2}{n} \sum_{i=1}^{n} b_n(\hat{\theta}, \hat{\phi}_1, h_2) b_n(\hat{\theta}, \hat{\phi}_1, h_2)^T, \tag{3.14}$$

$$\hat{H} = H_n(\hat{\theta}, \hat{\phi}_1, h_2), \tag{3.15}$$

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

$$b_n(\hat{\theta}, \hat{\phi}_1, h_2) = [2 \cdot 1(Y_i = 1) - 1] \left(\frac{\tilde{w}_{i1}}{h_2} \right) G' \left(\frac{w_{i1}^T \theta}{h_2} \right). \tag{3.16}$$

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 [12]: Given ν, choose any $h_2 \propto n^{-1/(2\nu+1)}$ and any $h_2^* \propto n^{-\delta/(2\nu+1)}$ for $0 < \delta < 1$. Obtain the smoothed maximum score estimator $\hat{\theta}$ based on h_2, and use $\hat{\theta}$ and h_2^* to compute $\hat{B}$, $\hat{D}$ and $\hat{H}$. Then compute optimal h_2 by the following formula:

$$h_2 = \left[\frac{\mathrm{Tr}(\hat{H}^{-1} \hat{H}^{-1} \hat{D})}{2n\nu \hat{B}^T \hat{H}^{-1} \hat{H}^{-1} \hat{B}} \right]^{\frac{1}{2\nu+1}}, \tag{3.17}$$

in which case

$$n^{\frac{\nu}{2\nu+1}} (\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N} \left(- \left(\frac{\mathrm{Tr}(H^{-1} H^{-1} D)}{2\nu B^T H^{-1} H^{-1} B} \right)^{\frac{\nu}{2\nu+1}} H^{-1} B, \left(\frac{\mathrm{Tr}(H^{-1} H^{-1} D)}{2\nu B^T H^{-1} H^{-1} B} \right)^{\frac{-1}{2\nu+1}} H^{-1} D H^{-1} \right).$$

3.5 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., [8], [10], [11], [72] 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 3.1.

Table 3.1: Descriptive Statistics of Key Variables

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

[4] See the Add Health website (http://www.cpc.unc.edu/projects/addhealth) for a detailed description of surveys and data.

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 alleviate the dimensionality problem, in the first stage estimation we use the smoothing method proposed in [73], 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 (3.10) and the second stage bandwidth h_2 is selected by the plug-in method proposed in section 3.4.2. The smoothing function $G(\cdot)$ is chosen to be the integral of a fourth-order kernel for nonparametric density estimation ([74]). 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

$$\gamma(H_{ij}) = \frac{H_{ij}^{-1}}{\sum_{l \in N} H_{il}^{-1}}. \tag{3.18}$$

It can be easily verified that (3.18) satisfies Assumption 3.2.1.

Our empirical results are presented in Table 3.2. The standard errors are computed using Theorem 3.4.3. 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. [8], [10] and [11] 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., [10] and [11]. 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.

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

$$\gamma(H_{ij}) = \frac{1}{n-1}. \tag{3.19}$$

[5] The negative effect of GPA on smoking has been confirmed by many previous literature, see, e.g., [52]

Table 3.2: Estimation Results (with homophily effect)

Variable	School 1	School 2	School 3	School 4	School 5	School 6	School 7
Age	2.588***	0.050**	0.753***	0.132*	2.352***	0.149**	0.060**
	(0.059)	(0.024)	(0.076)	(0.071)	(0.293)	(0.067)	(0.024)
Hispanic	4.866***	0.068	-9.962***	-3.162*	-6.354***	2.034	-5.047***
	(1.447)	(0.177)	(1.377)	(1.696)	(0.680)	(0.563)	(1.060)
White	-32.393***	0.338***	-7.383***	8.738***	-37.598***	-2.326***	1.801***
	(1.756)	(0.338)	(1.268)	(0.623)	(4.916)	(0.396)	(0.363)
Black	-14.921***	0.341**	-8.294***	-3.799**	-36.904***	12.274***	1.171***
	(0.927)	(0.174)	(1.269)	(1.635)	(4.813)	(2.170)	(0.333)
Asian	-23.795***	-4.850***	7.057***	-1.652	-10.406***	-2.185***	-80.839***
	(1.815)	(0.493)	(1.890)	(1.801)	(1.238)	(0.545)	(0.000)
Female	46.440***	3.064***	-14.499***	1.166***	-39.350***	-2.696***	-4.569***
	(0.000)	(0.411)	(1.417)	(0.396)	(5.165)	(0.416)	(0.935)
No club	2.528***	-0.039	4.797***	-4.142***	-0.214	-8.312***	-0.161
	(0.679)	(0.084)	(0.226)	(0.000)	(0.283)	(2.217)	(0.099)
Mother college	-7.638***	-2.681***	-0.142	-0.897**	-5.961***	15.000***	2.220***
	(1.130)	(0.466)	(0.128)	(0.364)	(0.796)	(2.432)	(0.287)
Father work	-1.052**	-5.268***	-10.296***	-10.914***	-36.661***	-19.213***	-4.116***
	(0.458)	(0.437)	(1.413)	(0.623)	(4.815)	(3.837)	(0.946)
Peer effects	6.794***	5.191***	12.686***	3.086***	7.865***	9.152***	2.346***
	(2.299)	(0.649)	(2.639)	(0.609)	(0.856)	(2.231)	(0.770)
Observations	805	818	846	973	855	1205	840

Standard errors in parentheses

* 10% significant, ** 5% significant, *** 1% significant.

The coefficient of GPA is normalized to -1.

Under this setup, our model incorporates a similar setting as a Manski-type linear-in-mean model. The results are listed in Table 3.3, 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.

3.6 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 existence of multiple equilibria across different networks and establish nonparametric identification of the model and propose a compu-

Table 3.3: Estimation Results (without homophily effect)

Variable	School 1	School 2	School 3	School 4	School 5	School 6	School 7
Age	1.096***	-0.556***	0.537**	0.516***	2.645***	0.007	0.063**
	(0.033)	(0.131)	(0.245)	(0.064)	(0.113)	(0.052)	(0.030)
Hispanic	-0.535	10.672	-10.690	-9.900***	-0.060	7.910	-6.611***
	(1.200)	(116.879)	(31.353)	(2.691)	(2.407)	(252.989)	(0.520)
White	-2.445***	1.829***	-1.981***	-1.709**	-0.117	-26.113	-0.093
	(0.403)	(0.242)	(0.607)	(0.712)	(2.640)	(84.410)	(0.256)
Black	-5.058***	12.359	-2.905***	-9.702***	-0.363	8.670	-0.453*
	(0.636)	(117.076)	(0.568)	(2.691)	(2.694)	(253.076)	(0.260)
Asian	-6.630***	9.642	-106.609***	-3.838***	-10.620***	-26.373	-4.751***
	(0.451)	(100.156)	(0.000)	(0.742)	(2.048)	(84.194)	(0.301)
Female	1.497***	8.646	-6.240***	3.351***	-5.865***	-0.054	-6.090***
	(0.000)	(100.301)	(1.145)	(0.200)	(2.065)	(0.169)	(0.317)
No club	0.128	11.356	3.940***	-4.936***	-0.112	-35.546	0.031
	(0.336)	(117.114)	(0.351)	(0.308)	(0.467)	(168.603)	(0.116)
Mother college	-11.433***	0.182	-0.183	-47.370***	0.355	1.865***	5.066***
	(1.708)	(16.615)	(0.160)	(0.000)	(0.429)	(0.214)	(0.272)
Father work	-0.516**	-1.230***	-2.666**	-6.605**	-5.848***	-0.778***	-4.879***
	(0.217)	(0.115)	(1.061)	(2.663)	(1.934)	(0.202)	(0.196)
Peer effects	0.832	-33.619	-3.694	-0.510	-62.560***	-10.762	1.055
	(3.757)	(578.515)	(6.041)	(5.883)	(8.815)	(441.697)	(0.871)
Observations	805	818	846	973	855	1205	840

Standard errors in parentheses

* 10% significant, ** 5% significant, *** 1% significant.

The coefficient of GPA is normalized to -1.

tationally 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. [75] 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.

4. A DATA-DRIVEN BANDWIDTH SELECTION METHOD FOR THE SMOOTHED MAXIMUM SCORE ESTIMATOR

4.1 Introduction

A binary response model is a regression model in which the dependent variable is a binary random variable. Binary response models are very useful for many economics and statistics applications.[1] In this paper, we consider a linear binary response model with the following form

$$y = \mathbb{1}(x^T\beta + u \geq 0), \tag{4.1}$$

where y is a scalar dependent variable, $\mathbb{1}(\cdot)$ is the indicator function, x is a $q \times 1$ vector of explanatory variables, u is an unobserved random variable, and β is a $q \times 1$ vector of parameters to be estimated using the observed data $\{y_i, x_i^T\}_{i=1}^n$.

In this model, we do not impose parametric assumptions on the distribution of u. Therefore, β cannot be estimated by maximum likelihood method that has been widely used for probit and logit models. If u and x are independent of each other, various semiparametric methods (e.g., [14], [15], [16] and [17]) can be used to obtain a consistent estimator of β. The maximum score estimator (MS) of [18, 19] allows for the dependence of the distribution of u on x in an unknown and general way (heteroskedasticity of an unknown form). Specifically, the maximum score estimator $\hat{\beta}_{MS}$ can be obtained by

$$\hat{\beta}_{MS} = \underset{\beta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} (2y_i - 1)\mathbb{1}(x_i^T\beta \geq 0), \tag{4.2}$$

where Θ is the parameter space. However, since the objective function is discontinuous, the convergence rate of the maximum score estimator is $n^{-1/3}$, and its limiting distribution is non-standard ([20]). [12] develops a smoothed version of Manski's maximum score estimator, which is asymptotically normal and has a faster convergence rate. The convergence rate could approach $n^{-1/2}$,

[1] See [13] for a review of econometric applications of binary response models.

depending on the strength of certain smoothness conditions.

The idea of Horowitz's smoothed maximum score estimator (SMS) is analogous to the nonparametric estimation of cumulative distribution function (CDF), and involves replacing the indicator function by a continuously differentiable function in the objective function of the maximum score estimation. The continuously differentiable function retains the essential features of an indicator function. Specifically,

$$\hat{\beta}_{SMS} = \underset{\beta \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} (2y_i - 1) G\left(\frac{x_i^T \beta}{h_n}\right), \tag{4.3}$$

where h_n is the smoothing parameter (bandwidth) that converges to zero as $n \to \infty$, and $G(v)$ is a continuous function satisfying $|G(v)| < M$, for $v \in \mathbb{R}$ and some $M < \infty$, $\lim_{v \to -\infty} G(v) = 0$, and $\lim_{v \to \infty} G(v) = 1$. The identification of β (up to scale) requires that at least one component of x to be absolutely continuous with respect to the Lebesgue measure conditional on the remaining components ([19]). We arrange x so that x_1 satisfies this condition. Under some technical conditions in [12], $\hat{\beta}_{SMS}$ can be shown to be uniformly consistent and asymptotically normal.

It is generally acknowledged that kernel smoothing method can be very sensitive to the selection of bandwidth. Different bandwidths can lead to completely different results. In terms of bandwidth selection, [12] proposes a method that is analogous to the plug-in method in kernel density estimation. The method requires initial "pilot" values of h_n to compute the SMS estimator $\hat{\beta}$, and then uses this estimator to obtain the optimal bandwidth. This method has the disadvantage of not being fully data-driven, since the estimated optimal bandwidth depends on the initial selection of h_n.

In this paper, we propose an alternative method to obtain the bandwidth. Unlike the conventional plug-in method, we choose the bandwidth by minimizing a cross-validated criterion function. It is completely data-driven and does not require the selection of the initial bandwidth.

This essay is organized as follows. In Section 4.2 we discuss existing bandwidth selection procedures and introduce our proposed method. In Section 4.3 we use simulations to examine the finite sample performance of our proposed method. Section 4.4 concludes.

4.2 Bandwidth Selection Procedures

In this section, we first discuss the bandwidth selection method in [12], and then introduce our bandwidth selection procedure.

Based on Section 2 in [12], the optimal bandwidth selection requires two "pilot" bandwidths, $h = c_1 n^{-1/(2s+1)}$ and $h^* = c_2 n^{-\delta/(2s+1)}$, where c_1, c_2 and δ are some constants, with $c_1 \in (0, \infty)$, $c_2 \in (0, \infty)$, and $\delta \in (0, 1)$, and s is the order of the kernel function $G'(\cdot)$. To obtain the optimal bandwidth, one needs to manually select the values of c_1, c_2 and δ. Nevertheless in practice there is little guidance on how to choose the three constants c_1, c_2 and δ. Different values of these constants may result in completely different estimates of β. Furthermore, as shown in the simulation studies of [12, 74], the empirical levels of t test based on first-order asymptotics are highly sensitive to the choice of bandwidth. Therefore, inappropriate choices of c_1, c_2 and δ may lead to a large gap between the empirical and nominal levels of hypothesis test, and hence invalidate the inference results. The alternative bandwidth selection method we propose in this paper avoids these problems.

Our bandwidth selection method is motivated by the cross-validation method in [76], who propose a bandwidth selection method in univariate CDF estimation by minimizing the following cross-validation function:

$$CV(h_n) = \frac{1}{n} \sum_{i=1}^{n} \int \left\{ \mathbb{1}(x_i < x) - \hat{F}_{-i}(x) \right\}^2 dx, \tag{4.4}$$

where

$$\hat{F}_{-i}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} G\left(\frac{x - x_j}{h_n} \right)$$

is the leave-one-out nonparametric estimator for the univariate CDF, $F(x)$. Based on (4.4), we propose to select h_n for the SMS estimator by minimizing the following criterion function[2]:

$$CV_{SMS}(h_n;\beta)=\frac{1}{n}\sum_{i=1}^{n}\left\{(2y_i-1)\mathbb{1}\left(x_i^T\beta\geq 0\right)-\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}(2y_j-1)G\left(\frac{x_j^T\beta}{h_n}\right)\right\}^2. \quad (4.5)$$

Note that (4.5) is analogous to (4.4) by setting $x=0$, replacing the indicator function and $\hat{F}_{-i}(x)$ in (4.4) with $(2y_i-1)1\left(x_i^T\beta\geq 0\right)$ and the "leave-one-out" SMS objective function, respectively. One problem with minimizing (4.5) is that β is unknown. We can obtain an initial value of β by the MS estimation in (4.2). Therefore, our bandwidth selection method is a two-step procedure:

- **Step 1**: Obtain the initial value of β by

$$\hat{\beta}_{MS}=\underset{\beta\in\Theta}{\operatorname{argmax}}\,\frac{1}{n}\sum_{i=1}^{n}(2y_i-1)\mathbb{1}(x_i^T\beta\geq 0).$$

  Note that β can only be identified up to scale. It is convenient to impose the normalization that $|\beta_1|=1$, where β_1 denotes the coefficient of x_{i1};

- **Step 2**: Obtain the estimated bandwidth $\widetilde{h}_n$ by

$$\widetilde{h}_n=\underset{h_n}{\operatorname{argmin}}\,CV_{SMS}(h_n;\hat{\beta}_{MS}).$$

This two-step procedure does not require us to manually select the values of c_1, c_2 and δ. It is thus completely data-driven. Note that since the convergence rate of $\hat{\beta}_{MS}$ is slower than $\hat{\beta}_{SMS}$, the finite sample performance of our proposed method can be improved by iterating the procedures above. The iteration is as follows.

- **Step 3**: Obtain a new value of β by

$$\hat{\beta}_{SMS}=\underset{\beta\in\Theta}{\operatorname{argmax}}\,\frac{1}{n}\sum_{i=1}^{n}(2y_i-1)G\left(\frac{x_i^T\beta}{\widetilde{h}_n}\right),$$

[2]For using the cross-validation methods to select smoothing parameters in general conditional distribution function and conditional mean function estimations, see [73] and [77].

where $\widetilde{h}_n$ is the bandwidth obtained in Step 2. Same normalization restriction as in Step 1 should also be applied in this step;

- **Step 4**: Obtain the estimated bandwidth $\widehat{h}_n$ by

$$\widehat{h}_n = \underset{h_n}{\operatorname{argmin}}\, CV_{SMS}(h_n; \hat{\beta}_{SMS});$$

- **Step 5**: Repeat Steps 3 and 4, with $\widetilde{h}_n = \widehat{h}_n$, until $\widehat{h}_n$ converges. The convergence criterion in practice could be $|\widehat{h}_n - \widehat{h}_{n,-1}| < \epsilon$, where $\widehat{h}_{n,-1}$ is the $\widehat{h}_n$ in the previous iteration, and ϵ is a small positive constant.

4.3 Monte Carlo Simulations

This section describes Monte Carlo investigation of the finite sample performance of our proposed method. Each Monte Carlo experiment is concerned with estimating the scalar parameter β in the model

$$y = \mathbb{1}(x_1 + \beta x_2 + u \geq 0),$$

where the true value of β is -1, $x_1 \sim N(0, 1)$, and $x_2 \sim N(1, 1)$. We consider two different distributions for u. One is the uniform distribution with median of 0 and variance of 1, the other is the Student's t distribution with 3 degrees of freedom normalized to have variance of 1. Note that the coefficient of x_1 has been normalized to 1 for the purpose of identification. We use the CDF of standard normal distribution as the smoothing function $G(\cdot)$.

The Monte Carlo experiments are conducted under three different scenarios. The first one is to use the plug-in method proposed in Section 2 of [12] to select the smoothing parameter h_n. For the manually selected constants, c_1, c_2 and δ, we use the following values: $c_1 = 1$, $c_2 = 1$ and $\delta = 0.2$. The second one is to use the cross-validation method proposed in this paper. The convergence criterion is $|\widehat{h}_n - \widehat{h}_{n,-1}| < 10^{-3}$. For the purpose of comparison, we also include the (non-smoothed) maximum score estimator $\hat{\beta}_{MS}$ in our experiments. The sample sizes we consider are $n = 500,\ 1000,$ and 1500. The number of replications is 1000. We compare the performance of the three methods in terms of mean squared errors (MSE), which is defined as

$$\frac{1}{m}\sum_{j=1}^{m}(\hat{\beta}_j-\beta)^2,$$

where m is the number of replications, and $\hat{\beta}_j$ is the estimate of β in j-th experiment.

Tables 4.1 and 4.2 report the simulation results for Student's t distribution and uniform distribution, respectively. In each table, the first column shows the sample sizes, and the second to fourth column correspond to the three methods in our experiments, i.e., cross-validation, non-smoothed, and plug-in, respectively. The upper block reports the mean of the MSEs, while the lower block reports the median of the MSEs. We see that in all of the cases, our cross-validation method performs the best, while non-smoothed method performs the worst. In the comparison between cross-validation method and plug-in method, we find that the MSEs of cross-validation method are about 10% less than those of plug-in method. These results indicate that our data-driven cross-validation method not only overcomes the disadvantages of plug-in method, but also improves the performance in estimation.

Table 4.1: MSE for Student's *t* Distribution

Method	CV	Non-smoothed	Plug-In
		Mean	
$n = 500$	0.0284	0.0350	0.0299
$n = 1000$	0.0146	0.0185	0.0158
$n = 1500$	0.0117	0.0142	0.0128
		Median	
$n = 500$	0.0110	0.0144	0.0132
$n = 1000$	0.0056	0.0072	0.0064
$n = 1500$	0.0049	0.0064	0.0056

Table 4.2: MSE for Uniform Distribution

Method	CV	Non-smoothed	Plug-In
		Mean	
$n = 500$	0.0890	0.0951	0.0910
$n = 1000$	0.0489	0.0574	0.0529
$n = 1500$	0.0426	0.0504	0.0464
		Median	
$n = 500$	0.0182	0.0240	0.0210
$n = 1000$	0.0121	0.0144	0.0144
$n = 1500$	0.0090	0.0110	0.0100

4.4 Conclusion

In this paper, we propose a new method of selecting smoothing parameters in the smoothed maximum score estimator. We select bandwidth by minimizing a cross-validated criterion function. It does not require the selection of initial values for bandwidth, and is hence completely data-driven. Simulation results show that our proposed method performs better than existing methods. Future extensions of this paper include deriving the asymptotic properties of the cross-validation method, and the application to the bandwidth selection of partially linear binary response models as in [78].

5. SUMMARY AND CONCLUSIONS

The first essay develops an econometric framework to nonparametrically identify CCPs and peer effects of a network game with incomplete information, allowing for the presence of measurement error in network connections. In particular, we show that under the large game setting, the CCPs are asymptotically equivalent to the ones that are conditional on players' own characteristics and a scalar valued function of their network structure. Hence the CCPs can be nonparametrically identified by applying the method in [2]. Then the payoff primitives are proved to be identified up to a monotone transformation. We also propose a semiparametric method to consistently estimate the peer effects. As an application of the proposed methods, we study the peer effects of adolescent alcohol drinking behaviors and find that the peer effects will be significantly underestimated when measurement are ignored.

In the second essay, we construct 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 existence 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.

In the third essay, we propose a new method of selecting smoothing parameters in the smoothed maximum score estimator. We select bandwidth by minimizing a cross-validated criterion function.

It does not require the selection of initial values for bandwidth, and is hence completely data-driven. Simulation results show that our proposed method performs better than existing methods. Specifically, in the comparison between cross-validation method and plug-in method, we find that the MSEs of cross-validation method are about 10% less than those of plug-in method. These results indicate that our data-driven cross-validation method not only overcomes the disadvantages of plug-in method, but also improves the performance in estimation.

REFERENCES

[1] P. V. Marsden, “Network data and measurement,” *Annual Review of Sociology*, vol. 16, no. 1, pp. 435–463, 1990.

[2] Y. Hu and S. M. Schennach, “Instrumental variable treatment of nonclassical measurement error models,” *Econometrica*, vol. 76, no. 1, pp. 195–216, 2008.

[3] F. Molinari, “Partial identification of probability distributions with misclassified data,” *Journal of Econometrics*, vol. 144, no. 1, pp. 81–117, 2008.

[4] R. L. Matzkin, “Nonparametric estimation of nonadditive random functions,” *Econometrica*, vol. 71, no. 5, pp. 1339–1375, 2003.

[5] C. F. Manski and E. Tamer, “Inference on regressions with interval data on a regressor or outcome,” *Econometrica*, vol. 70, no. 2, pp. 519–546, 2002.

[6] X. Shen *et al.*, “On methods of sieves and penalization,” *The Annals of Statistics*, vol. 25, no. 6, pp. 2555–2591, 1997.

[7] A. Calvó-Armengol, E. Patacchini, and Y. Zenou, “Peer effects and social networks in education,” *The Review of Economic Studies*, vol. 76, no. 4, pp. 1239–1267, 2009.

[8] A. Gaviria and S. Raphael, “School-based peer effects and juvenile behavior,” *Review of Economics and Statistics*, vol. 83, no. 2, pp. 257–268, 2001.

[9] R. L. Matzkin, “Nonparametric identification and estimation of polychotomous choice models,” *Journal of Econometrics*, vol. 58, no. 1-2, pp. 137–168, 1993.

[10] R. Nakajima, “Measuring peer effects on youth smoking behaviour,” *The Review of Economic Studies*, vol. 74, no. 3, pp. 897–935, 2007.

[11] A. R. Soetevent and P. Kooreman, “A discrete-choice model with social interactions: with an application to high school teen behavior,” *Journal of Applied Econometrics*, vol. 22, no. 3, pp. 599–624, 2007.

[12] J. L. Horowitz, “A smoothed maximum score estimator for the binary response model,” *Econometrica*, vol. 60, no. 3, pp. 505–531, 1992.

[13] J. L. Horowitz and N. Savin, “Binary response models: Logits, probits and semiparametrics,” *The Journal of Economic Perspectives*, vol. 15, no. 4, pp. 43–56, 2001.

[14] A. K. Han, “Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator,” *Journal of Econometrics*, vol. 35, no. 2-3, pp. 303–316, 1987.

[15] J. L. Powell, J. H. Stock, and T. M. Stoker, “Semiparametric estimation of index coefficients,” *Econometrica*, vol. 57, no. 6, pp. 1403–1430, 1989.

[16] H. Ichimura, “Semiparametric least squares (sls) and weighted sls estimation of single-index models,” *Journal of Econometrics*, vol. 58, no. 1-2, pp. 71–120, 1993.

[17] R. W. Klein and R. H. Spady, “An efficient semiparametric estimator for binary response models,” *Econometrica*, vol. 61, no. 2, pp. 387–421, 1993.

[18] C. F. Manski, “Maximum score estimation of the stochastic utility model of choice,” *Journal of Econometrics*, vol. 3, no. 3, pp. 205–228, 1975.

[19] C. F. Manski, “Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator,” *Journal of Econometrics*, vol. 27, no. 3, pp. 313–333, 1985.

[20] J. Kim and D. Pollard, “Cube root asymptotics,” *The Annals of Statistics*, vol. 18, no. 1, pp. 191–219, 1990.

[21] M. O. Jackson, *Social and economic networks*. Princeton university press, 2010.

[22] C. F. Manski, “Identification of endogenous social effects: The reflection problem,” *The Review of Economic Studies*, vol. 60, no. 3, pp. 531–542, 1993.

[23] W. A. Brock and S. N. Durlauf, “Discrete choice with social interactions,” *The Review of Economic Studies*, vol. 68, no. 2, pp. 235–260, 2001.

[24] W. A. Brock and S. N. Durlauf, “Identification of binary choice models with social interactions,” *Journal of Econometrics*, vol. 140, no. 1, pp. 52–75, 2007.

[25] B. S. Graham, "Identifying social interactions through conditional variance restrictions," *Econometrica*, vol. 76, no. 3, pp. 643–660, 2008.

[26] Y. Bramoullé, H. Djebbari, and B. Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, vol. 150, no. 1, pp. 41–55, 2009.

[27] H. Xu, "Social interactions in large networks: A game theoretic approach," *International Economic Review*, vol. 59, no. 1, pp. 257–284, 2018.

[28] B. Kline, "Identification of complete information games," *Journal of Econometrics*, vol. 189, no. 1, pp. 117–131, 2015.

[29] A. G. Chandrasekhar and R. Lewis, "Econometrics of sampled networks," *Working Paper*, 2016.

[30] D. J. Wang, X. Shi, D. A. McFarland, and J. Leskovec, "Measurement error in network data: A re-classification," *Social Networks*, vol. 34, no. 4, pp. 396–409, 2012.

[31] A. De Paula and X. Tang, "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, vol. 80, no. 1, pp. 143–172, 2012.

[32] A. Galeotti, S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv, "Network games," *The Review of Economic Studies*, vol. 77, no. 1, pp. 218–244, 2010.

[33] E. Tamer, "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, vol. 70, no. 1, pp. 147–165, 2003.

[34] M. Rosenbaum, A. B. Tsybakov, *et al.*, "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.

[35] A. Belloni, M. Rosenbaum, and A. B. Tsybakov, "Linear and conic programming estimators in high dimensional errors-in-variables models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, no. 3, pp. 939–956, 2017.

[36] M. P. Leung, "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, vol. 188, no. 1, pp. 182–195, 2015.

[37] G. Ridder and S. Sheng, "Estimation of large network formation games," *Working Paper*, 2017.

[38] P. Goldsmith-Pinkham and G. W. Imbens, "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, vol. 31, no. 3, pp. 253–264, 2013.

[39] E. Kalai, "Large robust games," *Econometrica*, vol. 72, no. 6, pp. 1631–1665, 2004.

[40] P. Bajari, H. Hong, J. Krainer, and D. Nekipelov, "Estimating static models of strategic interactions," *Journal of Business & Economic Statistics*, vol. 28, no. 4, pp. 469–482, 2010.

[41] X. D'Haultfoeuille, "On the completeness condition in nonparametric instrumental problems," *Econometric Theory*, vol. 27, no. 3, pp. 460–471, 2011.

[42] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.

[43] C. Ai and X. Chen, "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, vol. 71, no. 6, pp. 1795–1843, 2003.

[44] X. Chen, H. Hong, and E. Tamer, "Measurement error models with auxiliary data," *The Review of Economic Studies*, vol. 72, no. 2, pp. 343–366, 2005.

[45] W. K. Newey and J. L. Powell, "Instrumental variable estimation of nonparametric models," *Econometrica*, vol. 71, no. 5, pp. 1565–1578, 2003.

[46] A. R. Gallant and D. W. Nychka, "Semi-nonparametric maximum likelihood estimation," *Econometrica*, vol. 55, no. 2, pp. 363–390, 1987.

[47] X. Chen and X. Shen, "Sieve extremum estimates for weakly dependent data," *Econometrica*, vol. 66, no. 2, pp. 289–314, 1998.

[48] S. M. Schennach and Y. Hu, "Nonparametric identification and semiparametric estimation of classical measurement error models without side information," *Journal of the American Statistical Association*, vol. 108, no. 501, pp. 177–186, 2013.

[49] E. L. Glaeser, B. Sacerdote, and J. A. Scheinkman, "Crime and social interactions," *The Quarterly Journal of Economics*, vol. 111, no. 2, pp. 507–548, 1996.

[50] A. Calvo-Armengol and M. O. Jackson, “The effects of social networks on employment and inequality,” *American Economic Review*, vol. 94, no. 3, pp. 426–454, 2004.

[51] A. Badev, “Discrete games in endogenous networks: Theory and policy,” *Working Paper*, 2013.

[52] A. Bisin, A. Moro, and G. Topa, “The empirical content of models with multiple equilibria in economies with social interactions,” *Working Paper*, 2011.

[53] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a feather: Homophily in social networks,” *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[54] G. AKERLOF, “Social distance and social decisions,” *Econometrica*, vol. 65, no. 5, pp. 1005–1027, 1997.

[55] T. G. Conley and G. Topa, “Socio-economic distance and spatial patterns in unemployment,” *Journal of Applied Econometrics*, vol. 17, no. 4, pp. 303–327, 2002.

[56] J. T. Fox, “Semiparametric estimation of multinomial discrete-choice models using a subset of choices,” *The RAND Journal of Economics*, vol. 38, no. 4, pp. 1002–1019, 2007.

[57] N. Liu and H. Xu, “Semiparametric inference on social interactions with homophily,” *Working Paper*, 2015.

[58] A. Aradillas-Lopez, “Semiparametric estimation of a simultaneous game with incomplete information,” *Journal of Econometrics*, vol. 157, no. 2, pp. 409–431, 2010.

[59] Y. Wan and H. Xu, “Semiparametric identification of binary decision games of incomplete information with correlated private signals,” *Journal of Econometrics*, vol. 182, no. 2, pp. 235–246, 2014.

[60] A. Lewbel and X. Tang, “Identification and estimation of games with incomplete information using excluded regressors,” *Journal of Econometrics*, vol. 189, no. 1, pp. 229–244, 2015.

[61] D. Pollard, *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

[62] A. W. Van Der Vaart and J. A. Wellner, *Weak convergence*. Springer, 1996.

[63] D. W. Andrews, "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica*, vol. 62, no. 1, pp. 43–72, 1994.

[64] W. K. Newey, "The asymptotic variance of semiparametric estimations," *Econometrica*, vol. 62, no. 6, pp. 1349–1382, 1994.

[65] H. Ichimura and S. Lee, "Characterization of the asymptotic distribution of semiparametric m-estimators," *Journal of Econometrics*, vol. 159, no. 2, pp. 252–266, 2010.

[66] K. Menzel, "Inference for games with many players," *The Review of Economic Studies*, vol. 83, no. 1, pp. 306–337, 2015.

[67] R. L. Matzkin, "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models," *Econometrica*, vol. 60, no. 2, pp. 262–270, 1992.

[68] A. Lewbel, "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, vol. 97, no. 1, pp. 145–177, 2000.

[69] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, vol. 4, pp. 2111–2245, 1994.

[70] L.-Y. Chen, S. Lee, and M. J. Sung, "Maximum score estimation with nonparametrically generated regressors," *The Econometrics Journal*, vol. 17, no. 3, pp. 271–300, 2014.

[71] Q. Li and J. S. Racine, *Nonparametric econometrics: theory and practice*. Princeton University Press, 2007.

[72] B. V. Krauth, "Peer and selection effects on youth smoking in california," *Journal of Business & Economic Statistics*, vol. 25, no. 3, pp. 288–298, 2007.

[73] P. Hall, J. Racine, and Q. Li, "Cross-validation and the estimation of conditional probability densities," *Journal of the American Statistical Association*, vol. 99, no. 468, pp. 1015–1026, 2004.

[74] J. L. Horowitz, “Bootstrap critical values for tests based on the smoothed maximum score estimator,” *Journal of Econometrics*, vol. 111, no. 2, pp. 141–167, 2002.

[75] U. Horst and J. A. Scheinkman, “Equilibria in systems of social interactions,” *Journal of Economic Theory*, vol. 130, no. 1, pp. 44–77, 2006.

[76] A. Bowman, P. Hall, and T. Prvan, “Bandwidth selection for the smoothing of distribution functions,” *Biometrika*, vol. 85, no. 4, pp. 799–808, 1998.

[77] P. Hall, Q. Li, and J. S. Racine, “Nonparametric estimation of regression functions in the presence of irrelevant regressors,” *The Review of Economics and Statistics*, vol. 89, no. 4, pp. 784–789, 2007.

[78] J. M. Krief, “An integrated kernel-weighted smoothed maximum score estimator for the partially linear binary response model,” *Econometric Theory*, vol. 30, no. 3, pp. 647–675, 2014.

[79] J. L. Horowitz and C. F. Manski, “Identification and robustness with contaminated and corrupted data,” *Econometrica*, vol. 63, no. 2, pp. 281–302, 1995.

APPENDIX A

IDENTIFICATION AND ESTIMATION OF PEER EFFECTS IN MIS-MEASURED SOCIAL NETWORKS

A.1 Proofs

Proof of Lemma 2.2.1

By the law of iterated expectation, we know $\mathbb{E}[W_i^* - \mathbb{E}(W_i^*|X^c, G^*)] = 0$. Therefore in order to show that $W_i^* - \mathbb{E}(W_i^*|X^c, G^*) = o_p(1)$, we need to show that

$$\mathbb{E}[W_i^* - \mathbb{E}(W_i^*|X^c, G^*)]^2 = o(1). \tag{A.1}$$

Since

$$
\begin{aligned}
&\mathbb{E}[W_i^* - \mathbb{E}(W_i^*|X^c, G^*)]^2 = \mathbb{E}\left[\frac{\sum_{j\in N_i^*} Y_j}{|N_i^*|} - \frac{\sum_{j\in N_i^*} \mathbb{E}(Y_j|X^c, G^*)}{|N_i^*|}\right]^2 \\
&= \mathbb{E}\left[\frac{1}{|N_i^*|}\sum_{j\in N_i^*}(Y_j - \mathbb{E}(Y_j|X^c, G^*))\right]^2 \\
&= \sum_{j\in N_i^*} \mathbb{E}\left[\frac{1}{|N_i^*|}(Y_j - \mathbb{E}(Y_j|X^c, G^*))\right]^2 \\
&+ \sum_{j\in N_i^*}\sum_{k\neq j, k\in N_i^*} \mathbb{E}\left\{\frac{1}{|N_i^*|}\left[Y_j - \mathbb{E}(Y_j|X^c, G^*)\right]\left[Y_k - \mathbb{E}(Y_k|X^c, G^*)\right]\right\} \\
&\equiv A_1 + A_2.
\end{aligned}
$$

Since Y_j is binary, $A_1 = O(1/|N_i^*|) = o(1)$ by Assumption 2.2.2. By law of iterated expectation and conditional independence of $\{Y_i\}_{i\in N}$, we have

$$\begin{aligned}
A_2 &= \sum_{j\in N_i^*}\sum_{k\neq j,k\in N_i^*}\mathbb{E}\left\{\frac{1}{|N_i^*|}\mathbb{E}\{[Y_j-\mathbb{E}(Y_j|X^c,G^*)]\,[Y_k-\mathbb{E}(Y_k|X^c,G^*)]\,|X^c,G^*\}\right\}\\
&= \sum_{j\in N_i^*}\sum_{k\neq j,k\in N_i^*}\mathbb{E}\left\{\frac{1}{|N_i^*|}\mathbb{E}\left[Y_j-\mathbb{E}(Y_j|X^c,G^*)|X^c,G^*\right]\mathbb{E}\left[Y_k-\mathbb{E}(Y_k|X^c,G^*)|X^c,G^*\right]\right\}\\
&= 0.
\end{aligned}$$

Therefore $\mathbb{E}[W_i^*-\mathbb{E}(W_i^*|X^c,G^*)]^2=o(1)$ and the lemma is proved. □

Proof of Proposition 2.2.1

Let's consider the correspondence $\Gamma:[0,1]^n\mapsto[0,1]^n$ with each coordinate-function component given by Definition 2.2.1. It follows from Heine-Borel Theorem and Tychonoff Theorem that $[0,1]^n$ is a compact space for sufficiently large n. It is obvious that $[0,1]^n$ is nonempty and convex.

Note from Definition 2.2.1 that

$$\begin{aligned}
p_i(X^c,G^*) &= Pr\{\epsilon_i\leq\alpha(X_i)+\beta(X_i)\mathbb{E}[W_i^*|X^c,G^*]|X^c,G^*\}\\
&= Pr\left\{\epsilon_i\leq\alpha(X_i)+\sum_{y_{N_i^*}\in\{0,1\}^{|N_i|}}W(y_{N_i^*},G_i^*)\beta(X_i)Pr(Y_{N_i^*}=y_{N_i^*}X^c,G^*)\middle|X^c,G^*\right\}
\end{aligned}$$
$$= Pr\left\{\epsilon_i\leq\alpha(X_i)+\textstyle\sum_{y_{N_i^*}\in\{0,1\}^{|N_i|}}[W(y_{N_i^*},G_i^*)\beta(X_i)Pr(Y_{N_i^*}=y_{N_i^*}|X^c,G^*)+0\cdot Pr(Y_{N/N_i^*}=y_{N/N_i^*}|X^c,G^*)]\middle|X^c,G^*\right\},$$

where Y_{N/N_i^*} denotes the equilibrium actions of all players other than those connected with i and y_{N/N_i^*} denotes the realized values for Y_{N/N_i^*}. Note from equation (2.3) that both $Pr(Y_{N_i^*}=y_{N_i^*}|X^c,G^*)$ and $Pr(Y_{N/N_i^*}=y_{N/N_i^*}|X^c,G^*)$ are continuous functions with respect to $p_j(\cdot)$ for $\forall j\in N$, we thus get that each coordinate function $p_i(X^c,G^*)$ of the correspondence Γ is a continuous function of $p_j(\cdot)$ for $\forall j\in N$. As a result, Γ per se is continuous. Since it is obvious that $[0,1]^n$ is a compact Hausdorff space, we get by applying the Closed-Graph Theorem that Γ has a closed-graph and hence is said to be sequentially upper hemicontinuous. Given that $[0,1]^n$ is compact, Γ is also compact-valued upper hemicontinuous.

We now show that the correspondence Γ is also convex-valued. For this purpose, we consider two alternative Bayesian strategies $p_i'(X^c, G^*) \in [0,1]$ and $p_i''(X^c, G^*) \in [0,1]$ for any given agent $i \in N$. By definition, for any given best responses $p_{-i}(\cdot)$ of the other players, we have

$$p_i'(X^c, G^*)U_{i1}(p_{-i}(\cdot)) + (1 - p_i'(X^c, G^*))U_{i0}(p_{-i}(\cdot)) \geq p_i(\cdot)U_{i1}(p_{-i}(\cdot)) + (1 - p_i(\cdot))U_{i0}(p_{-i}(\cdot))$$

and

$$p_i''(X^c, G^*)U_{i1}(p_{-i}(\cdot)) + (1 - p_i''(X^c, G^*))U_{i0}(p_{-i}(\cdot)) \geq p_i(\cdot)U_{i1}(p_{-i}(\cdot)) + (1 - p_i(\cdot))U_{i0}(p_{-i}(\cdot))$$

for $\forall p_i(\cdot) \in [0,1]$. Since we have $U_{i0}(p_{-i}(\cdot)) = 0$ for any $p_{-i}(\cdot)$, we thus have

$$p_i'(X^c, G^*)U_{i1}(p_{-i}(\cdot)) \geq p_i(\cdot)U_{i1}(p_{-i}(\cdot)),$$

$$p_i''(X^c, G^*)U_{i1}(p_{-i}(\cdot)) \geq p_i(\cdot)U_{i1}(p_{-i}(\cdot)),$$

therefore, for any coefficient $\lambda \in [0,1]$, we obtain

$$[\lambda p_i'(X^c, G^*) + (1 - \lambda)p_i''(X^c, G^*)]U_{i1}(p_{-i}(\cdot)) \geq p_i(\cdot)U_{i1}(p_{-i}(\cdot))$$

for any $p_{-i}(\cdot) \in [0,1]^{n-1}$. Thus we have shown that Γ is also convex-valued.

For each individual, the maximization problem is written as

$$\max_{p_i \in [0,1]} \{0, p_i U_{i1}(p_{-i})\},$$

for any $p_{-i}(\cdot) \in [0,1]^{n-1}$ of the remaining players. Given the linear property of expected utility and the compactness of domain $[0,1]$, it follows from Weierstrass Theorem that Γ is a nonempty-valued correspondence. Therefore, an application of the Kakutani Fixed-point Theorem shows that a Bayesian Nash equilibrium always exists.

Finally, noting that if $X_i = X_j$, then it is straightforward that $\alpha(X_i) = \alpha(X_j)$, $\beta(X_i) = \beta(X_j)$. By Lemma 2.2.1, $W_i^* = W_j^*$ implies that $\mathbb{E}(W_i^*|X^c, G^*) - \mathbb{E}(W_j^*|X^c, G^*) = o_p(1)$. In consequence, it follows from Assumption 2.2.1 that these two players actually face the same decision problem in the game for sufficiently large n, therefore their equilibrium strategies must be the same. The proof is therefore complete. □

Proof of Proposition 2.3.1

Let Υ denote the support of ϵ_i. Then for all $i \in N$, we have

$$\begin{aligned}
&|p_i(X^c, G^*) - p_i(X_i, W_i^*)| \\
=&|Pr\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]|X^c, G^*\} - Pr\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*|X_i, W_i^*\}| \\
=& \left|\int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]\} f_{\epsilon|X^c,G^*}(\epsilon_i|X^c, G^*)d\epsilon_i - \int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\} f_{\epsilon|X,W^*}(\epsilon_i|X_i, W_i^*)d\epsilon_i\right| \\
=&\left|\int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]\} f_\epsilon(\epsilon_i)d\epsilon_i - \int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\} f_{\epsilon|W^*}(\epsilon_i|W_i^*)d\epsilon_i\right| \\
=& \left|\int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]\} f_\epsilon(\epsilon_i)d\epsilon_i - \int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\}[f_\epsilon(\epsilon_i) - f_\epsilon(\epsilon_i) + f_{\epsilon|W^*}(\epsilon_i|W_i^*)]d\epsilon_i\right| \\
\leq& \left|\int_\Upsilon [\mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]\} - \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\}] f_\epsilon(\epsilon_i)d\epsilon_i\right| \\
&+ \left|\int_\Upsilon \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\}(f_\epsilon(\epsilon_i) - f_{\epsilon|W^*}(\epsilon_i|W_i^*))d\epsilon_i\right| \\
\leq& \int_\Upsilon |[\mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*]\} - \mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\}]|\, dF_\epsilon(\epsilon_i) \\
&+ \int_\Upsilon \left|\mathbb{1}\{\epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\}(f_\epsilon(\epsilon_i) - f_{\epsilon|W^*}(\epsilon_i|W^*))\right| d\epsilon_i \\
\equiv& B_1 + B2,
\end{aligned}$$

where the second equality is by the definition of conditional expectation, the third equality is because ϵ_i is independent with X^c and G^*, the first inequality is by triangular inequality and the second inequality is by the fact that $|\mathbb{E}(A)| \leq \mathbb{E}|A|$. We now prove that both B_1 and B_2 are $o_p(1)$.

Since the maximum value of $|\mathbb{1}(\cdot) - \mathbb{1}(\cdot)|$ is 1, we have

$$\begin{aligned}B_1 \leq &\int_{\Upsilon} \mathbb{1}\{\alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*] \geq \epsilon_i \geq \alpha(X_i) + \beta(X_i)W_i^*\}dF_\epsilon(\epsilon_i) \\ &+ \int_{\Upsilon} \mathbb{1}\{\alpha(X_i) + \beta(X_i)\mathbb{E}[W_i^*|X^c, G^*] \leq \epsilon_i \leq \alpha(X_i) + \beta(X_i)W_i^*\}dF_\epsilon(\epsilon_i) \\ \equiv &C_1 + C_2.\end{aligned}$$

Since C_1 and C_2 are similar, without loss of generality it suffices to show that C_1 is $o_p(1)$. By mean value theorem,

$$\begin{aligned}C_1 &= F_\epsilon[\alpha(X_i) + \beta(X_i)\mathbb{E}(W_i^*|X^c, G^*)] - F_\epsilon(\alpha(X_i) + \beta(X_i)W_i^*) \\ &= f_\epsilon(\eta)[\alpha(X_i) + \beta(X_i)\mathbb{E}(W_i^*|X^c, G^*) - \alpha(X_i) - \beta(X_i)W_i^*] \\ &= f_\epsilon(\eta)[\beta(X_i)\mathbb{E}(W_i^*|X^c, G^*) - \beta(X_i)W_i^*] \\ &= f_\epsilon(\eta)[\beta(X_i)(\mathbb{E}(W_i^*|X^c, G^*) - W_i^*)],\end{aligned}$$

where $\eta \in [\beta(X_i)W_i^*, \beta(X_i)\mathbb{E}(W_i^*|X^c, G^*)]$. Since by Lemma 2.2.1 and Assumption 2.2.2, $\beta(X_i)(\mathbb{E}(W_i^*|X^c, G^*) - W_i^*) = o_p(1)$, we know $C_1 = o_p(1)$ and hence $B_1 = o_p(1)$. Also note that by Lemma 2.2.1, the dependence between W_i^* and ϵ_i disappears as $n \to \infty$, we have

$$f_\epsilon(\epsilon_i) - f_{\epsilon|W^*}(\epsilon_i|W^*) = o_p(1).$$

Therefore by dominated convergence theorem, $B_2 = o_p(1)$ and the desired result follows. $\square$

Proof of Proposition 2.3.2

By the definition of conditional densities and Assumption 2.3.2,

$$
\begin{aligned}
f_{YW|X,W'}(y,w|x,w') &= \int_{\mathcal{W}^*} f_{YWW^*|W',X}(y,w,w^*|w',x)dw^* \\
&= \int_{\mathcal{W}^*} f_{Y|XWW^*W'}(y|x,w,w^*,w')f_{WW^*|XW'}(w,w^*|x,w')dw^* \\
&= \int_{\mathcal{W}^*} f_{Y|X,W^*}(y|x,w^*)f_{WW^*|XW'}(w,w^*|x,w')dw^* \\
&= \int_{\mathcal{W}^*} f_{Y|X,W^*}(y|x,w^*)f_{W|W^*XW'}(w|w^*,x,w')f_{W^*|XW'}(w^*|x,w')dw^* \\
&= \int_{\mathcal{W}^*} f_{Y|X,W^*}(y|x,w^*)f_{W|W^*}(w|w^*)f_{W^*|W'}(w^*|w')dw^*.
\end{aligned}
$$

This establishes (2.6), and then we can follow the proof of Theorem 1 in [2] to show the uniqueness of the solution. □

Proof of Lemma 2.3.1

If v and v' are observationally equivalent, there exists $F_{\epsilon|X,W^*}$ and $F'_{\epsilon|X,W^*}$ in $\mathcal{F}$ such that for all $(X,W^*) \in \mathcal{X} \times \mathcal{W}^*$, $F_{\epsilon|X,W^*}[v(X,W^*)] = F'_{\epsilon|X,W^*}[v'(X,W^*)]$. Since $F'_{\epsilon|X,W^*}$ is strictly increasing, $v'(X,W^*) = (F'_{\epsilon|X,W^*})^{-1} \circ F_{\epsilon|X,W^*}[v(X,W^*)]$. Let $g = (F'_{\epsilon|X,W^*})^{-1} \circ F_{\epsilon|X,W^*}$, and then g is strictly increasing.

On the other hand, suppose that $v' = g \circ v$ for some strictly increasing function g. Let $F'_{\epsilon|X,W^*} = F_{\epsilon|X,W^*} \circ g^{-1}$. Then for all $X \in \mathcal{X}$ and $W^* \in \mathcal{W}^*$, $F_{\epsilon|X,W^*}[v(X,W^*)] = F'_{\epsilon|X,W^*}[v'(X,W^*)]$. Hence v and v' are observationally equivalent. This completes the proof. □

Proof of Proposition 2.4.1

By Bayes theorem,

$$
\mathcal{P}(W_d^* = j|X, W_d = i) = \frac{\mathcal{P}(W_d = i|X, W_d^* = j)\mathcal{P}(W_d^* = j|X)}{\mathcal{P}(W_d = i|X)}.
$$

Therefore, the identification region of $\mathcal{P}(W_d^* = j|X, W_d = i)$ can be characterized as

$$H[\mathcal{P}(W_d^* = j|X, W_d = i)] = \left\{ \frac{\Xi_{ij} \mathcal{P}_j^{W_d}(\Xi)}{\mathcal{P}_i^{W_d^*}}, \Xi \in H(\Xi^*) \right\}, \ i, j \in \mathcal{M}, \tag{A.2}$$

where $\mathcal{P}_j^W(\Xi)$ is the jth element of the matrix on the left hand side of (2.19). Then by (2.18) and following [79], for any $\Xi \in H(\Xi^*)$, the sharp identification region for $Pr(Y = 1|X, W_d = i, W_d^* = j)$ is given by

$$H(Pr(Y = 1|X, W_d = i, W_d^* = j)) = \Psi_p \cap \left\{ \frac{Pr(Y = 1|X, W_d = i) - \psi[1 - \varsigma_{ji}(\Xi)]}{\varsigma_{ji}(\Xi)}, \psi \in \Psi_p \right\}, \tag{A.3}$$

where Ψ_p denotes the space of all probability distributions on the measurable space $(\mathcal{Y}, \Omega)$ and $\varsigma_{ji}(\Xi) \in H[\mathcal{P}(W_d^* = j|X, W_d = i)]$. Based on (A.3) and without any other restrictions on $Pr(Y = 1|X, W_d = i, W_d^* = j)$, the sharp bounds of $Pr(Y = 1|X, W_d = i, W_d^* = j)$ can be characterized as the smallest and largest values in the identification region, which are respectively,

$$L_1(\Xi) = \frac{Pr(Y = 1|X, W_d = i) - [1 - \varsigma_{ji}(\Xi)]}{\varsigma_{ji}(\Xi)}$$

and

$$U_1(\Xi) = \frac{Pr(Y = 1|X, W_d = i)}{\varsigma_{ji}(\Xi)}.$$

Hence by the law of total probability, for a given value of $\Xi \in H(\Xi^*)$, the smallest and largest values in the identification region of $p(X, W_d^* = j)$ are

$$L_2(\Xi) = \sum_{i=1}^{M} \frac{Pr(Y = 1|X, W_d = i) - [1 - \varsigma_{ji}(\Xi)]}{\varsigma_{ji}(\Xi)} \cdot \Xi_{ij} \tag{A.4}$$

and

$$U_2(\Xi) = \sum_{i=1}^{M} \frac{Pr(Y = 1|X, W_d = i)}{\varsigma_{ji}(\Xi)} \cdot \Xi_{ij} \tag{A.5}$$

respectively. The sharp bounds of the discretized CCP $p(X, W_d^* = j), \ j \in \mathcal{M}$ can be characterized accordingly. $\square$

Proof of Theorem 2.5.1

The proof proceeds by checking conditions of Theorem 4.1 in [45]. Following the proof of Lemma 2 in [2], we only need to verify the pointwise convergence of $\frac{1}{n}\sum_{i=1}^{n} Q_i(\gamma)$ to $\mathbb{E}[Q_i(\gamma)]$ for all $\gamma \in \mathit{\Gamma}_n$, where

$$Q_i(\gamma) \equiv \ln \int_{\mathcal{W}^*} f_{Y|X,W^*}(y_i^*|x_i, w^*; \theta) f_1(w_i|w^*) f_2(w^*|w_i') dw^*$$

By LIE,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))\right] = \mathbb{E}\left\{\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))\right)\middle| X^c, G^*\right]\right\} = 0.$$

Furthermore we have

$$\begin{aligned}
&\mathrm{Var}\left\{\left[\frac{1}{n}\sum_{i=1}^{n} Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))\right]\middle| X^c, G^*\right\} \\
&= \frac{1}{n^2}\sum_{j=1}^{n} \mathbb{E}\left\{[Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))]^2 \middle| X^c, G^*\right\} \\
&+ \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n} \mathbb{E}\left\{[Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))][Q_j(\gamma) - \mathbb{E}(Q_j(\gamma))]|X^c, G^*\right\} \\
&\equiv D_1 + D_2.
\end{aligned}$$

By Assumption 2.5.5 and following a similar argument as in the proof of Lemma 2.2.1, we know

$$D_1 = O\left(\frac{1}{n}\right) = o(1).$$

Furthermore by Assumption 2.5.2,

$$D_2 \leq \frac{2m}{(n-1)} \max_{j \neq i, j \in N} \mathbb{E}\left\{[Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))][Q_j(\gamma) - \mathbb{E}(Q_j(\gamma))] | X^c, G^*\right\}$$
$$= o(1).$$

Then by law of total variance and dominated convergence theorem,

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} Q_i(\gamma) - \mathbb{E}(Q_i(\gamma))\right] = o(1).$$

Therefore, we can conclude that $\frac{1}{n}\sum_{i=1}^{n} Q_i(\gamma) - \mathbb{E}(Q_i(\gamma)) = o_p(1)$. By Lemma 2 in [2], all the conditions in Theorem 4.1 of [45] are satisfied, hence we know

$$\|\hat{\gamma} - \gamma_0\|_s = o_p(1).$$

□

Proof of Theorem 2.5.2

We prove this theorem by checking conditions of Theorem 2 in [47]. Following the proof of Theorem 3 in [2], Conditions B.1-B.4 are verified and we only need to verify Condition B.5, i.e.,

$$\sqrt{n}\mu_n\left(\frac{d \ln f_{YW|W'X}(D, \gamma_0)}{d\gamma}[v_n^*]\right) \xrightarrow{d} N(0, \sigma^2_{v^*}), \tag{A.6}$$

where $\mu_n(g) \equiv n^{-1}\sum_{i=1}^{\infty}[g(\gamma, D_i) - \mathbb{E}g(\gamma, D_i)]$ denotes the empirical measure induced by g. We have

$$
\begin{aligned}
&\sqrt{n}\mu_n\left(\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]\right)\\
=&\sqrt{n}\mu_n\Bigg\{\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]-\mathbb{E}\left(\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]\middle|X^c,G^*\right)\\
&+\mathbb{E}\left(\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]\middle|X^c,G^*\right)\Bigg\}\\
=&\sqrt{n}\mu_n\Bigg\{\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]-\mathbb{E}\left(\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]\middle|X^c,G^*\right)\Bigg\}\\
&+\sqrt{n}\mu_n\left\{\mathbb{E}\left(\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]\middle|X^c,G^*\right)\right\}\\
\equiv&\sqrt{n}H_{n1}+\sqrt{n}H_{n2}.
\end{aligned}
$$

We first show that $H_{n1}=o_p(1)$, which follows a similar argument as in the proof of Theorem 2.5.1. Specifically, by law of iterated expectation,

$$
\mathbb{E}(H_{n1})=0, \tag{A.7}
$$

and by law of total variance, we have

$$
\mathrm{Var}(H_{n1})=\mathbb{E}\left[\mathrm{Var}(H_{n1}|X^c,G^*)\right].
$$

By a similar argument as in the proof of Proposition 2.3.1,

$$
\begin{aligned}
&\mathrm{Var}\left(H_{n1}|X^c,G^*\right)\\
=&\mathrm{Var}\left\{\mu_n\Bigg\{\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]-\mathbb{E}\left(\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*]\middle|X^c,G^*\right)\Bigg\}\middle|X^c,G^*\right\}\\
=&O\left(\frac{1}{n}\right).
\end{aligned}
$$

Hence by dominated convergence theorem,

$$\mathrm{Var}(H_{n1}) = o(1). \tag{A.8}$$

(A.7) and (A.8) together implies that

$$H_{n1} = o_p(1). \tag{A.9}$$

By applying a classical finite-dimensional CLT for strong mixing process, we have

$$H_{n2} \xrightarrow{d} N(0, \sigma^2_{v^*}), \tag{A.10}$$

where

$$\begin{aligned}\sigma^2_{v^*} =& \mathrm{Var}\left\{\frac{d\ln f_{YW|W'X}(D_1,\gamma_0)}{d\gamma}[v_n^*]\right\} \\ &+ 2\sum_{j=1}^{m}\mathrm{Cov}\left\{\frac{d\ln f_{YW|W'X}(D_1,\gamma_0)}{d\gamma}[v_n^*], \frac{d\ln f_{YW|W'X}(D_j,\gamma_0)}{d\gamma}[v_n^*]\right\}.\end{aligned}$$

Following the results in [2],

$$\frac{d\ln f_{YW|W'X}(D,\gamma_0)}{d\gamma}[v_n^*] = G_{m^*}(D)(\mathbb{E}\{G_{m^*}(D)G_{m^*}(D)^T\})^{-1}\lambda, \tag{A.11}$$

Consequently,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V) \tag{A.12}$$

where

$$\begin{aligned}V =& \left[\mathbb{E}\{G_{m^*}(D_1)G_{m^*}(D_1)^T\}\right]^{-1} \\ &+ 2\sum_{j=1}^{m}\mathbb{E}\left\{\left[\mathbb{E}\{G_{m^*}(D_1)G_{m^*}(D_1)^T\}\right]^{-1}G_{m^*}(D_1)G_{m^*}(D_j)^T\left[\mathbb{E}\{G_{m^*}(D_j)G_{m^*}(D_j)^T\}\right]^{-1}\right\}.\end{aligned}$$

$\square$

A.2 Additional Simulation Results

In this appendix, we provide additional simulation results, specifically we change the sample size n, the DGP for the IV G', and the matrix $\mathbf{P}_{G_{ij}|G^*_{ij}}$ to examine the performance of our estimator.

In Table A.1 and A.2, we change the sample size n to 500 and 2000 respectively, and it is not surprising to see that our method still works very well in terms of correcting the bias and reducing the MSE of the estimated peer effects caused by measurement errors. In Tables A.3-A.5 the IV G' is generated as $G'_{ij} = 1(0.6\eta_{g^*} + 0.4\eta_z > 0.2)$, and we increase the probability of correct reporting to 0.8, specifically the matrix $\mathbf{P}_{G_{ij}|G^*_{ij}}$ is we used here is

$$\mathbf{P}_{G_{ij}|G^*_{ij}} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}.$$

We can see that the bias of α_0 is relatively small if we ignore the presence of measurement errors. However, the bias in β is still fairly large and the Sieve MLE can reduce the bias and MSE in this case. The results are robust to different values of the smoothing parameters.

Table A.1: Simulation Results ($n = 500$)

	Parameter(=True Value)								
	$\alpha_0 = 1$			$\alpha_1 = 1$			$\beta = 1$		
	Mean	Std.dev	MSE	Mean	Std.dev	MSE	Mean	Std.dev	MSE
Ignoring meas. error	1.4380	1.1374	1.4829	0.9139	0.1020	0.0178	0.0144	1.3068	2.6756
Accurate data	1.0022	0.1316	0.0173	1.0209	0.1156	0.0138	1.0378	0.1940	0.0390
Sieve MLE	1.0040	0.3106	0.0963	0.9177	0.1274	0.0230	1.0297	0.6577	0.4326
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 2, j_n = 3$ in f_2.									
Ignoring meas. error	1.3981	1.1156	1.4006	0.9155	0.1026	0.0177	0.0611	1.2827	2.5237
Accurate data	1.0047	0.1316	0.0173	1.0220	0.1155	0.0138	1.0330	0.1927	0.0381
Sieve MLE	0.9963	0.3095	0.0956	0.9114	0.1230	0.0230	1.0300	0.6590	0.4344
Smoothing parameters: $i_n = 3, j_n = 4$ in f_1; $i_n = 3, j_n = 4$ in f_2.									
Ignoring meas. error	1.4020	1.1357	1.4489	0.9171	0.1021	0.0173	0.0547	1.3060	2.5957
Accurate data	1.0042	0.1302	0.0169	1.0238	0.1149	0.0137	1.0324	0.1903	0.0372
Sieve MLE	0.9847	0.3267	0.1068	0.9183	0.1398	0.0262	1.0546	0.6680	0.4483
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 6, j_n = 4$ in f_2.									

Notes: Simulation results for $G'_{ij} = \mathbb{1}(0.6\eta_{g^*} + 0.4\eta_z > 0.2)$ and $\mathcal{P}(G_{ij} = k|G^*_{ij} = k) = 0.2$.

Table A.2: Simulation Results ($n = 2000$)

	Parameter(=True Value)								
	$\alpha_0 = 1$			$\alpha_1 = 1$			$\beta = 1$		
	Mean	Std.dev	MSE	Mean	Std.dev	MSE	Mean	Std.dev	MSE
Ignoring meas. error	1.3964	1.2575	1.7352	0.9046	0.0534	0.0120	0.0564	1.4609	3.0203
Accurate data	1.0040	0.0659	0.0044	1.0069	0.0576	0.0034	1.0063	0.0936	0.088
Sieve MLE	0.9974	0.2121	0.0449	0.9042	0.0761	0.0150	1.0444	0.4512	0.2052
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 2, j_n = 3$ in f_2.									
Ignoring meas. error	1.3648	1.2651	1.7304	0.9053	0.0538	0.0119	0.0961	1.4687	2.9697
Accurate data	1.0060	0.0658	0.0044	1.0075	0.0576	0.0034	1.0061	0.0926	0.0086
Sieve MLE	0.9994	0.2374	0.0563	0.9088	0.1044	0.0192	1.0603	0.4655	0.2199
Smoothing parameters: $i_n = 3, j_n = 4$ in f_1; $i_n = 3, j_n = 4$ in f_2.									
Ignoring meas. error	1.3679	1.2584	1.7158	0.9044	0.0538	0.0120	0.0906	1.4633	2.9639
Accurate data	1.0052	0.0661	0.0044	1.0066	0.0578	0.0034	1.0051	0.0928	0.0086
Sieve MLE	1.0126	0.2133	0.0455	0.9070	0.0935	0.0174	1.0247	0.4594	0.2112
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 6, j_n = 4$ in f_2.									

Notes: Simulation results for $G'_{ij} = \mathbb{1}(0.6\eta_{g^*} + 0.4\eta_z > 0.2)$ and $\mathcal{P}(G_{ij} = k | G^*_{ij} = k) = 0.2$.

Table A.3: Simulation Results ($n = 500$)

	Parameter(=True Value)								
	$\alpha_0 = 1$			$\alpha_1 = 1$			$\beta = 1$		
	Mean	Std.dev	MSE	Mean	Std.dev	MSE	Mean	Std.dev	MSE
Ignoring meas. error	1.0835	0.4254	0.1876	0.9164	0.1030	0.0176	0.4300	0.4847	0.5594
Accurate data	1.0041	0.1329	0.0177	1.0219	0.1149	0.0137	1.0350	0.1898	0.0372
Sieve MLE	1.0348	0.3597	0.1303	0.9105	0.2037	0.0494	0.8536	0.6288	0.4160
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 2, j_n = 3$ in f_2.									
Ignoring meas. error	1.0854	0.4268	0.1891	0.9151	0.1016	0.0175	0.4257	0.4894	0.5689
Accurate data	1.0031	0.1298	0.0168	1.0207	0.1158	0.0138	1.0363	0.1918	0.0380
Sieve MLE	1.0503	0.3536	0.1273	0.9067	0.1770	0.0400	0.8477	0.6212	0.4084
Smoothing parameters: $i_n = 3, j_n = 4$ in f_1; $i_n = 3, j_n = 4$ in f_2.									
Ignoring meas. error	1.0952	0.4308	0.1942	0.9186	0.1015	0.0169	0.4162	0.4933	0.5837
Accurate data	1.0061	0.1294	0.0168	1.0232	0.1149	0.0137	1.0303	0.1918	0.0376
Sieve MLE	1.0376	0.3840	0.1486	0.9201	0.1892	0.0421	0.8610	0.6389	0.4266
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 6, j_n = 4$ in f_2.									

Notes: Simulation results for $G'_{ij} = \mathbb{1}(0.6\eta_{g^*} + 0.4\eta_z > 0.5)$ and $\mathcal{P}(G_{ij} = k | G^*_{ij} = k) = 0.8$.

Table A.4: Simulation Results ($n = 1000$)

	Parameter(=True Value)								
	$\alpha_0 = 1$			$\alpha_1 = 1$			$\beta = 1$		
	Mean	Std.dev	MSE	Mean	Std.dev	MSE	Mean	Std.dev	MSE
Ignoring meas. error	1.0552	0.4811	0.2341	0.9036	0.0779	0.0153	0.4519	0.5536	0.6062
Accurate data	1.0013	0.0937	0.0088	1.0037	0.0848	0.0072	1.0087	0.1366	0.0187
Sieve MLE	1.0226	0.2668	0.0715	0.9136	0.1223	0.0224	0.9750	0.5470	0.2993
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 2, j_n = 3$ in f_2.									
Ignoring meas. error	1.0546	0.4735	0.2267	0.9063	0.0797	0.0151	0.4553	0.5430	0.5909
Accurate data	1.0027	0.0934	0.0087	1.0076	0.0870	0.0076	1.0115	0.1369	0.0188
Sieve MLE	1.0363	0.2664	0.0721	0.9048	0.1378	0.0280	0.9375	0.5611	0.3181
Smoothing parameters: $i_n = 3, j_n = 4$ in f_1; $i_n = 3, j_n = 4$ in f_2.									
Ignoring meas. error	1.0497	0.4716	0.2244	0.9059	0.0772	0.0148	0.4585	0.5471	0.5919
Accurate data	1.0002	0.0917	0.0084	1.0062	0.0846	0.0072	1.0103	0.1327	0.0177
Sieve MLE	1.0214	0.2698	0.0731	0.9016	0.1166	0.0233	0.9558	0.5595	0.3144
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 6, j_n = 4$ in f_2.									

Notes: Simulation results for $G'_{ij} = \mathbb{1}(0.6\eta_{g^*} + 0.4\eta_z > 0.5)$ and $\mathcal{P}(G_{ij} = k|G^*_{ij} = k) = 0.8$.

Table A.5: Simulation Results ($n = 2000$)

	Parameter(=True Value)								
	$\alpha_0 = 1$			$\alpha_1 = 1$			$\beta = 1$		
	Mean	Std.dev	MSE	Mean	Std.dev	MSE	Mean	Std.dev	MSE
Ignoring meas. error	1.0154	0.5111	0.2609	0.9054	0.0545	0.0119	0.5033	0.5900	0.5941
Accurate data	1.0070	0.0655	0.0043	1.0062	0.0598	0.0036	1.0027	0.0932	0.0087
Sieve MLE	0.9950	0.2093	0.0437	0.9051	0.0707	0.0140	1.0484	0.4610	0.2144
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 2, j_n = 3$ in f_2.									
Ignoring meas. error	1.0204	0.5122	0.2623	0.9043	0.0542	0.0121	0.4965	0.5905	0.6015
Accurate data	1.0055	0.0657	0.0043	1.0051	0.0597	0.0036	1.0034	0.0958	0.0092
Sieve MLE	1.0213	0.2067	0.0431	0.9077	0.0960	0.0177	0.9890	0.4652	0.2161
Smoothing parameters: $i_n = 3, j_n = 4$ in f_1; $i_n = 3, j_n = 4$ in f_2.									
Ignoring meas. error	1.0071	0.5088	0.2584	0.9045	0.0535	0.0120	0.5110	0.5882	0.5844
Accurate data	1.0035	0.0654	0.0043	1.0062	0.0578	0.0034	1.0069	0.0940	0.0089
Sieve MLE	1.0071	0.2272	0.0516	0.9112	0.0963	0.0171	1.0328	0.4708	0.2223
Smoothing parameters: $i_n = 2, j_n = 3$ in f_1; $i_n = 6, j_n = 4$ in f_2.									

Notes: Simulation results for $G'_{ij} = \mathbb{1}(0.6\eta_{g^*} + 0.4\eta_z > 0.5)$ and $\mathcal{P}(G_{ij} = k|G^*_{ij} = k) = 0.8$.

APPENDIX B

NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF ADDITIVE SOCIAL INTERACTION MODELS WITH HOMOPHILY

B.1 Equilibrium and Identification

Proof of Theorem 3.2.1: This proof mainly relies on an application of Schauder Fixed Point Theorem and Arzelà-Ascoli Theorem. We use the same approach developed by [36] in proving Theorem 1 and we shall complete it in 5 steps.

Step 1. Let Δ 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 Δ with the norm $\|f\| \equiv \sup_{S \in \mathcal{S}} |f(S)|$, in which $\mathcal{S} \subseteq \mathbb{R}^{n[d+q(K+1)]}$ is a compact Hausdorff space, so that Δ is a subset of the Banach space $C(\mathcal{S}, \|\cdot\|)$. Pick Σ as the subset consisting of such functions satisfying properties: (1) They are symmetric everywhere on $\mathcal{S}$; (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 Σ.

Step 3. For any $k \in A$, let $U_{ik}(\sigma_{-i}, S, \epsilon_{ik})$ be the payoff function at the true parameter θ. To emphasize the dependence of Γ on S, we w.o.l.g. rewrite $\Gamma^{(S;\theta)}(\sigma)$ defined above as $\Gamma^{(\theta)}(\sigma, S)$, for any $\sigma \in \Sigma$. By making use of Assumption 3.2.4, we are led to

$$\begin{aligned}
&\Gamma^{(\theta)}_{\pi(i)k}(\sigma, \pi(S)) \\
&\quad = Pr\left(U_{\pi(i)k}\left(\sigma_{-\pi(i)}, \pi(S), \epsilon_{\pi(i)k}\right) > U_{\pi(i)h}\left(\sigma_{-\pi(i)}, \pi(S), \epsilon_{\pi(i)h}\right), \forall\, h \in A, h \neq k | S, \sigma\right) \\
&\quad = Pr\left(U_{ik}\left(\sigma_{-i}, S, \epsilon_{ik}\right) > U_{ih}\left(\sigma_{-i}, S, \epsilon_{ih}\right), \forall\, h \in A, h \neq k | S, \sigma\right) \\
&\quad = \Gamma^{(\theta)}_{ik}(\sigma, S).
\end{aligned}$$

Hence, we can claim that $\Gamma^{(\theta)}_{ik}(\cdot, S)$ maps symmetric functions to symmetric functions. Also, notice from (5) that

$$\Gamma^{(\theta)}_{ik}(\sigma, S) = \int_{\epsilon\in\mathbb{R}} \left[\prod_{h\neq k} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S))\right] f_{\epsilon_{ik}|S}(\epsilon)d\epsilon,\ \forall i, k,$$

using Assumption 3.2.2 and 3.2.5 assures that $\Gamma^{(\theta)}(\sigma, S)$ is continuous in S and σ and also preserves equicontinuity. Also, by using the definition of $\Gamma^{(\theta)}(\sigma, S)$, it is straightforward to show that the columns have a sum of one.

Step 4. We show that $\Gamma^{(\theta)}(\Sigma, S)$ is a subset of a compact space. And we just need to show that Σ is compact. Noting that $\mathcal{S}$ is compact, Σ is uniformly bounded, and Σ is equicontinuous by our construction, we apply the Arzelà-Ascoli Theorem to obtain that Σ is relatively compact. It hence suffices to show that Σ is also closed, and we prove it by means of contradiction. Letting $\{\sigma^m\}_m$ be a sequence in Σ that converges to a limit written as σ^*. Suppose that $\sigma^* \notin \Sigma$. Then, for any $i \in N$ and $k \in A$, we should have $\sigma^*_{ik}(S;\theta) \neq \sigma^*_{\pi(i)k}(\pi(S);\theta)$. However, $\sigma^m \in \Sigma$ for any m implies that $\sigma^m_{ik}(S;\theta) = \sigma^m_{\pi(i)k}(\pi(S);\theta)$ 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 3.3.1: The proof is a modification of the argument in [9], 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 3.3.2, $\exists\, l \in A$ and Z_{il} process an everywhere positive Lebesgue density conditional on $S\backslash\{Z_{il}\}$ and V_i and V_i' are strictly increasing with respect to Z_{il}. Then by Assumption 3.3.1 and the argument in Matzkin (1993), there exist a set $\tilde{\mathcal{S}} \subset \mathcal{S}$ with positive Lebesgue measure such that $\forall S \in \tilde{\mathcal{S}}$, 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 3.2.2, $F_{\epsilon_{ih}|S}(\cdot)$ 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

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

Hence,

$$\begin{aligned}
\sigma_{ik}(S;\theta) - \sigma_{il}(S;\theta) &= \int_{\epsilon\in\mathbb{R}} \left[\prod_{h\neq k} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{ik}, S) - V_i(X_i, Z_{ih}, S))\right] f_{\epsilon_{ik}|S}(\epsilon)d\epsilon \\
&\quad - \int_{\epsilon\in\mathbb{R}} \left[\prod_{h\neq l} F_{\epsilon_{ih}|S}(\epsilon + V_i(X_i, Z_{il}, S) - V_i(X_i, Z_{ih}, S))\right] f_{\epsilon_{ih}|S}(\epsilon)d\epsilon \\
&> 0.
\end{aligned}$$

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 $i \in N$. $\square$

B.2 Nonparametric estimation

Proof of Theorem 3.4.1:

$$
\begin{aligned}
&\hat{\phi}_{ik}(S) - \phi_{ik}(S) = \sum_{j\neq i}^{n} \hat{\sigma}_{jk}(S)\gamma(H_{ij}) - \sum_{j\neq i}^{n} \sigma_{jk}(S)\gamma(H_{ij}) = \sum_{j\neq i}^{n} [\hat{\sigma}_{jk}(S) - \sigma_{jk}(S)]\gamma(H_{ij}) \\
&= \sum_{j\neq i}^{n} \left\{ \frac{\sum_{l=1}^{n} 1(Y_l = k)K(\frac{S_l - S_j}{h_1})}{\sum_{l=1}^{n} K\left(\frac{S_l - S_j}{h_1}\right)} - \mathbb{E}[1(Y_j = k)|S] \right\} \gamma(H_{ij}) \\
&= \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \left\{ \frac{1(Y_l = k) - \mathbb{E}[1(Y_j = k)|S]}{\hat{f}(S_l)} \right\} K\left(\frac{S_l - S_j}{h_1}\right) \gamma(H_{ij}) \\
&= \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \left\{ \frac{1(Y_l = k) - \mathbb{E}[1(Y_l = k)|S]}{f(S_l)} \right\} K\left(\frac{S_l - S_j}{h_1}\right) \gamma(H_{ij}) + \\
&\sum_{j\neq i}^{n} \left\{ \frac{1}{nh_1} \sum_{l=1}^{n} \frac{\mathbb{E}[1(Y_l = k)|S]K\left(\frac{S_l - S_j}{h_1}\right)}{\hat{f}(S_l)} - \mathbb{E}[1(Y_j = k)|S] \right\} \gamma(H_{ij}) + (s.o.) \\
&\equiv A_{n1} + A_{n2} + (s.o.),
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathbb{E}(A_{n1}) = \mathbb{E}(\mathbb{E}(A_{n1}|S)) \\
&= \mathbb{E}\left[\frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \left\{ \frac{\mathbb{E}[1(Y_l = k)|S] - \mathbb{E}[1(Y_l = k)|S]}{f(S_l)} \right\} K\left(\frac{S_l - S_j}{h_1}\right) \gamma(H_{ij}) \right] \\
&= 0.
\end{aligned}
$$

since $\sum_{j\neq i}^{n} \gamma(H_{ij}) = 1$. To simplify notation define

$$
K_{lj} = K\left(\frac{S_l - S_j}{h_1}\right) \tag{B.1}
$$

and

$$D_{lji} = \left\{ \frac{1(Y_l = k) - \mathbb{E}[1(Y_l = k)|S]}{f(S_l)} \right\} K_{lj} \gamma(H_{ij}). \tag{B.2}$$

Then we have

$$\begin{aligned}
&Var(A_{n1}|S) = \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \sum_{m=1}^{n} \sum_{o \neq p}^{n} Cov\left[D_{lji}, D_{mop}|S\right] \\
&= \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} Var\left[D_{lij}|S\right] + \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \sum_{o \neq p}^{n} Cov\left[D_{lji}, D_{lop}|S\right] + \\
&\frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{m=1}^{n} \sum_{j \neq i,p}^{n} Cov\left[D_{lji}, D_{mjp}|S\right] + (s.o.) \\
&\equiv B_{n1} + B_{n2} + B_{n3} + (s.o.).
\end{aligned}$$

Since $\mathbb{E}(D_{lji}|S) = 0$,

$$\begin{aligned}
B_{n1} &= \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \mathbb{E}(D_{lji}^2|S) \\
&= \frac{1}{n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \frac{\mathbb{E}[1(Y_l = k)|S] \cdot \{1 - \mathbb{E}[1(Y_l = k)|S]\}}{f^2(S_l)} K_{lj}^2 \gamma^2(H_{ij}) \\
&\leq \frac{1}{4n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \frac{K_{lj}^2 \gamma^2(H_{ij})}{f^2(S_l)}
\end{aligned}$$

Since $f(\cdot)$ is bounded away from zero, we know for some $C < \infty$,

$$\begin{aligned}
\mathbb{E}(B_{n1}) &= \mathbb{E}[\mathbb{E}(B_{n1}|S_i, S_j)] \\
&\leq \frac{C}{4n^2 h_1^2} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \int \int \gamma^2(H_{ij}) \left(\int K_{lj}^2 f(S_l|S_i, S_j) dS_l \right) f(S_i) f(S_j) dS_i dS_j \\
&= \frac{C}{4n^2 h_1} \sum_{l=1}^{n} \sum_{j \neq i}^{n} \int \int \gamma^2(H_{ij}) \left(\int K^2(v) f(S_j + h_1 v) dv \right) f(S_i) f(S_j) dS_i dS_j \\
&= O\left(\frac{1}{n h_1} \right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
B_{n2} &= \frac{1}{n^2h_1^2}\sum_{l=1}^{n}\sum_{j\neq i}^{n}\sum_{o\neq p}^{n}\mathbb{E}(D_{lji}D_{lop}|S) \\
&\leq \frac{1}{4n^2h_1^2}\sum_{l=1}^{n}\sum_{j\neq i}^{n}\sum_{o\neq p}^{n}\frac{K_{lj}K_{lo}\gamma(H_{ij})\gamma(H_{po})}{f(S_j)f(S_o)}
\end{aligned}$$

and then

$$\begin{aligned}
&\mathbb{E}(B_{n2}) \\
&\leq \frac{C}{4n^2h_1^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{o\neq p}^{n}\int\int\int\int K_{lj}K_{lo}\gamma(H_{ij})\gamma(H_{po})f(S_l)f(S_j)f(S_o)f(S_p)dS_ldS_jdS_odS_p \\
&= O\left(\frac{1}{n}\right).
\end{aligned}$$

Following a similar argument we can show that $\mathbb{E}(B_{n3}) = O_p(1/n)$, hence by the law of total variance,

$$\begin{aligned}
Var(A_{n1}) &= \mathbb{E}[Var(A_{n1})] + Var[\mathbb{E}(A_{n1})] \\
&= \mathbb{E}(B_{n1}) + \mathbb{E}(B_{n2}) + \mathbb{E}(B_{n3}) \\
&= O\left(\frac{1}{nh_1}\right) + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\right) \\
&= O\left(\frac{1}{nh_1}\right).
\end{aligned}$$

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 sth order derivative of $\rho_{nk}(S_i, S_{-i})$ at S_i as

$$\rho_{nk}^{(\alpha)}(S_i) \equiv \frac{\partial^{|\alpha|}\rho_{nk}}{\partial S_{i1}^{\alpha_1}\partial S_{i2}^{\alpha_2}\cdots\partial S_{i(d+q(K+1))}^{\alpha_{d+q(K+1)}}},\ |\alpha| \leq s$$

where $|\alpha| = \sum_{i=1}^{d+q(K+1)} \alpha_i$ and S_{ij} denotes the jth 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 $\sigma_{lk}(S;\theta)$ at S_j,

$$\sigma_{lk}(S;\theta) = \rho_{nk}(S_l, S_{-l}) = \rho_{nk}(S_j, S_{-l}) + \sum_{1\leq|b|<s} \frac{\rho_{nk}^b(S_j)}{b!}(S_l - S_j)^b + \sum_{|b|=s} C_b(S_m)(S_l - S_j)^b$$

where $\lim_{S_m \to S_j} C_b(S_l) = 0$. By symmetry, $\rho_{nk}(S_j, S_{-l}) = \sigma_{jk}(S;\theta)$, hence

$$\begin{aligned}
A_{n2} = &\sum_{j\neq i}^{n} \left\{ \frac{1}{nh_1} \sum_{l=1}^{n} \frac{K\left(\frac{S_l - S_j}{h_1}\right)\gamma(H_{ij})}{\hat{f}(S_j)} \mathbb{E}[1(Y_j = k)|S] - \mathbb{E}[1(Y_j = k)|S] \right\} + \\
&\frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \sum_{1\leq|b|<s} \frac{\frac{\rho_{nk}^b(S_j)}{b!}(S_l - S_j)^b K\left(\frac{S_l - S_j}{h_1}\right)\gamma(H_{ij})}{\hat{f}(S_j)} + \\
&\frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \sum_{|b|=s} \frac{C_b(S_l)(S_l - S_j)^b K\left(\frac{S_l - S_j}{h_1}\right)\gamma(H_{ij})}{\hat{f}(S_j)} \\
&\leq \sum_{1\leq|b|\leq s} \frac{B_b}{b!} \left| \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \frac{(S_l - S_j)^b K\left(\frac{S_l - S_j}{h_1}\right)\gamma(H_{ij})}{\hat{f}(S_j)} \right| \\
&\leq \sum_{1\leq|b|\leq s} \frac{B_b}{b!} \left| \frac{1}{nh_1} \sum_{l=1}^{n} \sum_{j\neq i}^{n} \frac{(S_l - S_j)^b K\left(\frac{S_l - S_j}{h_1}\right)\gamma(H_{ij})}{\hat{f}(S_j)} - \mathbb{E}[(S_j - S_i)^b] \right| + \\
&\sum_{1\leq|b|\leq s} \frac{B_b}{b!} \left| \mathbb{E}[(S_j - S_i)^b] \right|
\end{aligned}$$

where the term B_b is the upper bound for $\rho_{nk}^b(S_i)$ and $\max_{l\in N} C_b(S_l) < B_b$. 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

$$\hat{\phi}_{ik}(S) - \phi_{ik}(S) = O_p\left(\frac{1}{\sqrt{nh_1}} + \sum_{r=1}^{d+q(K+1)} h_{1r}^v \right) = o_p(1). \tag{B.3}$$

□

Proof of Lemma 3.4.1: Let

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

Then by Triangle inequality,

$$\begin{aligned}
&\sup_{\theta \in \Theta} |Q_n(\theta, \hat{\phi}, h_2) - Q(\theta, \phi)| \\
&\leq \sup_{\theta \in \Theta} [|Q_n(\theta, \hat{\phi}, h_2) - Q_n(\theta, \phi, h_2)| + |Q_n(\theta, \phi, h_2) - Q_n^*(\theta, \phi)| + |Q_n^*(\theta, \phi) - Q(\theta, \phi)|] \\
&\leq \sup_{\theta \in \Theta} |Q_n(\theta, \hat{\phi}, h_2) - Q_n(\theta, \phi, h_2)| + \sup_{\theta \in \Theta} |Q_n(\theta, \phi, h_2) - Q_n^*(\theta, \phi)| \\
&+ \sup_{\theta \in \Theta} |Q_n^*(\theta, \phi) - Q(\theta, \phi)| \\
&\equiv A_{n1} + A_{n2} + A_{n3}.
\end{aligned}$$

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

$$\begin{aligned}
A_{n1} &= \sup_{\theta \in \Theta} \left| \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} G\left(\frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2}\right) - \tfrac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} 1(Y_i = k) \sum_{h \neq k}^{K} G\left(\frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2}\right) \right| \\
&\leq \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k}^{K} \sup_{\theta \in \Theta} \left| \frac{1}{n} 1(Y_i = k) \left[G\left(\frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2}\right) - G\left(\frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2}\right) \right] \right| \\
&\leq \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{h \neq k}^{K} \sup_{\theta \in \Theta} \left| \frac{1}{n} \left[G\left(\frac{\hat{\phi}_i^T \theta_k - \hat{\phi}_i^T \theta_h}{h_2}\right) - G\left(\frac{\phi_i^T \theta_k - \phi_i^T \theta_h}{h_2}\right) \right] \right|.
\end{aligned}$$

By condition $G3$,

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

by Theorem 3.4.1, Assumption 3.4.3 and Cauchy-Schwarz Inequality. Thus we know $A_{n1} = o_p(1)$

by Slutsky's theorem.

$$\begin{aligned}
A_{n2} &= \sup_{\theta\in\Theta}\left|\tfrac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}1(Y_i=k)\sum_{h\neq k}^{K}G\left(\frac{\phi_i^T\theta_k-\phi_i^T\theta_h}{h_2}\right)-\tfrac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}1(Y_i=k)\sum_{h\neq k}^{K}1\left(\phi_i^T\theta_k>\phi_i^T\theta_h\right)\right| \\
&\le \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{h\neq k}^{K}\sup_{\theta\in\Theta}\left|G\left(\frac{\phi_i^T\theta_k-\phi_i^T\theta_h}{h_2}\right)-1\left(\phi_i^T\theta_k>\phi_i^T\theta_h\right)\right| \\
&\equiv B_{n1}(a)+B_{n2}(a),
\end{aligned}$$

where

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

and

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

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

$$\begin{aligned}
B_{n2}(a) &\le \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{h\neq k}^{K}M\cdot\sup_{\theta\in\Theta}1(|\phi_i^T\theta_k-\phi_i^T\theta_h|\le a) \\
&= M\sum_{k=1}^{K}\sum_{h\neq k}^{K}\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}1(|\phi_i^T\theta_k-\phi_i^T\theta_h|\le a).
\end{aligned}$$

By Lemma 2.6.17 and 2.6.18 in [62], we know $\{1(|\phi_i^T\theta_k-\phi_i^T\theta_h|\le a):\theta\in\Theta\}$ is VC-subgraph given Assumption 3.4.1. Thus Glivenko-Cantelli Theorem (see, e.g., Theorem 2.4.3 in [62]) implies that

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}1(|\phi_i^T\theta_k-\phi_i^T\theta_h|\le a)-\mathbb{E}\left[1(|\phi_i^T\theta_k-\phi_i^T\theta_h|\le a)\right]\right|=o_p(1). \qquad \text{(B.4)}$$

Let $r(\cdot):\mathcal{Z}\mapsto\Theta$ be such that $r(Z_{il})=v_i(Z_{il},S\backslash\{Z_{il}\})$. Then $r^{-1}(\cdot)$ exists by Assumption

3.3.2, thus by Triangle Inequality and Law of Iterated Expectation,

$$\mathbb{E}\left[1(|\phi_i^T\theta_k - \phi_i^T\theta_h| \leq a)\right] \leq \int_{s\in\mathcal{S}} \left[\int_{r^{-1}(-a)}^{r^{-1}(a)} |\phi_i^T\theta_k| + |\phi_i^T\theta_h| f_{Z_{il}|S}(z)dz\right] f_S(s)ds, \tag{B.5}$$

where $f_{Z_{il|S}}(\cdot)$ denotes the conditional density function of Z_{il} given S and $f_S(\cdot)$ is the density function of S. By Assumption 3.3.2, the integral in brackets of $(B.5)$ 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 $\mathbb{E}\left[1(|\phi_i^T\theta_k - \phi_i^T\theta_h| \leq a)\right]$ converges to 0 uniformly over $\theta \in \Theta$. Thus by (B.4),

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

Again by Slutsky's Theorem, $B_{n2}(a) = o_p(1)$, so $A_{n2} = o_p(1)$ as well.

$$\begin{aligned}
A_{n3} &= \sup_{\theta\in\Theta}\left|\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{K} 1(Y_i = k)\sum_{h\neq k}^{K} 1\left(\phi_i^T\theta_k > \phi_i^T\theta_h\right) - \mathbb{E}\left[\sum_{k=1}^{K} 1(Y_i = k)\sum_{h\neq k}^{K} 1(\phi_i^T\theta_k > \phi_i^T\theta_h)\right]\right| \\
&\leq \sum_{k=1}^{K}\sum_{h\neq k}^{K}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n} 1(Y_i = k)\cdot 1\left(\phi_i^T\theta_k > \phi_i^T\theta_h\right) - \mathbb{E}\left[1(Y_i = k)\cdot 1\left(\phi_i^T\theta_k > \phi_i^T\theta_h\right)\right]\right| \\
&= o_p(1)
\end{aligned}$$

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 3.4.2: By Law of Iterated Expectation and Assumption 3.3.2,

$$
\begin{aligned}
Q(\theta,\phi) &= \mathbb{E}\left[\sum_{k=1}^{K} 1(Y_i = k)\sum_{h\neq k}^{K} 1(\phi_i^T\theta_k > \phi_i^T\theta_h)\right] \\
&= \sum_{k=1}^{K}\sum_{h\neq k}^{K}\int_{s\in\mathcal{S}}\left[\int_{r^{-1}(\phi_i^T\theta_h)}^{\infty} 1(Y_i = k) f_{Z_{il}|S}(z)dz\right] f_S(s)ds.
\end{aligned}
$$

Since $r(\cdot) : \mathcal{Z} \mapsto \Theta$ is continuous and strictly increasing by Assumption 3.2.5, $r^{-1}(\cdot) : \Theta \mapsto \mathcal{Z}$ is also continuous. By Assumption 3.4.2, Θ 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 Θ and $\|\theta_{nh} - \theta_h\|_\Theta \to 0$ as $n \to \infty$. Then by Assumption 3.4.2,

$$
\sup_{\theta\in\Theta}\|\theta_{nh} - \theta_h\|_\Theta = o(1). \tag{B.6}
$$

By definition of uniform continuity, $\exists\, \delta > 0$ such that if $\|\theta_{nh} - \theta_h\|_\Theta < \delta$, then

$$
|r^{-1}(\phi_i^T\theta_{nh}) - r^{-1}(\phi_i^T\theta_h)| < \epsilon \tag{B.7}
$$

for all $\epsilon > 0$. By (B.6), $\exists\, n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$
\sup_{\theta\in\Theta}\|\theta_{nh} - \theta_h\|_\Theta < \delta,
$$

then (B.7) will hold. Define $r_n(Z_{il}) = \phi_i^T\theta_{nh}$, and then by Triangle Inequality and (B.7),

$$
\begin{aligned}
&\sup_{\theta\in\Theta}|r_n^{-1}(\phi_i^T\theta_{nh}) - r^{-1}(\phi_i^T\theta_h)| \\
&\leq \sup_{\theta\in\Theta}|r_n^{-1}(\phi_i^T\theta_{nh}) - r^{-1}(\phi_i^T\theta_{nh})| + \sup_{\theta\in\Theta}|r^{-1}(\phi_i^T\theta_{nh}) - r^{-1}(\phi_i^T\theta_h)| \\
&\leq \sup_{\theta\in\Theta}|r^{-1}(\phi_i^T\theta_{nh}) - r^{-1}(\phi_i^T\theta_h)| \\
&< \epsilon.
\end{aligned}
$$

Since $\epsilon > 0$ is arbitrary, we conclude that $r_n^{-1}(\phi_i^T \theta_{nh}) \to r^{-1}(\phi_i^T \theta_h)$ uniformly over $\theta \in \Theta$. Hence by Lebesgue Dominated Convergence Theorem,

$$\int_{r_n^{-1}(\phi_i^T \theta_{nh})}^{\infty} 1(Y_i = k) f_{Z_{il}|S}(z) dz - \int_{r^{-1}(\phi_i^T \theta_h)}^{\infty} 1(Y_i = k) f_{Z_{il}|S}(z) dz = o(1).$$

Thus the integral in brackets is continuous in θ. Since this integral is also bounded by 1, again by Dominated Convergence Theorem, we conclude that $Q(\theta, \phi)$ is continuous in θ. □

Proof of Lemma 3.4.3: Using a similar argument as in the proof of Lemma 3.4.2, we know

$$Pr(\phi_i^T \theta_k = \phi_i^T \theta_h) = \int_{s \in \mathcal{S}} \left[\int_{r^{-1}(\phi_i^T \theta_h)}^{r^{-1}(\phi_i^T \theta_h)} f_{Z_{il}|S}(z) dz \right] f_S(s) ds = 0. \tag{B.8}$$

Using Law of Iterated Expectation, rewrite $Q(\theta, \phi)$ as

$$Q(\theta, \phi) = \sum_{k=1}^{K} \sum_{h \neq k}^{K} \mathbb{E} \left[\sigma_{ik}(S; \theta) 1(\phi_i^T \theta_k > \phi_i^T \theta_h) \right]. \tag{B.9}$$

By Theorem 3.3.1, we know θ^* is identified, thus by the proof of Theorem 3.3.1, we can get

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

for any $(k, h) \in A \times A$ such that $k \neq h$. Therefore, $Q(\theta, \phi)$ will be globally maximized by θ^* since there is no tie in choice probability. Now we need to prove that $Q(\theta, \phi)$ is uniquely maximzed by θ^*, suppose by contradication there exists another $\theta' \in \Theta$ such that $\theta' \neq \theta^*$ and θ' maximizes $Q(\theta, \phi)$. Then by the proof of Theorem 3.3.1, we know there will exist some $(k, l) \in A \times A$ such that

$$\phi_i^T \theta_k^* > \phi_i^T \theta_l^* \text{ and } \phi_i^T \theta_k' < \phi_i^T \theta_l'$$

or

$$\phi_i^T \theta_k^* < \phi_i^T \theta_l^* \text{ and } \phi_i^T \theta_k' > \phi_i^T \theta_l'.$$

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

B.3 Semiparametric estimation

Proof of Theorem 3.4.3: Under our setting, $\{Y_i\}_{i \in N}$ is not an independent random sequence. Therefore, the results in [12] 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,

$$B_n(\theta, \hat{\phi}_1, h_2) + H_n(\tilde{\theta}, \hat{\phi}_1, h_2)(\hat{\theta} - \theta) = 0, \tag{B.10}$$

where $\tilde{\theta}$ lies between θ and $\hat{\theta}$. Therefore,

$$\sqrt{nh_2} B_n(\theta, \hat{\phi}_1, h_2) + H_n(\tilde{\theta}, \hat{\phi}_1, h_2)\sqrt{nh_2}(\hat{\theta} - \theta) = 0. \tag{B.11}$$

Then we have

$$\begin{aligned}
&\sqrt{nh_2}(\hat{\theta} - \theta) = -H_n(\tilde{\theta}, \hat{\phi}_1, h_2)^{-1}\sqrt{nh_2} B_n(\theta, \hat{\phi}_1, h_2) \\
&= -H_n(\tilde{\theta}, \hat{\phi}_1, h_2)^{-1}\sqrt{nh_2}\{B_n(\theta, \hat{\phi}_1, h_2) - \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S] + \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S]\} \\
&\equiv -H_n(\tilde{\theta}, \hat{\phi}_1, h_2)^{-1}\sqrt{nh_2}\{C_n + \mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S]\}.
\end{aligned} \tag{B.12}$$

We first show that $\sqrt{nh_2} C_n = o_p(1)$, Let

$$b_i(\theta, \hat{\phi}_1, h_2) = [2 \cdot 1(Y_i = 1) - 1]\left(\frac{\tilde{w}_{i1}}{h_2}\right) G'\left(\frac{w_{i1}^T \theta}{h_2}\right). \tag{B.13}$$

Then $\sqrt{nh_2}B_n(\theta,\hat{\phi}_1,h_2) = (\sqrt{h_2/n})\sum_{i=1}^n b_i(\theta,\hat{\phi}_1,h_2)$. By Law of Iterated Expectation

$$\mathbb{E}(\sqrt{nh_2}C_n) = 0. \tag{B.14}$$

$$\begin{aligned}
&\mathbb{E}[(\sqrt{nh_2}C_n)^2] = \mathbb{E}\{\mathbb{E}[(\sqrt{nh_2}C_n)^2|S]\} \\
&= \frac{h_2}{n}\mathbb{E}\left\{\mathbb{E}\left[\left(\sum_{i=1}^n b_i(\theta,\hat{\phi}_1,h_2) - \mathbb{E}(b_i(\theta,\hat{\phi}_1,h_2)|S)\right)^2\middle|S\right]\right\} \\
&= \frac{h_2}{n}\sum_{i=1}^n\mathbb{E}\left\{\mathbb{E}\left[\left(b_i(\theta,\hat{\phi}_1,h_2) - \mathbb{E}(b_i(\theta,\hat{\phi}_1,h_2)|S)\right)^2\middle|S\right]\right\} \\
&= o(1),
\end{aligned}$$

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 $Var(b_i(\theta,\hat{\phi}_1,h_2))$ is bounded by Assumption 3.4.4. Hence

$$\sqrt{nh_2}C_n = o_p(1). \tag{B.15}$$

By Law of Iterated Expectation and Mean Value Theorem,

$$\begin{aligned}
&\sqrt{nh_2}\mathbb{E}\{\mathbb{E}[B_n(\theta,\hat{\phi}_1,h_2)|S]\} \\
&= \sqrt{nh_2}\mathbb{E}[B_n(\theta,\hat{\phi}_1,h_2)] \\
&= \sqrt{nh_2}\mathbb{E}[B_n(\theta,\phi_1,h_2)] + \sqrt{nh_2}\frac{\partial\mathbb{E}[B_n(\theta,\tilde{\phi}_1,h_2)]}{\partial\phi_1}(\hat{\phi}_1 - \phi_1),
\end{aligned}$$

where $\tilde{\phi}_1$ is between $\hat{\phi}_1$ and ϕ_1. By Assumption 3.4.5, $\sqrt{nh_2}(\hat{\phi}_1 - \phi_1) = o_p(1)$ and $\frac{\partial\mathbb{E}[B_n(\theta,\tilde{\phi}_1,h_2)]}{\partial\phi_1}$ is bounded by Condition G4 and Assumption 3.4.4, hence we know

$$\sqrt{nh_2}\mathbb{E}[B_n(\theta,\hat{\phi}_1,h_2)] = \sqrt{nh_2}\mathbb{E}[B_n(\theta,\phi_1,h_2)] + o_p(1). \tag{B.16}$$

By Lemma 5 in [12],

$$\lim_{n\to\infty} \sqrt{nh_2}\mathbb{E}[B_n(\theta, \phi_1, h_2)] = \sqrt{\lambda}B.$$

Therefore,

$$\lim_{n\to\infty} \sqrt{nh_2}\mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)] = \sqrt{\lambda}B + o_p(1). \tag{B.17}$$

Consequently, by Lebegesue Dominated Convergence Theorem and Lemma 5 in Horowitz (1992),

$$\lim_{n\to\infty} Var\{\sqrt{nh_2}\mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S]\} = D + o_p(1). \tag{B.18}$$

By (B.17) and (B.18) and apply Lindeberg-Feller's Central Limit Theorem, we have

$$D^{-\frac{1}{2}}\sqrt{nh_2}[\mathbb{E}[B_n(\theta, \hat{\phi}_1, h_2)|S] - \mathbb{E}(B_n(\theta, \hat{\phi}_1, h_2))] \xrightarrow{d} \mathcal{N}(0, I_{d+q}), \tag{B.19}$$

where I_{d+q} is an identify matrix with dimension $d + q$. Furthermore it is easy to verify that if $\hat{\phi}_1 - \phi_1 = o_p(1)$. Then

$$H_n(\tilde{\theta}, \hat{\phi}_1, h_2) - H_n(\tilde{\theta}, \phi_1, h_2) = o_p(1)$$

by condition G4. Hence by Law of Iterated Expectation and Lemma 8 and Lemma 9 in [12], the stochastic equicontinuity of $H_n(\theta, \hat{\phi}_1, h_2)$ holds at θ and then we know

$$H_n(\tilde{\theta}, \hat{\phi}_1, h_2) = H + o_p(1) \tag{B.20}$$

since $\hat{\theta}$ is a consistent estimator for θ. By Slutsky's Theorem, (B.12), (B.19), (B.20) together imply that

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

□