ESSAYS ON THE ECONOMETRICS OF DISCRETE GAMES OF COMPLETE INFORMATION

By

Li Zhao

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

May, 2016

Nashville, Tennessee

Approved:

Tong Li, Ph.D.

Yanqin Fan, Ph.D.

Federico H. Gutierrez, Ph.D.

Alejandro Molnar, Ph.D.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude towards my advisor and mentor, Professor Tong Li, for his invaluable encouragement, guidance, academic feedback and moral support. This dissertation is based on extensive discussion with Professor Li. His wealth of knowledge and keen insight greatly deepen my understanding about economics and research. With his wisdom, enthusiasm, hard-working and kindness, Professor Li has been an inspiration and an incredible role model.

I extend my profound thanks to professors Yanqin Fan and Shiu-Yuen Cheng, whose extraordinary classes in econometrics and game theory sparked my own excitement in the these two areas. A big thank you goes to my committee, professors Alejandro Molnar, Yanqin Fan and Federico Gutierrez, for the constant guidance and support. Their constructive suggestions and advices improved my approach to research, writing and presenting.

I would also like to thank faculty, staff and students at Vanderbilt Economics Department for their support throughout my graduate study. In particular, I thank professors Andrew Daughety, William Collins, Eun Jeong Heo, Dong-Hyuk Kim, Jennifer Reinganum, John Weymark, and Myrna Wooders for their invaluable comments and discussions. I am grateful to Ms. Kathleen Finn for her extraordinary assistance to the graduate program. I am also thankful for the academic inspiration and close friendship of Craig Benedict, Nayana Bose, Brantly Callaway, Dong Chen, Heng Chen, Kan Chen, Kate Fritzdixon, Difei Geng, Matthew Gentry, Zhengfeng Guo, Ziyi Guo, Peter Griffin, Alper Hayri, Kai Hong, Zhongzhong Hu, Michael Mathes, Michael Moody, Sebastian Tello Trillo, Nam Tuan Vu and Ying Zheng. Additionally, I gratefully acknowledge financial support from the Economics Department throughout my study.

Finally, I would like to thank my family for their endless love and encouragement. To my parents and my husband, thank you for being extremely supportive of my academic endeavors. To my grandpa, your love is always with me. And to my daughter, Emily, thank you for being a sweet girl and for bringing me joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Equilibrium Selection in Discrete Games of Complete Information: An Application to the Home Improvement Industry

1.1 Introduction

Although game theory has been developed for a long time as a powerful means for modeling the interaction of agents, empirical studies on games did not emerge until the last decades. The most challenging issue in estimating games with discrete outcomes is the existence of multiple equilibria: if the model primitives are not able to provide a unique outcome, it will be very difficult to use outcomes to infer the underling game structure.

Building on the benchmark model of discrete game of complete information formulated by Bresnahan and Reiss (1991a), this chapter proposes an econometric framework that achieves point identification without relying on pre-specified equilibrium selection. We show equilibrium selection can be revealed from the data if an additional equation is added to characterize how equilibrium is selected. We discuss identification of the model and propose an estimation strategy based on the MCMC algorithm. The framework is then implemented using a Monte Carlo experiment and a real-world application.

The additional equation added in our framework models the selection of equilibrium as a discrete choice problem. Witnessing the significant difference in game outcomes across geographic regions or groups of people, we believe the selection of equilibrium is not a universal rule, but rather an empirical question that is worthy to be addressed. Therefore, we allow the equation to have covariates that may include some features of the game, the characteristics of the players, or other relevant information. To the best of our knowledge, Bajari, Hong and Ryan (2010, henceforth BHR) is the only paper before ours that includes equilibrium selection in a structural model. We further deepen the understanding of identification and estimation of discrete games by discussing the consequence of neglecting variables in the selection equation and the effect of endogenous selection.

Adding a selection equation is desirable because it allows for point estimates of the parameters. Ciliberto and Tamer (2009) propose a partial identification approach that gives set identification of the model. From a policy point of view, it is much easier to conduct counterfactual analysis using the point estimates of the parameters rather than using the interval estimates.

Compared with other approaches, the point estimates of our framework is achieved by utilizing more information from the data. Most existing methods either pre-specify a selection mechanism or allow for any form of equilibrium selection, yet they neglect the fact that equilibrium selection can be estimated. Intuitively, some outcomes are more likely to be a unique equilibrium for some games, hence their frequencies provide more information about the payoff functions; the frequencies of other outcomes that are more affected by the selection mechanism reveal how equilibrium is selected.

Including equilibrium selection also improves efficiency and generality. The original work by Bresnahan and Reiss (1991a) discusses a way (which we later refer as the "grouping method") of avoiding the issue of multiple equilibria by grouping outcomes together such that the likelihood of the new event is uniquely determined regardless of how the equilibrium is selected. In our framework, because the probabilities of individual outcomes rather than their combination are considered, more information is extracted hence efficiency is improved. Moreover, modeling equilibrium selection allows a more general framework of the model. The "grouping method" only works in some cases, while our framework allows for asymmetry (e.g. individual specific coefficients) and a greater number of players.

In addition to proposing an econometric framework, we also provide a computationally attractive procedure to estimate the model. When the number of parameters increases, making maximizing the likelihood function and minimizing the distance function difficult, the Bayesian method can be more convenient. Moreover, there has been recent work on identification of random coefficient games. The Bayesian method is well-known for its flexibility in estimating random coefficient models. To the best of our knowledge, little attention has been paid to conducting Bayesian analysis of economic games. One exception is the work by Narayanan (2013), which uses a Bayesian model selection technique to evaluate a set of models under various assumptions of equilibrium selection. His method is essentially a Metropolis-Hastings algorithm that requires simulating the likelihood function at every iteration. By taking advantage of a data augmentation technique and the properties of the Gibbs sampler, we propose an algorithm that requires sampling from standard distributions only.

A Monte Carlo experiment is conducted to investigate the finite sample performance of the proposed estimation method. Various approaches are implemented and compared. As predicted, adding selection equation helps to point identify game primitives. Furthermore, a degenerate selection equation also helps to identify the model.

The last part of the chapter applies the framework to study entry competition in the home improvement industry, which is led by two large firms: Home Depot and Lowe's. Population is shown to be the biggest influencing factor of market structure while the effect of income is much smaller. There seems a strong similarity between the two firms in response to the market characteristics or opponent's entry. Empirical findings suggest an equilibrium selection favoring Lowe's. This may because Lowe's has a longer history than Home Depot therefore it has a first-move advantage.

This chapter is most related to empirical games of complete information. We adopt a similar parametric setup to Bresnahan and Reiss (1990), Bresnahan and Reiss (1991a), Tamer (2003), Krauth (2006), Ciliberto and Tamer (2009) and BHR (2010). Within the complete information paradigm, recent work has made much progress in relaxing many assumptions of the game: for example, Kline (2014, 2015), Kline and Tamer (2012) study non-parametric and semi-parametric identification of the game; Dunker, Hoderlein and Kaido (2013) study identification and estimation of a random coefficient model; Kline and

Tamer (2012) allow players to have other levels of rationality rather than playing Nash equilibrium. There are also extensive studies on games with other information structures, e.g. the incomplete information games (e.g. Bajari et. al. 2010, Aradillas-Lopez 2010) or games with both private and common information (e.g. Greico 2014). This chapter is also related to the entry literature, such as the work by Berry (1992), Seim (2006), Jia (2008) and Holmes (2011), etc. Because we propose an MCMC algorithm for estimation, this paper is related to the Bayesian literature, including Bayesian algorithms for games (e.g. Narayanan 2013, Hartmann 2010), Bayesian algorithms for selection models (e.g. van Hasselt 2011) and the literature on Gibbs sampling and data augmentation (Albert and Chib 1993).

The chapter proceeds as follows: Section 2 describes the model setup, the econometric complication and our empirical strategy. Section 3 proposes our estimation procedure. Section 4 uses a Monte Carlo example to show the benefits of our modeling approach. Section 5 applies the framework to study entry competition in the home improvement industry. The last section concludes.

1.2 Model

1.2.1 Model Setup

We consider a simultaneous-move game with complete information. There are N players, each has two choices, $y_i \in \{0,1\}$. The utility of action $y_i = 0$ is normalized to be 0 for identification. The utility of action $y_i = 1$ is a function of player i's characteristics X_i, the actions of the rest of players y_{-i}, and a shock ε_i. In a complete information game, individual payoff functions (including shocks) are observed by all players but not the econometrician.

Players choose the action that gives the higher payoff. The action $y_i = 1$ will be chosen if and only if the utility of action 1 is positive:

$$y_i = 1(\beta_i X_i + \gamma_i y_{-i} + \varepsilon_i > 0). \tag{1.1}$$

Assume $\varepsilon_i \sim N(0,1)$. Individual shocks can be correlated.

Many economic activities can be characterized as games of complete information. For example, this framework has been used to study labor force participation of household members (Bjorn and Vuong 1984), the initiation of risky behavior in a friend group (Krauth 2006, Card and Guiliano 2013), the entry decision among competing firms (Berry 1992, Jia 2008, Ciliberto and Tamer 2009), among others.

The complete information model is favored over incomplete information models if agents have capabilities to coordinate. Complete games are used to study the fertility decision among groups of co-workers, the labor force participation in a household, plans for holiday for a couple, etc. In these cases, co-workers or family members could coordinate with each other and ensure no one wants to deviate after the coordination.

Another situation where complete information games are used is when we believe what observed are the outcomes of long run equilibrium. Social interaction and firm competition are two examples. Incomplete information makes more sense to model games with "ex-post regret." In a social interaction game, individuals respond to friend's action, such as whether or not to take a healthy or risky behavior. If time is long enough for them to make adjustment, the final choice individuals make can be described as a game of complete information. Likewise, when firms are competing with each other in a technology adoption game or in an entry game, if time is long enough, firms take an action only if the utility of doing so is higher than the utility of the alternative. Therefore, firm competition can be modeled as a game of complete information.

In the rest of the chapter, we refer to the game as a social interaction game if $\gamma > 0$. When $\gamma > 0$, player i is more likely to choose action 1 if the other participants choose action

1, as in the case of social interaction where there exists conformity within a group. Also, we refer to the game as an entry game if $\gamma < 0$. When $\gamma < 0$, an agent is less likely to choose an action if the opponent takes the action. Entry games fit this pattern.

1.2.2 Empirical Complications

Discrete games often have multiple equilibria. A classic example from the entry literature is that the size of a market may be large enough for one entrant but not two. The multiple equilibria then consist of all the market structures with only one entrant, of which there are as many as the number of firms. An example from the social interaction literature is that teenagers have the tendency to imitate each other, therefore the multiple equilibria may consist of a group of people having the same action, regardless of which action it is.

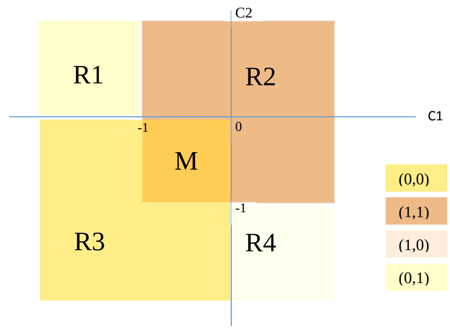
A game empiricist's objective is to quantify various factors that influce decisions. In this subsection we use a 2-person social interaction game to illustrate how the existence of multiple equilibria challenges empirical studies.

Consider a 2-person symmetric social interaction game ($\gamma > 0$):

$$\begin{aligned} y_1 &= 1(\beta X_1 + \gamma y_2 + \varepsilon_1 > 0); \\ y_2 &= 1(\beta X_2 + \gamma y_1 + \varepsilon_2 > 0). \end{aligned}$$

The solution of the game can be obtained by enumerating all outcomes and checking if each player plays a best response given the actions of others. This games has four possible outcomes $y_i \in \{0,1\}$, $i = 1,2$. $(0,0)$ is an equilibrium if $c_1 = \frac{\beta X_1 + \varepsilon_1}{\gamma} < 0$ and $c_2 = \frac{\beta X_2 + \varepsilon_2}{\gamma} < 0$, similarly $(1,1)$ is an equilibrium if $c_1 > -1$ and $c_2 > -1$. Repeating this process for all the outcomes, we plot a figure showing the equilibrium of the game characterized by c.

Figure 1.1: Equilibrium of the 2-Player Social Interaction Game



Region M in Figure 1.1 have multiple equilibria: (0,0) and (1,1). If we think of the action as whether or not to smoke, in this region, (smoke, smoke) and (not smoke, not smoke) are the two equilibria. This makes sense because Region M represents the case where $c = \frac{\beta x+\varepsilon}{\gamma} \in [-1,0]$. A game will fall into this region if the utility of being the single smoker $(\beta x+\varepsilon)$ is not too high or too low, or if the peer effect λ is very strong. In such game, best response highly depends on the actions of friends, therefore (smoke, smoke) and (not smoke, not smoke) are the two possible outcomes.

Multiple equilibria are prevalent in discrete games and they become more complicated in games other than the one illustrated in Figure 1.1. In general, the number of regions having multiple equilibria grows much faster than the number of players. For example, in a 4-person game, there are a total of 17 regions having multiple equilibria. In addition, the composition of multiple equilibria differs by region,[1] making it difficult to group outcomes together such that the probability of the "group" is unique predicted. Moreover, if asymmetry is allowed, the multiple equilibria in an entry game do not have the same number of entrants.[2] Given the variety of multiple equilibria in different games, empirical strategies need to be general enough to account for all those variations.

[1]For example, in 4-person social interaction game, the outcomes (0,0,0,0) and (0,0,1,1) are the multiple equilibria in some games and the outcomes (0,0,1,1) and (1,1,1,1) are the multiple equilibria in some other games.

[2]If firm 1 is a big firm and firms 2 and 3 are smaller, the multiple equilibria include the outcome where the big firm enters as a monopoly (1,0,0) and the outcome where the two small firms enter as the oligopolist (0,1,1).

Complications arise if some regions admit two or more outcomes as equilibria. If the game does not predict a unique outcome, as in the case of multiple equilibria, it will be very difficult to recover the underlying data generation process using observations.

1.2.3 Identification and Estimation in the Previous Literature

A number of methods have been proposed in the entry literature to address the issue of multiple equilibria. One simple way is just to assume a selection mechanism.[3] By doing so, the outcome is uniquely predicted so all the complications caused by multiple equilibria no longer exist. At the expense, however, the assumed equilibrium may not be the true underlying mechanism and the estimators will be affected.

The second approach, which we later refer to as the "grouping method," circumvents the problem by defining probabilities on events that are not affected by equilibrium selection (Bresnahan and Reiss 1991a, Berry 1992). For example, in the 2-person entry game where outcomes (1,0) and (0,1) are the multiple equilibria, one may define a new event called "monopoly" that consists of the two outcomes. This approach's advantage in robustness comes at a cost. Information is lost by aggregation, affecting the efficiency of the estimator. Moreover, this approach is not feasible universally. As illustrated at the end of the previous subsection, when the number of players grows, or a game is asymmetric, we are not able to find a simple way to group outcomes together into events that are free from equilibrium selection.[4]

The third approach, proposed by Ciliberto and Tamer (2009), constructs bounds for the probabilities. Specifically, the probability of observing an outcome is no less than the probability of playing a game where this outcome is a unique equilibrium; likewise, the

[3] For example, Jia (2008) studies entry competition between Kmart and Walmart and assumes the equilibrium favors Kmart. She then uses different assumptions of equilibrium selection as robustness checks.

[4] For example, in a 4-person social interaction game, one region admits (0,0,0,0) and (0,0,1,1) as multiple equilibria, but we cannot group the two outcomes together because in another region, (0,0,0,0) and (0,1,1,0) are the multiple equilibria. We cannot group the three outcomes together because in a third region, (0,1,1,0) and (1,1,1,1) are the multiple equilibria. This process could repeat.

probability of observing this outcome is no greater than the probability of being in a game where the outcome is an equilibrium. By constructing both lower and upper bounds for the probability of observing the outcome, the "partial approach" provides set estimation of the game-theoretic model.

1.2.4 Equilibrium Selection and Identification

We want to argue that equilibrium selection can be revealed from the data. By exploring the selection of equilibrium using the data, the payoff functions can be identified.

Assume equilibrium is selected according to a Probit function:

$$S = 1(\lambda Z + u > 0). \tag{1.2}$$

Assume $u \sim N(0,1)$. For the purposes of the present discussion, we assume $u \perp \varepsilon$. Later we will allow selection to be endogenous.

Here are a few examples of how we follow Equation 1.2 to model equilibrium selection. In a two-person entry game, $(1,0)$ and $(0,1)$ are the multiple equilibria. We could model the selection of outcome $(1,0)$ against $(0,1)$ as S_{01}. In a two-person social interaction game, we could model S_{11}. In the four person game, even though different regions have different equilibria, the multiple equilibria could be ranked by the total number of 1s. The selection of the "high equilibrium" S_{high} can be modeled by a Probit function.

To capture heterogeneous equilibrium selection among games, our model adds covariate Z in the selection equation. The covariate Z may include features of the game, characteristics of the players, or other factors that are believed to affect equilibrium selection. When studying the rate of smoking, researchers find very different rates among peers of different social groups, so we may wonder if the selection of equilibrium varies by social background. In an entry game, some researchers believe the equilibrium selection depends on a firm's distance to its headquarters, so Z could account for that. In later discussion, we will shows that even if researchers are not able to consider all covariates in the selection equation, modeling equilibrium is still desirable, because the neglected variables can be absorbed into the error term (u) and our model allows for endogenous selection.

To the best of our knowledge, BHR is the only paper besides ours that explicitly models equilibrium selection. They assume a Logit equation of equilibrium selection that depends on the properties of the equilibrium (for example, whether the outcome is dominated). In many games (such as social interaction games, or 2-person entry games), our setup and theirs are similar with the exception that ours allows for covariates and endogenous selection. When many features are considered, a multinomial Logit equation is needed. The main idea of our procedure that involves iteratively sampling the parameters from the utility and equilibrium selection works for the Logit case, though the specific sampling distribution needs to be modified to accommodate the different distributional assumption of the selection mechanism.

Identification When Selection is Modeled Explicitly

Identification of the game follows BHR. They proposes two identification approaches. The first one is based on identification at infinity. If the support of the covariates is large enough, there is a positive chance that all players but one play a certain action for sure. The decision for the remaining player reduces to a single-agent problem therefore the utility function is identified. The second identification approach proposed by BHR uses an exclusion restriction. If there exist some covariates that shift the utility of one player but the utility of the other players is not affected, the model is also identified.

We use a 2-person social interaction game as an example to provide intuition why the model can be identified. Recall that in this game, two persons decide whether or not to take an action. The two possible equilibria are (0,0) and (1,1).

Equilibrium selection is identified by the frequencies of the outcomes that are highly affected by the selection of equilibrium. For example, in the social interaction game, the frequencies of (0,0) and (1,1) depend on equilibrium selection but the frequencies of (0,1), (1,0) and $y_1 = y_2$ do not. Suppose we have two samples with similar covariates, but the first sample has more outcome (1,1) than the second sample has. This suggests that the (1,1) equilibrium is more often selected in sample 1. By exploring the frequencies of these outcomes, parameters in the selection equation can be identified.

Payoff functions are identified as well. Coefficients in the utility function can be identified by varying the exogenous variable X. Correlation between the error terms can be identified by checking if the two people choose the same action or not. If it is more often to observe (0,0) or (1,1) than observing (0,1) or (1,0), it is likely that payoffs are highly correlated.

Discussion: Neglected Heterogeneity in Selection Equation

If the true selection equation depends on covariates, a natural question is what happen if some of the factors influencing the selection are neglected? Suppose the true selection equation depends on W, but it is not observed in the data:

$$S = 1(\lambda Z + \delta W + \eta > 0). \tag{1.3}$$

Assume $W \perp \eta$, $\eta \sim N(0,1)$ and $W \sim N(0, \sigma_W^2)$.

We want to illustrate that missing covariates in selection equation in not a problem in terms of consistency as long as the neglected variable is not correlated with existing covariates. The intuition is as follows. In a standard Probit model $S = 1(\lambda Z + \delta W + \eta > 0)$, the conditional probability $E(S|Z)$ is consistently estimated if the neglected variable W is independent of the controlled variable Z.[5] In the game-theoretic model, the likelihood of the outcome is the expectation of the probability that the outcome is an equilibrium and is

[5] A sketch of proof: Consider a Probit model $S = 1(\lambda Z + \delta W + \eta > 0)$ where $Z \perp W$. The true probability of observing S given Z is given by $E(S|Z) = P(\delta W + \eta > -\lambda Z|Z) = \Phi(\frac{\lambda Z}{\sigma})$ where $\sigma^2 = 1 + \delta^2 \sigma_w^2$. If Probit estimation of S on Z is run instead, the estimator we get is $\tilde{\lambda}$, which is a consistent estimator of $\frac{\lambda}{\sigma}$. We then get $\hat{E}(S|Z) = \Phi(\tilde{\lambda} Z) \xrightarrow{p} \Phi(\frac{\lambda}{\sigma} Z) = E(S|Z)$ because $\tilde{\lambda} \xrightarrow{p} \frac{\lambda}{\sigma}$.

selected. Applying this logic, if the missing covariate W in the selection equation is independent of Z, the probability of selecting a given outcome conditional on Z is consistently estimated, thus the likelihood function is not affected by neglecting W.

A corollary of the proposition is that a degenerate equilibrium selection equation that only contains a constant but no covariates may still be considered. Even if we do not have much information about how the equilibrium is selected, we may avoid using the partial approach by assuming the selection mechanism follows a binomial distribution.

Discussion: Endogenous Equilibrium Selection

In some cases, one may worry that equilibrium selection may be endogenous. Our model could account for endogenous selection by allowing correlation between selection equation (u) and payoff function (ε). For example, in a social interaction game, we may believe the selection of the high equilibrium is correlated with taking the action, so error terms are correlated:

$$\varepsilon_i = \rho u + \nu_i,$$

where ν_i is independent across individuals.

An endogenous selection model is identified because model predictions are different for different degrees of endogeneity. For example, consider the social interaction game again. If selection of the high equilibrium and payoffs are positively correlated, the chance of observing (1,1) increases for two reasons: because it is selected more often, and because payoff of taking the action is high. As a consequence we will observe a lot of (1,1). Conversely, if the correlation between selection and payoffs is negative, one mechanism increases the chance of observing (1,1) and the other mechanism decreases the chance. Therefore we will not observe as many (1,1) as if the correlation between selection and payoff were positive.

1.3 Estimation

This section discusses the estimation procedure for the game. In principle, the system of equations could be estimated by MSL or MSM. These two approaches rely on calculating the likelihood function, which can be approximated using simulation. We propose an alternative procedure strategy: the Bayesian approach.

The Bayesian method does not require doing maximization, thus it is attractive when the parameter space is of large dimension. As will be shown shortly, we develop an algorithm that require random number generation from standard distributions only. To the best of our knowledge, little work has been done on estimating game-theoretic models using the Bayesian approach. One exception is the paper by Narayanan (2013) which uses a Bayesian method to compare multiple equilibrium selection mechanisms. For a given mechanism, he uses a Metropolis-Hastings algorithm that samples all the parameters at the same time. Our procedure takes advantage of the Gibbs Sampling algorithm and draws a subset of parameters at a time.

This section starts with a brief introduction to the Bayesian method and the two techniques used to develop the sampling algorithm. Then we describe the Bayesian algorithm. The algorithm is an iteration of four steps. The first two steps of the algorithm can be used for games with given selection mechanisms. In the end of this section we discuss potential ways of generalizing our procedure to other setups.

1.3.1 A Quick Review of Bayesian Inference, Gibbs Sampling and Data Augmentation

The Bayesian approach thinks of parameters of interest as a random vector and uses Bayes' rule to update the distribution of the random vector based prior beliefs and the likelihood function. The posterior distribution $p(\theta|X)$ is usually approximated by the MCMC algorithm that samples parameters from the density function $p(\theta|X) = \frac{f(X|\theta)f(\theta)}{f(X)} \propto f(X|\theta)f(\theta)$, where $f(\theta)$ is the prior belief for the parameters and $f(X|\theta)$ is the likelihood function.

Gibbs sampling is an MCMC algorithm for drawing a sequence of random vectors. It is appealing in the case where the joint distribution of the parameters is difficult to be sampled from, but the distribution of one set of parameters conditional on the rest of parameters is much easier to deal with. Instead of drawing all entries of the random vector simultaneously, Gibbs sampling allows one to divide the entries into blocks and sequentially draw one block conditional on the rest of the blocks. The final sequence follows the same distribution as if the blocks are draw simultaneously from their joint distribution. Concretely, if we divide a random vector θ into three subvectors and denote them as $(\theta_1, \theta_2, \theta_3)$, let the subscript denote the iteration, Gibbs sampling says one can draw $\theta_1^{(r)} | \theta_2^{(r-1)}, \theta_3^{(r-1)}$, $\theta_2^{(r)} | \theta_1^{(r)}, \theta_3^{(r-1)}$, $\theta_3^{(r)} | \theta_1^{(r)}, \theta_2^{(r)}$ and $\theta_1^{(r+1)} | \theta_2^{(r)}, \theta_3^{(r)}$, so on and so forth. In our algorithm, our iteration includes four steps. The first two steps are for the utility function and the last two steps are for the selection equation.

Data augmentation refers to the method that introduces additional variables (such as latent variables in a discrete model) as the augmented variables. The augmented variables connects observables with the underlying data generating process, making the posterior distribution of parameters easier to be sampled from. For example, in the Probit model $y = 1(x\beta + \varepsilon > 0)$, this method introduces the augmented data $y^* = x\beta + \varepsilon$. The original problem of calculating the posterior distribution of β conditional on binary variable y is difficult, while the posterior distribution of β conditional on continuous y^* is mush easier and reduces to a simple linear model. In our sampling algorithm, the left hand side of the utility function and selection equation are all discrete. We borrow the idea of data augmentation and introduce the latent variables as the augmented data.

1.3.2 Bayesian Algorithm for the Game-Theoretic Model

In what follows, we present the Bayesian algorithm for our model. The Gibbs Sampling algorithm iteratively samples four blocks of parameter. The first two blocks are the latent variable and the parameters in the utility function, and the last two blocks are the latent variables and the parameters in the selection equation. If the selection mechanism is known, one can just repeat the first two steps to estimate the utility function.

Let $y_i^* = \beta x_i + \gamma y_{-i} + \varepsilon_i$, and $s^* = \lambda Z + \delta W + \eta$ represent the latent variables of the utility function and selection equation respectively. We are interested in the posterior distribution of y^*, $\theta_U \equiv \{\beta, \gamma\}$, s^*, and $\theta_S \equiv \{\lambda\}$. The following algorithm describes the general steps of Bayesian procedure for the structural model:

Algorithm I (General Model)

For given starting values of s^*, θ_U and θ_S, the Gibbs sampler involves repeating the following 4 steps iteratively:

1. Sample $y^*|_{s^*,\theta_U} \sim TN$, where TN stands for truncated normal distribution;

2. Sample $\theta_U|_{y^*}$;

3. Sample $s^*|_{y^*,\theta_S} \sim TN$;

4. Sample $\theta_S|_{s^*} \sim N$.

Note that Gibbs sampling requires sampling one subset of parameters conditional on the rest of the parameters. We drop the parameters that do not affect the conditional distribution. For example, since y^* is independent of θ_S given s^*, θ_S can be dropped in step 1. The argument applies for the rest of the steps.

An easy way to understand Algorithm I is to relate it to Bayesian estimation of the Probit model. Consider a Probit model where the latent variable is expressed as a linear function. In the first step, the latent variable follows a normal distribution truncated from below at 0 if the outcome is 1 and follows a normal distribution truncated from above at zero if the outcome is 0. In the second step, the parameter is sampled from a normal distribution as in the linear model of the latent variable. Our sampling procedure "repeats" Porbit two times: one for utility functions and one for selection equation. The first two steps of our sampling algorithm focus on the utility function, where we first draw the latent utility and then draw the parameters. We then do Probit again, draw the latent value for the selection equation and then draw the parameters of the selection equation.

A few remarks about the algorithm: First, by using Gibbs sampling, data augmentation and reparameterization, all the steps require sampling from standard distributions only. Second, the starting value will not affect the limiting distribution that the Markov Chain will finally converge to. In practice, we use the naive Probit estimator as the initial value for θ_U , set the initial value in the selection equation to be 0 and randomly draw the initial value of s^* from a Bernoulli distribution. Third, the algorithm assumes away mixed strategies. For applications such as social interaction and entry competition, assuming only pure strategies are played is not controversy. In the entry game, if mixed strategies are being played, there is a positive chance that both firms enter. In the region allowing for mixed strategies, firms get negative profit if both enter. In the social interaction game, if the mixed strategy is being played, there is a positive chance that only one takes the action. After observing such case, both individuals want to change their actions. If firms and individuals have time to adjust their actions, they will finally reach to a state where none of them wants to deviate. Such state can be characterized by a complete information game where only pure strategies are played.

Example: 2-Person Game of Strategic Complements This subsection uses the 2-person social interaction game as an example to illustrate how the sampling algorithm works. We make a slight modification. Instead of using $y_i^* = \beta x_i + \gamma y_{-j} + \varepsilon_i$, we sample $c_i = \frac{\beta x_i + \varepsilon_i}{\gamma}$ which has a one-to-one mapping to y^* given the parameter θ_U and the data y_{-j}. We can stick to the algorithm discussed in the previous subsection. Here we sample c instead of y because the relationship between c and the outcome is more direct and has been discussed in Figure 1.1. The next algorithm summarizes the procedure.

Algorithm II (2-Person Game)

In the 2-person social interaction game, for given starting values of s^*, θ_U and θ_S, the

Gibbs sampler involves repeating the following 4 steps iteratively:

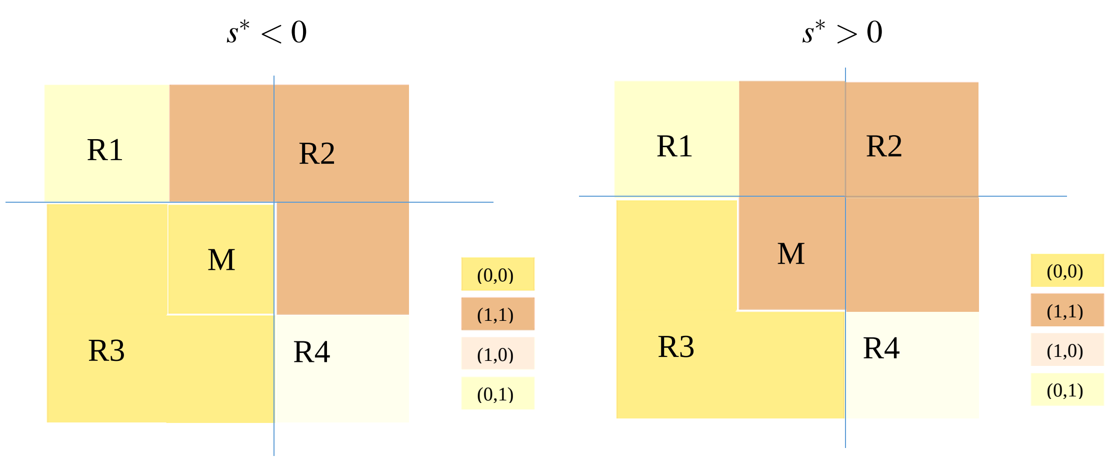
1. Sample $c|_{s^*,\theta_U} \sim TN(\frac{\beta x}{\gamma}, \frac{1}{\gamma}I)$;

2. Sample $\hat{\beta}|_c \sim N(b, \tilde{\sigma}^2(X'X)^{-1})$ *and* $\tilde{\sigma}_i \sim IG(\frac{n}{2}, \frac{v \cdot s^2}{2})$ *, and recover* (β, γ) *from* $(\tilde{\beta}, \tilde{\gamma})$*, where* $\tilde{\beta} = \frac{\beta}{\gamma}$ *,* $\tilde{\sigma} = \frac{1}{\gamma}$ *,* v *,* s^2*,* b *are the degrees of freedom, the sum of square errors and the coefficient of the OLS estimation of* c_i *on* X_i *;*

3. Sample $s^*|_{c,\theta_S} \sim TN(\lambda Z, 1)$;

4. Sample $\theta_S|_{s^*} \sim N(\lambda Z, 1)$ *.*

Figure 1.2: s^* and c



Step 1: The sampling distribution of c depends on s^* drawn in the last iteration and the data. Figure 1.2 shows the relationship between the sign of s^* and the outcomes of the game. Note $c = \frac{\beta}{\gamma}x + \frac{1}{\gamma}\varepsilon$. c follows a truncated multivariate normal distribution $TN(\frac{\beta}{\gamma}X, \frac{1}{\gamma}I)$ where the truncation depends on s^* and the data. The truncation values are listed in Table 1.1.

Table 1.1: Truncation of c Given s^* and the Data

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$s^* < 0$	$c \in R_3 \cup M$	$c \in R_1$	$c \in R_4$	$c \in R_2$
$s^* > 0$	$c \in R_3$	$c \in R_1$	$c \in R_4$	$c \in R_2 \cup M$

Step 2: Let $\tilde{\beta} = \frac{\beta}{\gamma}$ and $\tilde{\sigma} = \frac{1}{\gamma}$, so c can been expressed using the reduced form parameters: $c = \frac{\beta}{\gamma}x + \frac{1}{\gamma}\varepsilon = \tilde{\beta}x + \tilde{\sigma}\varepsilon$. The posterior distribution of $(\tilde{\beta}, \tilde{\sigma})$ can be derived using the results from linear model. Once the reduced form parameters $\tilde{\beta}, \tilde{\sigma}$ are sampled, the structural parameters β, γ can been recovered.

Step 3: The truncation of s^* depends on c drawn in the first step. If c falls into a region where only a unique equilibrium is predicted ($c \in R_1 \cup R_2 \cup R_3 \cup R_4$), it doesn't provide any information about the selection of the equilibium, so s^* is not truncated. If $c \in M$ and the data is $(0,0)$, it must be the case that the low equilibrium is selected, so s^* is truncated from above at 0 ($s^* \in (\infty, 0]$); similarly, if $c \in M$ and the data is $(1,1)$, $s^*$is truncated from below at 0 ($s^* \in [0, \infty)$).

Step 4: This step is a linear problem of $s^* = \lambda Z + u$.

1.3.3 Extensions

Note that a few assumptions are made in the previous example: we assume $\varepsilon \sim N(0, I)$ and $\varepsilon \perp u$. In this subsection, we relax the assumptions in three ways: allow correlation in payoffs, allow endogenous selection, and allow alternative distributional assumptions about ε.

Correlation in ε

The first two steps of our sampling algorithm needs to be modified slightly if we allow for correlation between payoff functions. Let Σ represent the variance and covariance matrix of vector ε, Σ has 1 in its diagonal entries and ρ in its off-diagonal entries. In the first step of Gibbs sampling, y^* will be drawn from a truncated normal distribution with covariance matrix. In the second step, once y^* is augmented, we have a systems of payoffs which form a seemingly unrelated regressions (SUR) model. The posterior distribution of Σ follows a Wishard distribution, and the posterior distribution of (β, γ) follows normal distribution.

Endogenous Selection

Suppose selection is endogenous according to the formula $\varepsilon = \rho u + v$. In the second step of the previous algorithm, $y^*|_{s^*,\theta_U,\theta_S}$ is the same as $y^*|_{s^*,\theta_U}$because payoff functions depend on the selection equation only through equilibrium selection s^*. If ε and u are correlated, the conditional distribution of ε given u follows $\varepsilon|_u \sim N(\rho u, \sigma_v^2)$. In Step 1 of our sampling algorithm, we need to first recover $u^{(r-1)}$ from the previous interaction using $s^{*(r-1)} = \lambda Z + u$, then draw $y^* = \beta x + \gamma y + \varepsilon$ conditional on $u^{(r-1)}$. Similarly, the conditional distribution of u given ε is $u|_\varepsilon \sim N(\frac{\rho}{1+\rho^2}(\varepsilon_1 + \varepsilon_2), \frac{1-\rho^2}{1+\rho^2})$. Therefore in Step 3 we need to recover ε first and then draw u conditional on ε.

Logit Selection Equation

Using data augmentation, a Logit model for the selection equation can be estimated using the Bayesian method. A more complicated issue is when the selection of multiple equilibria is characterized by a multinomial Logit model. There are mature algorithms for estimating multinomial Logit using data augmentation and MCMC. Because our main message regarding identification and estimation can be delivered using a selection equation of a Probit form, we leave a formal distribution about game-theoretic models with multinomial Logit selection function for future research.

1.4 Monte Carlo Study

In the Monte Carlo experiment, we consider the following game:

$$y_i = 1(\beta_0 + \beta_1 x_i + \gamma y_j + \varepsilon_i > 0).$$

When $\gamma > 0$, the two outcome $(0,0)$ and $(1,1)$ could be the multiple equilibria for some game. The high equilibrium $(1,1)$ is selected according to a Probit function

$$S_{11} = 1(\lambda_0 + \lambda_1 Z + u > 0).$$

We set $x \sim N(0,1)$, $z \sim U(0,1)$ and $u \sim N(0,1)$. The parameters are $\beta = (-1.5,1)$, $\gamma = 3$, and $\lambda = (-0.5,2)$. The sample size equals 2000. The parameters are chosen such that half of the observations comes from the game with multiple equilibria. Table 1.2 shows the percentage of each outcome in our sample:

Table 1.2: Frequency of the Outcomes in the Monte Carlo Experiment

	(0,0)	(0,1)	(1,0)	(1,1)
Single	21.65%	1.9%	1.3%	22.80%
Multiple	17.65%			34.70%
Total	39.30%	1.90%	1.30%	57.50%

In the sample, 52.35% of games have multiple equilibria. Among them 67.2% of the games select the high equilibrium. We know this because it is generated data. In the real world, researchers are not able to tell how many games have multiple equilibria, neither do they know if the observed outcome comes from the region of multiple equilibria or not. For example, the observation $(0,0)$ may be from a game where it is the unique equilibrium, or it may come from a game where $(0,0)$ and $(1,1)$ are both equilibrium and players coordinate to play $(0,0)$.

Table 1.3 collects coefficients and standard errors estimated by MLE approaches. The MLE_low approach assumes $(0,0)$ is always selected if the game has multiple equilibria and the MLE_high approach assumes $(1,1)$ is always selected. The MLE_group approach uses the "grouping method" that calculates the likelihood function of the event "$(0,0)$ or $(1,1)$". The MLE_p uses a probabilistic equation for the selection but neglects the true covariates Z. It assumes the high equilibrium $(1,1)$ is selected with probability p. The MLE_z approach uses the information on Z. It includes Z in the selection equation. Likewise, the Bayesian_p approach assumes the selection follows a Bernoulli distribution with mean p, and the Bayesian_z uses a Probit selection equation with covariate Z.

Many features of estimating games are illustrated in Table 1.3. First, the bias of Probit regression highlights the already well-known importance of using a structural model that is robust to strategic interactions between agents. A shock to player i's utility affects the action of the player i, which in turn affects the action of player j, therefore the magnitude of the strategic interaction (in absolute value) is upward biased if Probit regression is used.

Table 1.3: Probit and MLE of the Experiment

	β_0	β_1	γ	λ_1	λ_2	p
True	-1.5	1	3	-0.5	2	(0.7)
Probit	-1.958	0.892	4.147			
	(0.073)	(0.057)	(0.120)			
MLE_low	-0.397	0.649	2.267			
	(0.026)	(0.032)	(0.071)			
MLE_high	-1.670	0.689	2.485			
	(0.068)	(0.033)	(0.077)			
MLE_group	-1.286	0.952	2.944			
	(0.165)	(0.078)	(0.136)			
MLE_p	-1.478	1.010	3.041			0.640
	(0.097)	(0.057)	(0.127)			(0.040)
MLE_z	-1.513	1.006	3.031	-0.558	2.053	
	(0.089)	(0.056)	(0.125)	(0.132)	(0.251)	
Bayesian_p	-1.480	1.008	3.138			0.616
	(0.127)	(0.077)	(0.174)			(0.045)
Bayesian_z	-1.477	0.982	3.046	-0.644	2.189	
	(0.192)	(0.088)	(0.204)	(0.410)	(0.703)	

Second, different assumptions of the selection of the equilibrium lead to different estimators. In this example, we set the parameters such that half of the sample falls into the region of multiple equilibria. Having a correct specification of the equilibrium selection mechanism is especially important in this case. In our experiment, the high equilibrium

$(1,1)$ is selected with probability approximately equal to 0.7. Neither the assumption used in MLE_low or MLE_high is correct, therefore neither MLE_low nor MLE_high are consistent..

Third, the "grouping method" is consistent, but the variance of the estimator is larger than the estimators using other approaches. The parentheses in Table 1.3 collect the standard errors of the MLE estimators. The standard error of the MLE_group estimator is larger than standard error of the other MLE estimators. In the next section, we will see the efficiency loss of the "grouping method" is so strong that we are not able reject the null hypothesis that the parameter equals zero.

Fourth, payoff functions are correctly estimated if an equilibrium selection equation is included in the model. MLE_z uses a Probit selection equation with covariate Z. The model is correctly specified. MLE_z estimates coefficients that are very closed to the true data generating process. In addition, even if the selection equation is degenerate, the payoff functions are still well estimated. The MLE_p approach assumes the probability of selecting the high equilibrium follows a Bernoulli distribution with mean p. It gives a decent estimation of the utility function even though the variable Z is neglected.

The last few rows of Table 1.3 collect the results from the Bayesian method. The Bayesian_p approach neglects the covariate Z. The posterior distribution of a Bernoulli coefficient follows a beta distribution, so we modify Step 4 of the Bayesian procedure accordingly. The Bayesian_z approach includes Z in the selection equation. In Figure A.1 in the appendix, we plot the draws from the posterior distribution. It can been seen from Table 1.3 and Figure A.1 that the Bayesian method is able to a provide a good estimation of the model.

1.5 Application

This section applies the econometric framework to study entry competition in the home improvement industry. Entry competition is a classic example of a discrete game of complete information; therefore, we would like to study a real-world entry game to discuss new insights this chapter provides relative to the previous literature. We first provide a brief overview of the industry and the two key players. After that we present the empirical findings of the competition effect and equilibrium selection. We find an interesting result that equilibrium selection always favors one firm.

1.5.1 Industry Overview

The U.S. home improvement retail industry comprises retailers that sell appliances, building materials, hardware, lawn and garden products, and home supplies. The industry is highly concentrated. The annual sales of the industry are about 300 billion dollars, half of which is generated by the two largest players, Home Depot (HD) and Lowe's (LOW). Both Home Depot and Lowe's have around 2000 stores across the U.S. The third largest firm in the industry, Menard's, has less than 400 stores and does not do business nationwide. Because of the high concentration, studying entry competition is interesting because consumer welfare is affected by the number and the identity of the firm in the market.

Lowe's was founded in the 1940s. It has expanded since the 1950s through opening throughout North Carolina. Home Depot opened its first store in 1978 in Georgia. Home Depot's proposition was to build home improvement superstores so it adopted the big-box format since its start. Facing strong competition from Home Depot, Lowe's switched to the big-box format in the 1980s. Since then, the two companies have grown rapidly and expanded nationally.

Today, the home improvement industry is a mature industry. Both Home Depot and Lowe's have stores in all U.S. states. The number of stores has been stable for the past few years. Home Depot is currently the largest retailer in the industry with more than 2200 stores nationwide. Lowe's is the second largest retailer and has around 1800 stores. Despite some minor differences, the two companies have very similar business structures. The products and services provided by these two companies are very similar as well. Figure A.2 in the appendix shows two maps of all stores of the two firms in the contiguous United States. It can be seen that the two stores have very similar national-wide geographic distributions.

1.5.2 Data and Summary Statistics

The market is defined as a Core Based Statistical Area (CBSA). CBSA is a commonly used notion of geographic area. It is defined as an urban core whose population exceeds 10 thousand. Our data comes from two sources. First, we collect the addresses of all stores owned by Home Depot or Lowe's in the U.S. We use the addresses to determine which market each store belongs to. Second, we refer to the U.S. census to collect information on the characteristics of the market, such as population and income.

Our sample consists of 919 CBSA areas. Table 1.4 shows the descriptive statistics of the data. We provide the descriptive statistics of the number of stores of each type, the dummy variable indicating whether the firm enters, and market conditions including log population and log per capita income. There is skewness in the data. Most CBSAs have only 0 or 1 store opened by each company, but the mean number of stores is much larger because a few CBSAs have 20 or more stores. Table 1.4 also shows significant variation in population and strong correlation between population and market structure. The population of the CBSA at the 90% percentile is about 20 times larger than the population of the CBSA at the 10th percentile. The variation in per capita income is much smaller. In the empirical analysis, we focus on the subsample of CBSAs whose population does not exceed 100 thousand. It is very rare that a firm opens more than two stores in a market of this size. We limit our attention to areas of small population because these are the places where the entry decision is strategic.

Table 1.4: Descriptive Statitics

	Whole Sample					Pop<100k	
	Mean	Stdev	10%	Median	90%	Mean	Stdev
Obs	919					538	
# of HD	1.402	4.021	0	0	3	0.193	0.400
# of LOW	1.308	2.838	0	0	4	0.242	0.433
Entry_{HD}	0.442	0.497	0	0	1	0.191	0.394
Entry_{LOW}	0.486	0.500	0	1	1	0.240	0.427
Log population	11.524	1.251	10.235	11.234	13.286	10.692	0.456
Log pcincome	10.517	0.186	10.307	10.493	10.744	10.483	0.189

In order to take a closer look at the entry pattern, we divide the markets into population brackets and count the frequency of different market structures in each group. Table 1.5 shows three features. First, as expected, the number of stores grows with population. Second, it is very rare that one firm opens two stores but the other firm does not enter. This suggests that the two firms do act strategically. Third, in most of the population brackets, there are more markets having LOW as the single entrant than markets having HD as the single entrant. Even though Home Depot has more stores nationwide, it does not enter as many small markets as Lowe's does.

Table 1.5: Population and Entry Patterm

(HD, LOW)	$[10,30)$	$[30,40)$	$[40,60)$	$[60,90)$	$[90,120)$	$[120,200)$	$[200,400)$	≥ 400
Obs	113	114	160	119	78	118	85	132
$(0,0)$	91.2	73.7	60.6	33.6	25.6	12.7	1.2	9.1
$(0,1)$	0.9	14.0	22.5	32.8	30.8	14.4	2.4	0.0
$(1,0)$	7.1	11.4	16.9	21.0	5.1	9.3	4.7	0.0
$(1,1)$	0.9	0.9	0.0	11.8	30.8	39.0	20.0	0.8
$(\geq 2,0)$	0.0	0.0	0.0	0.8	1.3	1.7	1.2	2.3
$(0,\geq 2)$	0.0	0.0	0.0	0.0	1.3	3.4	1.2	0.0
$(\geq 1,\geq 1)\backslash(1,1)$	0.0	0.0	0.0	0.0	5.1	19.5	69.4	87.9

To sum up, the summary statistics show the close connection between population and entry. We do find evidence of strategic action between Home Depot and Lowe's, and Lowe's is more likely to be the only firm in the market. Next, we would like to explore more about the magnitude of the entry effect and the selection of equilibrium using the structural approach.

1.5.3 Empirical Finding on Entry

We first regress the entry of one firm on the entry of the other firm using Probit. Probit regression shows negative effect from opponent's entry, confirming that the entry game is indeed a game of strategic substitutes. The magnitude of the entry coefficient in the regression of HD and LOW are around 0.4, which is upward biased if there exists strategic interaction between the two firms.

Table 1.6: Probit Regression on Entry

	Entry_i
Constant	-18.912
	(3.055)
L	0.111
	(0.105)
Entry_{-i}	-0.487
	(0.160)
$\text{L}\cdot\text{Entry}_{-i}$	0.145
	(0.222)
Log population	1.167
	(0.174)
Log pcincome	0.878
	(0.166)

Probit Regression provides evidence of symmetry between the two firms. The coefficients for the identity dummy are not significant, suggesting that the two firms' responses to opponent's entry are similar. Given the moderate sample size of the data and the similarities between the two Probit regressions, we adopt a symmetric version of the model in our structural analysis and assume the two firms have the same utility function.

Specifically, we consider an entry game with the following utility function:

$$y_i = 1(\beta_0 + \beta_1 \ln(population) + \beta_2 \ln(pcincome) - \gamma y_{-i} + \varepsilon_i > 0).$$

Table 1.7 collects our main results from MLE using different assumptions about equilibrium. The four methods from top to bottom are the grouping method, a deterministic rule assuming LOW is always selected, a deterministic rule assuming HD is always selected, and a probabilistic rule assuming HD is selected with probability p.

A few findings should be noted in Table 1.7. First, as expected, bias of the competition effect has been corrected using the structural approach. Compared with the result from Probit estimation, the competition effect estimated by MLE drops from 0.37 to 0.20-0.25. Second, though in principle the grouping method is robust to equilibrium selection, its efficiency loss is a big concern. In our application, summary statistics show clearly how population affects market structure, but the standard error in the grouping method is so large that the coefficient for population is not significant. Third, the results from two extreme assumptions about equilibrium selection differ slightly. In our application, the issue of multiple equilibria is not as strong as in the Monte Carlo experiment. Finally, when estimating equilibrium selection empirically, our finding suggests the probability of selecting LOW is approaching one, meaning that LOW always enters the market that admits multiple equilibria.

Table 1.7: MLE on Entry

	β_0	β_1	β_2	γ	p_{01}
MLE_group	-17.957	1.362	0.241	0.226	
	(3.017)	(2.185)	(3.180)	(0.090)	
MLE_01	-18.045	1.371	0.240	0.250	
	(3.020)	(0.122)	(0.256)	(0.094)	
MLE_10	-17.867	1.353	0.240	0.204	
	(3.013)	(0.121)	(0.256)	(0.085)	
MLE_p	-18.043	1.371	0.240	0.250	0.999
	(2.892)	(0.125)	(0.250)	(0.100)	(0.984)

Because MLE suggests an extremely high probability of selecting LOW, we conduct Bayesian estimation by assuming a deterministic rule such that LOW enters if the market has multiple equilibria. Table 1.8 reports the sample mean and sample standard deviation of sampling from the posterior distribution. The Bayesian approach gives the same conclusion that the entry effect is around 0.2, which is lower than the Probit estimator.

Table 1.8: Bayesian Estimation of the Entry Game

	β_0	β_1	β_2	γ
Bayesian_01	-17.902	1.358	0.239	0.206
	(2.978)	(0.119)	(0.255)	(0.062)

One surprising result of the model is that one firm dominates. Our finding suggests that if either of the firms could enter a market but not both, Lowe's is the firm that enters. This may because Lowe's has a longer history and may have had a first-move advantage in many markets.[6] The first Lowe's store opened 20 years before the start of Home Depot. Even though the rapid expansion of Lowe's did not began until Home Depot was established, as a more experienced firm, Lowe's may have more knowledge about the potential profitability of the market, thus it quickly entered the markets where either firm could be the monopoly.

1.6 Conclusion

As discussed above, this chapter illustrates the feasibility and benefits of including equilibrium selection in empirical games. Our econometric framework achieves point identification by using more information from the data than is common using a structural approach. We consider both exogenous and endogenous selection mechanisms, and we analyze the consequences of having neglected variables in the selection equation. In addition, we develop an estimation procedure that uses an MCMC algorithm rather than MLE or MSM. Doing this may not be straightforward at first glance, but it overcomes the computational issues of searching for a global maximizer or minimizer of a simulated objective function.

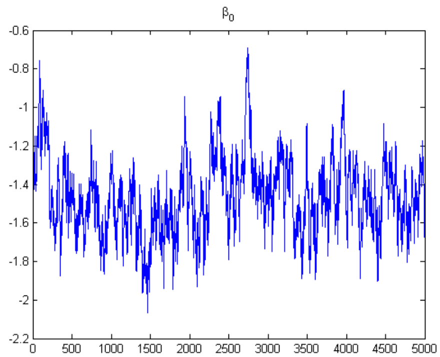
[6] Jia (2008) studies the entry competition between Kmart and Walmart. She assumes the selection of equilibrium favors Kmart because Kmart was derived from an established entity. Our finding provides support of the assumption used in her paper, because we find the equilibrium selection favors the firm that has a long history.

A Monte Carlo experiment illustrates the performance of our modeling strategy and estimation procedure. Finally, using the framework proposed in this chapter, we estimate the entry effect in the competition between Home Depot and Lowe's and find an interesting equilibrium selection mechanism. Our framework could be adapted to more complicated models. For example, we may allow random coefficients in the payoff function or consider generalizations on the selection equation. We believe adding a selection equation is fruitful in many of these generalizations. More studies on identification and estimation need to be done along this direction.

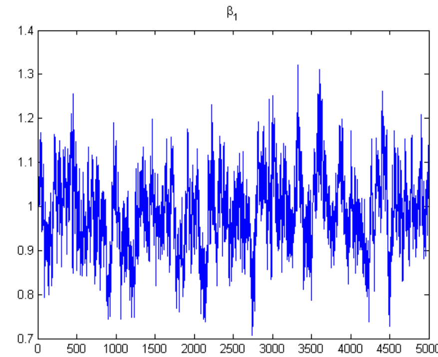
Appendix

Figure A.1: Sampling of Posterior Distribution Using the Bayesian_z Approach

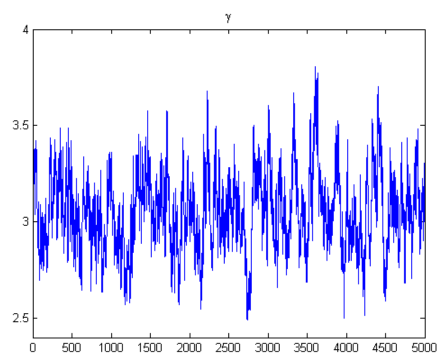
$\beta_0 = -1.5$



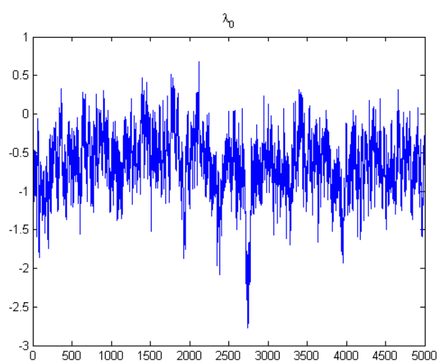
$\beta_1 = 1$



$\gamma = 3$



$\lambda_0 = -0.45$



$\lambda_1 = 2$

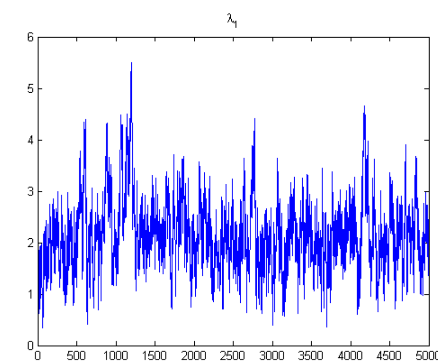
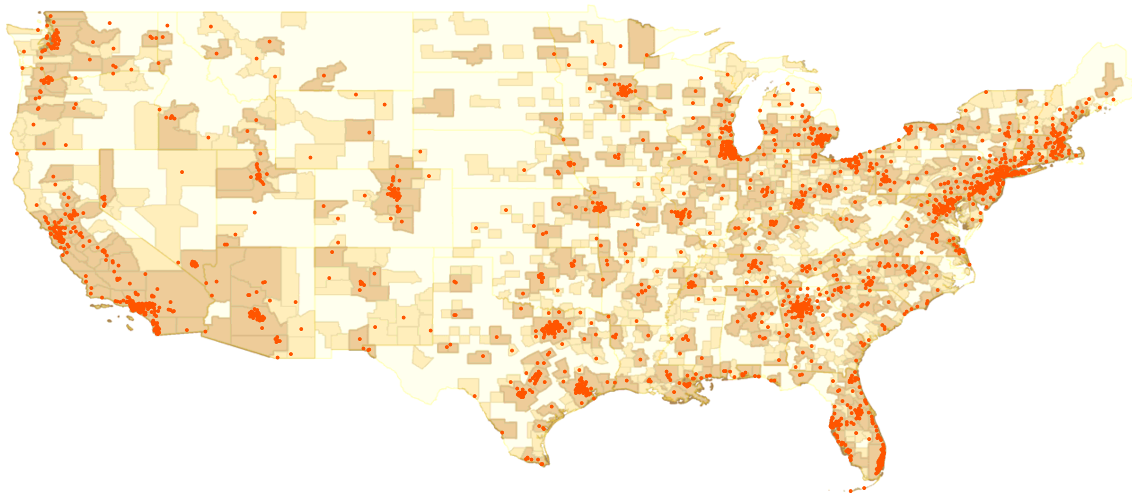


Figure A.2: Maps of HD and LOW Stores in the Contiguous United States

Home Depot Stores



Lowe's Stores

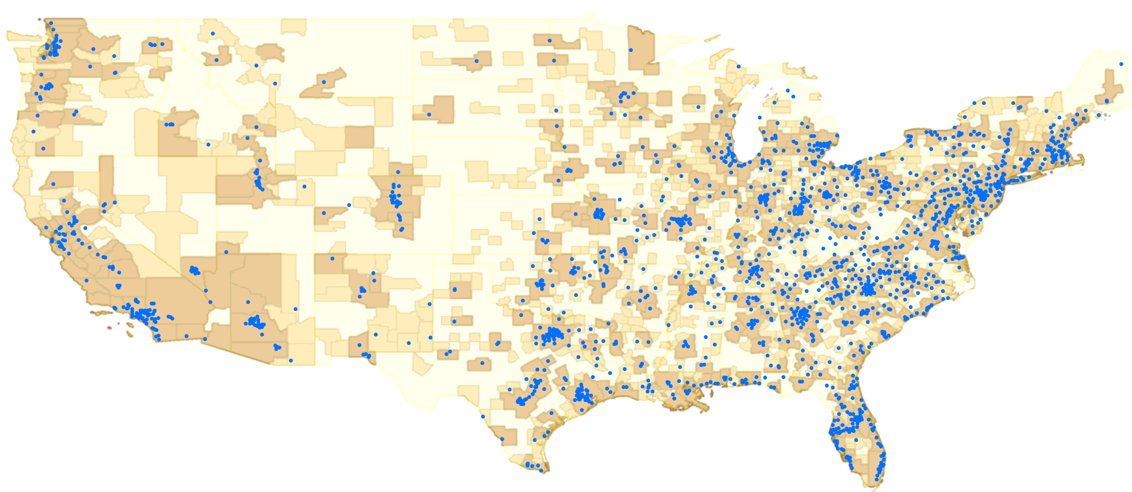
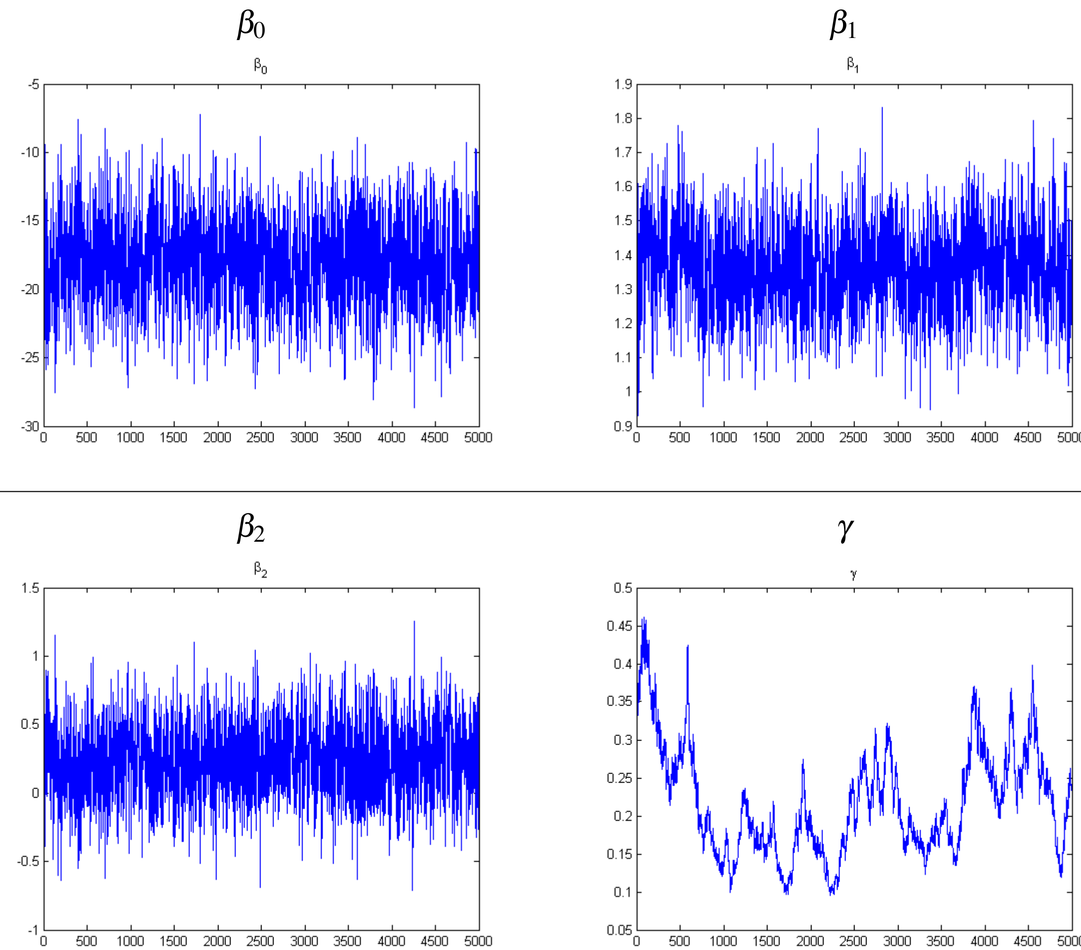


Figure A.3: Sampling of Posterior Distribution in the Entry Game



Chapter 2

A Unifying View of Partial Identification Approaches in Discrete Games of Complete Information

2.1 Introduction

Games of complete information have wide applications in empirical industrial organization, labor and health economics, among others. In recent years, several partial identification approaches have been proposed to estimate games of complete information. Partial identification approaches are attractive because they do not rely on additional assumptions of equilibrium selection mechanism. To the best of our knowledge, however, there exists no discussion about the relationship among various partial identification approaches.

The econometric framework we consider dates back to the work of Bjorn and Vuong (1984), Jovanovic (1989), and Bresnahan and Reiss (1991a, 1991b). Identification could pose challenging problems due to the lack of information of how game outcomes are selected when a game has multiple equilibria. Point identification could be achieved in certain ways with strong assumptions or/and restrictive structures. Examples include making assumptions of an equilibrium selection mechanism (Bjorn and Vuong 1984, Jia 2008), looking for a statistic that is invariant to equilibrium selection rules (Berry 1992), estimating equilibrium selection (Bajari, Hong and Ryan 2010, Narayanan 2013), among others. If no assumption of the equilibrium selection mechanism is imposed, the game is generally not point identified. Partial identification approaches have recently drawn much attention because of their flexibility in dealing with a variety of game-theoretic models without making strong assumptions on the equilibrium selection mechanism. However, the various partial identification methods were proposed independently and little is known about their relationship.

The goal of this chapter is to study the relationship among the existing partial identification approaches for discrete games of complete information. We divide the existing methods into three categories and aim to present them in a unifying framework. Specifically, we consider the three most influential approaches in the literature. The first is the sensitivity analysis approach as in Grieco (2014). This approach treats the equilibrium selection mechanism as an infinite dimensional nuisance parameter, following a general sensitivity analysis approach proposed in Chen, Tamer and Torgovitsky (2011). The identified set collects parameters such that there is at least one valid equilibrium selection mechanism that could predict the moments of the data. The second category is the bound estimation approach proposed by Ciliberto and Tamer (2009) and Andrews, Berry and Jia (2004). They construct upper and lower bounds of choice probabilities based on sufficient and necessary conditions of the model. The identification set is the set of parameters that satisfy these bounds. The third category establishes sharp identification of the model. Galichon and Henry (2011) introduce the concept of generalized likelihood function defined on sets of outcomes. The generalized likelihood differs from the standard likelihood because its integration exceeds one. The identified set is the set of parameters whose generalized likelihood is no less than the probability of observing the set of outcomes. Beresteanu, Molchanov and Molinari (2011) use the random set theory to make set inference and show their method has the same identified set as in Galichon and Henry (2011) within the complete information framework.

We first propose a modified version of bound estimation. This modified approach is a natural extension of Ciliberto and Tamer (2009), who construct bounds of the probability of individual outcomes. Andrews, Berry and Jia (2004) discuss the idea of constructing probability bounds on combinations of outcomes. We further the idea in Andrews, Berry and Jia (2004) by exploring restrictions on the probability of all combinations of outcomes. Compared with Ciliberto and Tamer (2009), the modified approach explores more predictions from the data, therefore its identification set is tighter.

We then establish the equivalence of the identified sets of the aforementioned three approaches using the modified bound estimation approach as a bridge. We show that not all bounds in modified bound estimation are needed because imposing upper bound on a set is the same as imposing lower bound on its complement. This naturally leads to the equivalence between the modified bound estimation method and the sharp identification approach proposed in Galichon and Henry (2011). We then show that the identified set from the sensitivity analysis is neither larger nor smaller than the identified set from the modified bound estimation method, therefore these two sets are identical. As a result, the three partial identification approaches are proven to be equivalent in the sense of drawing inference on the same identified set.

The chapter is organized as follows. Section 2 outlines the econometric framework. Several examples are included to illustrate the econometric problems. In Section 3 we describe in detail the sensitivity analysis approach, the bound approach, and the sharp identification approach, in sequence. Section 4 is devoted to comparing these three approaches and establishing our main results. Section 5 concludes.

2.2 Econometric Framework

2.2.1 Model

We consider a static game of complete information. There are N players, each has a finite set of actions. For player i, the utility of playing action y_i depends on the observed explanatory variable x_i, the actions of other players y_{-i}, and a random shock ε_i:

$$u_i(y_i; x_i, \varepsilon_i, y_{-i}, \theta). \tag{2.1}$$

The functional form of the utility function is given. It does not necessarily need to be linear in all its inputs. Under the complete information framework, shocks are common knowledge to all players.

In many circumstances, the game may have more than one Nash equilibrium. Let

$\mathscr{E}(\varepsilon,x;\theta)$ denote the set of equilibria of the game described by $(\varepsilon,x;\theta)$, where letters without subscripts denote vectors of all game players. $\mathscr{E}(\varepsilon,x;\theta)$ may contain multiple outcomes depending on where individual characteristics and shocks locate.

Discrete games of complete information have a wide range of applications. For example, Bjorn and Vuong (1984) use this framework to study husband-wife labor force participation. The decision of whether or not to work depends on the decision of an individual's spouse. This framework is especially popular in the entry literature (e.g. Bresnahan and Reiss 1991a, 1991b, Berry 1992, Ciliberto and Tamer 2009). Entry decision is strategic because the underlying profit of one potential entrant depends on the actions of competing firms. This framework could also be used to model peer effects (Krauth 2006), because the utility of taking an action is affected by the actions of peers. In all of these cases, it is reasonable to assume complete information and pure strategy. If individuals and firms have enough time to adjust their actions, the realized action of one player is the best response to the realized action of all other players. Such feature can be captured by a model of complete information game with pure strategy Nash equilibria.

2.2.2 Examples

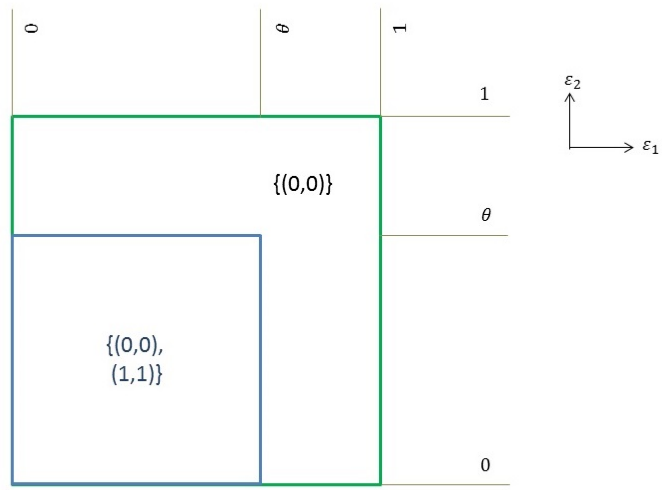
Example 1: Jovanovic (1989) The first example we discuss is from Jovanovic (1989). This model assumes a simple version of (2.1) that does not have an explanatory variable x . Jovanovic (1989) considers a 2-person game with payoff functions

$$\begin{aligned} u_1 &= (\theta y_2 - \varepsilon_1)y_1, \\ u_2 &= (\theta y_1 - \varepsilon_2)y_2, \end{aligned}$$

where $y_i \in \{0,1\}$ is player i's action, ε_1 and ε_2 are assumed to independently follow a uniform distribution $U[0,1]$. The remaining structural parameter is θ , which is assumed to be in the interval $(0,1]$.

Given the restrictions on the support of ε and θ, $y = (0,1)$ and $y = (1,0)$ can never be an equilibrium of a game. $y = (0,0)$ is an equilibrium if $\varepsilon \in [0,1]^2$. $y = (1,1)$ is an equilibrium if $\varepsilon \in [0,\theta]^2$. A game has multiple equilibria $\mathscr{E}(\varepsilon, x; \theta) = \{(0,0),(1,1)\}$ if $\varepsilon \in (0,\theta]^2$, and has unique equilibrium otherwise.

Figure 2.1: Equilibria of Jovanovic (1989)



This game nests in a broader set of models that only impose a positive sign restriction on θ but no restrictions on the support of θ and ε. These models are popular in the literature of social interactions. In social interactions, people gain extra utility when taking the same activity as the rest of peers do; this justifies the use of game-theoretic models. For social interaction games, multiple equilibria differ by having a set of players switch from one action to another simultaneously. The combination of multiple equilibria becomes difficult to tract when the number of players grows.[1] It is hard to find a reasonable assumption of equilibrium selection for all combinations of multiple equilibria.

Example 2: 2-Person Entry Game The second example is from Bresnahan and Reiss (1991a). It is a classic model in the entry literature.

[1]For example, in the game discussed above, $y = (0,0)$ and $y = (1,1)$ are the multiple equilibria. In a 3-person social interaction game, there are 7 different combinations of multiple equilibria depending on where ε is.

Two firms simultaneously decide whether or not to enter a market. Their profit functions are

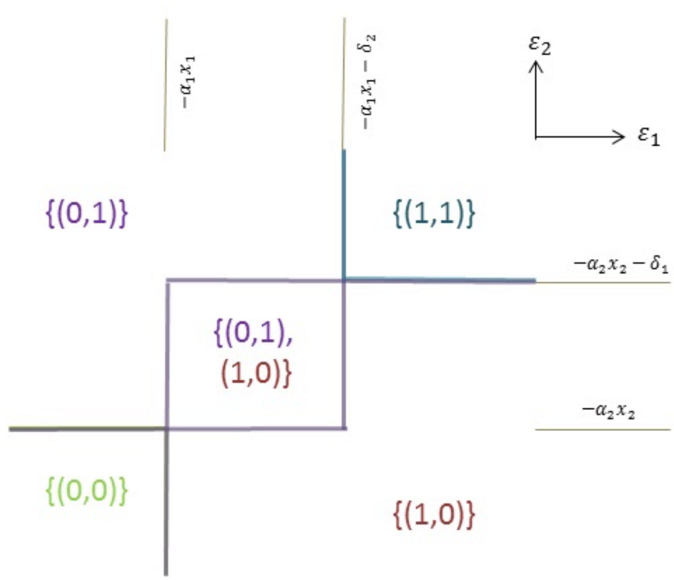
$$
\begin{aligned}
y_1 &= 1(\alpha_1 x_1 + \delta_2 y_2 + \varepsilon_1 \geq 0); \\
y_2 &= 1(\alpha_2 x_2 + \delta_1 y_1 + \varepsilon_2 \geq 0).
\end{aligned}
$$

where ε_i follows independently a standard normal distribution. Further assume $\delta_i < 0$ for $i = 1, 2$.

In entry games, observables include market conditions and specific characteristics. Unobservables ε_i and ε_j represent fixed costs that are known to both firms but not to the researcher. A negative coefficient δ captures the feature that profit decreases with the entry of other firms.

In this game, $y = (0, 1)$ is an equilibrium of games such that $\varepsilon_1 < -\alpha_1 x_1 - \delta_2$ and $\varepsilon_2 > -\alpha_2 x_2$; $y = (1, 0)$ is an equilibrium of games such that $\varepsilon_1 > -\alpha_1 x_1$ and $\varepsilon_2 < -\alpha_2 x_2 - \delta_2$. These two regions are overlapping. If neither firm has too big or too small fixed costs, the game has two equilibria, $y = (0, 1)$ and $y = (1, 0)$.

Figure 2.2: Equilibria of the 2-Person Entry Game



Example 2 can be generalized to allow for more players. We say an entry game is symmetric if entry effect δ_i is the same for all firms. In a symmetric entry game, the total number of entrants is the same among multiple equilibria, therefore the likelihood on the number of entrants could be used to estimate the model. When a game is asymmetric, as discussed in the next example, the number of entrants varies across multiple equilibria, therefore the likelihood approach based on the number of entrants cannot be used.

Example 3: Entry Game with 2 Types of Players The third example is taken from Berry and Tamer (2007) and Galichon and Henry (2011). There are two types of firms in the market. Each type has two firms. The profits of Type 1 firms depend on the total number of firms in the market whereas the profits of Type 2 firms depend both on the number and on the type of firms present in the market:

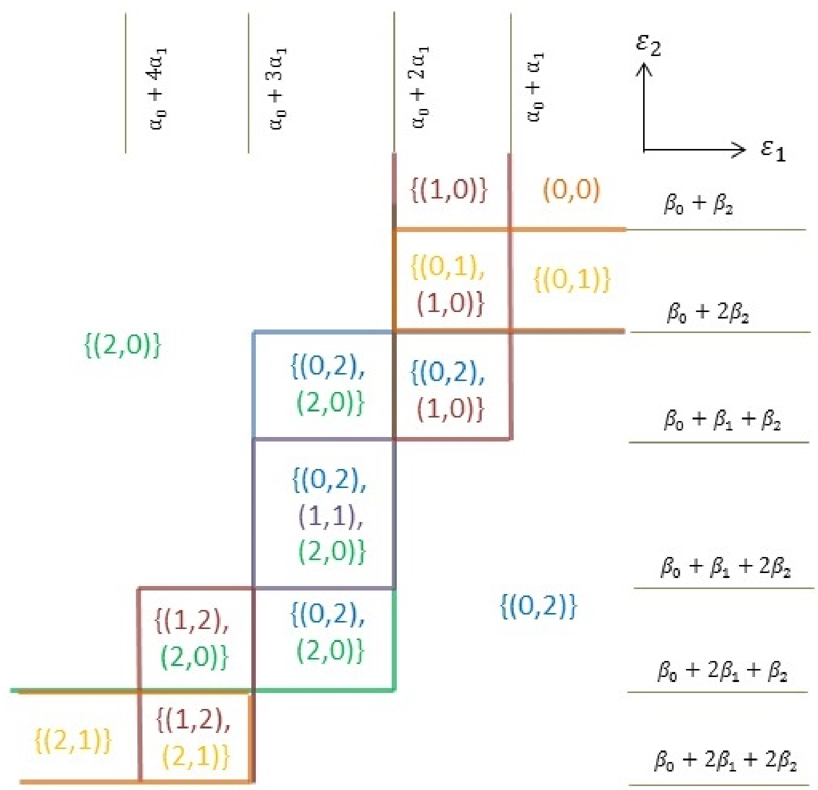
$$
\begin{aligned}
u_1 &= \alpha_0 + \alpha_1(y_1 + y_2) + \alpha_2 x_1 - \varepsilon_1, \\
u_2 &= \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 x_2 - \varepsilon_2.
\end{aligned}
$$

where the latent variable ε_i is assumed to be uniformly distributed over $[0,1]$. Further assume $\alpha_1, \beta_1, \beta_2 < 0$ and $\beta_2 > \beta_1$.

Each type of firms could have zero, one or two entrants, so this game has 9 potential outcomes. $Y = \{(i,j) : i,j = 0,1,2\}$. The following graph, taken from Galichon and Henry (2011), shows the set of equilibria of different games.

In this case, competition effects $(\alpha_1, \beta_1, \beta_2)$ are different depending on firm types. As noted before, when a game allows for asymmetric competition effects, multiple equilibria do not necessarily share the same number of entrants. For example, there are games whose equilibria are (1,2) and (2,0). Also, the game is complicated in the sense that the composition of multiple equilibria differs significantly when the unobservable ε changes. These features make identification difficult.

Figure 2.3: Equilibria of the Asymmetric Entry Game



2.2.3 Empirical Challenges of Estimating Games with Multiple Equilibria

The objective of the empirical analysis of game-theoretic models is to infer on the underlying utility function from actions chosen by game players. Understanding how individuals interact with each other is important in conducting counterfactual analyses.

A data set usually includes the outcome of each game, together with attributes of players. A usual way of estimating a discrete game is to find parameters that can predict the same probability distribution of the game outcomes. The probability of observing an outcome is an integration function over shocks, which are not observed by the econometrician:

$$\Pr(y|x;\theta) = \int \Pr(y|\varepsilon,x;\theta)dF(\varepsilon;\theta).$$

The main challenge of obtaining the moment equation for $\Pr(y|x;\theta)$ is the lack of knowledge of $\Pr(y|\varepsilon,x;\theta)$. These are two cases that are easy to handle. For games where y is not the equilibrium, the probability of observing y is zero. For games where y is the unique equilibrium, the probability of observing y is 1. However, if both y and other outcomes are equilibria of a game, it is not known which outcome will be observed. Estimation methods such as maximum likelihood or method of moments are therefore not applicable, because $\Pr(y|x;\theta)$ is not well-defined.

Though the value of $\Pr(y|\varepsilon,x;\theta)$ is not known, in practice there exists a probability distribution over multiple equilibria. An equilibrium selection mechanism is a conditional probability over the equilibria of a game.[2] Equilibrium outcomes are played with a non-negative probability; non-equilibrium outcomes are never played. A formal definition of equilibrium selection is as follows.

Definition 2.1 *let Y be the set of all possible outcomes of the game, an **equilibrium selection mechanism** $\lambda(y|\varepsilon,x;\theta)$ is a conditional probability distribution on $y \in Y$ such that*

1) $\lambda(y|\varepsilon,x;\theta) \geq 0$ *for* $\forall y \in Y$*;*

2) $\lambda(y|\varepsilon,x;\theta) = 0$ *for* $\forall y \notin \mathscr{E}(\varepsilon,x;\theta)$*;*

3) $\sum_{y\in\mathscr{E}(\varepsilon,x;\theta)} \lambda(y|\varepsilon,x;\theta) = 1$.

Lacking knowledge of equilibrium selection is the main reason why point identification is difficult. Point identification could be achieved in a few circumstances. The simplest way is to make an assumption of the equilibrium selection rule (Bjorn and Vuong 1984, Jia 2008, Krauth 2006). Alternatively, one could make parametric assumptions of the equilibrium selection mechanism (Bajari, Hong and Ryan 2010, Narayanan 2013). In some special cases, such as the symmetric entry game in Example 2, point identification could be achieved because there are some statistics (e.g. the number of entrants) that are not affected by equilibrium selection mechanism.

[2]Note that we only consider a pure strategy equilirbium. It is an assumption maintained in the three partial identification approaches discussed in this chapter. Galichon and Henry (2012) also discuss mixed strategy equilirbium as an extension. We leave mixed strategy equilibrium for future research.

If no assumption of equilibrium selection mechanism is made, the model is generally not point identified. Partial identification approaches attract a growing attention because they do not rely on assumptions of the selection of multiple equilibria. The next section reviews the three partial identification approaches that have been developed in the literature. The notations used in this chapter are slightly different from those in the original papers as we want to provide a unifying framework for all the three approaches.

2.3 Partial Identification Approaches

2.3.1 The Sensitivity Analysis Approach

Sensitivity analysis generally refers to the study of how uncertainty in outputs is affected by uncertainty in inputs. In the context of discrete games, the output - the parameter of interest - is affected by the input - equilibrium selection mechanism. The sensitivity analysis approach is used by Grieco (2014) in a more general model that allows both private and public shocks. We apply the same idea to games of complete information where all shocks are public and common knowledge to the players.

Sensitivity analysis treats equilibrium selection mechanism λ as an infinite dimensional nuisance parameters. It identifies the set of parameters that can be rationalized by at least one equilibrium selection mechanism. The identified set defined in the sensitivity analysis approach is

$$\Theta_I^{SA} = \{\theta \in \Theta : \forall x, \exists \lambda \in \Lambda \; s.t. \; P(y|x) = \int \lambda(y|\varepsilon, x; \theta) dF(\varepsilon; \theta) \; for \; \forall y\}, \tag{2.2}$$

where,

$$\begin{aligned} \Lambda \;\; = \;\; & \{\lambda : \forall \varepsilon, x; \theta, \; \lambda(e|\varepsilon, x; \theta) \geq 0 \; \forall e \in Y; \; \lambda(e|\varepsilon, x; \theta) = 0 \; for \; \forall e \notin (\varepsilon, x; \theta) \; and \\ & \sum_{e \in \mathscr{E}(\varepsilon, x; \theta)} \lambda(e|\varepsilon, x; \theta) = 1\} \end{aligned}$$

is the set of valid equilibrium selection mechanisms. Inference of the identified set can be made by using the profiled sieve likelihood ratio method discussed in Chen, Tamer and Torgovitsky (2011).

2.3.2 Bound Estimation

Bound estimation is a method that collects parameters such that the moments of the data are inside the lower and upper bounds predicted by the parameters. Lower and upper bounds are calculated based on sufficient and necessary conditions of the model. The parameter of true data generating process satisfies these bounds. See Ciliberto and Tamer (2009) and Andrews, Berry and Jia (2004).

Before presenting the bound estimation approach, let us first introduce a few notations.

Let A be a subset of Y, and define three sets

$$R(A|x;\theta) := \{\varepsilon : \mathscr{E}(\varepsilon, x;\theta) = A\},$$

$$S(A|x;\theta) := \{\varepsilon : \mathscr{E}(\varepsilon, x;\theta) \subseteq A\},$$

and

$$H(A|x;\theta) := \{\varepsilon : \mathscr{E}(\varepsilon, x;\theta) \cap A \neq \emptyset\}.$$

$R(A|x;\theta)$ denotes the set of games whose equilibrium set is the same as set A. $S(A|x;\theta)$ denotes the set of games whose equilibrium set equals A or is a subset of A. $H(A|x;\theta)$ is the set of games that contains at least one equilibrium in A. $H(A|x;\theta)$ may contain equilibria outside A. Clearly, $R(A|x;\theta) \subseteq S(A|x;\theta) \subseteq H(A|x;\theta)$. When A contains only one outcome, $R(A|x;\theta) = S(A|x;\theta)$. For simplicity we write $R(A)$ instead of $R(A|x;\theta)$ from now on. We use the same notation simplification for $S(A|x;\theta)$ and $H(A|x;\theta)$.

When constructing bounds for outcome y, the support of unobservables is partitioned into three regions. In the first region, $R(y)$, outcome y is the unique equilibrium of the game. The probability of observing y is one. In the second region, $H(y)\backslash R(y)$, y is one of the multiple equilibria. The likelihood of playing y is determined by the equilibrium selection rule. The last region, $\mathbb{R}^N\backslash H(y)$, denotes the regions where y is not the equilibrium of the game. The probability of observing y is zero. The original problem of calculating the probability of observing y is an integration function over these three regions:

$$\begin{aligned}
\Pr(y|x;\theta) &= \int \Pr(y|\varepsilon,x;\theta)dF(\varepsilon;\theta) \\
&= \int_{R(y)} \lambda(y|\varepsilon,x;\theta)dF(\varepsilon;\theta) + \int_{H(y)\backslash R(y)} \lambda(y|\varepsilon,x;\theta)dF(\varepsilon;\theta) \\
&\quad + \int_{\mathbb{R}^{\mathbb{N}}\backslash H(y)} \lambda(y|\varepsilon,x;\theta)dF(\varepsilon;\theta) \\
&= \int_{R(y)} dF(\varepsilon;\theta) + \int_{H(y)\backslash R(y)} \lambda(y|\varepsilon,x;\theta)dF(\varepsilon;\theta).
\end{aligned}$$

where $H(y)\backslash R(y)$ denotes the case where y is one of the multiple equilibria of the game. Ciliberto and Tamer (2009) construct lower bounds of $\Pr(y|x;\theta)$ by assuming $\lambda(y|\varepsilon,x;\theta) = 0$ for all games in this region; they construct upper bounds of $\Pr(y|x;\theta)$ by assuming $\lambda(y|\varepsilon,x;\theta) = 1$. Mathematically, $\Pr(y|x;\theta)$ is bounded by:

$$\int_{R(y)} dF(\varepsilon;\theta) \leq \Pr(y|x;\theta) \leq \int_{R(y)} dF(\varepsilon;\theta) + \int_{H(y)\backslash R(y)} dF(\varepsilon;\theta) = \int_{H(y)} dF(\varepsilon;\theta). \quad (2.3)$$

The identified set of bound estimation is the set of parameters such that the inequalities in (2.3) hold for all outcomes :

$$\Theta_I^{BE} = \{\theta \in \Theta : \forall y, x, \ \int_{R(y)} dF(\varepsilon;\theta) \leq \Pr(y|x;\theta) \leq \int_{H(y)} dF(\varepsilon;\theta).\} \quad (2.4)$$

Let $\Pr(Y|x;\theta)$ denote the vector of $\Pr(y|x;\theta)$ ($y \in Y$) and let $H_1(\theta,x)$ and $H_2(\theta,x)$ be the vector of lower and upper bounds for all outcomes. The inequalities could be written in the following vector notation,

$$H_1(\theta,x) \leq \Pr(Y|x;\theta) \leq H_2(\theta,x).$$

Inference is based on the criterion function

$$Q(\theta) = \int [\| (P(Y|x) - H_1(\theta,x))_- || + \| (P(Y|x) - H_2(\theta,x))_+ ||] dFx,$$

where $(\cdot)_-$ keeps the negative part of a vector, $(\cdot)_+$ keeps the positive part, and $||\cdot||$ is a distance measure, $P(Y|x)$ is a vector of conditional choice probabilities. The confidence region of the identified set is calculated via the subsampling method, as in Chernozhukov, Hong and Tamer (2007).

Andrews, Berry and Jia (2004) construct bounds in a similar way. In addition to constructing bounds on an individual outcome y, they also mention that there could be additional bounds on collections of outcomes. We will first explain why the identified set based on individual outcomes is not sharp. In the next section we will introduce a modified version of bound estimation that extends Ciliberto and Tamer (2009) and Andrews, Berry and Jia (2004).

Non-sharpness of bound estimation based on individual outcomes

Identified set of bound estimation is not sharp, meaning that parameters in the identified set may not satisfy all the restrictions predicted by the model. To see this, consider outcome $y_1 = (0,2)$ and $y_2 = (2,2)$ in Example 3. By the discussion above, the left bounds for these two outcomes are

$$\int_{R(2,0)} dF(\varepsilon;\theta) \leq P((0,2)|x); \tag{2.5}$$

$$\int_{R(0,2)} dF(\varepsilon;\theta) \leq P((2,0)|x). \tag{2.6}$$

For simplicity, in the rest of the paper, we write $dF\varepsilon$ instead of $dF(\varepsilon,\theta)$, and write $\lambda(y)$

instead of $\lambda(y|\varepsilon,x;\theta)$.

Consider set $A = \{(0,2),(2,0)\}$ that consists of two outcomes:

$$\begin{aligned}
\Pr(A|x;\theta) &= \Pr((0,2)|x;\theta) + \Pr((2,0)|x;\theta) \\
&= \int_{R(2,0)} dF\varepsilon + \int_{R(A)} \lambda(0,2)\cdot dF\varepsilon + \int_{\mathbb{R}^2\setminus R(2,0)\setminus R(A)} \lambda(0,2)\cdot dF\varepsilon \\
&\quad + \int_{R(0,2)} dF\varepsilon + \int_{R(A)} \lambda(2,0)\cdot dF\varepsilon + \int_{\mathbb{R}^2\setminus R(0,2)\setminus R(A)} \lambda(0,2)\cdot dF\varepsilon \\
&\geq \int_{R(2,0)} dF\varepsilon + \int_{R(0,2)} dF(\varepsilon;\theta) + \int_{R(A)} (\lambda(0,2)+\lambda(2,0))\cdot dF\varepsilon \\
&= \int_{R(2,0)} dF\varepsilon + \int_{R(0,2)} dF\varepsilon + \int_{R(A)} dF\varepsilon \\
&> \int_{R(2,0)} dF\varepsilon + \int_{R(0,2)} dF\varepsilon. \qquad (2.7)
\end{aligned}$$

Compared with inequalities (2.5) and (2.6), inequality (2.7) is more restrictive. This suggests that parameters satisfying inequalities of individual outcomes (2.5) and (2.6) do not necessarily satisfy the inequalities of collections of outcomes. There may exist parameters in the identified set Θ_I^{BE} that cannot be the true data generating process.

2.3.3 Modified Bound Estimation

The above example also sheds light on how a simple modification can give more constraints on the identified set. The identified set will shrink if moment conditions are imposed on combinations of outcomes.

Let $A \subset Y$ denote a set of outcomes. For each outcome y in A,

$$\Pr(y|x;\theta) = \int_{S(A)} \lambda(y)\cdot dF\varepsilon + \int_{H(A)\setminus S(A)} \lambda(y)\cdot dF\varepsilon + \int_{\mathbb{R}^N\setminus H(A)} \lambda(y)\cdot dF\varepsilon.$$

The probability of event A is

$$\begin{aligned}
\Pr(A|x;\theta) &= \sum_{y\in A}\Pr(y|x;\theta) \\
=\sum_{y\in A}(&\int_{S(A)}\lambda(y)\cdot dF\varepsilon+\int_{H(A)\backslash S(A)}\lambda(y)\cdot dF\varepsilon+\int_{\mathbb{R}^N\backslash H(A)}\lambda\cdot dF\varepsilon) \\
&= \int_{S(A)}dF\varepsilon+\sum_{y\in A}\int_{H(A)\backslash S(A)}\lambda(y)\cdot dF\varepsilon.
\end{aligned}$$

The last line follows from the fact that $\sum_{e\in\mathscr{E}(\varepsilon,x;\theta)}\lambda(e|\varepsilon,x;\theta)=1$ for $\forall\varepsilon\in S(A)$ and $\lambda(y|\varepsilon,x;\theta)=0$ for $\varepsilon\in\mathbb{R}^N\backslash H(A)$.

Let M_A denote an indicator vector of the same size of Y_i. The ith element of M_A takes value 1 if y is in A. For example, in a 2-person entry game such that $Y=\{(0,0),(0,1),(1,0),(1,1)\}$, $M_{\{(0,0),(1,1)\}}=[1,0,0,1]$. The probability of observing event A is the sum of probabilities of observing each of A's elements,

$$\Pr(A|x;\theta)=M_A\cdot\Pr(Y|x;\theta).$$

The modified bound estimation imposes moment inequalities on event A:

$$\int_{S(A)}dF\varepsilon\le M_A\cdot\Pr(Y|x;\theta)\le\int_{S(A)}dF\varepsilon+\int_{H(A)\backslash S(A)}dF\varepsilon=\int_{H(A)}dF\varepsilon. \tag{2.8}$$

The identified set of modified bound estimation is

$$\Theta_I^{MBE}=\{\theta\in\Theta:\forall A,x,\ \int_{S(A|x;\theta)}dF\varepsilon\le M_A\cdot P(Y|x)\le\int_{H(A|x;\theta)}dF\varepsilon\}. \tag{2.9}$$

Inference can follow Ciliberto and Tamer (2009) or Andrews, Berry and Jia (2004).

2.3.4 Momoment Inequalities with Sharp Identification

Galichon and Henry (2011) and Beresteanu, Molchanov and Molinari (2011) show that sharp identification of a complete information game can be achieved by defining a collection of finite inequalities.

Let A be an event, Galichon and Henry (2011) interpret the generalized likelihood function

$$\mathscr{L}(A|x;\theta) = \int_{H(A|x;\theta)} dF(\varepsilon;\theta)$$

as a Choquet capacity function, which is a notion in set theory. The identified set equals the set of parameters such that the probability of events do not exceed their generalized likelihood functions,

$$\Theta_I^{GH} = \{\theta \in \Theta : \forall A, x,\ P(A|x) \leq \mathscr{L}(A|x;\theta)\}.$$

In the case of pure strategy equilibria, Galichon and Henry (2011) show that the generalized likelihood is a submodular function. The original problem of checking whether a distribution is less than the generalized likelihood function is equivalent to minimizing a submodular function. Galichon and Henry (2011) recommend to use the algorithms in the optimal transformation literature to solve the minimization problem.

Beresteanu, Molchanov and Molinari (2011) characterize the identified set by the random set theory. In the case of pure strategy equilibrium, they prove that the random set theory and the approach in Galichon and Henry (2011) impose the same set of moment inequalities and the two methods coincide.

2.4 Equivalence among the Partial Identification Approaches

2.4.1 Equivalence between Θ_I^{MBE} and Θ_I^{GH}

With the introduction of notations $R(A)$, $S(A)$ and $R(A)$, the identification set of modified bound estimation is very similar to that of Galichon and Henry (2011) except the former has both upper and lower restrictions.

Consider the complement of A, denoted as A^c. Because

$$\{\varepsilon : \mathscr{E}(\varepsilon, x; \theta) \cap A^c \neq \emptyset\} = \{\varepsilon : \mathscr{E}(\varepsilon, x; \theta) \nsubseteq A\},$$

A^c and A are related by

$$\int_{H(A^c)} dF\varepsilon = 1 - \int_{S(A)} dF\varepsilon.$$

The lower bound of event A and the upper bound of its complement have the following relationship:

$$\begin{aligned} & \int_{S(A)} dF\varepsilon \leq \Pr(A|x;\theta) \\ \iff \quad & 1 - \int_{H(A^c)} dF\varepsilon \leq 1 - \Pr(A^c|x;\theta) \\ \iff \quad & \Pr(A^c|x;\theta) \leq \int_{H(A^c)} dF\varepsilon. \end{aligned} \tag{2.10}$$

Note that $\int_{S(A)} dF\varepsilon$ is the lower bound of A and $\int_{H(A^c)} dF\varepsilon$ is the upper bound of A^c. This finding is summarized in Proposition 2.1.

Proposition 2.1 *In modified bound estimation, the lower bound of set A is equivalent to the upper bound of its complement.*

Proposition 2.1 suggests that half of the moment inequalities in modified bound estimation could be dropped. Note that the identified set Θ_I^{GH} used in Galichon and Henry (2011) includes upper bounds only. By Proposition 2.1, the entire lower bounds can be removed as long as all upper bounds are kept. This suggests the equivalence between modified bound estimation and optimal transport approach.

Theorem 2.1 *The modified bound estimation and the sharp identification approach have the same identified set.* ($\Theta_I^{MBE} = \Theta_I^{GH}$).

Define another bound estimation that imposes lower bound restrictions only:

$$\Theta_I^{MBE-L} = \{\theta \in \Theta : \forall A, x, \int_{S(A)} dF\varepsilon \leq M_A \cdot Pr(Y|x;\theta)\} \tag{2.11}$$

By Proposition 2.1, $\Theta_I^{MBE-L} = \Theta_I^{MBE}$.

2.4.2 Equivalence between Θ_I^{SA} and Θ_I^{MBE}

In the previous discussion we show that the upper bound of an event is the same as the lower bound of its complement. In what follows, we drop all upper bounds in the modified bound estimation approach because Θ_I^{MBE-L} and Θ_I^{MBE} are the same. The objective of this subsection is to show the equivalence between the identified set of sensitive analysis (Θ_I^{SA}) and the identified set of modified bound estimation imposing lower bounds only (Θ_I^{MBE-L}). The equivalence of the two sets is achieved by proving that one set is a subset of the other and the vice versa. We start with showing all elements in Θ_I^{SA} belong to Θ_I^{MBE-L}.

Proposition 2.2 *If* $\theta \in \Theta_I^{SA}$, $P(A|x) \geq \int_{S(A)} dF\varepsilon$ *for* $\forall(y,x)$ *and* $\forall A$. *Therefore,* $\Theta_I^{SA} \subseteq \Theta_I^{MBE-L}$.

Intuitively, if parameter θ is in the identified set of sensitivity analysis, there exists an equilibrium selection mechanism that rationalizes parameter θ. If all the equilibria of a game belong to event A ($\varepsilon \in S(A)$), by the second criterion in the definition of equilibrium selection mechanism, event A will be observed with probability 1. The chance of having a game whose equilibria all belong to A is therefore no greater than the probability of observing A. This means $\int_{S(A)} dF\varepsilon \leq \Pr(A|x)$. A formal proof is in Appendix.

The proof of the reverse relationship is less obvious. We proceed as follows: For a fixed θ in identified set Θ_I^{MBE-L}, the set of probability distributions satisfying moment inequalities is a convex polytope. The observed choice probabilities $P(Y|x)$, which is inside the polytope, can be represented by a convex combinations of the vertexes of the polytope. We first prove by construction that each of the vertexes can be rationalized by an equilibrium selection mechanism that takes value 0 and 1 only. We then show $P(Y|x)$, a convex combination of vertexes, can be rationalized by a valid equilibrium selection mechanism.

To start, for a fixed $\theta \in \Theta_I^{MBE-L}$, consider the set of T such that

$$T := \{t \in \mathbb{R}^N : \forall A,\, M_A \cdot t \geq \int_{S(A|x;\theta)} dF\varepsilon\}. \tag{2.12}$$

Set T is not empty, as it has at least one element $t = P(Y|x)$. Given θ, the above inequalities impose linear restrictions on t, therefore T is a polytope. Let the vertexes of T be V.

For each $v \in V$, some of the moment inequalities are binding. Denote the set of A that are binding as $\mathscr{A}^B$. Because v is uniquely determined by the binding constraints, there is only one solution to

$$M_A \cdot t = \int_{S(A|x;\theta)} dF\varepsilon \tag{2.13}$$

for $\forall A \in \mathscr{A}^B$.

Not every combination of lower bounds can be achieved simultaneously. Next proposition discusses a necessary condition under which inequality constraints can be binding at the same time. This property enables us to construct equilibrium selection mechanism for v.

Lemma 2.1 *If inequalities of events $A_{r_1}, A_{r_2}, ...A_{r_k}$ are binding, the set $B := \{\varepsilon : A_i \subset \mathscr{E}(\varepsilon, x; \theta)$ and $\mathscr{E}(\varepsilon, x; \theta) = \cup_{i=r_1}^{r_k} A_{r_i},\ \forall i = r_1, r_2, ..r_k\}$ has zero measure.*

An implication of Lemma 2.1 is that if outcomes y_1 and y_2 are multiple equilibria for the, the lower bounds for y_1 and y_2 cannot be achieved simultaneously. This can be shown in Eq (2.7). If the lower bounds of $y_1 = (0,2)$ and $y_2 = (2,0)$ are binding and there are games with positive measure such that y_1 and y_2 are both equilibria, the inequality of event $A = \{y_1, y_2\}$ is not satisfied. A necessary condition for lower bounds for y_1 and y_2 to achieve simultaneously is when no game has an equilibrium set identical to $A = \{y_1, y_2\}$. Lemma 2.1 generalizes this discussion to the case of multiple events.

Now we are ready to construct equilibrium selection mechanism for v .

Algorithm 1: Procedure of Constructing Equilibrium for Vertex v

---

Step 1: For each ε, calculate its equilibrium set $\mathscr{E}(\varepsilon, x; \theta)$;

Step 2: For all $A \in \mathscr{A}^B$ such that $\mathscr{E}(\varepsilon, x; \theta) \nsubseteq A$, set $\lambda_v(y|\varepsilon, x; \theta) = 0$ for $\forall y \in (\mathscr{E}(\varepsilon, x; \theta) \cap A))$;

Step 3: After repeating the process for all $A \subset \mathscr{E}(\varepsilon, x; \theta)$, set $\lambda_v(y|\varepsilon, x; \theta) = 1$ for one of the remaining y in $\mathscr{E}(\varepsilon, x; \theta)$. If there is no outcome left after Step 2, assign $\lambda_v(y|\varepsilon, x; \theta) = 1$ for one $y \in \mathscr{E}(\varepsilon, x; \theta)$;

Step 4: Assign $\lambda_v(y|\varepsilon, x; \theta) = 0$ for the rest of $y \in Y$.

By construction, function $\lambda_v(y|\varepsilon, x; \theta)$ is non-negative. In addition, $\lambda_v(y|\varepsilon, x; \theta) = 1$ for only one $y \in \mathscr{E}(\varepsilon, x; \theta)$, thus the sum of probabilities of outcomes in $\mathscr{E}(\varepsilon, x; \theta)$ equals 1. By definition, $\lambda_v(\varepsilon, x; \theta)$ is a valid equilibrium selection mechanism.

Furthermore, Lemma 2.1 guarantees that with probability 1 that there will be one remaining outcome after Step 2. Therefore, except for a case with zero probability, outcomes in A will not be selected as long as the equilibrium set contains elements outside A. The next lemma shows that the probability of observing $A \in \mathscr{A}^B$ achieves its lower bound if equilibrium is selected according to λ_v.

Lemma 2.2 *The equilibrium selection mechanism $\lambda_v(y|\varepsilon, x; \theta)$ constructed by Algorithm* 1 *satisfies*

$$\sum_{y \in A} \int \lambda_v(y|\varepsilon, x; \theta) dF\varepsilon = \int_{S(A)} dF\varepsilon, \forall A \in \mathscr{A}^B.$$

Let $\lambda_v(Y|\varepsilon,x;\theta)$ denote the vector of $\lambda_v(y|\varepsilon,x;\theta)$ for $y \in Y$. By the definition of M_A, we have $M_A \cdot \lambda_v(Y|\varepsilon,x;\theta) = \sum_{y\in A} \lambda(y|\varepsilon,x;\theta)$. Therefore

$$M_A \cdot \int \lambda_v(Y|\varepsilon,x;\theta) \cdot dF\varepsilon = \sum_{y\in A} \int \lambda_v(y|\varepsilon,x;\theta) dF\varepsilon = \int_{S(A)} dF\varepsilon, \forall A \in \mathscr{A}^B.$$

Because there is a unique solution to (2.13) and both v and $\int \lambda_v(Y|\varepsilon,x;\theta) \cdot dF\varepsilon$ are the solution of (2.13) ,

$$v = \int \lambda_v(Y|\varepsilon,x;\theta) \cdot dF\varepsilon. \tag{2.14}$$

This suggests that each v can be rationalized by a equilibrium selection mechanism $\lambda_v(y|\varepsilon,x;\theta)$.

Since $P(Y|x)$ belongs to set T, it can be represented as a convex combination of vertexes of T:

$$P(Y|x) = \sum_{i=1}^{N_V} \mu_i v_i,$$

where N_V is the number of vertexes, $\sum_{i=1}^{N_V} \mu_i = 1$ and $\mu_i \geq 0$ $(i = 1,2,...N_V)$. Substituting v_i with Eq. (2.14), we get

$$\begin{aligned} P(Y|x) &= \sum_{i=1}^{N_V} \mu_i v_i \\ &= \sum_{i=1}^{N_V} \mu_i \int \lambda_i(y|\varepsilon,x;\theta) dF\varepsilon \\ &= \int [\sum_{i=1}^{N_V} \mu_i \lambda_i(y|\varepsilon,x;\theta)] dF\varepsilon. \end{aligned} \tag{2.15}$$

Define $\lambda_Y(y|\varepsilon,x;\theta) = \sum_{i=1}^{N_V} \mu_i \lambda_i(y|\varepsilon,x;\theta)$ for $\forall y$. Eq. (2.15) can be written as $P(Y|x) = \int \lambda_Y(y|\varepsilon,x;\theta) dF\varepsilon$. Lemma 2.3 further shows that $\lambda_Y(y|\varepsilon,x;\theta)$ satisfies the definition of equilibrium selection mechanism, therefore is a valid equilibrium selection mechanism.

Lemma 2.3 *If $\lambda_i(y|\varepsilon,x;\theta)$ $(\forall i = 1,2,...N_v)$ satisfy the properties in the definition of equi-*

librium selection mechanism, $\sum_{i=1}^{N_V} \mu_i = 1$ *and* $\mu_i \geq 0$*, then*

$$\lambda_Y(y|\varepsilon,x;\theta) = \sum_{i=1}^{N_V} \mu_i \lambda_i(y|\varepsilon,x;\theta)$$

satisfies these properties as well.

Proposition 2.3 summarizes above findings and shows that all parameters in Θ_I^{MBE-L} belong to Θ_I^{SA}.

Proposition 2.3 *For each* $\theta \in \Theta_I^{MBE-L}$*, there exist a valid equilibrium selection* $\lambda(y|\varepsilon,x;\theta)$ *such that* $P(y|x) = \int \lambda(y|\varepsilon,x;\theta)dF\varepsilon$*. In other words,* $\Theta_I^{MBE-L} \subset \Theta_I^{SA}$*.*

Based on Proposition 2.2 and Proposition 2.3, the two sets Θ_I^{MBE-L} and Θ_I^{SA} are equal. Because Θ_I^{MBE-L} and Θ_I^{MBE-L} are identical, we further conclude that Θ_I^{MBE} and Θ_I^{SA} are identical.

Theorem 2.2 *The modified bound estimation and the sensitivity analysis approach have the same identified set.* $(\Theta_I^{MBE} = \Theta_I^{SA})$.

2.5 Conclusion

This chapter provides a unifying framework for partial identification approaches in discrete games of complete information. These approaches were developed independently; little was known about the relationship among them.

We review the three main approaches, namely, the sensitivity approach as in Grieco (2014), the bound estimation approach in Ciliberto and Tamer (2009) and Andrews, Berry and Jia (2004), and the sharp identification appraoch in Galichon and Henry (2011) and Beresteanu, Molchanov and Molinari (2011). One novelty of this chapter is to introduce concepts such as $R(A|x;\theta)$, $S(A|x;\theta)$ and $H(A|x;\theta)$. Even though the original approaches in Ciliberto and Tamer (2009) and Galichon and Henry (2011) look very different, with the newly introduced concepts, we characterize the identified sets of these two approaches using the same framework.

A key idea in our paper is to propose a modified version of Ciliberto and Tamer (2009), which we call modified bound estimation. The benefits of introducing the modified approach are three folds. First, by comparing the moment inequalities in the new and original approaches, we show why the identified set in the original approach is not sharp. The model imposes further restrictions through collections of outcomes that are multiple equilibria, but this information is not used if moment conditions are constructed on individual outcomes only.

The modified approach also demonstrates the computational cost of achieving sharp identification. Suppose a game has k outcomes, the bound estimation approach checks $2k$ upper and lower bounds combined, and the modified approach, even dropping all upper bounds, need to check k^2 outcomes. As a result, sharp identification may not always be more desirable, because the number of moments needed in the modified bound estimation is much greater than the number of moments checked in Ciliberto and Tamer (2009).

The third benefit of introducing a modified version of bound estimation is to establish the relationship among the three existing partial identification approaches in the literature. We show that the modified bound estimation, the sensitive approach, and the sharp identification approach, draw inference on the same identification set.

A takeaway of this chapter is that, because the identified sets are equal, it is not the case that one partial identification approach is superior to the other. In practice, the choice of approach should be based on computational concerns rather than the size of the identified set. The modified bound estimation method and the approaches in Galichon and Henry (2011) and Beresteanu, Molchanov and Molinari (2011) are suitable for the case where the number of outcomes is small, because these approaches check moments of all combinations of outcomes. The sensitivity approach is feasible in the case where the multiple equilibria are relatively easy to characterize, as it searches over the space of selection mechanism to find parameters that can predict choice probabilities of the model. If the number of outcomes is large and the characterization of multiple equilibria is more involved, it may be more practical to check fewer moment inequalities as in Ciliberto and Tamer (2009) and Andrews, Berry and Jia (2004) because the cost of conducting the other approaches is high.

Proof of Proposition 2.2 If $\theta \in \Theta_I^{SA}$, $\exists \lambda \in \Lambda$ such that

$$P(y|x) = \int \lambda(y|\varepsilon, x; \theta) \cdot dF\varepsilon$$

holds for $\forall (y,x)$.

Consider an event A. Let the space of unobservables be divided into $S(A)$ and its complement:

$$\begin{aligned}
P(A|X) &= \sum_{y\in A} P(y|x;\theta) \\
&= \sum_{y\in A} \left(\int_{S(A)} \lambda(y|\varepsilon,x;\theta)\cdot dF\varepsilon + \int_{\mathbb{R}^N\setminus S(A)} \lambda(y|\varepsilon,x;\theta)\cdot dF\varepsilon\right) \\
&= \int_{S(A)} \left(\sum_{y\in A} \lambda(y|\varepsilon,x;\theta)\right)\cdot dF\varepsilon + \int_{\mathbb{R}^N\setminus S(A)} \left(\sum_{y\in A} \lambda(y|\varepsilon,x;\theta)\right)\cdot dF\varepsilon. \qquad (2.16)
\end{aligned}$$

If $\varepsilon \in S(A)$, $\mathscr{E}(\varepsilon,x;\theta) \subseteq A$. Because $\sum_{y\in\mathscr{E}(\varepsilon,x;\theta)} \lambda(y|\varepsilon,x;\theta)) = 1$ and $\lambda(y|x;\theta) = 0$ for $y \notin \mathscr{E}(\varepsilon,x;\theta)$,

$$\begin{aligned}
\sum_{y\in s(A)} \lambda(y|\varepsilon,x;\theta) &= \sum_{y\in\mathscr{E}(\varepsilon,x;\theta)} \lambda(y|\varepsilon,x;\theta) + \sum_{y\in A\setminus\mathscr{E}(\varepsilon,x;\theta)} \lambda(y|\varepsilon,x;\theta) \\
&= 1+0 \\
&= 1.
\end{aligned}$$

Equation (2.16) becomes

$$\begin{aligned}
P(A|x) &= \int_{S(A)} 1\cdot dF\varepsilon + \int_{\mathbb{R}^N\setminus S(A)} \left(\sum_{y\in A} \lambda(y|\varepsilon,x;\theta)\right)\cdot dF\varepsilon \\
&\geq \int_{S(A)} dF\varepsilon.
\end{aligned}$$

Above discussion holds for any event $A \subset Y$, therefore $\theta \in \Theta_I^{MCT-L}$.

Proof of Lemma 2.1 We first prove a special case of Lemma 2.1 with two sets. Suppose the lower bounds of A_i and A_j are binding.

Define

$$g(\varepsilon) = 1(\varepsilon \in S(A_i \cup A_j)) - 1(\varepsilon \in S(A_i)) - 1(\varepsilon \in S(A_j)) + 1(\varepsilon \in S(A_i \cap A_j)).$$

Consider the following exclusive and exhaustive partition of ε:

a.) If $\mathscr{E}(\varepsilon) \nsubseteq (A_i \cup A_j)$:

$$g(\varepsilon) = 0 + 0 + 0 + 0 = 0;$$

b.) If $\mathscr{E}(\varepsilon) \subseteq (A_i \cup A_j)$ and $\mathscr{E}(\varepsilon) \subseteq (A_i \cap A_j)$,

$$g(\varepsilon) = 1 - 1 - 1 + 1 = 0;$$

c.) If $\mathscr{E}(\varepsilon) \subseteq (A_i \cup A_j)$ but $\mathscr{E}(\varepsilon) \nsubseteq (A_i \cap A_j)$, $1(\varepsilon \in S(A_i)) + 1(\varepsilon \in S(A_j)) \leq 1$ because otherwise $\mathscr{E}(\varepsilon) \subseteq (A_i \cap A_j)$. Therefore

$$g(\varepsilon) \geq 1 - 1 + 0 = 0.$$

(a), (b) and (c) suggest that

$$\begin{aligned}
& \int_{S(A_i \cup A_j)} dF\varepsilon - \int_{S(A_i)} dF\varepsilon - \int_{S(A_j)} dF\varepsilon + \int_{S(A_i \cap A_j)} dF\varepsilon \\
= & \int [1(\varepsilon \in S(A_i \cup A_j)) - 1(\varepsilon \in S(A_i)) - 1(\varepsilon \in S(A_j)) + 1(\varepsilon \in S(A_i \cap A_j))] \cdot dF\varepsilon \\
= & \int g(\varepsilon) \cdot dF\varepsilon \geq 0. \qquad (2.17)
\end{aligned}$$

Equivalently,

$$\int_{S(A_i \cup A_j)} dF\varepsilon \geq \int_{S(A_i)} dF\varepsilon + \int_{S(A_j)} dF\varepsilon - \int_{S(A_i \cap A_j)} dF\varepsilon. \qquad (2.18)$$

On the other hand, because the lower bound of A_i and A_j are binding,

$$\begin{aligned}\int_{S(A_i)} dF\varepsilon &= M_{A_i}\cdot v,\\ \int_{S(A_j)} dF\varepsilon &= M_{A_j}\cdot v.\end{aligned}$$

The lower bound of $A_i \cap A_j$ may or may not be binding, therefore

$$\int_{S(A_i\cap A_j)} dF\varepsilon \leq M_{A_i\cap A_j}\cdot v.$$

Combining these three equalities and inequalities together, we get

$$\begin{aligned}\int_{S(A_i)} dF\varepsilon + \int_{S(A_j)} dF\varepsilon - \int_{S(A_i\cap A_j)} dF\varepsilon &\geq M_{A_i}\cdot v + M_{A_j}\cdot v - M_{A_i\cap A_j}\cdot v\\ &= M_{A_i\cup A_j}\cdot v. \qquad (2.19)\end{aligned}$$

(2.18) and (2.19) suggest that

$$\int_{S(A_i\cup A_j)} dF\varepsilon \geq \int_{S(A_i)} dF\varepsilon + \int_{S(A_j)} dF\varepsilon - \int_{S(A_i\cap A_j)} dF\varepsilon \geq M_{A_i\cup A_j}\cdot v. \qquad (2.20)$$

However, because the lower bound of $A_i \cup A_j$ holds,

$$\int_{S(A_i\cup A_j)} dF\varepsilon \leq M_{A_i\cup A_j}\cdot v,$$

it must be the case that all the inequalities in (2.20) are equalities,

$$\int_{S(A_i\cup A_j)} dF\varepsilon = \int_{S(A_i)} dF\varepsilon + \int_{S(A_j)} dF\varepsilon - \int_{S(A_i\cap A_j)} dF\varepsilon = M_{A_i\cup A_j}\cdot v. \qquad (2.21)$$

This proves part (1) of Lemma 2.1 for $k = 2$ that $A^* = A_i \cup A_j$ achieves its lower bound.

Furthermore, in order to achieve (2.21), by expression (2.17) it must be the case that $\int g(\varepsilon)\cdot dF\varepsilon = 0$.

Let

$$B_2 := \{\varepsilon : \mathscr{E}(\varepsilon, x; \theta) \subseteq (A_i \cup A_j),\ \mathscr{E}(\varepsilon, x; \theta) \nsubseteq A_i,\ and\ \mathscr{E}(\varepsilon, x; \theta) \nsubseteq A_j\}.$$

If $\varepsilon \in B_2$,

$$\begin{aligned} g(\varepsilon) &= 1(\varepsilon \in S(A_i \cup A_j)) - 1(\varepsilon \in S(A_i)) - 1(\varepsilon \in S(A_j)) + 1(\varepsilon \in S(A_i \cap A_j)) \\ &= 1 - 0 - 0 + 0 \\ &= 1. \end{aligned}$$

In order to get $\int g(\varepsilon) \cdot dF\varepsilon = 0$, it must be the case such that $\int_{\varepsilon \in B_2} dF = 0$. This proves part (2) of Lemma 2.3 for $k = 2$.

Having proved the lemma at $k = 2$, we now prove it for the rest of k by induction.

Suppose Lemma 2.1 holds for case $k = t$ $(t \geq 2)$.

Let $A_p^* = \cup_{i=1}^{t} A_i$. The set

$$B_t := \{\varepsilon : \mathscr{E}(\varepsilon, x; \theta) \subseteq A_p^*,\ and\ \mathscr{E}(\varepsilon, x; \theta) \nsubseteq A_i\ \forall i = 1, 2, ...t\}$$

satisfies

$$\int_{B_t} dF\varepsilon = 0.$$

Because the lower bound of A_p^* is binding by the lemma at $k = t$, by part (1) at $k = 2$, the lower bound of $(A_p^* \cup A_{t+1})$ is achieved. This proves part (1) of the lemma at $k = t + 1$.

Part (2) of the lemma at $k = 2$ also implies that for two events A_p^* and A_{t+1} whose lower bounds are binding,

$$B_{p,(t+1)} := \{\varepsilon : \mathscr{E}(\varepsilon, x; \theta) \subseteq (A_p^* \cup A_{t+1}), \mathscr{E}(\varepsilon, x; \theta) \nsubseteq A_p^*,\ and\ \mathscr{E}(\varepsilon, x; \theta) \nsubseteq A_{t+1}\},$$

satisfies

$$\int_{B_{p,(t+1)}} dF\varepsilon = 0.$$

Let B_{t+1} be the case of $k = t+1$,

$$B_{t+1} := \{\varepsilon : \mathscr{E}(\varepsilon,x;\theta) \subseteq (A_p^*) \cup A_{t+1},\ \mathscr{E}(\varepsilon,x;\theta) \nsubseteq A_i,\ \forall i = 1,2,...t,\ and\ \mathscr{E}(\varepsilon,x;\theta) \nsubseteq A_{t+1}\}.$$

Consider $B_{t+1} \backslash B_{p,(t+1)}$:

$$\begin{aligned} B_{t+1} \backslash B_{p,t+1} &= \{\varepsilon : \mathscr{E}(\varepsilon,x;\theta) \subseteq (A_p^*) \cup A_{t+1},\ \mathscr{E}(\varepsilon,x;\theta) \subseteq A_p^*,\ \mathscr{E}(\varepsilon,x;\theta) \nsubseteq A_i,\ \forall i = 1,2,...t, \\ &\quad and\ \mathscr{E}(\varepsilon,x;\theta) \nsubseteq A_{t+1}\}. \end{aligned}$$

Because $B_{t+1} \backslash B_{p,t+1} \subseteq B_t$ and $\int_{B_t} dF\varepsilon = 0$, we have

$$\int_{B_{t+1} \backslash B_{p,t+1}} dF\varepsilon = 0.$$

Further because $\int_{B_{p,(t+1)}} dF\varepsilon = 0$,

$$\int_{B_{t+1}} dF\varepsilon = \int_{B_{t+1} \backslash B_{p,t+1}} dF\varepsilon + \int_{B_{p,(t+1)}} dF\varepsilon = 0.$$

This proves part (2) of the lemma at $k = t+1$.

By induction, Lemma 2.1 holds for any $k \geq 2$.

Proof of Lemma 2.2 Consider

$$\begin{aligned} \int (\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta)) \cdot dF\varepsilon &= \int_{S(A)} (\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta)) \cdot dF\varepsilon + \int_{B_k} (\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta)) \cdot dF\varepsilon + \\ &\quad \int_{\mathbb{R}^N \backslash S(A) \backslash B_k} (\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta)) \cdot dF\varepsilon. \end{aligned} \tag{2.22}$$

Let's consider the following three cases one by one:

a.) $\varepsilon \in S(A)$ is equivalent to $\mathscr{E}(\varepsilon,x;\theta) \subseteq A$. By construction, $\lambda_v(y|\varepsilon,x;\theta) = 1$ for one element in $\mathscr{E}(\varepsilon,x;\theta)$ and $\lambda_v(y|\varepsilon,x;\theta) = 0$ for other elements in Y. Therefore $\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta) = 1$ if $\varepsilon \in S(A)$.

b.) If $\varepsilon \notin S(A)$, for the majority of the case, $\lambda_v(y|\varepsilon,x;\theta) = 0$ for $\forall y \in A$. The only exception is when all elements are assigned in Step 1. When there is no outcome left after the first step, $\lambda_v(y|\varepsilon,x;\theta) = 1$ may be assigned to one of $y \in A$. There will be no outcome left if there exist a set of binding constraints for $A_{1,}...A_k$ such that

$$\mathscr{E}(\varepsilon,x;\theta) \nsubseteq A_i,$$

for $\forall i = 1,2,...k$ and

$$\cup_{i=1}^{k}(\mathscr{E}(\varepsilon,x;\theta) \cap A_i) = \mathscr{E}(\varepsilon,x;\theta).$$

Or equivalently, it will occur if $\mathscr{E}(\varepsilon,x;\theta) \nsubseteq A_i$ and $\cup_{i=1}^{k} A_i \supseteq \mathscr{E}(\varepsilon,x;\theta)$.

By Lemma 2.1,

$$B_k := \{\varepsilon : A_i \subseteq \mathscr{E}(\varepsilon,x;\theta) \textit{ and } \mathscr{E}(\varepsilon,x;\theta) = \cup_{i=1}^{k} A_i,\ \forall i = 1,2,...k\}$$

satisfies

$$\int_{B_k} dF\varepsilon = 0.$$

By the algorithm of constructing λ_v, $\lambda_v(y|\varepsilon,x;\theta) = 1$ for at most one element in A, so $0 \leq \sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta) \leq 1$ for $\varepsilon \in B_k$. Because $\int_{B_k} dF\varepsilon = 0$,

$$0 \leq \int_{B_k} (\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta)) \cdot dF\varepsilon \leq \int_{B_k} 1 \cdot dF\varepsilon = 0,$$

which suggests

$$\int_{B_k} (\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta)) dF\varepsilon = 0.$$

c.) If $\varepsilon \notin S(A)$ and $\varepsilon \notin B_k$, there is at least one element left after Step 1, therefore $\lambda_v(y|\varepsilon,x;\theta) = 1$ for one $y \in \mathscr{E}(\varepsilon,x;\theta) \backslash A$ and $\lambda_v(y|\varepsilon,x;\theta) = 0$ for the rest of outcomes. This suggests that

$$\sum_{y\in A} \lambda_v(y|\varepsilon,x;\theta) = 0$$

for $\varepsilon \in \mathbb{R}^N \backslash S(A) \backslash B_k$

By the above discussion, (2.22) can be written as

$$\begin{aligned}
\int (\sum_{y \in A} \lambda_v(y|\varepsilon, x; \theta)) \cdot dF\varepsilon &= \int_{S(A)} (\sum_{y \in A} \lambda_v(y|\varepsilon, x; \theta)) \cdot dF\varepsilon + \int_{B_k} (\sum_{y \in A} \lambda_v(y|\varepsilon, x; \theta)) \cdot dF\varepsilon + \\
&\quad \int_{\mathbb{R}^N \backslash S(A) \backslash B_k} (\sum_{y \in A} \lambda_v(y|\varepsilon, x; \theta)) \cdot dF\varepsilon \\
&= \int_{S(A)} dF\varepsilon + 0 + \int_{\mathbb{R}^N \backslash S(A) \backslash B_k} 0 \cdot dF\varepsilon \\
&= \int_{\varepsilon \in S(A)} dF\varepsilon.
\end{aligned}$$

This proves Lemma 2.2.

Proof of Lemma 2.3 a.) If $\lambda_i(y|\varepsilon, x; \theta)$ ($\forall i = 1, 2, ...N_V$) is a valid equilibrium selection mechanism, then $\lambda_i(y|\varepsilon, x; \theta) \geq 0$ for $\forall y \in Y$ and $\forall i = 1, 2, ..N_V$. Further because $\mu_i \geq 0$,

$$\lambda_Y(y|\varepsilon, x; \theta) = \sum_{i=1}^{N_V} \mu_i \lambda_i(y|\varepsilon, x; \theta) \geq 0$$

for $\forall y \in Y$ and $\forall i$.

$\lambda_Y(y|\varepsilon, x; \theta)$ satisfies the first requirement of equilibrium selection mechanism.

b.) If $\lambda_i(y|\varepsilon, x; \theta) = 0$ for $\forall y \notin \mathscr{E}(\varepsilon, x; \theta)$, $\forall i$,

$$\lambda_Y(y|\varepsilon, x; \theta) = \sum_{i=1}^{N_V} \mu_i \lambda_i(y|\varepsilon, x; \theta) = 0$$

holds for $\forall y \notin \mathscr{E}(\varepsilon, x; \theta)$.

$\lambda_Y(y|\varepsilon, x; \theta)$ satisfies the second requirement of equilibrium selection mechanism.

c.) If $\sum_{y \in \mathscr{E}(\varepsilon, x; \theta)} \lambda_i(y|\varepsilon, x; \theta) = 1$ holds for $\forall i = 1, 2, ...N_V$,

$$
\begin{aligned}
\sum_{y \in \mathscr{E}(\varepsilon,x;\theta)} \lambda_Y(y|\varepsilon,x;\theta) &= \sum_{y \in \mathscr{E}(\varepsilon,x;\theta)} \sum_{i=1}^{N_V} \mu_i \lambda_i(y|\varepsilon,x;\theta) \\
&= \sum_{i=1}^{N_V} \mu_i \sum_{y \in \mathscr{E}(\varepsilon,x;\theta)} \lambda_i(y|\varepsilon,x;\theta) \\
&= \sum_{i=1}^{N_V} \mu_i \cdot 1 \\
&= 1.
\end{aligned}
$$

Therefore $\lambda_Y(y|\varepsilon,x;\theta)$ satisfies the last properties of equilibrium selection mechanism.

Combing (a), (b) and (c), we can conclude that $\lambda_Y(y|\varepsilon,x;\theta)$ is a valid equilibrium selection mechanism.

Chapter 3

A Partial Identification Subnetwork Approach to Discrete Games in Large Networks: An Application to Quantifying Peer Effects

3.1 Introduction

Peer effects play a central role in influencing individual behaviors. In recent years, there is an exploding interest in studying interactions in social networks. For example, there is empirical evidence of peer effects on educational achievements (Zimmerman 2003, Calvó-Armengol, Patacchini and Zenou 2009), employment (Calvo-Armengol and Jackson 2004), health outcomes (Cohen-Cole and Fletcher 2008, Krauth 2006, Nakajima 2007, Badev 2013), risky behavior taking (Gaviria and Raphael 2001, Clark and Loheac 2007), adoption of new technology (Conley and Udry 2010), among others. Social interaction can be modeled as a system of equations where each equation is a regression of one person's action on the actions of his or her peers. This framework is widely used in studying peer effects on a continuous outcome. Whereas the identification of peer effects model with continuous outcomes has been studied by Manski (1993) and Bramoullé, Djebbari and Fortin (2009), identification and estimation issues of the model with discrete outcomes are not well addressed.

This chapter develops an empirical method to study peer effects on discrete choices in large social networks. Our framework extends the linear network model in Bramoullé, Djebbari and Fortin (2009) to the case of binary outcomes. Our model belongs to a large and growing literature on discrete games of complete information, which includes entry game as a special case. It has been well known in the entry literature that due to the presence of multiple equilibria, estimating strategic interaction of discrete outcomes requires either strong assumptions or special econometric tools (Bjorn and Vuong 1984, Bresnahan and Reiss 1991a, Berry 1992, Tamer 2003, Ciliberto and Tamer 2009, Andrews, Berry and

Jia 2004). While both peer effects model and entry model study strategic interaction of discrete choices, existing methods for entry games is not suitable to estimate games in networks because the peer effects model is different from the entry model in a number of ways.

One empirical challenge that is new to games in networks is due to the large number of agents in a network. To the best of our knowledge, all applications in discrete games of complete information study strategic interaction among a handful of agents (Bjorn and Vuong 1984, Bajari, Hong and Ryan 2010, Jia 2008, Krauth 2006). There are three reasons why identification and inference of large games are difficult. First, point identification relies on the knowledge of all the equilibria of a game. In practice the set of equilibria is usually calculated by enumerating all outcomes of the game and checking whether each of them is an equilibrium. The number of outcomes grows at an exponential rate of the number of agents. Therefore, obtaining the set of equilibrium for games in networks is computationally demanding. Second, when a game is played by many agents, the sets of equilibria vary significantly across games, making it harder to find a reasonable assumption of equilibrium selection mechanism that is needed for point identification. Third, existing partial identification approaches could not handle large games as they check moment conditions of all outcomes of the game. Each moment condition needs to be consistently estimated by a large number of networks of the same outcome. Since the number of outcomes is enormous, the number of networks needed for constructing moment conditions are enormous. Our data could allow one single network that connects everyone. We don't assume a large number of networks of same size and outcome are observed in data.

The variation in the network structure adds further challenges in estimating peer effects. An individual's action depends on whom he or she connects with. This is in contrast to entry games, in which the "network" is fixed in the sense that every firm interacts with all others firms in a market. The peer effects model has an additional variation in friend relationship. The moment conditions, in the partial identification approach, need to be constructed for each network structure. For the same reason as above, there may not be enough observation of the same friend relation to calculate empirical probabilities, therefore partial identification approaches based on moment conditions of full networks are not feasible in practice.

The novelty of this chapter is to address computational and consistency issues by partially identifying peer effects via subnetworks. Though the number of outcomes of full network is enormous, the number of outcomes in subnetworks is tractable. There is one additional issue we need to address. Because people in a subnetwork interact with people outside the subnetwork, moment conditions of subnetworks need to consider these potential interactions. Moment conditions developed in this chapter are constructed by requiring individuals inside subnetwork to play best response to whatever players outside the subnetwork act. Since moment conditions do not rely on the information outside the subnetwork, the number of moments needed to check depends on the features of the subnetwork only.

In the Monte Carlo study, we demonstrate that moment conditions of subnetworks are not only computable, but also informative. The subnetwork approach successfully excludes parameter values that are far from the true parameters of the data generating process. By using the Monte Carlo examples, we also illustrate the factors that influence the performance of our subnetwork approach. Generally, the subnetwork approach performs better if the number of links connected to individual inside and outside networks is small. This is because our approach cuts the dependence between agents in and outside the subnetwork in exchange for computation tractability. Since many of the real world applications have sparse networks, our approach will be well suited for these applications.

The final part of this chapter studies peer effects on smoking using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is a nation-wide survey of health related questions. This data set also contains information on friend nomination, from which we could form friend networks. The friend network revealed in Add Health data is very sparse. Using our econometric method, we find significant and positive peer effects of smoking.

This chapter contributes to the peer effects literature by proposing a computationally feasible way of estimating peer effects on discrete outcomes. Our framework is closely related to Manski (1993) and Bramoullé, Djebbari and Fortin (2009) except that we consider discrete actions. Multiple equilibria is not an issue if individuals choose continuous actions. As described in Bramoullé, Djebbari and Fortin (2009), the action of friends' friend could be served as an instrumental variable for the peer effect in the linear model. Unfortunately, the instrumental variable approach could not be extended to the case of discrete choices. In the peer effects literature, most of the empirical studies on discrete choices follow Brock and Durlauf (2001) and Brock and Durlauf (2007), who model individual behavior as a best response to the expected behavior of peer group. We assume individuals play best response to the realized action of their peer group instead.

This chapter also adds to the growing literature of identification and inference of discrete games of complete information. For the best of knowledge, all discussions of discrete games focus on small number of agents. For tractable number of players, point identification could be achieved in symmetric entry games (Berry 1992), by assuming or estimating equilibrium selection mechanism (Bjorn and Vuong 1984, Bajari, Hong and Ryan 2010), or by a large support condition (Tamer 2003). Partial identification approaches are discussed in Andrews, Berry and Jia (2004), Ciliberto and Tamer (2009), Galichon and Henry (2011), Beresteanu, Molchanov and Molinari (2011) and Henry, Meango and Queyranne (2015). We contribute to this literature by considering games in large but sparse networks. Though sharp identification is achievable in Galichon and Henry (2011), Beresteanu, Molchanov and Molinari (2011) and Henry, Meango and Queyranne (2015), because of the variation in network structure and the large size of the network, sharp identification is extremely difficult in the case we consider. Therefore, this chapter seeks necessary but not sufficient conditions of the model as in Ciliberto and Tamer (2009).

Last but at least, our work contributes to the very new literature on the econometrics of networks. One topic in this area focuses on network formation; some studies model network formation as a complete information game (Sheng 2012, Uetake 2012) and others model network as an incomplete information game (Leung 2015). Badev (2013) studies both network formation and interactions in networks. This chapter contributes to the literature by studying interactions in networks. Our work is most related to Sheng (2012) in the sense that both papers explore information from subnetwork to conduct inference. The objective of our paper and Sheng (2012) are very different because her work considers network formation while we study interactions in networks.

The rest of chapter is organized as follows. Section 2 describes our econometric framework and discusses the empirical challenges caused by the network features. Section 3 starts with an example of constructing moment conditions for a 2-person network game when the full network is of size 4. Then we discuss identification and inference in general cases. Section 4 is devoted to studying the performance of the subnetwork approach. Section 5 conducts an empirical exercise of peer effects on smoking. Section 6 concludes.

3.2 Model

3.2.1 Model Setup

Network A network can be described as a graph of nodes and edges. Each node represents an agent, which can be a person or a firm. Each edge connects one pairs of nodes and represents a relationship, such as friendship.

Let $V = \{1, 2, ..N\}$ be the set of agents in the network. Links are non-directional.[1] Let $g_{ij} = g_{ji} = 1$ if i and j are connected, $g_{ij} = g_{ji} = 0$ otherwise. The collection of links forms a $n \times n$ matrix called G, which filled with zeros and ones. We study interaction between agents, taking the formation of network as given.[2]

[1]Our model can easily extend to a directional network. Because our focus is to study peer effects in a network, the assumptions of non-directional friend relationship is more appropriate.

[2]Estimating network formation is computationally intensive and often relies on partial identification approach. See Sheng (2012) and Uetake (2012). In this chapter, we assume network is exogeneouly determined.

Figure 3.1: A Graph of 4-Person Network

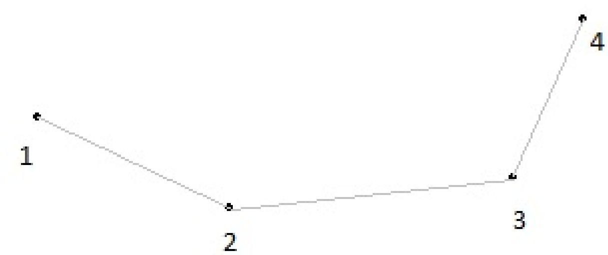


Figure 3.1 illustrates a friend network of 4 individuals. Each link denotes a friend relationship. In this example $g_{12} = g_{21} = 1$ because person 1 and 2 are connected. $g_{13} = g_{31} = 0$ because person 1 and 3 are not connected.

In the discussion later, we will develop our identification strategy using the information about subnetworks. A subnetwork consists of a subset of agents and the links associated with these agents. The subnetwork A contains three types of information: 1) a set of players A; 2) the links between agents in A: $G_A = \{g_{ij}\}$, $(i,j) \in A$; and 3) the number of links connecting to agents outside the subnetwork $n_A = \{n_{A,i}\}$ for $\forall i \in A$, where $n_{A,i} = \sum g_{ij} \cdot 1(j \notin A)$.

For example, we may be interested in the subnetwork that consists of agents 2 and 3. Let $A = \{2,3\}$ denote the set of agents inside the subnetwork. $G_A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ reveals that agents 2 and 3 are connected. $n_A = [1,\ 1]$ because both agents 2 and 3 have one link connecting to agents outside the subnetwork. In inference, we will use (A, G_A, n_A) to construct moment inequalities.

Utility Function Agents play a simultaneous game of complete information. Each individual i chooses a binary action $y_i \in \{0,1\}$. Normalize the utility of action 0 to 0. The utility of the alternative action is affected by person i's characteristics $\mathbf{x}_i$, individual shock ε_i, and the average actions taken by individuals that are connected with i:

$$u(y_i, \mathbf{x}_i, y_{-i}; \beta, \gamma) = \beta \mathbf{x}_i + \gamma \frac{\sum_{j \in V} g_{ij} y_j}{\sum_{j \in V} g_{ij}} + \varepsilon_i. \tag{3.1}$$

Assume $\gamma > 0$. Under the framework of games of complete information, $\varepsilon = (\varepsilon_1, \varepsilon_2, ... \varepsilon_N)$ is observed by all game players. As in Probit models, each component of ε follows a standard normal distribution independently.

Agent i chooses the action that has a higher utility, hence

$$y_i = 1(u(y_i, \mathbf{x}_i, y_{-i}; \beta, \gamma) > 0).$$

In this chapter, we adopt the standard form of utility function that assumes a linear functional form and homogeneous effects. This is the model used in early studies, e.g. Bresnahan and Reiss (1991a) and Berry (1992). Our identification result can extend to non-linear utility functions and cases where the interaction effects are heterogeneous. For example, we can let γ vary across i or j as in Ciliberto and Tamer (2009). In addition, our framework could be extended to the case of negative strategic interactions, only by a minor change in moment conditions. It is also feasible to allow for a correlation between shocks, if we consider the correlation as one additional parameter to be estimated. For the illustrative purpose, in what follows, we keep the simple form of utility function with a positive and homogeneous interaction effect γ. We will also discuss how to extend our identification strategy to more general cases.

It should also be noted that we focus on games of complete information. Actions are made in response to actual actions of others rather than to the belief of actions derived from the distribution of other players' types. Complete information is a reasonable assumption in applications such as the peer effects model, because an individual plays best response to the realized action of his or her peers. The second reason why we focus on complete information is because the equilibrium solutions to complete and incomplete games differ from each other significantly. To the best of our knowledge, the bound estimation approach has not yet be fully developed to estimate games of incomplete information.

Equilibrium We focus on pure strategy Nash equilibrium. An outcome is a Nash equilibrium if all players play best response to each other. Let $\mathbf{x}$ be the matrix of observed characteristics of all agents and let ε be the vector of unobserved characteristics. The possible Nash equilibrium of a game is determined by the utility function, which is a function of individual characteristics $(\mathbf{x}, \varepsilon)$ and the set of parameters $\theta = \{\beta, \gamma\}$. Let $\mathscr{E}(\varepsilon, \mathbf{x}; \theta)$ denote the set of Nash equilibria, $\mathscr{E}(\varepsilon, \mathbf{x}; \theta)$ is defined as

$$\mathscr{E}(\varepsilon, \mathbf{x}; \theta) = \{y \in \{0,1\}^N : \ y_i = 1(\beta \mathbf{x}_i + \gamma \frac{\sum_{j \in V} g_{ij} y_j}{\sum_{j \in V} g_{ij}} + \varepsilon_i > 0), \ \forall i \in V),$$

$\mathscr{E}(\varepsilon, \mathbf{x}; \theta)$ may contain one or more outcomes, depending on the realization of ε and $\mathbf{x}$. In the example above, both $y_1 = (0,0,0,0)$ and $y_2 = (1,1,1,1)$ are equilibria of games such that $-(\beta \mathbf{x}_i + \gamma) < \varepsilon_i < -\beta x_i$, $\forall i \in V$. There are many other combinations of multiple equilibria for games with different utility profiles.[3]

The presence of multiple equilibria is not a special feature of games in network, but rather a common feature in discrete games. Various approaches have been proposed to estimate discrete games (Bjorn and Vuong 1984, Bresnahan and Reiss 1991a, Berry 1992, Tamer 2003, Andrews, Berry and Jia 2004, Ciliberto and Tamer 2009, Galichon and Henry 2011, and Beresteanu, Molchanov and Molinari 2011). An entry game can be thought of as a special case of games in networks, where all agents connect to each other. When agents are not necessarily linked to all other agents, the property of multiple equilibria is much harder to describe, because the set of equilibria depends on the network structure as well. More importantly, in applications, we usually deal with networks with a large number of nodes and varying sizes. In the next subsection we elaborate on the reasons that make the

[3]For example, both $y_1 = (0,0,0,0)$ and $y_3 = (0,1,1,0)$ are equilibria of games such that $\varepsilon_i < -(\beta \mathbf{x}_i + \frac{1}{2})$, $\forall i \in \{1,4\}$ and $-(\beta \mathbf{x}_i + \frac{1}{2}) < \varepsilon_i < -\beta \mathbf{x}_i$, $\forall i \in \{2,3\}$.

estimation of games in network more challenging.

3.2.2 Empirical Challenges of Estimating Discrete Games in Large Networks

As discussed in Bresnahan and Reiss (1991a), Tamer (2003) and Ciliberto and Tamer (2009), when a game has multiple equilibria, the probability of outcomes are not well-defined without information on equilibrium selection. Traditional methods such as the likelihood approach and the method of moments cannot be used. Games in networks have other features that complicate identification and estimation. The network we consider contains more than a few agents. Data may contain networks of different sizes and structures. In this subsection we will explain why existing approaches for estimating discrete games are not easily extendable for games in networks.

In principle, point identification could be achieved if equilibrium selection mechanism is given (Bjorn and Vuong 1984, Jia 2008), or estimated (Bajari, Hong and Ryan 2010, Narayanan 2013). Point identification requires the calculation of all the equilibria of a game. A standard way of obtaining the set of equilibria is to enumerate all the possible outcomes of the game and check if each of them is an equilibrium. This method is not feasible in practice if the number of agents is large, because the number of possible outcomes that need to be checked grows at an exponential rate. For example, in a binary game with 20 agents, the number of possible outcomes is $2^{20} \approx 10^6$. In the data we consider, the majority of friend networks have sizes range from 70 to 90. It will be computationally costly to check such large number of outcomes for a network. Moreover, even if we are able to find ways to calculate all the possible equilibria of all games, point identification is still questionable because the composition of multiple equilibria varies significantly across games. It is hard to justify that the assumed selection mechanism is the true selection mechanism of the data generating process.

The partial identification approach does not rely on assumptions on equilibrium selection or calculation of all equilibria of the model. Thus it is more suitable for the model that is considered in this chapter. Following the idea of Ciliberto and Tamer (2009), moment conditions could be constructed by imposing bounds on the probability of the game outcomes. In the 4 person example, $y = \{0,0,0,0\}$ will be observed only if every players play best response. The upper bound of observing y is the probability that $u_i < 0$, $\forall i$, therefore $\Pr(y) \leq \Pr(u_i < 0, \forall i)$. In a game with N players, there are a total of 2^N upper bounds that could be used to construct moment conditions.

There are a number of concerns that can arise from constructing bounds like this. First, because the number of possible outcomes increases exponentially, the number of moment inequalities that we need to check grows at the exponential rate as well. In a network with 70 agents, the total number of outcomes of full network is more than 10^{21}. It is computationally infeasible to check so many number of moment conditions. More seriously, each individual moment condition cannot be consistently estimated if the number of outcomes is enormous. We consider the case where the number of agents is large in data, but we don't require the number of networks to be large. There will not be enough networks of the same outcome available in data to construct the empirical probability of each outcome, therefore moment conditions cannot be verified.

Another problem arising from games of network is the variation of structure among networks. In an entry game, each market has the same number of potential entrants, whose decision is affected by all the rest of players in the market. In the peer effects model, each disconnected network is an analogue of a market in an entry game. However, the number of people and the friendship relationship among them are generally different across disconnected networks. The outcome of a network depends on how players are linked. If a moment inequality is placed on the outcome of the full network, there may not be enough number of networks of the same size and structure to construct moment conditions of the game.

In this chapter, instead of constructing moment conditions of outcomes of full networks, we address the computation and consistency issues by exploring properties of subnetworks.

Because there are links that connect agents inside and outside the subnetwork, we have to consider the interaction between agents inside and outside the subnetwork as well. The novelty of our method is to find conditions of subnetwork that are satisfied regardless of which actions people outside the network take. By this additional relaxation in constructing upper bounds, the moment conditions can be easily verified. We detail our identification strategy in the next section.

3.3 Identification and Inference via Subnetworks

3.3.1 An Illustration

In the 4-person peer effects model in Figure 3.1, assuming no control variable **x**, agents' actions are characterized by the following set of equations:

$$\begin{aligned} y_1 &= 1(\gamma \cdot y_2 + \varepsilon_1 > 0), \\ y_2 &= 1(\gamma \cdot \frac{y_1 + y_3}{2} + \varepsilon_2 > 0), \\ y_3 &= 1(\gamma \cdot \frac{y_2 + y_4}{2} + \varepsilon_3 > 0), \\ y_4 &= 1(\gamma \cdot y_3 + \varepsilon_4 > 0). \end{aligned}$$

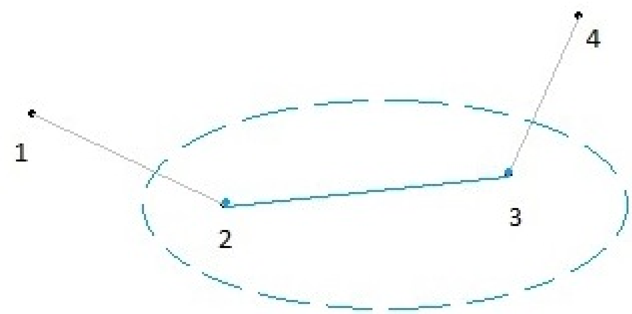
As before, assume $\varepsilon \overset{i.i.d.}{\sim} N(0,1)$. We temporarily assume $\gamma > 0$.

Consider a subnetwork that consists of agents 2 and 3. Our goal is to find moment inequalities for outcomes of the subnetwork, for example, the probability of observing the event

$$A := (y_2 = 0, y_3 = 1).$$

$(y_2 = 0, y_3 = 1)$ is observed if and only if one of the following outcomes of the full network is observed.

Figure 3.2: An Illustration of Subnetwork



$$
\begin{aligned}
B_1 &:= \quad (y_1 = \quad 0, y_2 = 0, y_3 = 1, y_4 = 0), \\
B_2 &:= \quad (y_1 = \quad 0, y_2 = 0, y_3 = 1, y_4 = 1), \\
B_3 &:= \quad (y_1 = \quad 1, y_2 = 0, y_3 = 1, y_4 = 0), \\
B_4 &:= \quad (y_1 = \quad 1, y_2 = 0, y_3 = 1, y_4 = 1).
\end{aligned}
$$

By checking the best response functions for each agents, it is verifiable that B_1 is an Nash equilibrium of a game if and only if $\varepsilon \in R_1$, where

$$R_1 := \{\varepsilon \in \mathbb{R}^4 : \gamma \cdot 0 + \varepsilon_1 < 0; \gamma \cdot \frac{1}{2} + \varepsilon_2 < 0; \gamma \cdot 0 + \varepsilon_3 > 0; \gamma \cdot 1 + \varepsilon_4 < 0\}.$$

If $\varepsilon \notin R_1$, B_1 cannot be observed because it is not a Nash equilibrium of the game. If $\varepsilon \in R_1$, the game may have other equilibria in addition to B_1. Whether or not observing B_1 depends on how multiple equilibria are selected. Therefore $\varepsilon \in R_1$ is a necessary condition of observing B_1.

Similarly, $\varepsilon \in R_2$ is a necessary condition of observing B_2, where

$$R_2 := \{\varepsilon \in \mathbb{R}^4 : \gamma \cdot 0 + \varepsilon_1 < 0; \gamma \cdot \frac{1}{2} + \varepsilon_2 < 0; \gamma \cdot \frac{1}{2} + \varepsilon_3 > 0; \gamma \cdot 1 + \varepsilon_4 > 0\};$$

$\varepsilon \in R_3$ is a necessary condition of observing B_3, where

$$R_3 := \{\varepsilon \in \mathbb{R}^4 : \gamma\cdot 0+\varepsilon_1 > 0; \gamma\cdot 1+\varepsilon_2 < 0; \gamma\cdot 0+\varepsilon_3 > 0; \gamma\cdot 1+\varepsilon_4 < 0\};$$

And $\varepsilon \in R_4$ is a necessary condition of observing B_4, where

$$R_4 := \{\varepsilon \in \mathbb{R}^4 : \gamma\cdot 0+\varepsilon_1 > 0; \gamma\cdot 1+\varepsilon_2 < 0; \gamma\cdot \frac{1}{2}+\varepsilon_3 > 0; \gamma\cdot 1+\varepsilon_4 > 0\}.$$

Define H as

$$H := \{\varepsilon \in \mathbb{R}^4 : \gamma\cdot \frac{1}{2}+\varepsilon_2 < 0; \gamma\cdot \frac{1}{2}+\varepsilon_3 > 0\}.$$

It is easy to check that $R_i \subset H, \forall i = 1,2,3,4$.

Putting these together, we get

$$\begin{aligned}
\Pr(A) &= \Pr(B_1 \textit{ or } B_2 \textit{ or } B_3 \textit{ or } B_4) \\
&\leq \Pr(\varepsilon \in (R_1 \cup R_2 \cup R_3 \cup R_4)) \\
&\leq \Pr(\varepsilon \in H). \qquad (3.2)
\end{aligned}$$

By replacing H and A with their expressions, the following inequality for the probability of the outcome in subnetwork $A = \{2,3\}$ is satisfied :

$$\Pr(y_2 = 0, y_3 = 1) \leq \Pr(\gamma\cdot \frac{1}{2}+\varepsilon_2 < 0; \gamma\cdot \frac{1}{2}+\varepsilon_3 > 0). \qquad (3.3)$$

Note that because $y_2 = 1(\gamma\cdot \frac{y_1+y_3}{2}+\varepsilon_2 > 0)$ and $y_3 = 1$, $\gamma\cdot \frac{1}{2}+\varepsilon_2 < 0$ is a necessary condition for $y_2 = 0$ when $y_1 = 0$. Similarly, because $y_3 = 1(\gamma\cdot \frac{y_2+y_4}{2}+\varepsilon_3 > 0)$ and $y_2 = 0$, $\gamma\cdot \frac{1}{2}+\varepsilon_3 > 0$ is a necessary condition for $y_3 = 1$ when $y_4 = 1$. The upper bound is constructed by requiring each player i to play best response to the hypothetical scenario such that player i's all friends outside subnetwork A take the same action as player i does.

3.3.2 Moment Conditions in General Cases

For the full network consists of notes $V = \{1, 2, ...N\}$, let y_V denote the outcome of all players in V. y_V is a Nash equilibrium of a game if each individual in V plays best response given the actions of all other players in the network. Player i chooses action 1 if the utility of choosing action 1 is positive, because the utility of the alternative is normalized to 0. Player i chooses the alternative if the utility of action 1 is negative.

Let

$$\begin{aligned} R(y_V; x; \theta) \quad := \quad & \{\varepsilon \in \mathbb{R}^N : u(y_i, x_i, y_{-i}; \theta) \geq 0, \ \forall y_i = 1, i \in V; \\ & u(y_i, x_i, y_{-i}; \theta) \leq 0, \ \forall y_i = 0, i \in V\}, \end{aligned} \tag{3.4}$$

denote the set of games of which y_V is a Nash equilibrium. If the utility function takes the form as in Eq. (3.1) , $R(y_V; x; \theta)$ is equivalent to

$$R(y_V; x; \theta) = \{\varepsilon \in \mathbb{R}^N : (2 \cdot y_i - 1) \cdot (\beta x_i + \gamma \frac{\sum_{j \in V} g_{ij} y_j}{\sum_{j \in V} g_{ij}} + \varepsilon_i) \geq 0, \ \forall i \in V\}. \tag{3.5}$$

y_V will only be observed if it is an equilibrium of the game. Hence

$$\Pr(y_V | \mathbf{x}) \leq \Pr(\varepsilon \in R(y_V; \mathbf{x}, \theta)). \tag{3.6}$$

This is an example of moment inequalities of full network. As discussed in the previous section, moment conditions of full network cannot be consistently estimated if the size of network is large, or if the interaction matrix varies across networks. Therefore we need to seek alternative moment conditions.

Recall that a subnetwork contains three types of information: the list of agents A, the connects among agents in the subnetwork G_A, and the number of connections to agents outside the subnetwork n_A. Theorem 3.1 shows how moments of subnetwork are bounded above by moments predicted by the model. Moment inequalities like this could be used to make partial identification of the original model.

Theorem 3.1 *Consider a simultaneous game of complete information in network $V = \{1,2,...N\}$ with the utility function*

$$y_i = 1(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij} y_j}{\sum_{j\in V} g_{ij}} + \varepsilon_i > 0),$$

where $\gamma > 0$ and $\varepsilon_i \overset{i.i.d.}{\sim} N(0,1)$, $\forall i \in V$. Let A be a subset of V. Define

$$\begin{aligned} & H(y_A;\mathbf{x},G_A,n_A,\theta) \\ := \quad & \{\varepsilon \in \mathbb{R}^N : (2\cdot y_i - 1)\cdot(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}[1(j\in A)\cdot y_j + 1(j\notin A)\cdot y_i]}{\sum_{j\in V} g_{ij}} + \varepsilon_i) \geq 0,\ \forall i \in A\} \\ = \quad & \{\varepsilon \in \mathbb{R}^N : (2\cdot y_i - 1)\cdot(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in A} g_{ij}\cdot y_j + n_{A,i}\cdot y_i}{\sum_{j\in A} g_{ij} + n_{A,i}} + \varepsilon_i) \geq 0,\ \forall i \in A\} \end{aligned}$$

The following inequality holds for any $A \subset V$:

$$\Pr(y_A|\mathbf{x},G_A,n_A,\theta) \leq \Pr(\varepsilon \in H(y_A;\mathbf{x},G_A,n_A,\theta)). \tag{3.7}$$

$H(y_A;\mathbf{x},G_A,n_A,\theta)$ is the key innovation of our subnetwork approach. Besides that the conditions of ε are placed on the agents in the subnetwork rather than all agents, what is special about $H(y_A;\mathbf{x},G_A,n_A,\theta)$ is that y_j in $R(y_V,\mathbf{x};\theta)$ is replaced by $1(j\in A)\cdot y_j + (j \notin A)\cdot y_i$. Mathematically, this new term takes value y_j for agent j inside the subnetwork and takes value y_i for j outside the subnetwork. In other words, for individual i, we assume all of his or her friends outside the subnetwork takes the same value as i takes, regardless of what their true actions are. This is because when we focus on actions in the subnetwork, we look for conditions that will be satisfied regardless of what action agents outside the subnetwork take. When we construct the upper bound for $\Pr(y_A)$, we seek actions that will make y_A most likely to happen. When $\gamma > 0$, individual i will gain extra utility of taking an action if a larger percent of his or her friends take the same action. Following this intuition, for $i \in A$ who has friends outside the subnetwork, $y_i = 0$ is more likely to occur if i's "outside friends" all take action 0; alternatively, $y_i = 1$ is more likely to occur if i's "outside friends" all take action 1, that is why we replace the actions of agents outside the network by the action of player i in the definition of $H(y_A; \mathbf{x}, G_A, n_A, \theta)$.

As an extension of Theorem 3.1, we can also consider the case where the interaction effect is negative. In this case, we replace the actions of agents outside the network by the opposite of player i'th action. This is summarized in the next corollary:

Corollary 3.1 *Consider a simultaneous game of complete information in network $V = \{1, 2, ...N\}$ with the utility function*

$$y_i = 1(\beta \mathbf{x}_i - \gamma \frac{\sum_{j \in V} g_{ij} y_j}{\sum_{j \in V} g_{ij}} + \varepsilon_i > 0), \tag{3.8}$$

where $\gamma > 0$ and $\varepsilon_i \overset{i.i.d.}{\sim} N(0,1)$, $\forall i \in V$. Let A be a subset of V. Define

$$\begin{aligned} & \tilde{H}(y_A; \mathbf{x}, G_A, n_A, \theta) \\ := \quad & \{\varepsilon \in \mathbb{R}^N : (2 \cdot y_i - 1) \cdot (\beta \mathbf{x}_i - \gamma \frac{\sum_{j \in V} g_{ij}[1(j \in A) \cdot y_j + 1(j \notin A) \cdot (1 - y_i)]}{\sum_{j \in V} g_{ij}} + \varepsilon_i) \geq 0, \ \forall i \in A\} \\ = \quad & \{\varepsilon \in \mathbb{R}^N : (2 \cdot y_i - 1) \cdot (\beta \mathbf{x}_i - \gamma \frac{\sum_{j \in A} g_{ij} \cdot y_j + n_{A,i} \cdot (1 - y_i)}{\sum_{j \in A} g_{ij} + n_{A,i}} + \varepsilon_i) \geq 0, \ \forall i \in A\} \end{aligned}$$

The following inequality holds for any $A \subset V$:

$$\Pr(y_A | \mathbf{x}, G_A, n_A, \theta) \leq \Pr(\varepsilon \in \tilde{H}(y_A; \mathbf{x}, G_A, n_A, \theta)). \tag{3.9}$$

3.3.3 Inference and Estimation

Our inference procedure is based on subnetworks. For each subnetwork A, the model predicts moment inequality (3.7) if the interaction effect is positive.

Suppose we are interested in subnetwork A, let y_A denote all the possible outcomes of A. The identified set is

$$\Theta_I = \{\theta : P(y_A|\mathbf{x}, G_A, n_A) \leq \Pr(\varepsilon \in H(y_A; \mathbf{x}, G_A, n_A, \theta))\ \forall y_A \in Y_A\},$$

where $P(y_A|\mathbf{x}, G_A, n_A)$ is the choice probability of the data.

Let $m(y_A, \mathbf{x}, G_A, n_A; \theta) = P(y_A|\mathbf{x}, G_A, n_A) - \Pr(\varepsilon \in H(y_A; \mathbf{x}, G_A, n_A, \theta))$, the model predicts $m(y_A, \mathbf{x}, G_A, n_A; \theta) \leq 0$. Let $\mathbf{m}(\mathbf{x}, G_A, n_A; \theta)$ be the vector of of these moment conditions for all outcomes of subnetwork A. Our inference procedure uses the objective function

$$Q(\theta) = E_{(\mathbf{x}, G_A, n_A)}||(\mathbf{m}(\mathbf{x}, G_A, n_A; \theta))_+||$$

where $(\cdot)_+$ takes the positive part of the vector and $||\cdot||$ is the Euclidean norm.

To make inference, we use the sample analogue of $Q(\theta)$. We randomly select T ($T \to \infty$.) subnetworks of the same size as A. The sample analogue of criterion function is

$$\Theta_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} ||(\mathbf{m}_t(\mathbf{x}, G_A, n_A; \theta))_+||,$$

where $\mathbf{m}_t(\mathbf{x}, G_A, n_A; \theta)$ is an empirical analogue of $\mathbf{m}(\mathbf{x}, G_A, n_A; \theta)$ evaluated at observation t. The elements in $\mathbf{m}_t(\mathbf{x}, G_A, n_A; \theta)$ are

$$m_t(y_A, \mathbf{x}, G_A, n_A; \theta) = P(y_A|\mathbf{x}, G_A, n_A) - \Pr(\varepsilon \in H(y_A; \mathbf{x}, G_A, n_A, \theta)),$$

where $y_A \in Y_A$, $P(y_A|\mathbf{x}, G_A, n_A)$ is an empirical conditional distribution of y_A given $\mathbf{x}$, G_A and n_A of observation t, $\Pr(\varepsilon \in H(Y_A; \mathbf{x}_t, G_{A,t}, n_{A,t}, \theta)$ is calculated using the expression in Theorem 3.1.

Inference could be made by subsampling as discussed in Chernozhukov, Hong and Tamer (2007) and Ciliberto and Tamer (2009). The confidence interval is

$$\hat{\Theta}_I = \{\theta : T \cdot \Theta_T(\theta) \leq c_\tau\}.$$

3.4 Monte Carlo Study

In this section, we conduct a sequence of Monte Carlo experiments to study finite sample properties of our subnetwork approach. For simplicity, we make inference on one parameter. The utility function contains the interaction effect only,

$$u(y_i, \mathbf{x}_i, y_{-i}; \beta, \gamma) = \gamma \frac{\sum_{j \in V} g_{ij} y_j}{\sum_{j \in V} g_{ij}} + \varepsilon_i.$$

The true parameter is $\gamma_0 = 1$. We set the number of networks to be1000. For each network, we generate individual shocks ε and calculate its set of equilibria. If the game has multiple equilibria, each one is selected with equal probability. Inference is based on a 80% confidence interval of 200 re-samples. Upper bounds are approximated by 1000 simulations.

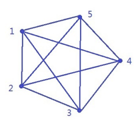
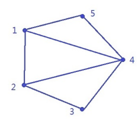
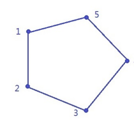
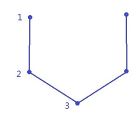
First, we find that upper bounds on outcomes of subnetwork are informative. Table 3.1 collects the empirical probabilities of outcomes of subnetwork $\{1,2\}$, and upper bounds calculated by different hypothetical values of parameter γ in a 4-person network described in Figure 3.1. The numbers in blue show cases where moment inequalities are violated. When γ is too low, the upper bound predicts too small share of outcome $(y_1, y_2) = (1,1)$; when γ is too large, it predicts too small share of outcomes $(y_1, y_2) = (0,1)$ and $(y_1, y_2) = (1,0)$. Only when the value is near the true value $\gamma_0 = 1$ all the moment inequalities are satisfied.

The second sets of example show the impact of the structure of full network on the performance of our approach. Table 3.2 collects the identified set of four networks using four choices of subnetworks. From the left to the right, we generate full networks with decreasing number of edges. The performance of our approach improves when there are

Table 3.1: Upper Bounds and Parameter Values

Outcome	Emp. Prob.	Upper Bound						
(y_1,y_2)	$P(y_A)$	$\gamma=0.1$	$\gamma=0.5$	$\gamma=1$	$\gamma=1.5$	$\gamma=2$	$\gamma=3$	$\gamma=5$
$(1,1)$	0.642	0.291	0.478	0.708	0.870	0.955	0.997	1
$(0,1)$	0.102	0.240	0.201	0.154	0.113	0.080	0.033	0.003
$(1,0)$	0.100	0.239	0.185	0.110	0.052	0.020	0.001	0
$(0,0)$	0.155	0.250	0.250	0.250	0.250	0.250	0.250	0.250

Table 3.2: Identified Set of Varying Full and Sub Networks

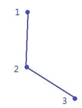
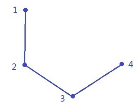
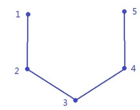
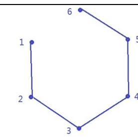
Subnetwork	Identification Set			
				
$\{1,2\}$	$(0.834,3.001)$	$(0.716,3.001)$	$(0.680,2.669)$	$(0.771,1.601)$
$\{1,2,3\}$	$(0.823,2.987)$	$(0.737,3.001)$	$(0.758,1.659)$	$(0.780,1.337)$
$\{1,2,3,4\}$	$(0.830,1.861)$	$(0.832,1.578)$	$(0.812,1.612)$	$(0.877,1.240)$
$\{1,2,3,4,5\}$	$(1.027,1.037)$	$(0.900,0.903)$	$(0.827,1.210)$	$(0.911,1.001)$

fewer links connecting individuals inside and outside subnetworks. For example, in the first network, agent 1 is connected with all others in the network, but in the last network, agent 1 is connected to agent 2 only. When we construct upper bounds for subnetwork $\{1,2\}$, we relax our upper bounds more in the first network because we ignore many connections. Our upper bounds are tighter if the network is sparse.

Table 3.2 also illustrates the relationship between the choice of subnetwork and the size of the confidence interval. From the top to the bottom, we increase the size of subnetworks. When the size of subnetwork increases, the confidence interval shrinks. This is because more information is used to construct moment conditions. However, it is not the case that the performance is the best if we make inference based on the full-subnetwork, which has the largest number of agents. This is because with a large number of outcomes, the probability of each individual outcome may not be precisely estimated for a fixed sample size.

The last sets of examples show the performance of our approach when the size of full

Table 3.3: Full Network Size and Confidence Interval

Subnetwork	Identification Set			
				
$\{1,2\}$	$(0.743, 1.276)$	$(0.712, 1.605)$	$(0.771, 1.601)$	$(0.734, 1.678)$
$\{1,2,3\}$	$(0.899, 1.081)$	$(0.834, 1.410)$	$(0.780, 1.337)$	$(0.846, 1.391)$
$\{1,2,3,4\}$		$(0.865, 1.136)$	$(0.877, 1.240)$	$(0.904, 1.350)$

network increases. As shown in Table 3.3, for a given choice of subnetwork, confidence interval increases slightly with the size of full network, but the magnitude is small. This is because for a given choice of subnetwork, the links whose end points are outside the subnetwork do not affect moment conditions. The moment conditions are still informative even if the network is of large size.

3.5 Application

In this application, we study peer effects on smoking using the data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is a nation-wide longitudinal survey of adolescent health. We use Wave I in Home Survey, which collects social, economic and physical information of teenagers from grades 7 to 12 in year 1993 and 1994. Add Health is one of the most commonly used data set in studying peer effects (e.g. Gaviria and Raphael 2001 on juvenile behavior, Calvó-Armengol, Patacchini and Zenou 2009 on education, Trogdon, Nonnemaker and Pais 2008 on overweight, Cohen-Cole and Fletcher 2008 on obesity). What is special about Add Health is that respondents are asked to nominate their friends. The information on family, social background, activities together with friend nomination provides a unique opportunity to study interactions among friends controlling for other social and economic influences.

Table 3.4 reports summary statistics of our key variables. There are a total of 20745

Table 3.4: Summary Statistics

	Full Sample		Regression Sample	
	Mean	Stdev	Mean	Stdev
Obs	20,745		8,865	
Smoke	0.26	0.44	0.26	0.44
Grade	9.67	1.64	9.73	1.63
log(Income)	45.73	51.62	3.57	0.82
Gender	0.49	0.50	0.49	0.50
Race	0.72	0.45	0.62	0.49
Parent_smoke	0.26	0.44	0.24	0.43
School_program	0.92	0.27	0.93	0.26
NumMF	0.80	1.12		
NumFF	0.85	1.13		

observations in the survey. The key dependent variable, smoke, is an indicator of whether a correspondent smokes on a regularly basis. On average 26% of the respondents are smokers. The control variables include age, gender, race, parents' smoking behavior and household income. We further include a dummy variable indicating whether a person's school has prevention programs for smoking.

The goal of our study is to identify factors that influence smoking decision, especially the peer effects. Our model is as follows

$$y_i = 1(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij} y_j}{\sum_{j\in V} g_{ij}} + \varepsilon_i),$$

where V is the set of all individuals in the data set, $\varepsilon_i \sim i.i.d.\, N(0,1)$. $y_i = 1$ if person i smokes, $y_i = 0$ otherwise. $\mathbf{x}_i$ are control variables. We set $g_{ij} = g_{ji} = 1$ for $\{i,j\} \in V$ as long as one of person i and j nominates the other as a friend.

Our estimation starts with a naive Probit estimation of smoking. The sample contains individuals that have at least one friend. Coefficients, standard errors and marginal effects at mean are reported in Table 3.5. People are more likely to smoke if their peers or parents smoke. The rate of smoking also differs by gender and race. Because in the Probit regression, the coefficient for grade, parents' smoking, and race are significant, we include these

Table 3.5: Probit Estimation of Smoking

	(1)			(2)		
	Coef.	Std.Err.	Marg.Eff.	Coef.	Std.Err.	Marg.Eff.
Obs	6,421			8,865		
Peer_effect	1.259***	0.057	0.358	1.186***	0.048	0.338
Birth_year	−0.013	0.010	−0.004			
Grade	0.064***	0.015	0.020	0.076***	0.010	0.022
School_program	0.043	0.070	0.012			
Gender	0.022	0.036	0.006			
Log_income	−0.004	0.023	−0.001			
Race	0.420***	0.041	0.119	0.430***	0.033	0.122
Parent_smoke	0.248	0.042	0.070	0.240***	0.035	0.068
Constant	−0.930	0.937		−2.022***	0.101	

three control variables in our partial identification approach.

Results from the Probit estimation show that the marginal effect of peer effects is 30%, which means if half of a person's friends smoke, that person's probability of smoking would increase by 15%. However, Probit estimation could be misleading because it treats friend's smoking behavior exogenous. If there is indeed peer effects on smoking, an unobserved factor that influences person i's behavior can be correlated with the behavior of his/her friends', because his/her friends' actions are affected by his/her behavior via peer effects. To deal with the endogeneity problem of friends' actions, we next turn to our partial identification approach.

We make inference using subnetwork of size 2 for computational tractability. The sparsity of the data set also supports our choice of using small subnetworks. The last two rows of Table 3.4 show that the average number of male and female friends nominated by an individual is 0.80 and 0.85, respectively. Table 3.5 reports more information about friend nomination. About $2/3$ of respondents nominate zero of one friend. 83% of people nominate two friends or fewer. This suggests that each individual only connects to a few people. Inference based on small subnetworks is therefore informative.

The moment conditions are tested by 11,492 pairs of agents that are friends. For sub-

Table 3.6: Number of Nominated Friends

	Obs	Frequency
NumF = 0	6045	29.13%
NumF = 1	6738	32.48%
NumF = 2	4453	21.47%
NumF $\leq$2	17206	83.09%

Table 3.7: Partial Identification of Peer Effects in Friend Network

	Confidence Interval
Peer_effect	(0.872,1.493)
Grade	(-0.018,0.104)
Parent_smoke	(-0.150,0.290)
Race	(-0.001,0.391)
Constant	(-2.145,-1.027)

networks of 2 agents, there are two adjacency matrix G_A, either the two connects or disconnects. The upper bounds are higher if agents are not connected because the percentage of fiends outside subnetwork is larger. Moment inequalities are more likely to be violated for pairs of agents who are connected. Due to computational concerns, we only check inequalities for pairs of agents that are friends. This is equivalent to checking the conditional moment inequalities given a fixed G_A.

The final result is reported in Table . The confidence interval of peer effects estimated by the structural approach is (0.872, 1.493). This suggests positive and significant peer effects on smoking. Though the subnetwork approach concludes larger confidence intervals as compared to Probit estimation, it produces more convincing results because it takes into account the endogeneity of friends actions due to social interaction. It should also be noted that the criterion function evaluated at the Probit estimator is 0.016, while the minimum of criterion function is 0.003. Probit estimator lies outside the confidence set, and is therefore unlikely to be the true parameter of the data generating process.

3.6 Conclusion

This chapter studies identification and estimation of peer effects on binary choices in social networks. Our framework belongs to discrete games of complete information. As is well-known in the entry literature, identifying discrete games is difficult in general because the model often yields multiple equilibria. The inherited network feature of our model makes identification and estimation even more challenging.

The existing econometric methods that rely on the choice probabilities of the full game are not feasible in our case because we consider networks that are large and have varying friendship relationships. Not only does the number of outcomes grow exponentially, the number of observations of the same network structure is not sufficient to construct moment conditions. Therefore, we seek alternative moment conditions that can be consistently estimated.

The novelty of our identification strategy is the use of subnetworks. A subnetwork is a collection of agents, whose actions depend on actions of their friends, both inside and outside the subnetwork. Because we seek conditions that hold regardless the behavior of the agents outside the subnetwork, the bound of an outcome of a subnetwork is constructed to require each agents i inside the subnetwork to play best response to their friends' actions, assuming all "outside friends" of agent i take the same action as agent i takes. Because peer effects are positive, such bound gives the highest probability of observing an outcome. We therefore get a set of moment inequalities that could be used to partially identify the model.

Our estimation strategy is closely related to Ciliberto and Tamer (2009). The criterion function penalizes a parameter if upper bounds predicted by the parameters are less than the empirical choice probabilities of the subnetwork. The identified set is the set of parameters whose criterion function is less than a threshold, calculated by subsampling.

The Monte Carlo examples presented in this chapter study the performance of our approach. The subnetwork approach is able to provide an informative inference on parameters of the model, especially when the network is sparse. We apply our identification and estimation strategy to study peer effects on smoking using the data from the National Longitudinal Study of Adolescent to Adult Health. The identified set suggests positive and significant peer effects on smoking.

A limitation of our approach is that we treat the network formation as given. This is a concern in peer effects models because friend relationships are endogenous. A possible approach to address this concern is to allow correlation among individual unobserved characteristics. Furthermore, extension of our approach to structural model taking into account the endogenous network formulation is left for future research.

Another extension of this chapter concerns with sharper identification. The moment conditions used in this chapter are only a small part of conditions implied by the model. The identification set could shrink further if more informative and easily verifiable conditions are considered.

Appendix

Proof of Theorem 3.1: For a given subnetwork A, let y_A and y_{-A} denote the actions of all players inside and outside A. Define

$$R(y_A, y_{-A}|\mathbf{x}, G_A, n_A, \theta) := \{\varepsilon \in \mathbb{R}^N : \ (2\cdot y_i - 1)\cdot(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij} y_j}{\sum_{j\in V} g_{ij}} + \varepsilon_i) \geq 0 \ , \forall i \in A\}$$

A necessary condition for (y_A, y_{-A}) being an outcome of the game is that $\varepsilon \in R(y_A, y_{-A}|\mathbf{x}, G_A, n_A, \theta)$. Otherwise players in A don't play best response to the actions of other players, therefore (y_A, y_{-A}) is not a Nash equilibrium of the game.

On the other hand, for $i \in A$ such that $y_i = 0$, we have $y_i \leq y_j$ for $\forall j \in V$, therefore

$$\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}[1(j\in A)\cdot y_j + 1(j\notin A)\cdot y_i]}{\sum_{j\in V} g_{ij}} + \varepsilon_i \leq \beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}\cdot y_j}{\sum_{j\in V} g_{ij}} + \varepsilon_i.$$

When $y_i = 0$, $2\cdot y_i - 1 = -1$. The above inequality is equivalent to

$$(2\cdot y_i - 1)\cdot(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}[1(j\in A)\cdot y_j + 1(j\notin A)\cdot y_i]}{\sum_{j\in V} g_{ij}} + \varepsilon_i) \geq (2\cdot y_i - 1)(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}\cdot y_j}{\sum_{j\in V} g_{ij}} + \varepsilon_i). \tag{3.10}$$

Similarly, for $i \in A$ such that $y_i = 1$, we have $y_i \geq y_j$ for $\forall j$, therefore

$$\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}[1(j\in A)\cdot y_j + 1(j\notin A)\cdot y_i]}{\sum_{j\in V} g_{ij}} + \varepsilon_i \geq \beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}\cdot y_j}{\sum_{j\in V\cdot} g_{ij}} + \varepsilon_i,$$

which is equivalent to

$$(2\cdot y_i - 1)\cdot(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij}[1(j\in A)\cdot y_j + 1(j\notin A)\cdot y_i]}{\sum_{j\in V} g_{ij}} + \varepsilon_i) \geq (2\cdot y_i - 1)(\beta \mathbf{x}_i + \gamma \frac{\sum_{j\in V} g_{ij} y_j}{\sum_{j\in V} g_{ij}} + \varepsilon_i). \tag{3.11}$$

Note the left hand sides of inequalities (3.10) and (3.11) appear in the definition of $H(y_A;\mathbf{x},G_A,n_A,\theta)$. The right hand sides appear in $R(y_A,y_{-A}|\mathbf{x},G_A,n_A,\theta)$. If $\varepsilon \in R(y_A,y_{-A}|\mathbf{x},G_A,n_A,\theta)$, the left hands of inequalities (3.10) and (3.11) are greater than 0, therefore the left hands are greater than 0. Hence $\varepsilon \in H(y_A;\mathbf{x},G_A,n_A,\theta)$. In other words,

$$R(y_A,y_{-A}|\mathbf{x},G_A,n_A,\theta) \subset H(y_A;\mathbf{x},G_A,n_A,\theta).$$

If $\varepsilon \notin H(y_A;\mathbf{x},G_A,n_A,\theta)$, $\varepsilon \notin R(y_A,y_{-A}|\mathbf{x},G_A,n_A,,\theta)$ for $\forall y_{-A}$. As a consequence, there doesn't exist y_{-A} such that (y_A,y_{-A}) is an outcome of the game. Therefore, y_A cannot be observed. This suggests that

$$\Pr(\varepsilon \notin H(y_A;\mathbf{x},G_A,n_A,\theta)) \leq 1-\Pr(y_A\mathbf{x},G_A,n_A,\theta),$$

or equivalently,

$$\Pr(y_A|\mathbf{x},G_A,n_A,\theta) \leq \Pr(\varepsilon \in H(y_A;\mathbf{x},G_A,n_A,\theta)).$$

BIBLIOGRAPHY

[1] Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*(422), 669–679.

[2] Andrews, D. W., S. Berry, and P. Jia (2004). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. *Working Paper, Yale University*.

[3] Aradillas-Lopez, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics 157*(2), 409–431.

[4] Badev, A. (2013). Discrete games in endogenous networks: Theory and policy. *Working Paper*.

[5] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010). Estimating static models of strategic interactions. *Journal of Business & Economic Statistics 28*(4), 469–482.

[6] Bajari, P., H. Hong, and S. P. Ryan (2010). Identification and estimation of discrete games of complete information. *Econometrica 78*(5), 1529–1568.

[7] Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica 79*(6), 1785–1821.

[8] Berry, S. and P. Reiss (2007). Empirical models of entry and market structure. *Handbook of Industrial Organization 3*, 1845–1886.

[9] Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica 60*(4), 889–917.

[10] Bjorn, P. A. and Q. H. Vuong (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. *Working Paper, California Institute of Technology*.

[11] Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

[12] Bresnahan, T. F. and P. C. Reiss (1990). Entry in monopoly market. *The Review of Economic Studies 57*(4), 531–553.

[13] Bresnahan, T. F. and P. C. Reiss (1991a). Empirical models of discrete games. *Journal of Econometrics 48*(1), 57–81.

[14] Bresnahan, T. F. and P. C. Reiss (1991b). Entry and competition in concentrated markets. *Journal of Political Economy 99*(5), 977–1009.

[15] Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

[16] Brock, W. A. and S. N. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*(1), 52–75.

[17] Calvó-Armengol, A. and M. O. Jackson (2004). The effects of social networks on employment and inequality. *American Economic Review 94*(3), 426–454.

[18] Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

[19] Card, D. and L. Giuliano (2013). Peer effects and multiple equilibria in the risky behavior of friends. *Review of Economics and Statistics 95*(4), 1130–1149.

[20] Chen, X., E. T. Tamer, and A. Torgovitsky (2011). Sensitivity analysis in semiparametric likelihood models. *Cowles Foundation Discussion Paper*.

[21] Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models1. *Econometrica 75*(5), 1243–1284.

[22] Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

[23] Clark, A. E. and Y. Loheac (2007). It wasn't me, it was them! social influence in risky behavior by adolescents. *Journal of Health Economics 26*(4), 763–784.

[24] Cohen-Cole, E. and J. M. Fletcher (2008). Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics 27*(5), 1382–1387.

[25] Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review 100*(1), 35–69.

[26] Dunker, F., S. Hoderlein, and H. Kaido (2013). Random coefficients in static games of complete information. *Working Paper, Centre for Microdata Methods and Practice*.

[27] Galichon, A. and M. Henry (2011). Set identification in models with multiple equilibria. *The Review of Economic Studies 78*(4), 1264–1298.

[28] Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics 83*(2), 257–268.

[29] Grieco, P. L. (2014). Discrete games with flexible information structures: An application to local grocery markets. *The RAND Journal of Economics 45*(2), 303–340.

[30] Hartmann, W. R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing science 29*(4), 585–601.

[31] Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society 47*(1), 153–161.

[32] Henry, M., R. Méango, and M. Queyranne (2015). Combinatorial approach to inference in partially identified incomplete structural models. *Quantitative Economics 6*(2), 499–529.

[33] Holmes, T. J. (2011). The diffusion of wal-mart and economies of density. *Econometrica 79*(1), 253–302.

[34] Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

[35] Jovanovic, B. (1989). Observable implications of models with multiple equilibria. *Econometrica 57*(6), 1431–1437.

[36] Judd, K. L. (1985). Credible spatial preemption. *The RAND Journal of Economics 16*(2), 153–166.

[37] Kline, B. (2014). The empirical content of games with bounded regressors. *working paper*.

[38] Kline, B. (2015). Identification of complete information games. *Journal of Econometrics 189*(1), 117–131.

[39] Kline, B. and E. Tamer (2012). Bounds for best response functions in binary games. *Journal of Econometrics 166*(1), 92–105.

[40] Krauth, B. V. (2006). Simulation-based estimation of peer effects. *Journal of Econometrics 133*(1), 243–271.

[41] Leung, M. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

[42] Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of Economic Studies 60*(3), 531–542.

[43] Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies 74*(3), 897–935.

[44] Narayanan, S. (2013). Bayesian estimation of discrete games of complete information. *Quantitative Marketing and Economics 11*(1), 39–81.

[45] Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics 37*(3), 619–640.

[46] Sheng, S. (2012). Identification and estimation of network formation games. *Working Paper*.

[47] Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

[48] Trogdon, J. G., J. Nonnemaker, and J. Pais (2008). Peer effects in adolescent overweight. *Journal of Health Economics 27*(5), 1388–1399.

[49] Uetake, K. (2012). Strategic network formation and performance in the venture capital industry. *Working Paper*.

[50] van Hasselt, M. (2011). Bayesian inference in a sample selection model. *Journal of Econometrics 165*(2), 221–232.

[51] Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.