The Pennsylvania State University
The Graduate School
Department of Economics

DISCRETE CHOICE MODELS WITH ENDOGENEITY

A Dissertation in
Economics
by
Haiqing Xu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2011

The dissertation of Haiqing Xu was reviewed and approved* by the following:

Joris Pinkse
Professor of Economics
Dissertation Co-Advisor and Co-Chair of Committee

Sung Jae Jun
Assistant Professor of Economics
Dissertation Co-Advisor and Co-Chair of Committee

Herman J. Bierens
Professor of Economics

Runze Li
Professor of Statistics

Barry W. Ickes
Professor of Economics
Acting Head of the Department of Economics

*Signatures are on file in the Graduate School.

ABSTRACT

CHAPTER 1: Discrete Choice Models With Local Interactions: A Game Theoretical Approach

Consider observations from a single equilibrium of a local interaction game in which each player, a firm, has a finite number of actions (discrete choice) and is subject to interactions that are local–for example, the two surrounding neighbors in a 'linear' Hotelling model. Asymptotics in this setting is studied by assuming that all the players are located in a single market and that the number of them grows. All observations are potentially dependent on each other because they are interpreted as arising from a single equilibrium of settings where players interact directly or indirectly. Simple assumptions about the structure are made that ensure that the game with a fixed number of players has a unique equilibrium and the equilibrium has a stability property. The formulation of this stability property is new and is the basis for consistency. I introduce an estimation procedure called (sieve) maximum approximated likelihood. This estimator has the same asymptotic properties as the corresponding maximum likelihood estimator, but is easier to compute.

CHAPTER 2: Estimation of Bayesian Nash Equilibrium in Static Discrete Games with Correlated Private Signals

This paper studies a two by two static game of incomplete information. I allow players' private signals to be correlated, which adds complexity to Bayesian Nash Equilibrium (BNE) solutions of the game. Further, the econometric structure of this model is 'incomplete' due to the existence of multiple equilibria (Tamer (2003)). I therefore focus on a nontrivial subset of the support of public information variables (regressors), where a unique Monotone Strategy Bayesian Equilibrium (MSBE) exists. I propose a four–step procedure to estimate the payoff structure. In the first step I estimate a set of parameters containing the underlying parameter of interest, θ_0. I then obtain a point estimator of θ_0 in the second step and prove its consistency. The third and fourth step estimators are $\sqrt{n}$–consistent for θ_0; the fourth step estimation is more efficient.

CHAPTER 3: Semiparametric estimation of binary decision games of incomplete information with correlated private signals (with Yuanyuan Wan)

This paper studies the identification and estimation of a semiparametric binary decision game of incomplete information. We make no parametric assumptions on the joint distribution of private signals and allow them to be correlated. Focusing on Monotone Strategy Bayesian Equilibrium, we show that the equilibrium strategies can be represented by a binary choice model with an unobserved regressor, of which we find (estimable) upper and lower bounds. We show that the parameters of interest can be point–identified subject to a scale normalization under mild support requirements for the regressors (publicly observed states) and errors (private signals). Following Manski and Tamer (2002), we use the modified maximum score estimator with estimated bounds and show its consistency.

CHAPTER 4: Tighter Bounds in Triangular System (with Sung Jae Jun and Joris Pinkse)

We study a nonparametric triangular system with (potentially discrete) endogenous regressors and nonseparable errors. Like in other work in this area, the parameter of interest is the structural function evaluated at particular values. We impose a global exclusion and exogeneity condition, in contrast to Chesher (2005), but develop a rank condition which is weaker than Chesher's. The alternative rank condition can be satisfied for binary endogenous regressors, and it often leads to a tighter identified interval than Chesher (2005)'s minimum length interval. We illustrate the potential of the new rank condition using the Angrist and Krueger (1991) data.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am deeply indebted to my advisor Joris Pinkse and co–advisor Sung Jae Jun for their invaluable advice, guidance, and encouragement. I can't describe how much I learn from Joris and Sung Jae over the past years. The memorable reading group initiated by them greatly inspired my interest on Econometrics.

I would also like to thank Herman Bierens and Quang Vuong for teaching me Econometrics and helping me develop practical views on the research. I also benefited greatly discussing with Kalyan Chatterjee, Vijay Krishna, Isabelle Perrigne and Neil Wallace on empirical game. I would also like to thank Runze Li, the member of my committee, for his help. I am especially grateful to Halbert White, whose mini course in Penn State inspired me thinking about the link between Econometrics and Decision theory. I am also grateful to my classmate and coauthor, Yuanyuan Wan for the helpful discussions. Insightful suggestions and comments from Edward Coulson, Edward Green, Paul Grieco, and James Tybout are also gratefully acknowledged. The conversations with Han Hong helped me a lot to motivate this research project much better.

I am very indebted to Hongbin Cai, Liutang Gong, Yaping Wang and Li'an Zhou for their attention to help me come to Penn State University. Here, I am also fortunate to have wonderful friends and co-workers: Frank Erickson, Yosuke Igarashi, Herman Li, Nianqing Liu, Sergey Lychagin, Ayse Pehlivan, Lily Samkharadze, Van Anh Vuong, Yuanyuan Wan, and Jun Xiao.

I most grateful to my wife Xiao Meng, my parents and my brother for their support on my Ph.D study. To them this dissertation is dedicated.

Dedication

To my family.

Chapter 1

Discrete Choice Models With Local Interactions: A Game Theoretical Approach

1.1 Introduction

I consider the structural estimation of discrete choice games with Hotelling–type interactions. Most of the recent literature on discrete choice games assumes that the number of players is fixed and that the same game is played repeatedly across markets, e.g. Aguirregabiria and Mira (2007a), Bajari, Hong, Krainer, and Nekipelov (2010), Berry and Tamer (2006), and Pesendorfer and Schmidt-Dengler (2003a). In contrast, I consider the case in which there is a single market and the number of players grows large. Importantly, the interactions among these players are local and are determined by their locations. These Hotelling–type discrete choice models with an increasing number of players are applicable in many areas such as spatial competition and social networks. I first establish the existence of a unique Bayesian Nash Equilibrium and a stability property of the equilibrium. The formulation of this stability property is new and is the basis for the consistency of estimates. Next, I introduce an estimation procedure called (sieve) maximum approximated likelihood (MAL) for the structural parameters in both parametric and semiparametric setups. This estimator has the same asymptotic properties as the corresponding maximum likelihood estimator, but is easier to compute.

The structure of the model is as follows: there are N players with exogenously determined locations, be it in a geographic area, a product space, or a social network. Each player makes a choice from a finite set, $\mathcal{A} = \{0, 1, 2, \cdots, K\}$. Player i's payoff depends on not only her private information, commonly observed characteristics X_i, and her own choice, but also the choices of her neighbors. The players move simultaneously and the equilibrium concept is pure strategy Bayesian Nash Equilibrium (BNE). A BNE gives rise to a conditional distribution over actions induced by the distribution of private information, where the conditioning variables are the commonly observed characteristics of all players. I provide sufficient conditions under which there exists a unique BNE and, crucially, a stability property of the equilibrium obtains. I will discuss the notion of stability later in more detail.

A special case covered by my setup is the Hotelling (1929) model, in which each firm's choice is an output level from the set $\mathcal{A}$. Under each output level, a firm's costs depend on both its private shock and the factor input characteristics X_i. Instead of modeling the demand side, I assume that the firm's profit depends on its costs, its choice of output, and the output choices of its neighbors. In contrast to the existing discrete choice game literature (e.g. entry games), the spirit of these models is that the interactions are local and determined by their locations.

In this chapter, I make assumptions about the interaction strength between players in a structural model, which ensures the uniqueness and stability of the equilibrium. The (spatial) stability of the BNE solution requires that the dependence of each player's equilibrium strategy on the characteristics of other players should vanish as their distance increases. In other words, a player's existence does not have significant impact on players located far away. The first assumption I make is that the number of 'neighbors' (to be defined later) of each player is less than some constant M, which is independent of the number of players. In the one–dimensional Hotelling (1929) model and the Salop (1979) circle model, M can equal two. Second, I impose an upper bound on the amount of influence that a single player can have on her neighbor's payoff function. Similar to autoregression models in time series and in spatial econometrics, interactions that are too strong can make inference difficult to impossible.

Most of the existing empirical game literature, e.g. Aguirregabiria and Mira (2002, 2007a), Bajari et al (2009), Bajari et al (2010), Berry and Tamer (2006), Bresnahan and Reiss (1991a,b), Ellickson and Misra (2008), Pesendorfer and Schmidt-Dengler (2003a), and Seim (2006a), assumes that the number of players is fixed and that the same game is played repeatedly in a sequence of independent local markets, or at different points in time. The number of markets is moreover assumed to grow large. In contrast, I assume that there is only a single market and that the number of players increases. Therefore, I consider a sequence of games indexed by the number of players.

There are many potential applications of my model. For instance, a large number of firms can be engaged in discrete marketing decisions, e.g. whether and how to advertise, or whether and how to offer promotions, in one competitive environment. Further examples can be found in the social network literature, e.g. peer effects of smoking. However, most of the existing social network literature assumes that each player's payoff is not affected by her friends' current choices but the choices in the previous stage (see, e.g., Maxwell (2002) and Ennett and Bauman (1993)), which effectively assumes away the presence of 'endogenous interactions' (see Manski (2000)).

To estimate this discrete choice game of incomplete information, I develop a new estimation method instead of using the two–step method pioneered by Aguirregabiria and Mira (2002) and used in much of the literature on such games. Because the equilibrium conditional choice distribution functions depend on the X variables of all players, the two–step method would entail estimating a nonparametric function whose domain increases with the number of players, which is unlikely to work well.

In contrast, my MAL estimator does not have this problem. My method starts with the loglikelihood function, instead of a pseudo loglikelihood. To obtain the likelihood, one needs an expression of the equilibrium conditional distribution of each player's action. However, obtaining those distribution functions is costly, because it involves solving a BNE in an N–player game. Alternatively, I approximate the equilibrium conditional distribution of actions for each player using the equilibrium strategy in her 'h–neighborhood' game, where she and all rivals that are within distance h interacts strategically with each other. Obtaining a solution to this smaller–sized game is computationally less expensive. In addition, I choose h to grow with N at a polynomial rate, and show that my MAL estimator behaves asymptotically as well as the efficient maximum likelihood estimator.

Although the asymptotic analysis in my paper is close to the spatial econometrics literature, my approach differs from that literature in two ways. First, I derive the spatial dependence of players' discrete choices from a structural model in which dependence is caused by the strategic effects instead of imposing a predefined dependence pattern (examples can be found in Robinson (2008)). As a guide for empirical work, the existing game theory literature, e.g. spatial competition or network economics, can be applied to model the endogenous interactions, and by using a structural approach one can assess the effect of a counterfactual change in the spatial interactions.

Second, the dependent variable in my model is discrete. In the existing spatial econometrics literature, there are few papers dealing with discrete choice models; there are exceptions including Pinkse and Slade (1998, 2007), and Klier and McMillen (2008). In general, the discrete choice variables are mutually dependent under a nonlinear pattern derived from the model solution. In my paper, because the belief of rivals' decisions enters a player's decision rule in a nonlinear way, a closed–form solution of the equilibrium choice probabilities does not obtain. Instead, the solution is implicitly defined by a nonlinear equation system in which the number of equations is proportional to N.

Moreover, while Manski (1993, 2000), Brock and Durlauf (2001a,b) and others adopt game–theoretic models for analyzing social network data, my primary methodology is different. Most social interaction models assume that an individual's choice and payoff are affected by the average behavior of the group that she belongs to. All individuals in the same group are equally affected. In contrast, I specify a particular interaction pattern that depends on players' locations, in which the interactions are local and the strategic effect between any two individuals depends on their network distance. My spatial interaction pattern allows me to establish the spatial stability of the equilibrium solution, which provides the basis for my MAL estimator.

The organization of the paper is as follows. In the next section, I specify a static discrete choice game of incomplete information. The solution concept adopted is standard: the Bayesian Nash Equilibrium. In Section 1.3, I establish the identification in a semiparametric setup. In Section 4, I propose the MAL estimation approach in both parametric and semiparametric setups. Asymptotics properties of MAL estimators are established. Section 5 discusses an extension of the MAL estimator when there are missing data.

1.2 The Model

1.2.1 Game structure

I consider a simultaneous–move game of incomplete information. There are N players indexed by $i \in \mathcal{I} = \{1, \cdots, N\}$, with exogenously determined locations. Each player i

simultaneously chooses her action $Y_i \in \mathcal{A} = \{0, 1, 2, ..., K\}$.[1]

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

$$u_{ik} = \beta_0(X_i, k) + \sum_{j \neq i} \alpha_0(k, Y_j) g_{ij} + \epsilon_i(k) \tag{1.1}$$

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

As is standard for discrete choice models, only the differences of the choice–specific utility functions matter to decision–makers. It is therefore necessary to impose some normalization. Without loss of generality, let $\beta_0(x, 0) = \alpha_0(0, \ell) = 0$ for all $x \in \mathcal{X}$ and $\ell \in \mathcal{A}$. Let further $\beta_0 = (\beta_0(\cdot, 1), \beta_0(\cdot, 2), \cdots, \beta_0(\cdot, K))^T$ and $\alpha_0 = \left(\alpha_{01}^T, \alpha_{02}^T, \cdots, \alpha_{0K}^T\right)^T \in \mathbb{R}^{K(K+1)}$, where $\alpha_{0k} = (\alpha_0(k, 0), \cdots, \alpha_0(k, K))^T \in \mathbb{R}^{K+1}$ for $k \in \mathcal{A} \backslash \{0\}$. Hence, α_0 is a finite–dimensional parameter and β_0 is a vector of functions. Further, let $\theta_0 = (\alpha_0^T, \beta_0^T)^T \in \Theta = \mathbb{A} \times \mathcal{B}$, where $\mathbb{A} \subseteq \mathbb{R}^{K(K+1)}$ and $\mathcal{B}$ is a vector–valued function space.

Let $W_i = (X_i^T, G_i^T)^T$ and $S_N = (W_1^T, \cdots, W_N^T)^T$. Let further $\mathscr{S}_N$ and $\mathbb{R}^{K+1}$ be the support of S_N and ϵ_i respectively. Given the structural parameter value θ_0, a strategy for player i is a function $r_i(\epsilon_i, S_N; \theta_0)$ which maps from player i's private information ϵ_i and the public signal S_N to a discrete choice $Y_i = r_i(\epsilon_i, S_N; \theta_0)$. In BNE, each player's strategy maximizes her (conditional) expected utility given all the information available to her. Let $\{r_i^*\}_{i=1}^N$ be a BNE strategy profile. Thus player i's equilibrium strategy satisfies

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

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

1.2.2 Equilibrium characterization

I first make an assumption on the distribution of private signals.

Assumption 1 *The private shocks $\epsilon_i(k)$ are distributed i.i.d. across both actions and players. Moreover, $\epsilon_i(k)$ has an extreme value distribution with density*

$$f(t) = \exp(-t)\exp[-\exp(-t)].$$

Let $\sigma^*_{ik}(S_N;\theta_0) = \mathbb{P}[r^*_i(\epsilon_i, S_N;\theta_0) = k|S_N]$ be the conditional choice probability of player i choosing k in equilibrium. Note that $\mathbb{P}\{r^*_j(\epsilon_j, S_N;\theta_0) = \ell|\epsilon_i, S_N\} = \sigma^*_{j\ell}(S_N;\theta_0)$ for $j \neq i$ when ϵ_i and ϵ_j are independent of each other. Hence, under Assumption 1, obtaining an expression for $\sigma^*_{ik}(S_N;\theta_0)$ in terms of X_i, G_i and $\sigma^*_{j\ell}(S_N;\theta_0)$ ($j \neq i$, $\ell \in \mathcal{A}$) is straightforward. Now I arrive at the following lemma

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

$$\sigma^*_{ik}(S_N;\theta_0) = \frac{\exp\left[\beta_0(X_i,k) + \sum_{l=0}^K \left\{\alpha_0(k,\ell)\sum_{j\neq i} g_{ij}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}{1 + \sum_{q=1}^K \exp\left[\beta_0(X_i,q) + \sum_{l=0}^K \left\{\alpha_0(q,\ell)\sum_{j\neq i} g_{ij}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}. \tag{1.2}$$

proof: See Appendix 1.6. □

It is routine to apply Brouwer's fixed point theorem to prove the existence of a solution in (1.2), which therefore gives me the existence of a BNE solution in this game. Note that multiple solutions to equation (1.2) could exist and each of them would correspond to a BNE in the game. Later, I will impose further conditions to obtain a unique solution.

1.2.3 Neighbors

I assume that all players are exogenously located in a geographical area, a product space, or a social network. The locations are public information. Observables g_{ij} depend on players' locations. In this chapter, I take $g_{ij} = 1$ if j is i's 'neighbor' and zero otherwise.[2] Note that $g_{ii} = 0$ by definition, as mentioned at the beginning of this section. A similar setup can

[2] The concept of neighbor can be formally defined in many ways, for instance, two players are neighbors if they are two firms sharing a market boundary, or if they are friends in a social network.

be found in the spatial econometrics literature (see e.g. Case (1991) and Pinkse and Slade (1998)). Although restricting g_{ij} to be binary is a limitation, using this makes my model more tractable. In principle, It is possible to allow for g_{ij}'s that depend on the distance between i and j; I intend to pursue this possibility in future work.

Under the above specification of g_{ij}, equation (1.2) can be simplified as follows. Let $\mathcal{N}_i = \{j \in \mathcal{I} : g_{ij} = 1\}$ be the collection of player i's neighbors. For all $i \in \mathcal{I}$ and all $k \in \mathcal{A}$,

$$\sigma^*_{ik}(S_N;\theta_0) = \frac{\exp\left[\beta_0(X_i,k) + \sum_{\ell=0}^{K}\left\{\alpha_0(k,\ell)\sum_{j\in\mathcal{N}_i}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}{1+\sum_{q=1}^{K}\exp\left[\beta_0(X_i,q) + \sum_{\ell=0}^{K}\left\{\alpha_0(q,\ell)\sum_{j\in\mathcal{N}_i}\sigma^*_{j\ell}(S_N;\theta_0)\right\}\right]}. \tag{1.3}$$

Hereafter, I focus on equation (1.3) instead of (1.2) to discuss the equilibrium solution.

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

1.2.4 Interaction strength

To establish my results, I need to restrict the interaction strength between players in my structural model, which ensures that the BNE solution is unique and has a stability property. Similar to other spatial models, with too much interaction, there could be multiple equilibria, or adding a new player to the existing game could cause the equilibrium to change radically (see Pinkse and Slade (2010)). To ensure that a player's existence does not have significant impact on players located far away, I now make the following primitive assumptions.

Assumption 2 *There exists a constant $M \in \mathbb{N}$, which is independent of N, such that*

$$\max_{i\in\mathcal{I}} C(\mathcal{N}_i) \le M$$

with probability one, where $C(\mathcal{N}_i)$ is the cardinality of the set $\mathcal{N}_i$.

Assumption 2 restricts the number of neighbors of any player to be less than a constant M. In the one–dimensional Hotelling (1929) model and the Salop (1979) circle model, M can equal two. Note that M does not depend on N, which restricts the formation of players' neighbor–relationships in a growing spatial structure.

Assumption 3 $\lambda_0 = \max_{k,\ell,m\in\mathcal{A}} |\alpha_0(k,\ell) - \alpha_0(m,\ell)| \times \frac{MK}{K+1} < 1$.

Given the number of choices $(K+1)$ and the upper–bound of the number of neighbors (M), Assumption 3 restricts the scale of the strategic effect coefficient differences. In the Salop (1979) circle model, if players compete with each other in terms of choosing an output level and $\alpha_0(k,\ell) \leq \alpha_0(k',\ell') \leq 0$ for all $k' \leq k$ and $\ell' \leq \ell$, then Assumption 3 means that $\lambda_0 = -\alpha_0(K,K) \times \frac{MK}{K+1} < 1$. Assumption 3 plays a similar role as the requirement in autoregression models that all roots lie inside the unit circle. It should be noted that Assumption 1 has already implicitly imposed a normalization restriction on the scale of α_0 since the standard error of the private signals has been assumed to be one.

Under Assumptions 2 and 3, I establish the properties of the BNE in the next lemma. For any integer $h \geq 0$, let $\mathcal{N}_{(i,h)}$ be the h–neighborhood of i, which is defined inductively,

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

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

$$S_N^{(i,h)} = \left(\{X_j\}_{j \in \mathcal{N}_{(i,h)}}, \{g_{nj}\}_{n,j \in \mathcal{N}_{(i,h)}} \right). \tag{1.4}$$

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

$$\sigma_{jk}^{(i,h)}(S_N^{(i,h)};\theta_0) = \frac{\exp\left[\beta_0(X_j,k) + \sum_{\ell=0}^K \left\{ \alpha_0(k,\ell) \sum_{n \in \mathcal{N}_j \cap \mathcal{N}_{(i,h)}} \sigma_{n\ell}^{(i,h)}(S_N^{(i,h)};\theta_0) \right\} \right]}{1 + \sum_{q=1}^K \exp\left[\beta_0(X_j,q) + \sum_{\ell=0}^K \left\{ \alpha_0(q,\ell) \sum_{n \in \mathcal{N}_j \cap \mathcal{N}_{(i,h)}} \sigma_{n\ell}^{(i,h)}(S_N^{(i,h)};\theta_0) \right\} \right]}. \tag{1.5}$$

By definition, the solution of equation (1.5) can be viewed as a BNE solution to a smaller–sized game than the original one. Players outside i's h–neighborhood are not taken into account.

Lemma 2 *Suppose that Assumptions 1 through 3 hold. Then, for any $N \in \mathbb{N}$ and any realization $s \in \mathscr{S}_N$ of S_N, (i) there exists a unique BNE; (ii) the equilibrium satisfies*

$$\max_{i \in \mathcal{I}} \left\| \sigma_i^*(s;\theta_0) - \sigma_i^{(i,h)}(s^{(i,h)};\theta_0) \right\| \leq 2 \times \lambda_0^h \to 0, \tag{1.6}$$

as $h \to \infty$.

proof: See Appendix 1.6 $\square$

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

1.3 Identification

I now discuss identification in the sense of Hurwicz (1950) and Koopmans and Reiersol (1950). I provide conditions which allow me to show that there is a unique $\theta_0 \in \Theta$ which rationalizes the distribution of observables.

I take a constructive approach for identification. For the given $\mathbb{P}_{Y_1,\cdots,Y_N;S_N}$, I provide explicit formulas for both α_0 and β_0. My identification results are related to those of Bajari, Hong, Krainer, and Nekipelov (2010).

Note that $\sigma^*_{ik}(S_N;\theta_0) = \mathbb{P}(Y_i = k|S_N)$ when the equilibrium is unique under Assumptions 1 through 3 (see Lemma 2). Let $\Delta_{ik}(S_N) = \ln \mathbb{P}(Y_i = k|S_N) - \ln \mathbb{P}(Y_i = 0|S_N)$ for $k \in \mathcal{A}$. Both $\sigma^*_{ik}(S_N;\theta_0)$ and $\Delta_{ik}(S_N)$ are functions of $\mathbb{P}_{Y_1,\cdots,Y_N|S_N}$, and hence they are identified. Note also that equation (1.3) gives me

$$\Delta_{ik}(S_N) = \beta_0(X_i,k) + \sum_{\ell=0}^{K} \left\{ \alpha_0(k,\ell) \sum_{j \in \mathcal{N}_i} \mathbb{P}(Y_j = \ell|S_N) \right\}, \;\; i \in \mathcal{I},\; k \in \mathcal{A}. \tag{1.7}$$

I will derive from equation (1.7) an expression for α_0 and β_0 , respectively, in terms of $\{\mathbb{P}_{Y_j|S_N}\}_{j \in \mathcal{N}_i}$ and $\{\Delta_{ik}(S_N)\}_{k \in \mathcal{A}}$.

Recall that S_N consists of all public information in the game. In equation (1.7), if I hold X_i constant and vary $\sum_{j \in \mathcal{N}_i} \mathbb{P}(Y_j = \ell|S_N)$ by changing S_N, then α_0 can be identified if an additional rank condition is satisfied. Essentially, equation (1.7) can, for the purpose of identification, be thought of as a partial linear model (see Robinson (1988)) and the rank condition in Assumption 4 originates in that literature.

Assumption 4 *For $i = 1 \in \mathcal{I}$ and $k \in \mathcal{A}$, let $\phi_{i\ell}(S_N) = \sum_{j \in \mathcal{N}_i} \mathbb{P}(Y_j = \ell | S_N)$ and $\Phi_i(S_N) = (\phi_{i0}(S_N), \cdots, \phi_{iK}(S_N))^T$. Assume*

$$\left| \det \left(\mathbb{E}\left[\operatorname{Var}\left\{ \Phi_i(S_N) | X_i \right\} \right] \right) \right| > 0.^3 \tag{1.8}$$

Assumption 4 is not primitive, but testable if the choice probabilities can be consistently estimated. Assumption 4 fails to hold when the number of neighbors is the same for all i as in e.g. the Salop (1979) model. Indeed, then $(\phi_{i0}, \cdots, \phi_{iK})$ would be collinear, since $\sum_{k=0}^{K} \phi_{ik}$ would be a constant. In this case, a further normalization of the coefficients will be necessary, e.g. $\alpha_0(k, 0) = 0$ for all $k \in \mathcal{A}$. On the other hand, if $\mathbb{E}\left[\operatorname{Var}\left\{ C(\mathcal{N}_i) | X_i \right\} \right] > 0$, then $\alpha_0(k, 0)$ can be identified using the variation in $C(\mathcal{N}_i)$ while X_i is held fixed.

Under Assumption 4, I obtain an expression for (α_0, β_0) from equation (1.7), which gives me the identification results. Before that, I introduce some notation. Let $U_i(S_N) = \Phi_i(S_N) - \mathbb{E}\left\{ \Phi_i(S_N) | X_i \right\}$ and $V_{ik}(S_N) = \Delta_{ik}(S_N) - \mathbb{E}\left\{ \Delta_{ik}(S_N) | X_i \right\}$ for $k \in \mathcal{A}$.

Theorem 1 *Suppose that Assumptions 1 through 4 hold. Then the structural parameter θ_0 is identified, i.e. $\bar{\theta} \neq \theta_0 \Longrightarrow \mathbb{P}_{Y_1, \cdots, Y_I; S_N}(\bar{\theta}) \neq \mathbb{P}_{Y_1, \cdots, Y_I; S_N}(\theta_0)$. Moreover, for all $k \in \mathcal{A}$ and $x \in \mathcal{X}$,*

$$\alpha_{0k} = \left[\mathbb{E}\left\{ U_i(S_N) U_i^T(S_N) \right\} \right]^{-1} \mathbb{E}\left\{ U_i(S_N) V_{ik}(S_N) \right\}, \tag{1.9}$$

$$\beta_0(x, k) = \mathbb{E}\left\{ \Delta_{ik}(S_N) | X_i = x \right\} - \sum_{\ell=0}^{K} \alpha_0(k, \ell) \mathbb{E}\left\{ \phi_{i\ell}(S_N) | X_i = x \right\}. \tag{1.10}$$

proof: ee Appendix 1.6. □

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

Assumption 5 (Rank Condition)

$$\liminf_{N \to \infty} \left| \det \left(\mathbb{E}\left[\operatorname{Var}\left\{ \Phi_i(S_N) | X_i \right\} \right] \right) \right| > 0,$$

Theorem 2 *Suppose that Assumptions 1 through 3, and 5 hold, then θ_0 is identified when N is sufficiently large.*

proof: ee Appendix 1.6. □

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

Analogous identification conditions can be formulated in fully parametrized models, and such conditions are more straightforward than those used in Theorem 2. If one assumes that $\beta_0(X_i, k) = X_i^T \beta_{0k}$ for $\beta_{0k} \in \mathbb{R}^p$ and $k \in \mathcal{A}\backslash\{0\}$. Let $\overline{X}_{Ni} = \left(\Phi_i^T(S_N); X_i^T\right)^T$. Then the rank condition is directly derived from equation (1.7).

Assumption 6 (Rank Condition for linear–index setup)

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

Replace Assumption 5 with 6 in Theorem 2, then the identification of the coefficients α_0 and $(\beta_{01}, \cdots, \beta_{0K})$ is straightforward for the sufficiently large N. Moreover, for all $k \in \mathcal{A}\backslash\{0\}$,

$$\left(\alpha_0(k,0), \cdots, \alpha_0(k,K), \beta_{0k}^T\right)^T = \left[\mathbb{E}\left\{\overline{X}_{Ni}\overline{X}_{Ni}^T\right\}\right]^{-1} \mathbb{E}\left\{\overline{X}_{Ni}\Delta_{ik}(S_N)\right\}. \tag{1.11}$$

1.4 Estimation

In this section, I discuss estimation of the structural parameters. The identification arguments suggest an estimator of θ_0 based on nonparametrically estimated $\sigma_{ik}^*(S_N; \theta_0)$'s. However, I do not pursue this approach because the asymptotics involve a sequence of equilibria, in which the domain of the equilibrium choice probability functions increases with N. Consequently, one would be estimating a nonparametric function whose domain is increasing with the sample size.

In contrast, my MAL estimation procedure solves $\sigma_{ik}^*(S_N; \theta)$ using the best response equation system (1.3). I will first illustrate it in a parametric setup and then in a semiparametric framework.

From hereon, I add a subscript N to σ_i^* and $\sigma_i^{(i,h)}$ to emphasize the fact that the equilibrium solution depends on the number of players. As in the classical multinomial logit model, the likelihood function depends on the equilibrium choice probability of each action of each player. Formally, the likelihood function is

$$p_N(S_N) \prod_{i=1}^{N} \prod_{k=0}^{K} \left\{\sigma_{Nik}^*(S_N; \theta)\right\}^{\mathbf{1}(Y_i = k)}, \tag{1.12}$$

where p_N is the density function of S_N. Because σ_{Nik}^* does not have an analytic expression and its numerical calculation becomes costly as N increases, I use an approximation $\sigma_{Nik}^{(i,h)}$ instead, where h is an integer that depends on N and will be specified later.

In this section, for the sake of notational simplicity, I focus on the case where all players' choices are observed. An extension to the situation of missing observations is discussed in Section 5.

1.4.1 A Parametric Approach

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

For every N, θ_0 maximizes the expectation of the following loglikelihood function:

$$\mathbb{E}\left\{\sum_{k=0}^{K} 1\left(Y_1 = k\right) \ln \sigma_{N1k}^*(S_N;\theta)\right\},$$

which depends on N. In the proof of Theorem 3, I show that θ_0 also maximizes

$$L_0(\theta) = \lim_{N\to\infty} \mathbb{E}\left[\sum_{k=0}^{K} 1\left(Y_1 = k\right) \ln \sigma_{N1k}^*(S_N;\theta)\right]. \tag{1.13}$$

I define my MAL estimator $\hat{\theta}$ as the maximizer of

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

where $S_N^{(i,h)}$ and $\sigma_{Ni}^{(i,h)}$ are defined in equations (1.4) and (1.5).

If $h = 0$, then $\mathcal{N}_{(i,0)} = \{i\}$ and $\mathcal{N}_i \cap \mathcal{N}_{(i,0)} = \emptyset$, equation (1.5) becomes

$$\sigma_{ik}^{(i,0)}(S_N^{(i,0)},\theta) = \frac{\exp\left(X_i^T\beta_k\right)}{\sum_{q=0}^{K}\exp\left(X_i^T\beta_q\right)}.$$

which is the choice probability in the classical multinomial logit model, where there are no interactions. In the presence of spatial interactions, the consistency of the estimator requires that h increase with N. In fact, I will have h increase at a polynomial rate, i.e. $h = h_0 \times [N^\omega]$ for some constant $h_0 \in \mathbb{N}$ and $\omega > 0$, where $[a]$, for arbitrary $a \in \mathbb{R}$, is the largest integer which is no larger than a.

The following assumptions are needed for the consistency of $\hat{\theta}$.

Assumption 7 *$\mathcal{X}$ is bounded.*

Assumption 8 *Θ is compact.*

Assumption 9 *There exists a $\lambda < 1$ such that*

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

Assumption 7 ensures that choice probabilities are bounded away from zero so that the likelihood function is bounded. Note that Assumption 7 is more than necessary and could be relaxed (see Bierens (1996a), Theorem 4.2.1 for more detail). Assumption 8 is standard. Assumption 9 strengthens Assumption 3.

Theorem 3 *Suppose that Assumptions 1, 2, and 6 through 9 hold. Then, $\hat{\theta} \xrightarrow{p} \theta_0$.*

proof: See Appendix 1.7.1. □

Now that the consistency of $\hat{\theta}$ is established, I discuss its asymptotic normality. First, I introduce further notation. Let $Z_{Ni} = (Y_i, S_N)$ and $f_{Ni}(Z_{Ni}, \theta) = \prod_{k=0}^{K} \sigma^*_{Nik}(S_N; \theta)^{\mathbf{1}(Y_i = k)}$. Moreover, let $J_N(S_N, \theta_0) = N^{-1} \sum_{i=1}^{N} J_{Ni}(S_N, \theta_0)$, where

$$J_{Ni}(S_N, \theta_0) = \mathbb{E}\left\{ \frac{\partial}{\partial \theta} \ln f_{Ni}(Z_{Ni}, \theta_0) \frac{\partial}{\partial \theta^T} \ln f_{Ni}(Z_{Ni}, \theta_0) \Big| S_N \right\}.$$

Assumption 10 *θ_0 belongs to the interior of Θ.*

Assumption 10 is standard in the literature.

Theorem 4 *Suppose that Assumptions 1, 2, 6 through 10 hold. Then*

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, J^{-1}(\theta_0)\right),$$

where $J(\theta_0) = \lim_{N \to \infty} \mathbb{E} J_N(S_N, \theta_0)$.

proof: See Appendix 1.7.1.

$J(\theta_0)$ is the Fisher information matrix when the number of players goes to infinity. Theorem 4 implies that the MAL estimator behaves asymptotically as well as the maximum likelihood estimator. Note that, $J(\theta_0)$ can be estimated by

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

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

1.4.2 A Semiparametric Approach

I now consider a semiparametric setup in which the payoff function β_0 is an unknown element of an infinite–dimensional function space $\mathcal{B}$. Below I propose a sieve estimator which I show to be consistent and asymptotically normal.

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

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

Assumption 11 *For $q\in\mathbb{N}$, $K_0\in\mathbb{R}_+$ and $q+m>1/2$,*

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

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

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

$$\mathcal{B}_N=\Big\{\beta:\ \beta_k(x)=a_{k0}+\sum_{j=1}^{r_N}\left\{a_{kj}\cos(2\pi jx)+b_{kj}\sin(2\pi jx)\right\},$$

$$a_{k0}^2 + \sum_{j=1}^{r_N} j^{2(q+m)} \left(a_{kj}^2 + b_{kj}^2\right) \leq A_0 \ln N;\ a_{kj}, b_{kj} \in \mathbb{R} \text{ and } k \in \mathcal{A} \backslash \{0\} \Big\},$$

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

Assumption 12 *$\mathbb{A}$ is compact.*

Assumption 13 $r_N = O\left(N^{\frac{1}{2q+2m+1}}\right)$.

Under Assumption 13, it is known that $\rho(\pi_N \theta, \theta) = O\left(N^{-\frac{q+m}{2q+2m+1}}\right)$ (see Lorentz (1966)), where $\pi_N \theta$ is the projection of θ on the sieve space Θ_N. Assumption 12 restricts the space of parametric part to be compact in this semiparametric setup. In contrast, compactness is not required for the space of the nonparametric part.

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

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

where $\tilde{\sigma}_{Ni}^{(i,h)}(S_N; \theta)$ is defined in equation (1.5).

Theorem 5 *Suppose Assumptions 1, 2, 5, 9, and 10 through 13 hold, then*

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

proof: See Appendix 1.7.2 □

To derive the limiting distribution of a random process, I need to define the directional derivative in the functional space. Let $m(\cdot) : \mathbb{R}^{d_Z} \times \Theta \to \mathbb{R}$, where $d_Z \in \mathbb{N}_+$. Then, for any $\nu \in \Theta - \{\theta\}$,[4] let further

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

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

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

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

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

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

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

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

In the parametric case, where θ_0 is a finite–dimensional vector, $\xi(\nu) = \nu \times J^{-1}(\theta_0)$.

Theorem 6 *Suppose Assumptions 1, 2, 5, 9, and 10 through 13 hold. Let $\nu \in \Theta - \{\theta_0\}$. Then*

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

proof: See Appendix 1.7.2. □

When $\nu \in (\mathbb{A} - \alpha_0) \times \mathbf{0}_{\mathcal{B}}$, Theorem 6 provides the limiting distribution of $\tilde{\alpha}$ in sieve estimates.

1.5 Missing Observations

With spatial dependence, missing data cause difficulties to obtain consistent estimates, even if the data are missing for exogenous reasons (Pinkse and Slade (2010)). In my model, the presence of missing data makes it infeasible to obtain a BNE solution to the whole game. However, a player's h–neighborhood equilibrium strategy is still available, provided that all players within distance h are observed. Thus, an MAL estimator can be constructed using a trimming procedure, which leaves out observations without a complete set of data in their h–neighborhood. I now obtain asymptotic properties for the trimmed MAL estimator in my parametric setup.

Let $\mathcal{I} = \{1, 2, \cdots, I\}$ be the set of players and $\mathscr{O}_N$ be the collection of observed players in the sample, where $N = C(\mathscr{O}_N)$ is the sample size. Let further $\mathscr{O}_N^h = \{i \in \mathcal{N} : \mathcal{N}_{(i,h)} \subseteq \mathscr{O}_N\}$. Then, my objective function is

$$\widehat{Q}_N^h(\theta) = \sum_{i=1}^{N} \left[\mathbf{1}\left(i \in \mathscr{O}_N^h\right) \times \sum_{k=0}^{K} \left\{ \mathbf{1}\left(Y_i = k\right) \ln \sigma_{Nik}^{(i,h)}\left(S_N; \theta\right) \right\} \right].$$

Let $\hat{\theta}_m = \text{argmax}_{\theta\in\Theta}\, \widehat{Q}_N^h(\theta)$ be the trimmed MAL estimator.

Assumption 14 *The data are missing for exogenous reasons, i.e., $\mathscr{O}_N \perp (\epsilon_1, \cdots, \epsilon_N; S_N)$ for all $i \in \mathcal{I}$. Moreover,*

$$\sum_{i=1}^{N} \boldsymbol{1}\left(i \in \mathscr{O}_N^h\right) = O_p(N).$$

Assumption 14 requires the unavailable data should be missing for exogenous reasons (see Lee (2007)), and the proportion of the data trimmed out should be at most a fraction (< 1) of the sample.

Theorem 7 *Suppose that Assumptions 1, 2, 6 through 10, and 14 hold. Then there is*

$$\sqrt{N}\left(\widehat{\theta}_m - \theta_0\right) = O_p(1).$$

In addition,

$$\sqrt{\zeta_N}(\hat{\theta}_m - \theta_0) \xrightarrow{d} \mathscr{N}\left(0, J^{-1}(\theta_0)\right),$$

where $\zeta_N = \sum_{i=1}^N \boldsymbol{1}\left(i \in \mathscr{O}_N^h\right)$.

proof: See Appendix 1.7.2.

1.6 Appendix A

To begin with, I first introduce some notation for the discussion of proofs. Let

$$\bar{\Delta}(\alpha) = \max_{k,\ell,m\in\mathcal{A}} \left|\alpha(k,\ell) - \alpha(m,\ell)\right|.$$

As convention, $\|\cdot\|_p$ refers to the p–norm in a real vector space $\mathbb{R}^d$. In particular, $\|x\|_1 = \sum_{t=1}^{d} |x_t|$ for all $x \in \mathbb{R}^d$ and $\|\cdot\|_2$ is the usual Euclidean norm.

Let $\Sigma_N = \{\sigma_i\}_{i=1}^N$ denote an arbitrary choice probability profile, and $\Sigma_N^*(S_N;\theta) = \left\{\sigma_i^*\left(S_N;\theta\right)\right\}_{i=1}^N$ for any value of $\theta \in \Theta$. For any $i \in \mathcal{I}$ and $k \in \mathcal{A}$, let

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

Let $\Gamma_i = (\Gamma_{i0},\cdots,\Gamma_{iK})^T$ and $\Gamma(S_N,\Sigma_N;\theta) = \{\Gamma_i(W_i,\Sigma_N;\theta)\}_{i=1}^N$. Thus, equation system (1.3) could be written as

$$\Sigma_N^*(S_N;\theta_0) = \Gamma\left(S_N,\Sigma_N^*(S_N;\theta_0);\theta_0\right).$$

Moreover, Let $D(S_N,\Sigma_N;\theta)$ be an $N(K+1)$ by $N(K+1)$ matrix, such that the element at $(i-1)(K+1)+k+1$–th row and $(j-1)(K+1)+\ell+1$–th column is $\frac{\partial}{\partial\sigma_{j\ell}}\Gamma_{ik}(S_N,\Sigma_N;\theta)$ for all $i,j \in \mathcal{I}$ and $k,\ell \in \mathcal{A}$. A useful result is given by Lemma 1.10.1, that for any $\mu \in \mathbb{R}^{N(K+1)}$, there is

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

where $\mathbf{1}_{N(K+1)}$ refers to the $N(K+1)$ by $N(K+1)$ identity matrix.

In the proofs, I will suppress the arguments of Σ_N for notational convenience when it does not cause any confusion, i.e. $D(S_N,\Sigma_N^*;\theta) = D\left(S_N,\Sigma_N^*(S_N,\theta);\theta\right)$.

Proof of Lemma 1

By definition, the choice probability profile $\{\sigma_i^*(S_N;\theta_0)\}_{i=1}^N$ is derived from the equilibrium strategy profile $\{r_i^*(\cdot,S_N;\theta_0)\}_{i=1}^N$. So it suffices to show that $\{r_i^*(\cdot,S_N;\theta_0)\}_{i=1}^N$ can also be induced from $\{\sigma_i^*(S_N;\theta_0)\}_{i=1}^N$. To see this, note that equation (1.1) can also be written as

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

□

Proof of Lemma 2

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

$$\max_{j \in \mathcal{N}_{(i,h-q)}} \left\| \sigma_j^*(s;\theta_0) - \sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1 \leq 2\lambda_0^q. \tag{1.14}$$

Lemma 1.6 implies that (1.14) holds for $q = 1$. Moreover, if for any $q \leq q_0 \in \{1, \cdots, h-1\}$, (1.14) is satisfied. Then, I need to show that (1.14) also holds for $q = q_0 + 1$. Lemma 1.6 implies that for any $j \in \mathcal{N}_{(i,h-q)}$, there is

$$\left\| \sigma_j^*(s;\theta_0) - \sigma_j^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1 \leq \lambda_0 \times \max_{n \in \mathcal{N}_j} \left\| \sigma_n^*(s;\theta_0) - \sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1. \tag{1.15}$$

Given $q = q_0 + 1$, $n \in \mathcal{N}_j$ and $j \in \mathcal{N}_{(i,h-q)}$ implies that $n \in \mathcal{N}_{(i,h-q_0)}$. Thus, for any $j \in \mathcal{N}_{(i,h-q)}$,

$$\max_{n \in \mathcal{N}_j} \left\| \sigma_n^*(s;\theta_0) - \sigma_n^{(i,h)}\left(s^{(i,h)};\theta_0\right) \right\|_1 \leq 2\lambda_0^{q_0}. \tag{1.16}$$

Thus, equations (1.15) and (1.16) imply that (1.14) also holds for $q = q_0 + 1$. Under Assumption 3, $2\lambda_0^h \downarrow 0$ as $h \to \infty$.

I then establish (i) by contradiction. Suppose that there are two equilibria $\{\sigma_n^*(s;\theta_0)\}_{n=1}^N$ and $\{\overline{\sigma}_n(s;\theta_0)\}_{n=1}^N$ for some $N \in \mathbb{N}$ and $s \in \mathscr{S}_N$. Using a similar argument as that in Lemma 1.6, for any $i \in \mathcal{I}$

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

Hence,

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

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

Proof of Theorem 1

By definition,

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

Under Assumption 4,

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

Hence α_{0k} is identified. Therefore, $\beta_0(\cdot, k)$ is also identified by

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

□

Proof of Theorem 2

Note that when N is sufficient large, there is

$$\left|\det\left(\mathbb{E}\left[\,\mathrm{Var}\left\{\Phi_i(S_N)|X_i\right\}\,\right]\right)\right| > 0.$$

Then the identification results follow Theorem 1. □

Lemma 1.6

Suppose that Assumptions 1 through 3 hold. Thus, for any $h, N \in \mathbb{N}$, $s \in \mathscr{S}_N$, $i \in \mathcal{I}$ and $j \in \mathcal{N}_{(i,h-1)}$,

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

In particular

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

proof: First, fix arbitrary $h, N \in \mathbb{N}$ and $s \in \mathscr{S}_N$. For any $i \in \mathcal{I}$ and $j \in \mathcal{N}_{(i,h-1)}$,

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

where $\Sigma_N^\dagger$ is between $\Sigma_N^*(s;\theta_0)$ and $\Sigma_N^{(i,h)}\left(s^{(i,h)};\theta_0\right)$.

By definition of $\Gamma_{jk}(w_j, \Sigma_N;\theta)$, for any $n \in \mathcal{N}_j$

$$\frac{\partial \ln \Gamma_{jk}}{\partial\sigma_{n\ell}} = \alpha_0(k,\ell) - \sum_{q=0}^{K}\left\{\Gamma_{jq}\alpha_0(q,\ell)\right\},$$

which gives me $\partial\Gamma_{jk}/\partial\sigma_{n\ell} = \Gamma_{jk}\sum_{q\neq k}\left[\Gamma_{jq}\left\{\alpha_0(k,\ell) - \alpha_0(q,\ell)\right\}\right]$. By definition, (*i*) $0 \leq \Gamma_{jk} \leq 1$; (*ii*) $\sum_{k=0}^{K}\Gamma_{jk} = 1$. Thus

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

The last step comes from the fact that $\sum_{k=0}^{K}\left[\Gamma_{ik}\left(1-\Gamma_{ik}\right)\right] \leq K/(K+1)$ for any Γ_i satisfying (*i*) and (*ii*).

Hence,

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

In particular, for any $n\in\mathcal{N}_j$

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

□

1.7 Appendix B

In this section, I provide the proofs for asymptotics analysis. As mentioned in Section 1.4.1, for the sake of notation clarification, I add subscript N to $\sigma_i^*(S_N;\theta)$ and $\sigma_i^{(i,h)}(S_N^{(i,h)};\theta)$. Let $\widehat{L}_N(\theta)=\frac{1}{N}\sum_{i=1}^{N}\ln f_{Ni}(Z_{Ni},\theta)$, $L_N(\theta)=\mathbb{E}\ln f_{N1}(Z_{N1},\theta)$ and $L_0(\theta)=\limsup_{N\to\infty}L_N(\theta)$. Moreover, let $f_{Ni}^h(Z_{Ni},\theta)=\prod_{k=0}^{K}\sigma_{Nik}^{(i,h)}(S_N^{(i,h)};\theta)^{\mathbf{1}(Y_i=k)}$ and $\widehat{L}_N^h(\theta)=\frac{1}{N}\sum_{i=1}^{N}\ln f_{Ni}^h(Z_{Ni},\theta)$. Let further $\widehat{G}_N(\theta)=\partial\widehat{L}_N(\theta)/\partial\theta$ and $\widehat{G}_N^h(\theta)=\partial\widehat{L}_N^h(\theta)/\partial\theta$. Let $S_\infty=(W_1,\cdots,W_\infty)$.

1.7.1 Lemma 1.7.1

Assume (*i*) $L_0(\theta)$ is uniquely maximized at θ_0; (*ii*) Θ is compact; (*iii*) $L_0(\theta)$ is continuous in θ; (*iv*) $\widehat{L}_N(\theta)$ converges uniformly in probability to $L_N(\theta)$, i.e.

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

(*v*) $L_N(\theta)$ converges uniformly to $L_0(\theta)$, i.e.

$$\sup_{\theta\in\Theta}\left|L_N(\theta)-L_0(\theta)\right|\to 0;$$

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

proof: ote that from (*iv*) and (*v*), there is

$$\sup_{\theta\in\Theta}\left|\widehat{L}_N(\theta)-L_0(\theta)\right|\xrightarrow{p}0.$$

Then the rest proofs basically follow the proof in Newey and McFadden (1994), Theorem 2.1. Note that condition (*vi*) is not a trivial statement here, since by definition, $\hat{\theta}$ is a maximizer of $\widehat{L}_N^h$, instead of $\widehat{L}_N$.□

Proof of Theorem 3

In this proof, I'll check the conditions in Lemma 1.7.1. First, there is $\left|\ln\sigma^*_{N1k}(S_N;\theta)\right|\le-\ln\underline{\sigma}$ by Lemma 1.8.1, then

$$\mathbb{E}\left|\ln f_{N1}(Z_{Ni},\theta)\right|\le\sum_{k=0}^{K}\mathbb{E}\left|\ln\sigma^*_{N1k}(S_N;\theta)\right|<\infty.$$

Then by Dominant Convergence Theorem and the fact $\sigma^*_{N1}(S_N;\theta)$ uniformly converges to a random process $\sigma_1(\theta)$ almost surely (see Lemma 1.8.2),

$$L_0(\theta)=\limsup_{N\to\infty}\mathbb{E}\left\{\ln f_{N1}(Z_{Ni},\theta)\right\}=\mathbb{E}\left[\sum_{k=0}^{K}\left\{\sigma_{1k}(\theta_0)\ln\sigma_{1k}(\theta)\right\}\right]<\infty.$$

By the identification argument, θ_0 is the unique maximizer of $L_0(\cdot)$, i.e. if both θ_0 and θ_1 maximize $L_0(\theta)$, then the difference of choice probabilities under these two parameter values could be arbitrary small when N is sufficient large, which contradicts the identification condition.

Second, the compactness of Θ is assumed in Assumption 8 and the continuity of $L_0(\theta)$ is given by Lemma 1.8.3.

Third, I will show $\widehat{L}_N(\hat{\theta})\ge\sup_{\theta\in\Theta}\widehat{L}_N(\theta)-o_p(1)$. By Lemma 1.8.4

$$\begin{aligned}\widehat{L}_N(\hat{\theta})&=\widehat{L}_N(\hat{\theta})-\widehat{L}_N^h(\hat{\theta})+\sup_{\theta}\widehat{L}_N^h(\theta)\\&\ge\widehat{L}_N(\hat{\theta})-\widehat{L}_N^h(\hat{\theta})+\sup_{\theta\in\Theta}\widehat{L}_N(\theta)-\sup_{\theta}\left\{\widehat{L}_N(\theta)-\widehat{L}_N^h(\theta)\right\}=\sup_{\theta\in\Theta}\widehat{L}_N(\theta)-o_p(1).\end{aligned}$$

Next, I show condition (*v*) in Lemma 1.7.1. By Lemma 1.8.2, there is

$$\mathbb{E}\sup_{\theta\in\Theta}\left|\sigma^*_{N1k}(S_N;\theta)-\sigma_{1k}(\theta)\right|\to 0.$$

Then

$$\mathbb{E}\sup_{\theta\in\Theta}\left|\ln\sigma^*_{N1k}(S_N;\theta)-\ln\sigma_{1k}(\theta)\right|=\mathbb{E}\sup_{\theta\in\Theta}\left|\frac{1}{\sigma^{\dagger}_{N1k}}\left\{\sigma^*_{N1k}(S_N;\theta)-\sigma_{1k}(\theta)\right\}\right|$$

$$\leq \frac{1}{\underline{\sigma}} \mathbb{E} \sup_{\theta \in \Theta} |\sigma^*_{N1k}(S_N;\theta) - \sigma_{1k}(\theta)| \to 0,$$

where $\sigma^{\dagger}_{Nik}$ is between $\sigma^*_{Nik}(S_N;\theta)$ and $\sigma_{1k}(\theta)$. Thus,

$$\mathbb{E} \sup_{\theta \in \Theta} |L_N(\theta) - L_0(\theta)| \leq \sum_{k=0}^{K} \mathbb{E} \sup_{\theta \in \Theta} |\ln \sigma^*_{N1k}(S_N;\theta) - \ln \sigma_{1k}(\theta)| \to 0.$$

To follow the argument in 1.7.1, it suffices to show the uniform convergence of $\widehat{L}_N(\theta)$ to $L_N(\theta)$, i.e.

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

By Lemma 1.8.1 and 1.8.5, $\ln f_{Ni}(Z_{Ni},\theta)$ is a bounded continuous function in θ. Since Θ is compact, then $\mathcal{F}_N = \{\ln f_{N1}(Z_{N1},\theta) : \theta \in \Theta\}$ can be covered by a finite number of ϵ–brackets. To apply the classical Glivenko-Cantelli argument, it suffices to show the pointwise LLN, i.e. for any $\theta \in \Theta$

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

Because

$$\begin{aligned}
\mathbb{E}\left\{\widehat{L}_N(\theta) - L_N(\theta)\right\}^2 &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left[\ln f_{Ni}(Z_{Ni},\theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2 \\
&= \frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\ln f_{Ni}(Z_{Ni},\theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2 \middle| S_N\right\}\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\ln f_{Ni}(Z_{Ni},\theta) - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}\right]\right)^2 \middle| S_N\right\}\right] \\
&\quad + \frac{1}{N^2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left[\ln f_{Ni}(Z_{Ni},\theta)|S_N\right] - \mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2 \middle| S_N\right\}\right]. \quad (1.17)
\end{aligned}$$

Note that I suppress a zero term in RHS of (1.17). Conditional on S_N, $\{Y_i\}_{i=1}^N$ is independent among each other. Then $\{f_{Ni}(Z_{Ni},\theta)\}_{i=1}^N$ is also conditionally independent, so

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

By Lemma 1.8.1, $\ln f_{Ni}(\cdot,\theta)$ is a bounded function uniformly in N, i and θ. Thus

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

Moreover, by Lemma 1.8.6

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

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

Proof of Theorem 4

proof: First, because Lemma 1.9.1 and the fact $\widehat{G}_N^h(\hat{\theta})=0$, there is $\widehat{G}_N(\hat{\theta})=o_p(1/\sqrt{N})$. It follows that

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

for some $\theta^{\ddagger}$ between θ_0 and $\hat{\theta}$.

Next, I am going to prove: (*i*) $\sqrt{N}\times\widehat{G}_N(\theta_0)\xrightarrow{d}N(0,J(\theta_0))$; (*ii*) $-\partial\widehat{G}_N(\theta^{\ddagger})/\partial\theta^T\xrightarrow{p}J(\theta_0)$ for $\theta^{\ddagger}\xrightarrow{p}\theta_0$.

First, I prove (*i*). Let $\varphi_{Ni}=\frac{\partial}{\partial\theta}\ln f_{Ni}(Z_{Ni},\theta_0)$. Because for any N and i, θ_0 maximizes the smooth function $\mathbb{E}\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\}$ almost surely, then $\mathbb{E}(\varphi_{Ni}|S_\infty)=0$. Since φ_{Ni} is conditionally independent across i. Then

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

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

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

Moreover, conditional on S_∞, $\|\varphi_{Ni}\|_1$ is bounded almost surely by Lemma 1.9.2. Hence, by the

Lindeberg-Feller Theorem (see Van der Vaart (2000), page 20),

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

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

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

where $\mathbf{1}_L$ is the L by L identity matrix. Since $N^{-1} J_N(S_N, \theta_0) \xrightarrow{p} J(\theta_0)$ by Lemma 1.7.2. Thus

$$N^{-1/2} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \ln f_{Ni}(Z_{Ni}, \theta_0) \to \mathcal{N}(0, J(\theta_0)).$$

Next, I prove (ii). First, I show that for arbitrary $m, m' = 1, \cdots, L$,

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta_m \partial \theta_{m'}} \ln f_{Ni}(Z_{Ni}, \theta^{\ddagger}) - V_{Nmm'}(\theta_0) \xrightarrow{p} 0$$

where $V_{Nmm'}(\theta) = \mathbb{E}\left\{\frac{\partial^2}{\partial \theta_m \partial \theta_{m'}} \ln f_{N1}(Z_{N1}, \theta)\right\}$.

By Lemma 1.9.3, there is $\left|\partial^2 \ln f_{Ni}(Z_{Ni}, \theta)/\partial \theta_m \partial \theta_{m'}\right| < \delta_3$ almost surely for some constant $\delta_3 \in \mathbb{R}_+$, which does not depend on N, i, m, m'. Then, pointwise LLN can be proved in a similar way as in Theorem 3, i.e.

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta_m \partial \theta_{m'}} \ln f_{Ni}(Z_{Ni}, \theta) - V_{Nmm'}(\theta) \xrightarrow{p} 0.$$

Hence, it suffices to show the UNLL. By Lemma 1.8.5 and 1.9.3, $\partial^2 \ln f_{Ni}(Z_{Ni}, \theta)/\partial \theta_m \partial \theta_{m'}$ is continuous and uniformly bounded (in θ), hence by a similar argument as that in Theorem 3

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta_m \partial \theta_{m'}} \ln f_{Ni}(Z_{Ni}, \theta^{\ddagger}) - V_{Nmm'}(\theta_0) \xrightarrow{p} 0,$$

which implies that

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta \partial \theta^T} \ln f_{Ni}(Z_{Ni}, \theta^{\ddagger}) - V_N(\theta_0) \xrightarrow{p} 0.$$

where $V_N(\theta) = \mathbb{E}\left\{\frac{\partial^2}{\partial \theta \partial \theta^T} \ln f_{N1}(Z_{N1}, \theta)\right\}$. By Lemma 1.7.2, $V_N(\theta_0) \to -J(\theta_0)$.

Under (i) and (ii), there is

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, J^{-1}(\theta_0)\right).$$

□

1.7.2 Lemma 1.7.2

Suppose that Assumptions 1, 2, 6 through 10 hold. Let $V_N(\theta)$ be defined as that in the proof of Theorem 1.7.1. Then there exists an L by L matrix $J(\theta_0)$, such that

$$J_N(S_N;\theta_0) \xrightarrow{p} J(\theta_0),$$

and

$$V_N(\theta_0) \to -J(\theta_0).$$

proof: By symmetry in the identity of players,

$$\begin{aligned}\mathbb{E}J_N(S_N;\theta_0) &= N^{-1}\mathbb{E}\left[\sum_{i=1}^{N}\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_{Ni}(Z_{Ni},\theta_0)\times\frac{\partial}{\partial\theta^T}\ln f_{Ni}(Z_{Ni},\theta_0)\Big|S_N\right\}\right]\\ &= \mathbb{E}\left[\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_{N1}(Z_{N1};\theta_0)\times\frac{\partial}{\partial\theta^T}\ln f_{N1}(Z_{N1};\theta_0)\Big|S_N\right\}\right]\\ &= \sum_{k=0}^{K}\sum_{\ell=0}^{K}\mathbb{E}\left\{\frac{\partial\sigma^*_{N1k}(S_N;\theta_0)}{\partial\theta}\times\frac{\partial\sigma^*_{N1\ell}(S_N;\theta_0)}{\partial\theta^T}\right\}.\end{aligned}$$

First, I will show that for arbitrary $m, m' = 1, \cdots, L$, the term

$$\mathbb{E}\left\{\frac{\partial\sigma^*_{N1k}(S_N;\theta_0)}{\partial\theta_m}\times\frac{\partial\sigma^*_{N1\ell}(S_N;\theta_0)}{\partial\theta_{m'}}\right\}$$

converges.

Under Lemma 1.10.2, I use a similar argument as Lemma 1.8.2, $\{\partial\sigma^*_{N1k}(S_N;\theta_0)/\partial\theta_m : N \in \mathbb{N}\}$ is a Cauchy sequence almost surely. Because $\partial\sigma^*_{N1k}(S_N;\theta_0)/\partial\theta_m$ is bounded for all k and m, then $\{\partial\sigma^*_{N1k}(S_N;\theta_0)/\partial\theta_m \times \partial\sigma^*_{N1\ell}(S_N;\theta_0)/\partial\theta_{m'} : N \in \mathbb{N}\}$ is also a Cauchy sequence almost surely. Then the existence of $J(\theta_0)$ obtains by the convergence property of Cauchy sequence in a complete space. Then, let

$$J(\theta_0) = \lim_{N\to\infty}\mathbb{E}J_N(S_N;\theta_0).$$

Now it suffices to show that $J_N(S_N;\theta_0) \xrightarrow{p} J(\theta_0)$, which is shown in Lemma 1.8.7.

For the second part, by Lemma 1.9.3,

$$\begin{aligned}V_N(\theta_0) &= \mathbb{E}\left\{\partial^2\ln f_{N1}(Z_{N1},\theta_0)/\partial\theta\partial\theta^T\right\}\\ &= -\mathbb{E}\left\{\frac{\partial\ln f_{N1}(Z_{N1},\theta_0)}{\partial\theta}\times\frac{\partial\ln f_{N1}(Z_{N1},\theta_0)}{\partial\theta^T}\right\}.\end{aligned}$$

Because

$$\mathbb{E}\left\{\frac{\partial \ln f_{N1}(Z_{N1},\theta_0)}{\partial\theta} \times \frac{\partial \ln f_{N1}(Z_{N1},\theta_0)}{\partial\theta^T}\right\}$$
$$= \mathbb{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln f_{Ni}(Z_{Ni},\theta_0)}{\partial\theta} \times \frac{\partial \ln f_{Ni}(Z_{Ni},\theta_0)}{\partial\theta^T}\right\},$$

and under the conditional independence argument similarly as that in the proof of Theorem 3,

$$\mathbb{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln f_{Ni}(Z_{Ni},\theta_0)}{\partial\theta} \times \frac{\partial \ln f_{Ni}(Z_{Ni},\theta_0)}{\partial\theta^T}\right\} = \mathbb{E}J_N(S_N,\theta_0) + o(1).$$

Hence,

$$V_N(\theta_0) = -\mathbb{E}J_N(S_N,\theta_0) + o_p(1) \to -J(\theta_0).$$

□

Proof of Theorem 5

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

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

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

The proof of convergence rate follows Shen and Wong (1994), Theorem 1. The conditions C1 and C2 in the theorem can be verified by choosing $\alpha = \beta = 1$ (in their notation), similarly as in Example 2, Shen and Wong (1994). For condition C3, similarly as in Example 3, pick $r_0 = \frac{1}{2(2q+2m+1)}$ and

$r = 0^+$. Thus by their Theorem 1, the converge rate is $N^{-\frac{q+m}{2q+2m+1}}$. Note that the extra $\ln n$ factor in their Theorem 1 for the case $r = r_0^+$ can be removed when the criterion difference is continuous in the Remark 4, which is exactly the case here. □

Proof of Theorem 6

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

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

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

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

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

Proof of Theorem 7

proof: Let $\widehat{Q}_N(\theta) = \frac{1}{\zeta_N} \sum_{i=1}^{N} \mathbf{1}(i \in \mathcal{O}_N) \ln f_{Ni}(Z_{Ni}, \theta)$, $Q_N(\theta) = \mathbb{E}\left\{\mathbf{1}(i \in \mathcal{O}_N) \ln f_{N1}(Z_{N1}, \theta)\right\}$ and $Q_0(\theta) = \limsup_{N\to\infty} Q_N(\theta)$. Moreover, $\widehat{Q}^h_N(\theta) = \frac{1}{\zeta_N} \sum_{i=1}^{N} \mathbf{1}(i \in \mathcal{O}_N) \ln f^h_{Ni}(Z_{Ni}, \theta)$.

Note that, under Assumption 14, $\mathbb{P}_{Y_1,\cdots,Y_I|S_I} = \mathbb{P}_{Y_1,\cdots,Y_I|S_I;\mathcal{O}_N}$. Thus, a similar argument could be established similarly as that in Theorem 3 and Theorem 4, by replacing $\left\{\widehat{L}_N(\theta), L_N(\theta), L_0(\theta), \widehat{L}^h_N(\theta)\right\}$ with $\left\{\widehat{Q}_N(\theta), Q_N(\theta), Q_0(\theta), \widehat{Q}^h_N(\theta)\right\}$. □

1.8 Appendix C

1.8.1 Lemma 1.8.1

Suppose Assumptions 1, 2, 7 and 8 hold, then

$$\inf_N \min_{i\in\mathcal{I}} \min_{k\in\mathcal{A}} \inf_{\theta\in\Theta} |\sigma^*_{Nik}(S_N; \theta)| \ge \underline{\sigma} \quad a.s.$$

for some $0 < \underline{\sigma} < 1$.

proof: Note that for arbitrary $N \in \mathbb{N}$, $i \in \mathcal{I}$, $k \in \mathcal{A}$ and $\theta \in \Theta$,

$$\sigma^*_{Nik}(S_N; \theta) = \Gamma_{ik}(W_i, \Sigma^*_N(S_N; \theta); \theta).$$

By definition, $\Gamma_{ik}(x, \Sigma_N; \theta)$ is a continuous and strictly positive function for any $x \in \mathbb{R}^p$, choice probability profile Σ_N, and $\theta \in \Theta$. Since $\mathcal{X}$ is bounded, $0 \leq \sum_{j \in \mathcal{N}_i} \sigma^*_{j\ell}(S_N; \theta) \leq M$ for all i and $\ell \in \mathcal{A}$, and θ is in a compact space Θ. Then $\Gamma_{ik}(x, \Sigma_N, \theta)$ is greater than some constant $0 < C_k < 1$, which is independent of i. Hence $\Gamma_{ik}(W_i, \Sigma^*_N(S_N; \theta); \theta) \geq C_k > 0$. Take $\underline{\sigma} = \min_k C_k$. Note that N, i, k and θ are arbitrary, thus

$$\inf_N \min_i \min_k \inf_{\theta \in \Theta} \Gamma_{ik}(W_i, \Sigma^*_N(S_N; \theta), \theta) \geq \underline{\sigma}.$$

Hence,

$$\inf_N \min_i \min_k \inf_{\theta \in \Theta} \sigma^*_{Nik}(S_N; \theta) \geq \underline{\sigma}.$$

□

1.8.2 Lemma 1.8.2

Suppose that Assumptions 1, 2 and 9 hold. Hence, for any $i \in \mathbb{N}$, there exists a random process $\sigma_i(\theta)$ indexed by θ such that

$$\lim_{N \to \infty} \sup_{\theta \in \Theta} \|\sigma^*_{Ni}(S_N; \theta) - \sigma_i(\theta)\|_1 = 0, \quad a.s.$$

proof: W.O.G.L, take $i = 1$. Note that by definition, $\sigma^*_{N1k}(S_N; \theta) = \sigma^*_{N1k}(W_1, \cdots, W_N; \theta)$. Consider an arbitrary realization of the sequence $\{w_1, \cdots, w_\infty\}$ and let $s_N = (w_1, \cdots, w_N)$. Under Assumption 2, given any $h \in \mathbb{N}$, there exists an $N_h \in \mathbb{N}$ such that for any $N_1, N_2 \geq N_h$, $s^{(1,h)}_{N_1} = s^{(1,h)}_{N_2}$ in two structures. By definition, $\sigma^{(1,h)}_{N_1 1k}\left(s^{(1,h)}_{N_1}; \theta\right) = \sigma^{(1,h)}_{N_2 1k}\left(s^{(1,h)}_{N_2}; \theta\right)$. Thus,

$$\begin{aligned}
&\sup_{\theta \in \Theta} \left|\sigma^*_{N_1 1k}(s_{N_1}; \theta) - \sigma^*_{N_2 1k}(s_{N_2}; \theta)\right| \\
&\qquad \leq \sup_{\theta \in \Theta} \left|\sigma^*_{N_1 1k}(s_{N_1}; \theta) - \sigma^{(1,h)}_{N_1 1k}\left(s^{(1,h)}_{N_1}; \theta\right)\right| \\
&\qquad\qquad + \sup_{\theta \in \Theta} \left|\sigma^*_{N_2 1k}(s_{N_2}; \theta) - \sigma^{(1,h)}_{N_2 1k}\left(s^{(1,h)}_{N_2}; \theta\right)\right| \leq 4\lambda^h.
\end{aligned} \tag{1.18}$$

The last step comes from Lemma 2. Since $\lambda < 1$ and h is arbitrary integer, hence $4\lambda^h$ can be arbitrarily small, which implies that $\{\sigma^*_{N1k}(s_N; \theta)\}_{N=1}^{\infty}$ is a uniform Cauchy sequence in $\mathbb{R}$, and it converges to a continuous function $\sigma_{ik}(\theta)$. Note that $\{w_1, \cdots, w_\infty\}$ is also arbitrarily chosen, then the convergence is almost surely. □

1.8.3 Lemma 1.8.3

Suppose that Assumptions 1 through 3 and 7 hold. Then $L_0(\theta)$ is continuous.

proof: Because

$$\left|L_0(\theta') - L_0(\theta)\right| = \left|\lim_{N \to \infty} \mathbb{E}\left\{\ln f_{N1}(Z_{N1}, \theta') - \ln f_{N1}(Z_{N1}, \theta)\right\}\right|$$

$$\leq \lim_{N\to\infty} \sum_{k=0}^{K} \mathbb{E} \left| \ln \sigma^*_{N1k}(S_N;\theta') - \ln \sigma^*_{N1k}(S_N;\theta) \right|$$
$$\leq \lim_{N\to\infty} \sum_{k=0}^{K} \mathbb{E} \left\{ \left\| \frac{\partial}{\partial\theta} \ln \sigma^*_{N1k}(S_N;\theta^\dagger) \right\|_1 \right\} \times \left\|\theta' - \theta\right\|_1 \leq \delta_1 \|\theta' - \theta\|_1,$$

for some constant $\delta_1 \in \mathbb{R}_+$ which does not dependent on N and θ. The last step comes from Lemma 1.9.2. □

1.8.4 Lemma 1.8.4

Suppose that Assumptions 1, 2, and 8 through 9 hold, then

$$\sup_\theta \left| \widehat{L}^h_N(\theta) - \widehat{L}_N(\theta) \right| = o_p(1).$$

proof: Because

$$\sup_\theta \left| \widehat{L}^h_N(\theta) - \widehat{L}_N(\theta) \right| \leq \sup_\theta \frac{1}{N} \sum_{i=1}^{N} \left| \ln f^h_{Ni}(Z_{Ni},\theta) - \ln f_{Ni}(Z_{Ni},\theta) \right|$$
$$\leq \sup_\theta \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{K} \left| \ln \sigma^{(i,h)}_{Nik}(S^{(i,h)}_N;\theta) - \ln \sigma^*_{Nik}(S_N;\theta) \right|$$
$$\leq \sup_\theta \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{K} \left(\frac{1}{\sigma^\dagger_{Nik}} \right) \times \left| \sigma^{(i,h)}_{Nik}(S^{(i,h)}_N;\theta) - \sigma^*_{Nik}(S_N;\theta) \right|$$
$$\leq \sup_\theta \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{K} \frac{1}{\underline{\sigma}} \left| \sigma^{(i,h)}_{Nik}(S^{(i,h)}_N;\theta) - \sigma^*_{Nik}(S_N;\theta) \right| \leq \frac{2(K+1)\lambda^h}{\underline{\sigma}},$$

where the last step comes from Lemma 2. By the choice of h and the fact that $\lambda < 1$,

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

□

1.8.5 Lemma 1.8.5

Suppose that Assumptions 1 through 3 hold. Hence, for all $N \in \mathbb{N}$, $i \in \mathcal{I}$ and $z \in \mathcal{A} \times \mathscr{S}_N$, $f_{Ni}(z,\cdot) \in \mathcal{C}^\infty(\Theta)$.

proof: By the definition of f_{Ni}, it is sufficient to show that for all $N \in \mathbb{N}$, $i \in \mathcal{I}$, $k \in \mathcal{A}$ and $s \in \mathscr{S}_N$, $\sigma^*_{Nik}(s;\theta) \in \mathcal{C}^\infty(\Theta)$. Note that $\{\sigma^*_{Ni}(s;\theta)\}_{i=1}^N$ is a solution in equation system (1.3) where $S_N = s$, and the solution is unique by Lemma 2. Because $\Gamma_i(x_i,\Sigma_N;\theta) \in \mathcal{C}^\infty\left(\mathbb{R}^{N(K+1)} \times \Theta; \mathbb{R}^{K+1}\right)$, then by implicit function theorem $\Sigma^*_N(s;\theta) \in \mathcal{C}^\infty(\Theta)$ for all N and s. □

1.8.6 Lemma 1.8.6

Suppose that Assumptions 1, 2, 7, 8 and 9 hold, then

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

proof: Given arbitrary $s\in\mathscr{S}_N$ and $h_0\in\mathbb{N}$. Then by Lemma 2, there is

$$\sup_\theta\left|\sigma^*_{Nik}(s;\theta)-\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta)\right|\leq 2\lambda^{h_0},$$

which implies that

$$\sup_\theta\left|\sum_{k=0}^{K}\sigma^*_{Nik}(s;\theta_0)\ln\sigma^*_{Nik}(s;\theta)-\sum_{k=0}^{K}\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta_0)\ln\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta)\right|$$
$$\leq(K+1)\times(-\ln\underline{\sigma})\times 2\lambda^{h_0}+\frac{(K+1)}{\underline{\sigma}}\times 2\lambda^{h_0}.$$

Thus,

$$\begin{aligned}
&\sup_\theta\left|\mathbb{E}\left\{f_{Ni}(Z_{ni}),\theta)|S_N\right\}-\mathbb{E}\left\{f_{Ni}(Z_{ni}),\theta)|S_N^{(i,h_0)}\right\}\right|\\
&=\sup_\theta\left|\sum_{k=0}^{K}\sigma^*_{Nik}(S_N;\theta_0)\ln\sigma^*_{Nik}(S_N;\theta)-\int\sum_{k=0}^{K}\sigma^*_{Nik}(S_N;\theta_0)\ln\sigma^*_{Nik}(S_N;\theta)d\mathbb{P}_{S_N|S_N^{(i,h_0)}}\right|\\
&\leq\sup_\theta\left|\sum_{k=0}^{K}\sigma^*_{Nik}(s;\theta_0)\ln\sigma^*_{Nik}(s;\theta)-\sum_{k=0}^{K}\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta_0)\ln\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta)\right|\\
&+\sup_\theta\int\left|\sum_{k=0}^{K}\sigma^*_{Nik}(s;\theta_0)\ln\sigma^*_{Nik}(s;\theta)-\sum_{k=0}^{K}\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta_0)\ln\sigma^{(i,h_0)}_{Nik}(s^{(i,h_0)};\theta)\right|\mathbb{P}_{S_N|S_N^{(i,h_0)}}\\
&=O_p\left(\lambda^{h_0}\right).
\end{aligned}\tag{1.19}$$

Because

$$\begin{aligned}
&\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2\right\}\\
&=\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h_0)}\right\}\right]\right)^2\right\}\\
&+\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h_0)}\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2\right\}.
\end{aligned}$$

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

By equation (1.19), and by Lemma 1.8.1, $\ln f_{Ni}(\cdot,\theta)$ is a bounded function uniformly in N, i and θ. Then,

$$\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h_0)}\right\}\right]\right)^2\right\}=O\left(N^2\lambda^{2h_0}\right),$$

and

$$\begin{aligned}&\mathbb{E}\left\{\left(\sum_{i=1}^{N}\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h_0)}\right\}-\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)\right\}\right]\right)^2\right\}\\&\quad=\sum_{i=1}\sum_{j\in\mathcal{N}_{(i,h)}}^{N}Cov\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h_0)}\right\},\mathbb{E}\left\{\ln f_{Nj}(Z_{Nj},\theta)|S_N^{(i,h_0)}\right\}\right]\\&\qquad+\sum_{i=1}^{N}Var\left[\mathbb{E}\left\{\ln f_{Ni}(Z_{Ni},\theta)|S_N^{(i,h_0)}\right\}\right]=O\left(NM^{h_0}\right)+O(N)\end{aligned}$$

Choose $h_0=\frac{b\ln N}{\ln M}$ for some $b\in(0,1)$. Then $h_0\to\infty$ as $N\to\infty$ and $M^{h_0}=o(N)$. Note that h_0 is not necessary to be h and the existence of such an h_0 is guaranteed by Assumption 2.

Thus

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

□

1.8.7 Lemma 1.8.7

Suppose that Assumptions 1, 2, 7, 8 and 9 hold, then

$$\mathbb{E}\left\{J_N(S_N;\theta_0)-\mathbb{E}J_N(S_N;\theta_0)\right\}^2\to 0.$$

proof: Because

$$\begin{aligned}&\mathbb{E}\left\{J_N(S_N;\theta_0)-\mathbb{E}J_N(S_N;\theta_0)\right\}^2\\&\quad=\mathbb{E}\left[J_N(S_N;\theta_0)-\mathbb{E}\left\{J_N(S_N;\theta_0)\middle|S_N^{(i,h_0')}\right\}\right]^2+\mathbb{E}\left[\mathbb{E}\left\{J_N(S_N;\theta_0)\middle|S_N^{(i,h_0')}\right\}-\mathbb{E}J_N(S_N;\theta_0)\right]^2\end{aligned}$$

for some $h_0'\in\mathbb{N}$. By Lemma 1.10.2 and use a similar argument as that in Lemma 1.8.6, there are

$$\mathbb{E}\left[J_N(S_N;\theta_0)-\mathbb{E}\left\{J_N(S_N;\theta_0)\middle|S_N^{(i,h_0')}\right\}\right]^2=O\left(h_0'^2\lambda^{2h_0'}\right)\to 0,$$

and

$$\mathbb{E}\left[\mathbb{E}\left\{J_N(S_N;\theta_0)|S_N^{(i,h_0')}\right\} - \mathbb{E}J_N(S_N;\theta_0)\right]^2 = O\left(N^{-1}M^{h_0'}\right) + O\left(N^{-1}\right).$$

Similarly as in Lemma 1.8.6, take $h_0' = h_0$. Then,

$$\mathbb{E}\left[\mathbb{E}\left\{J_N(S_N;\theta_0)|S_N^{(i,h_0')}\right\} - \mathbb{E}J_N(S_N;\theta_0)\right]^2 \to 0.$$

1.9 Appendix D

1.9.1 Lemma 1.9.1

Suppose that Assumptions 1 through 3, 7 and 8 hold.

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

proof: First, because Lemma 1.10.2,

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

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

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

□

1.9.2 Lemma 1.9.2

Suppose that Assumptions 1, 2, 7 8 and 9 hold. Then there exists a $C_1 \in \mathbb{R}_+$, such that

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

and

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

proof: Before I begin with, it should noted that without uniform over N, the results of this lemma can be easily established. Since $\ln\sigma_{Ni}(S_N;\theta)$ is infinitely smooth in both S_N and θ by a similar argument as in Lemma 1.8.5. Under Assumption 8 and 7, it is straightforward to see

$$\max_{i\in\mathcal{I}}\sup_{\theta\in\Theta}\left\|\frac{\partial\sigma^*_{Ni}(S_N;\theta)}{\partial\theta}\right\|_1 < \infty.$$

However, it does not imply a uniform bound over $N \in \mathbb{N}$.

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

$$\frac{\partial\sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m} = \frac{\partial\Gamma_{ik}(W_i,\Sigma^*_N(S_N,\theta);\theta)}{\partial\theta_m} + \sum_{j\in\mathcal{N}_i}\sum_{\ell=0}^{K}\left\{\frac{\partial\Gamma_{ik}(W_i,\Sigma^*_N(S_N;\theta);\theta)}{\partial\sigma^*_{j\ell}} \times \frac{\partial\sigma^*_{Nj\ell}(S_N;\theta)}{\partial\theta_m}\right\}. \quad (1.20)$$

Let $\chi(S_N,\theta)$ be an $N(K+1)$ dimensional vector, such that for all $i \in \mathcal{I}$ and $k \in \mathcal{A}$, there is $\chi_{(i-1)(K+1)+k+1} = \partial\sigma^*_{ik}(S_N;\theta)/\partial\theta_m$. Similarly I define $\tau(S_N,\theta)$ using $\partial\Gamma_{ik}/\partial\theta_m$ ($i \in \mathcal{I}$ and $k \in \mathcal{A}$). Thus, equation (1.20) becomes

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

Thus

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

and

$$\frac{\partial\sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m} = \iota'_{ik} \times \left\{\mathbf{1}_{N(K+1)} - D(S_N,\Sigma^*_N;\theta)\right\}^{-1} \times \tau(S_N,\theta)$$

where ι_{ik} is an $N(K+1)$ dimensional vector with value one only at the $[(i-1)(K+1)+k+1]$–th component and zero elsewhere.

Then, by Lemma 1.10.1

$$\frac{\partial\sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m} = \iota'_{ik} \times \left\{\sum_{t=0}^{\infty} D^t(S_N,\Sigma^*_N;\theta)\right\} \times \tau(S_N,\theta).$$

Hence

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

where the last step comes from Holder inequality, Lemma 1.8.1 and Lemma 1.10.2. Thus

$$\sum_{k=0}^{K}\left|\frac{\partial\ln\sigma^*_{Nik}(S_N;\theta)}{\partial\theta_m}\right| \le \frac{\delta_2(K+1)}{1-\lambda}. \quad (1.21)$$

Moreover, by Lemma 1.8.1,

$$\left|\frac{\partial \ln \sigma^*_{Nik}(S_N;\theta)}{\partial \theta_m}\right| \leq \frac{1}{\underline{\sigma}}\left|\frac{\partial \sigma^*_{Nik}(S_N;\theta)}{\partial \theta_m}\right| \leq \frac{\delta_2(K+1)}{\underline{\sigma}(1-\lambda)}. \tag{1.22}$$

To complete the proof, just note that the RHS's of (1.21) and (1.22) do not depend on N, i or θ. □

1.9.3 Lemma 1.9.3

Suppose that Assumptions 1, 2 , 7, 8 and 9 hold. Then there exists a $\delta_4 \in \mathbb{R}_+$, such that

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

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

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

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

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

which can also be written as

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

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

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

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

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

Now I am going to show

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

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

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

By Lemma 1.9.2 and Assumption 7,

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

Similarly, it can be shown that

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

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

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

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

$$\frac{\partial^2 \Gamma_{ik}(W_i, \Sigma_N^*(S_N;\theta), \theta)}{\partial\sigma_{j\ell}\partial\sigma_{j',\ell'}} = \sigma^*_{Nik} \left(\sum_{q\ne k} \left[\sigma^*_{Niq} \left\{ \alpha(k,\ell) - \alpha(q,\ell) \right\} \right] \right)^2$$

$$+\sigma^*_{Nik}\sum_{q\neq k}\left[\left(\sigma^*_{Niq}\sum_{q'\neq q}\left[\sigma^*_{Niq'}\left\{\alpha(q,\ell)-\alpha(q',\ell)\right\}\right]\right)\left\{\alpha(k,\ell)-\alpha(q,\ell)\right\}\right].$$

Thus

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

It follows that

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

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

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

Therefore,

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

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

1.10 Appendix E

1.10.1 Lemma 1.10.1

Suppose Assumptions 1, 2 and 9 hold, then for any choice probability profile Σ_N, $s\in\mathscr{S}_N$, and arbitrary real vector $\mu\in\mathbb{R}^{N(K+1)}$,

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

and

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

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

proof: By the definition of $D(s, \Sigma_N; \theta)$,

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

and from the proof in Lemma 1.6,

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

Thus

$$\begin{aligned} \|D(s, \Sigma_N; \theta)\mu\|_1 &\le \bar{\Delta}(\alpha) \sum_{i \in \mathcal{I}} \sum_{k=0}^{K} \sum_{j \in \mathcal{N}_i} \left\{ \left| \Gamma_{ik} \sum_{q \ne k} \Gamma_{iq} \right| \times \sum_{\ell=0}^{K} \left| \mu_{(j-1)(K+1)+\ell+1} \right| \right\} \\ &\le \frac{\bar{\Delta}(\alpha) K}{K+1} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i} \sum_{\ell=0}^{K} \left| \mu_{(j-1)(K+1)+\ell+1} \right| \le \frac{\bar{\Delta}(\alpha) MK}{K+1} |\mu|_1 = \lambda \|\mu\|_1. \end{aligned} \quad (1.23)$$

Next, I'll show $\left\| D^{T+1}(s, \Sigma_N; \theta)\mu \right\|_1 \to 0$ using equation (1.23). Because

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

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

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

which implies that

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

Moreover

$$\left\| \left\{ \mathbf{1}_{N(K+1)} - D(s, \Sigma_N; \theta) \right\}^{-1} \times \mu \right\|_1 = \left\| \sum_{t=0}^{\infty} D^t(s, \Sigma_N; \theta) \times \mu \right\|_1$$

$$\leq \sum_{t=0}^{\infty} \left\| D^t(s,\Sigma_N;\theta) \times \mu \right\|_1 \leq \sum_{t=0}^{\infty} \lambda^t \|\mu\|_1 = \frac{\|\mu\|_1}{1-\lambda}.$$

□

1.10.2 Lemma 1.10.2

Suppose Assumptions 1, 2, and 7 hold, then there exists $\delta_2 \in \mathbb{R}_+$ such that for all θ and $m = 1, \cdots, L$,

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

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

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

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

and $q \in \mathcal{A}/\{0\}$. Furthermore, by Assumption 7, there exists $C_0 \in \mathbb{R}_+$ such that $\|x\| \leq C_0$ for all $x \in \mathcal{X}$. Hence

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

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

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

which implies that

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

□

Lemma 1.10.2

Suppose that Assumptions 1 through 3 and 7 hold. Then

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

and

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

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

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

where $\Sigma^{(i,h)}_N = \Sigma^{(i,h)}_N(S^{(i,h)}_N;\theta)$. Furthermore, by Lemma 1.9.2 and 1.10.2,

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

By the proof in Lemma 1.6, the last term is bounded by

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

Hence,

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

where $C_4 = (K+1)\times(C_2+C_1C_3)$. Since $\left\|\sigma^*_{Nn}-\sigma^{(i,h)}_{Nn}\right\|_1 \le 2$ almost surely for all N, n and θ, and by Lemma 1.9.2, $\max_{n\in\mathcal{N}_j}\left\|\partial\sigma^*_{Nn}/\partial\theta-\partial\sigma^{(i,h)}_{Nn}/\partial\theta\right\|_1\leq 2C_1$, almost surely. Then for all $j\in\mathcal{N}_{(i,h)}$,

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

From the proof of 1.6, for all $j\in\mathcal{N}_{(i,h-1)}$, $\sup_{\theta\in\Theta}\max_{n\in\mathcal{N}_{(j,1)}}\left\|\sigma^*_{Nn}-\sigma^{(i,h)}_{Nn}\right\|_1\leq 2\lambda$ almost surely, then

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

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

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

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

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

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

□

Lemma 1.10.2

Suppose that Assumptions 1 through 3 and 7 hold. Then there exist $C_2,C_3\in\mathbb{R}_+$, such that for all N,i,k and arbitrary two strategy profile $\Sigma_N=\{\sigma_n\}_{n\in\mathcal{I}}$, $\Sigma'_N=\{\sigma'_n\}_{n\in\mathcal{I}}$.

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

and

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

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

proof: First, for any $q\in\mathcal{A}\backslash\{0\}$ and $\ell\in\mathcal{A}$,

$$\begin{aligned}&\left|\frac{\partial\Gamma_{ik}(W_i,\Sigma_N;\theta)}{\partial\alpha(q,\ell)}-\frac{\partial\Gamma_{ik}(W_i,\Sigma'_N;\theta)}{\partial\alpha(q,\ell)}\right|\\&\quad=\left|\Gamma_{ik}\left\{\mathbf{1}(q=k)-\Gamma_{iq}\right\}\sum_{j\in\mathcal{N}_i}\Gamma_{j\ell}-\Gamma'_{Nik}\left\{\mathbf{1}(q=k)-\Gamma'_{Niq}\right\}\sum_{j\in\mathcal{N}_i}\Gamma'_{Nj\ell}\right|\\&\quad\leq M\times\left|\Gamma_{ik}-\Gamma'_{Nik}\right|+M\times\left|\Gamma_{iq}-\Gamma'_{Niq}\right|+M\times\max_{j\in\mathcal{N}_i}\left|\Gamma_{j\ell}-\Gamma'_{Nj\ell}\right|,\end{aligned}$$

almost surely, where the last step is because $0\leq\Gamma_{jk}\leq1$ for all N,i,k and Assumption 2. Thus

$$\left|\frac{\partial\Gamma_{ik}(W_i,\Sigma_N;\theta)}{\partial\alpha(q,\ell)}-\frac{\partial\Gamma_{ik}(W_i,\Sigma'_N;\theta)}{\partial\alpha(q,\ell)}\right|\leq3M\times\max_{j\in\mathcal{N}_{(i,1)}}\left\|\Gamma_j-\Gamma'_{Nj}\right\|_1,\ \ a.s.$$

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

$$\begin{aligned}&\left\|\frac{\partial\Gamma_{ik}(W_i,\Sigma_N;\theta)}{\partial\beta_{\ell'}}-\frac{\partial\Gamma_{ik}(W_i,\Sigma'_N;\theta)}{\partial\beta_q}\right\|_1\\&\quad\leq\left|\Gamma_{ik}\left\{\mathbf{1}(q=k)-\Gamma_{iq}\right\}-\Gamma'_{Nik}\left\{\mathbf{1}(q=k)-\Gamma'_{Niq}\right\}\right|\times\|X_i\|_1\\&\quad\leq2C_0\times\max_i\|\Gamma_i-\Gamma_i\|_1,\end{aligned}$$

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

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

For the second part of the lemma,

$$\begin{aligned}&\sup_{\theta\in\Theta}\left|\sum_{j\in\mathcal{N}_i}\sum_{\ell\in\mathcal{A}}\left\{\frac{\partial\Gamma_{ik}(W_i,\Sigma_N;\theta)}{\partial\sigma_{Nj\ell}}-\frac{\partial\Gamma_{ik}(W_i,\Sigma'_N;\theta)}{\partial\sigma_{Nj\ell}}\right\}\right|\\&\quad=\sup_{\theta\in\Theta}\left|\sum_{j\in\mathcal{N}_i}\sum_{\ell\in\mathcal{A}}\sum_{q\neq k}\left[\left(\Gamma_{ik}\Gamma_{iq}-\Gamma'_{Nik}\Gamma'_{Niq}\right)\left\{\alpha(k,\ell)-\alpha(q,\ell)\right\}\right]\right|\\&\quad\leq MK(K+1)\bar{\Delta}(\alpha)\times\sup_{\theta\in\Theta}\|\Gamma_i-\Gamma'_{Ni}\|_1,\end{aligned}$$

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

Chapter 2

Estimation of Bayesian Nash Equilibria In Static Discrete Games With Correlated Private Signals

2.1 Introduction

This chapter studies a 2×2 static game of incomplete information. I allow players' private signals to be correlated, which adds complexity to *Bayesian Nash Equilibrium (BNE)* solutions of the game. Further, the econometric structure of this model is 'incomplete' due to the existence of multiple equilibria (Tamer (2003)). I therefore focus on a nontrivial subset of the support of public information variables (regressors),[1] where a unique *Monotone Strategy Bayesian Equilibrium (MSBE)* exists. I propose a four–step procedure to estimate the payoff structure. In the first step I estimate a set of parameters containing the underlying parameter of interest, θ_0. I then obtain a point estimator of θ_0 in the second step and prove its consistency. The third and fourth step estimators are $\sqrt{n}$–consistent for θ_0; the fourth step estimator is more efficient.

Static discrete games, like the one I study, are of interest because of their empirical applications, including models of entry (e.g. Bresnahan and Reiss (1990, 1991a)). Unlike Breshnahan and Reiss, I use an incomplete information structure and assume the error terms in the payoff functions are players' private signals. Departing from the literature,

[1]In this chapter, I use public information variables and regressors interchangeably; similarly for private signals and error terms.

notably Aguirregabiria and Mira (2007b), Bajari et al. (2004), Bajari et al. (2009), Pesendorfer and Schmidt-Dengler (2003b), and Seim (2006b), I allow for correlation in players' private signals. Allowing correlation is important for many reasons. For instance, one would expect that private shocks on the profitability of entering a given market are positively correlated with each other.

Dropping the independence assumption complicates the econometric analysis in several respects. First, there can be incompleteness in the sense of Tamer (2003) in the econometric model, caused by the existence of multiple equilibria. The correlation in private signals makes identification more difficult in the presence of incompleteness. Second, it is costly to compute BNEs. Because there is no closed form solution for the BNEs in this game, it is difficult to characterize BNEs in which the strategies are not monotone functions. To avoid such difficulties, I focus on a nontrivial subset of the support of regressors for which a unique MSBE — a special type of BNEs where the equilibrium strategies are monotonic functions of private signals — exists. In this 2×2 game, a monotone strategy is fully characterized by a cutoff point in the support of the private signal. Hence, an MSBE can be obtained by solving the best response equation system defined in terms of the unknown cutoff values. I compute this equilibrium and obtain the likelihood, then conduct maximum likelihood (ML) estimation.

Several alternative procedures have been proposed to estimate incomplete information games, see e.g. Aguirregabiria and Mira (2007b) and Hotz and Miller (1993). The correlation between private signals here makes those methods less appealing. Following the standard BNE concept (see e.g. Aumann (1964) and Harsanyi (1967–68)), a player's beliefs are a probability measure of how the other player behaves in equilibrium conditional on both the public information and her private signal. Because private signals in this game are error terms, which are unobserved to econometricians, a player's beliefs are a function of both observables and unobservables. The only exception arises when the errors are independent conditional on the regressors, in which case players' beliefs depend only on (observable) public information. In contrast, Aradillas-Lopez (2008) estimates the same game structure by assuming a player's beliefs are a conditional distribution of the other player's choice given the public information and outcome, but not her private signal.

Subject to a parametric assumption on the distribution of private signals, the parameters of interest, θ_0, are generally point–identified. Similar to Powell (1986), I use a parameter–dependent subset of the data, $\Pi_2(\theta_0)$, to estimate θ_0, but my procedure is different from

Powell's. A unique MSBE exists on $\Pi_2(\theta_0)$, which is unknown due to its dependence on θ_0. I first estimate a subset of $\Pi_2(\theta_0)$, $\Psi(\Lambda_0) = \bigcap_{\theta \in \Lambda_0} \Pi_2(\theta)$, where Λ_0 is an estimable set of parameters containing θ_0. Using the observations contained in $\Psi(\hat{\Lambda})$, I compute the likelihood based on the unique MSBE solution and obtain a consistent ML estimator $\tilde{\theta}$ of θ_0. I subsequently use $\tilde{\theta}$ to construct my third and fourth step ML estimators, $\bar{\theta}$ and $\hat{\theta}$, both of which converge at a $\sqrt{n}$ rate. The latter is more efficient, so I use it as my final estimator. It is possible to improve efficiency by imposing an equilibrium selection mechanism; this possibility is briefly discussed at the end of this paper.

Because Λ_0 is defined by conditional inequality moment constraints, it cannot be estimated by the set estimation strategies based on (finitely many) unconditional inequality moment constraints proposed in the recent literature, e.g. Chernozhukov et al. (2007). The moment conditions in this paper take the form $E(Y|X = x) - g(x, \theta) \geq 0$, where g is a known parametric function. I treat the conditional inequality restrictions as infinitely many unconditional inequalities, and estimate them non–parametrically using kernel estimation.

The remainder of the paper is organized as follows. Section 2 provides sufficient conditions for the existence of a unique MSBE. Those conditions are satisfied on a non–trivial subset of the support of the regressors, which I use for the purpose of estimation later. In section 3, I discuss the four–step estimation procedure under parametric assumptions, where asymptotic properties of the proposed estimators are established. Proofs are in the appendix.

2.2 The Model

Consider a two–player incomplete information simultaneous move game with the payoff matrix.

		PLAYER 2	
		$Y_2 = 1$	$Y_2 = 0$
PLAYER 1	$Y_1 = 1$	$X_1'\beta_{01} - \alpha_{01} - U_1,\ X_2'\beta_{02} - \alpha_{02} - U_2$	$X_1'\beta_{01} - U_1,\ 0$
	$Y_1 = 0$	$0,\ X_2'\beta_{02} - U_2$	$0,\ 0$

TABLE 1: Two–player incomplete information simultaneous move game

$X = (X_1, X_2) \in \mathscr{X} \subseteq \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ represents public information to both players. U_j $(j = 1, 2)$ is the private signal of player j and Y_j is the choice of player j. $U = (U_1, U_2)$ is assumed to be independent of X, and the joint distribution of U parametrized by ρ_0 is common knowledge to both players. Let $\theta_0 = (\alpha_{01}, \alpha_{02}, \beta_{01}', \beta_{02}', \rho_0)' \in \Theta$ be the parameters of interest. In this chapter, both players are assumed to be risk neutral.

The game with the same payoff structure but under a complete information setup is examined by Bajari et al. (2004), Bresnahan and Reiss (1990, 1991a), Tamer (2003); and references therein. I adopt the BNE solution concept in this incomplete information game. Since player j ($j = 1, 2$) does not observe u_{3-j}, she has to form beliefs about her opponent's move in equilibrium; player j maximizes her payoff by choosing $Y_j = 1$ if and only if $x_j'\beta_{0j} - \alpha_{0j}E\left(Y_{3-j}|X = x, U_j = u_j\right) - u_j \geq 0$, where $E(Y_{3-j}|X = x, U_j = u_j)$ is the beliefs of player j about her opponent's action in equilibrium.

This incomplete information game can be interpreted as an entry model, where two firms simultaneously decide whether to enter a market or not. Before they make their decisions, information $X = x$ is disclosed publicly, and the players each obtain a private signal. There are interactions between the players' strategies; α_0 is the size of the strategic effect. Hence, each player's payoff in this game is determined by the sum of a public information term, a private information term and a strategic effect term. Asymmetry in this game arises when $x_1'\beta_{01} \neq x_2'\beta_{02}$, which reflects the fact that one player may have a commonly known advantage or disadvantage.

2.2.1 Solution Concept

For this simultaneous move game with private information, the BNE solution is a strategy profile $S(u; x) = \{S_1(u_1; x), S_2(u_2; x)\}$, which satisfies

$$S_j = \begin{cases} 0 & x_j'\beta_{0j} - \alpha_{0j}\mu_j(x, u_j) - u_j < 0; \\ 1 & x_j'\beta_{0j} - \alpha_{0j}\mu_j(x, u_j) - u_j \geq 0, \end{cases} \tag{2.1}$$

for $j = 1, 2$, where $\mu_j(x, u_j)$ is the consistent equilibrium belief of player j, defined by

$$\mu_j\left(x, u_j\right) = \mathbb{E}\left[S_{3-j}|X = x, U_j = u_j\right]. \tag{2.2}$$

From the above solution concept, it follows that a BNE is a strategy profile mapping from private information U_1(or U_2) to a binary decision Y_1 (or Y_2), and that the mapping depends on the value of public information. When U_1 and U_2 are positively correlated, a closed form solution for the fixed point of the above mapping in function space does not generally obtain. In fact, for given x, it is not always clear that how many equilibria there are in this game or whether a pure strategy BNE exists.

2.2.2 The Existence and Uniqueness of MSBE

Given $X = x$, the existence of an MSBE implies that there exist two cutoff values, $u_1^*(x)$ and $u_2^*(x)$, for which

$$S_1(u_1; x) = 1[u_1 \leq u_1^*(x)]\,;\; S_2(u_2; x) = 1[u_2 \leq u_2^*(x)]\,,$$

where the cutoff values are given by

$$x_j'\beta_{0j} - \alpha_{0j}\mathbb{P}(U_{3-j} \leq u_{3-j}^* | U_j = u_j^*) - u_j^* = 0. \tag{2.3}$$

I now examine the set of regressors values that induce an MSBE and further a unique MSBE. Let $\kappa_x = [x_1'\beta_{01} - \alpha_{01}, x_1'\beta_{01}] \times [x_2'\beta_{02} - \alpha_{02}, x_2'\beta_{02}]$ and $h_j(u, \theta_0) = u_j + \alpha_{0j}\mathbb{P}(U_{3-j} \leq u_{3-j} | U_j = u_j)$ for $j = 1, 2$. Let further

$$\Pi_1(\theta_0) = \left\{x \in \mathscr{X} : h_j(u, \theta_0) \text{ is nondecreasing in } u_j, \text{for } u = (u_1, u_2) \in \kappa_x \text{ and } j = 1, 2.\right\}$$

For any $x \in \Pi_1(\theta_0)$, the *single crossing condition (SCC, Athey (2001))*, a sufficient condition for the existence of MSBE, is satisfied. Note that h_j is required to be nondecreasing in u_j only on an interval $[x_j'\beta_{0j} - \alpha_{0j}, x_j'\beta_{0j}]$ instead of the whole support. The reason is that for u_j outside of this interval, player j's decision will not be affected by the other player's move; then the fact that h_j is nondecreasing within the interval guarantees SCC to hold. Theorem 8 summarizes the discussion above.

Assumption 15 *The distribution function of (U_1, U_2) is absolutely continuous with respect to the Lebesgue measure.*

Theorem 8 *Under Assumption 15, for $x \in \Pi_1(\theta_0)$ there exists at least one MSBE in this game; i.e. there exists $(u_1^*(x), u_2^*(x))$, such that $S_1^*(u_1; x) = 1[u_1 \leq u_1^*(x)]$, $S_2^*(u_2; x) = 1[u_2 \leq u_2^*(x)]$, and $\{S_1^*, S_2^*\}$ constitutes an MSBE.*

Thus, with a parametric assumption on the distribution of (U_1, U_2), the MSBE solutions for any $x \in \Pi_1(\theta_0)$ can be found by solving (2.3), which provides a parametric expression for the conditional distribution of (Y_1, Y_2) given $X = x$ from which the likelihood obtains. There can be multiple MSBE here, like in Bresnahan and Reiss (1990, 1991a), and Tamer (2003). The possibility of multiple equilibria would require an equilibrium selection mechanism. In this chapter I avoid this issue by focusing on a subset $\Pi_2(\theta_0)$ of $\Pi_1(\theta_0)$, where there exists a unique MSBE.

Note that when U_1 and U_2 are independent, there is $\mu_1(x, u_1) = \mathbb{P}[S_2(U_2; x) = 1]$ and $\mu_2(x, u_2) = \mathbb{P}[S_1(U_1; x) = 1]$ for all u_1, u_2 and x, which means that the private signals are not informative about the strategic effects. In this special case, it can be shown that any BNE solution in this game must be MSBE, and $\Pi_1(\theta_0)$ is the whole support of regressors. An (conditional) independence assumption does not rule out the possibility of multiple equilibria.[2] The presence of multiple equilibria, however, does not complicate the identification strategy and estimation procedure significantly (see Aguirregabiria and Mira (2007b), Bajari et al. (2004), Bajari et al. (2009)).

Now I define a subset $\Pi_2(\theta_0)$ of $\Pi_1(\theta_0)$, which accommodates a unique MSBE:

$$\Pi_2(\theta_0) = \left\{ x \in \mathcal{X} : \forall (u_1, u_2) \in \kappa_x,\ \frac{\partial}{\partial u_j} h_j(u, \theta_0) > \frac{\partial}{\partial u_{3-j}} h_j(u, \theta_0) \text{ for } j = 1, 2. \right\}$$

Theorem 9 *Under Assumption 15, for $x \in \Pi_2(\theta_0)$ a unique MSBE exists; i.e. there exists a unique $(u_1^*(x), u_2^*(x))$ such that $S_1^*(u_1; x) = 1[u_1 \le u_1^*(x)]$, $S_2^*(u_2; x) = 1[u_2 \le u_2^*(x)]$, and $\{S_1^*, S_2^*\}$ constitutes a MSBE.*

Neither Theorem 8 nor Theorem 9 rules out the possible existence of non–MSBE solutions for $x \in \Pi_1(\theta_0)$ or $x \in \Pi_2(\theta_0)$. In this chapter, I assume that only MSBE equilibria are played when there exists at least one such solution; a similar assumption is adopted by Laffont et al. (1995).

2.3 Estimation

2.3.1 Outline of Estimation Strategy

This section gives a description of the estimation strategy. Suppose that $\{X_i, Y_i\}_{i=1}^n$ is an i.i.d. random sample of size n, where $X_i = (X_{1i}, X_{2i})$ and $Y_i = (Y_{1i}, Y_{2i})$. Note that if there

[2] Here I provide a simple example as follows: $\alpha_{01} = \alpha_{02} = 4$, $x_1'\beta_{01} = x_2'\beta_{02} = 2$, U are independent of X and conform to a joint normal distribution with unit variances and correlation $\rho_0 = 0$. Under above setup, three equilibria can be found for this game structure.

exists a unique MSBE for $X = x$, then I can use equation (2.3) to compute it. Under Theorem 9 and further parametric assumption on the distribution of private signals, I can derive the conditional distribution of Y_i given $X_i = x$ for $x \in \Pi_2(\theta_0)$. Because $\Pi_2(\theta_0)$ does not necessarily satisfy the regularity conditions I need in the parametric model concerned, I focus on a well–behaved, slightly smaller subset $\Pi_3(\theta_0)$ of $\Pi_2(\theta_0)$, which will be defined later.

In the first step of my estimation procedure, I use conditional moment inequalities to estimate a set $\Lambda_0 \subseteq \Theta$ that contains θ_0. Let $\Psi(\Lambda_0) = \bigcap_{\theta \in \Lambda_0} \Pi_3(\theta)$. Then, the second step estimator $\tilde{\theta}$ of θ_0 is defined by

$$\tilde{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} 1\left[X_i \in \Psi(\hat{\Lambda})\right] \log P_\theta(Y_i|X_i), \tag{2.4}$$

where $\hat{\Lambda}$ is the first step estimator of Λ_0. For the moment, let P_θ be the conditional probability function of Y given X, and the formal definition of P_θ will be discussed later. I then show that $\tilde{\theta}$ is a consistent estimator of θ_0.

$\tilde{\theta}$ allows me to exploit information in a larger region of regressor values to achieve greater efficiency and establish $\sqrt{n}$–consistency. For some $\epsilon > 0$, let $A_\epsilon(\theta) = \{\theta' \in \Theta : \|\theta' - \theta\|_\infty \leq \epsilon\}$, where $\|z\|_\infty = \max\{|z_1|, \ldots, |z_k|\}$ for $z \in \mathbb{R}^k$. Let further $B_\epsilon(\theta)$ be a collection of the extreme points of $A_\epsilon(\theta)$.[3] By definition, $B_\epsilon(\theta)$ contains a finite number of elements. Hence, $\Psi(B_\epsilon(\theta_0))$ is an "ϵ–smaller" subset of $\Pi_3(\theta_0)$ and $\Psi(B_\epsilon(\tilde{\theta}))$ is contained in $\Pi_3(\theta_0)$ with probability one when n goes to infinity. Then I define the third step estimator by

$$\bar{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} 1\left[X_i \in \Psi(B_\epsilon(\tilde{\theta}))\right] \log P_\theta(Y_i|X_i), \tag{2.5}$$

which is shown to be $\sqrt{n}$–consistent and asymptotically normally distributed. Similarly, I define $B_{\epsilon_n}(\theta)$ for some positive sequence $\epsilon_n \downarrow 0$, and $\Psi(B_{\epsilon_n}(\theta_0))$ converges to $\Pi_3(\theta_0)$. Then the fourth step estimator is defined by

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} 1\left[X_i \in \Psi(B_{\epsilon_n}(\bar{\theta}))\right] \log P_\theta(Y_i|X_i). \tag{2.6}$$

Under further assumptions, I show that $\hat{\theta}$ also $\sqrt{n}$–consistent and asymptotically normally distributed and is more efficient than $\bar{\theta}$, so I use it as the final estimator.

[3]For a convex subset C in a vector space, a point $x \in C$ is called an extreme point of C if it is not an interior point of any line segment in C. Intuitively, an extreme point is a "corner" of C.

Note that identification in this parametrized structure is implied by the above estimation procedure. Indeed, given regularity conditions on P_θ satisfied, if there exists an estimable set Λ_0 containing θ_0 for which $\mathbb{P}\left[X \in \Psi(\Lambda_0)\right] > 0$, then θ_0 is identified by

$$\theta_0 = \arg\max_{\theta \in \Theta} \mathbb{E}\big\{1\left[X \in \Psi(\Lambda_0)\right] \log P_\theta(Y|X)\big\}.$$

2.3.2 The Sets Λ_0 and $\Pi_3(\theta)$

From the above discussions, it follows that the estimation strategy requires me to find an estimable parameter set Λ_0 containing θ_0 and to define a class of subsets $\Pi_3(\theta)$ of the support of the regressors, which yields a unique MSBE when $x \in \Pi_3(\theta_0)$ and varies smoothly with θ. In this section, I will define these two sets using a following parametric assumption.

Assumption 16 *U_1, U_2 have a joint normal distribution with unit variances and correlation $0 \le \rho_0 < 1$.*

The structure of this game is not continuous at $\rho_0 = 1$ where it turns into a complete information game. This scenario is ruled out by above parametric assumption.

2.3.2.1 Definition of Λ_0

Let $c_j = (x'_{3-j}\beta_{3-j} - \rho x'_j\beta_j)/\sqrt{1-\rho^2}$, $g_j^L(x,\theta) = \Phi\left[x'_j\beta_j - \alpha_j\Phi\left(c_j + \rho\alpha_j/\sqrt{1-\rho^2}\right)\right]$, and $g_j^U(x,\theta) = \Phi\left[x'_j\beta_j - \alpha_j\Phi\left(c_j - \alpha_j/\sqrt{1-\rho^2}\right)\right]$, where $j = 1,2$. Let further

$$m_j(x,\theta) = \begin{pmatrix} \mathbb{E}(Y_j|x) - g_j^L(x,\theta) \\ g_j^U(x,\theta) - \mathbb{E}(Y_j|x) \end{pmatrix},$$

and $m(x,\theta) = (m'_1(x,\theta), m'_2(x,\theta))'$. Then I define Λ_0 by

$$\Lambda_0 = \{\theta \in \Theta : \mathbb{P}\left[m(X,\theta) \ge 0\right] = 1\} \tag{2.7}$$

A similar set of inequality constraints is examined by Aradillas-Lopez and Tamer (2008) [4]. The next lemma establishes that this partially identified set contains the parameters of interest.

[4] Aradillas-Lopez and Tamer (2008) obtain the inequality constraints by using "*level–k rationality*" restrictions, which are weaker than the BNE solution concept.

Lemma 3 *Under Assumption 16, $\theta_0 \in \Lambda_0$.*

2.3.2.2 A Subset Accommodating A Unique MSBE: $\Pi_3(\theta_0)$

I define $\Pi_3(\theta)$ as the collection of x's which satisfy the following constraints:

$$\begin{cases} \left(x_2'\beta_2 - \rho x_1'\beta_1 - \frac{\alpha_2-\rho\alpha_1}{2}\right)^2 \geq \left(\frac{\alpha_2+\rho\alpha_1}{2} + \nu_1(\theta)\right)^2 \\ \left(x_1'\beta_1 - \rho x_2'\beta_2 - \frac{\alpha_1-\rho\alpha_2}{2}\right)^2 \geq \left(\frac{\alpha_1+\rho\alpha_2}{2} + \nu_2(\theta)\right)^2 \end{cases} \tag{2.8}$$

where $\nu_j(\theta) = \sqrt{(1-\rho^2)\mu(\iota_j)}$, $\mu(t) = 2 \times 1\left[t \geq 1\right] \log t$, $\iota_j = \tau\alpha_j/\sqrt{2\pi}$, and $\tau = \frac{(1+\rho)}{\sqrt{1-\rho^2}}$ for $j = 1, 2$. From the above definitions, it follows that $\Pi_3(\theta_0)$ is a finite union of polyhedrons.

Lemma 4 *Under Assumption 16, $\Pi_3(\theta_0) \subseteq \Pi_2(\theta_0)$. In particular, there exists a unique MSBE for any $x \in \Pi_3(\theta_0)$.*

The following assumption is sufficient to guarantee that X belongs to $\Psi(\Lambda_0)$ with positive probability.

Assumption 17 *The support of $(X_1'\beta_{01}, X_2'\beta_{02})$ is $\mathbb{R}^2$.*

Assumption 17 requires that both of X_1 and X_2 contain at least one variable with a nonzero β–coefficient and that their joint support is $\mathbb{R}^2$. Similar assumptions can be found in other contexts, e.g. Manski (1985).

Theorem 10 *Under Assumptions 16 – 17, $\mathbb{P}\left[X \in \Psi(\Lambda_0)\right] > 0$.*

2.3.3 The Log–likelihood Contribution

As required from the estimation strategy, here I specify the function $P_\theta(y|x)$. Given $x \in \mathscr{X}$, let $\mathscr{S}_x = \{\theta \in \Theta : (x, \theta) \text{ satisfies (2.8)}\}$. By definition, if (x, θ) satisfies $\theta \in \mathscr{S}_x$, then $x \in \Pi_3(\theta)$, which implies that there is a unique solution $(u_1^*(x;\theta), u_2^*(x;\theta))$ to equation (2.3) by Theorem 9 and Lemma 4. Let $y = (y_1, y_2) \in \{0,1\} \times \{0,1\}$. For $\theta \in \mathscr{S}_x$, $P_\theta(y|x)$ is defined as the conditional probability of $Y = y$ given $X = x$, i.e.,

$$P_\theta(y|x) = \mathbb{P}\left[(1-2y_1) \times (U_1 - u_1^*(x;\theta)) \geq 0; (1-2y_2) \times (U_2 - u_2^*(x;\theta)) \geq 0\right]. \tag{2.9}$$

For $\theta \in \mathscr{S}_x^c$, there could be multiple solutions in equation (2.3), so I define $P_\theta(y|x)$ as a lower bound of the conditional probability of $Y = y$ given $X = x$,

$$P_\theta(y|x) = \mathbb{P}\left[(1-2y_1) \times (U_1 - x_1'\beta_1) \geq \alpha_1 y_1; (1-2y_2) \times (U_2 - x_2'\beta_2) \geq \alpha_2 y_2\right]. \tag{2.10}$$

By definition, for all $\theta \in \hat{\Lambda}$ and $x \in \Psi(\hat{\Lambda})$, $P_\theta(\cdot|x)$ is twice continuously differentiable in θ, and hence so is the objective function in (2.4). Note that, $P_\theta(\cdot|x)$ is not globally continuous and smooth in θ, therefore neither is the objective functions in (2.4). Asymptotically the maximizer of the objective function in (2.4), however, belongs to $\hat{\Lambda}$. Similar results hold for P_θ and the objective functions in equations (2.5) and (2.6).

To guarantee $P_\theta \neq P_{\theta_0}$ for $\theta \neq \theta_0$, here I assume a rank condition for the regressors.

Assumption 18 *There exists no proper linear subspace $\mathscr{L}$ of $\mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ for which $\mathbb{P}(X \in \mathscr{L}) = 1$.*

2.3.4 Asymptotic Properties of The Estimators

In this section I first propose a set estimator $\hat{\Lambda}$ and show its asymptotic properties, then I provide conditions to establish the asymptotic properties of the point estimators.

Let $Q(\theta) = \int \left\| \left(f(x)\,[m(x,\theta)]_-\right)' W^{1/2}(\theta) \right\|^2 dF$, where f and F are the p.d.f and c.d.f of X respectively, $[t]_- = (\max(-t_1, 0), ..., \max(-t_p, 0))'$ for $t \in \mathbb{R}^p$ and W is a diagonal matrix–valued function, which is continuous and positive definite for all $\theta \in \Theta$. In applications, W can be taken to be an identity matrix or chosen to weight the moments by the inverse of their variances, like a GMM weight matrix. Note that the inequality conditional moments have been multiplied by the density function $f(x)$ to avoid small values in the denominator of the estimator. By equation (2.7), Λ_0 is equivalent to the set $\{\theta \in \Theta : Q(\theta) = 0\}$.

Consider the sample analog of $Q(\theta)$,

$$\hat{Q}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \left\| \left[\hat{t}(X_i,\theta) + n^{-\gamma} e\right]_-' \hat{W}^{1/2}(\theta) \right\|^2,$$

in which γ is a strictly positive constant, $e = (1,1,1,1)'$, and $\hat{t}(X_i,\theta)$ is given by

$$\hat{t}(X_i,\theta) = \frac{1}{nh^p}\sum_{l=1}^{n} K\left(\frac{X_l - X_i}{h}\right) T_l(X_i,\theta),$$

where h is a bandwidth, $T_l(x,\theta) = \Big(Y_{1l} - g_1^L(x,\theta),\ g_1^U(x,\theta) - Y_{1l},\ Y_{2l} - g_2^L(x,\theta),\ g_2^U(x,\theta) - Y_{2l}\Big)'$, and $p = k_1 + k_2$.

Let $\hat{\Lambda} = \{\theta \in \Theta : \hat{Q}(\theta) = 0\}$. I make the following assumptions to establish the consistency of $\hat{\Lambda}$.

Assumption 19 *(1) $h \to 0$ as $n \to \infty$; (2) $nh^p \propto n^{\gamma_0}$, where $\gamma_0 > 0$; (3) $n^\gamma h^2 \to 0$ and $\gamma_0 - 2\gamma > 0$.*

Assumption 19 requires that h go to zero at a proper rate and γ not be chosen too large.

Assumption 20 *The kernel K is a symmetric Borel–measurable bounded real–valued function mapping from $\mathbb{R}^p$ to $\mathbb{R}$ such that (1) $\int K(u)du = 1$, (2) $0 < \left|\sum_{l=1}^{p}\sum_{k=1}^{p}\int_{\mathbb{R}^p} u_l u_k K(u)du\right| < \infty$, (3) $\int K^2(u)du < \infty$.*

Assumption 21 *The second–order derivatives of $\mathbb{E}(Y_j|X = x)$ are continuous in x and uniformly bounded for $j = 1, 2$. The density f_x is twice continuously differentiable with bounded second derivatives.*

Assumptions 20 – 21 are standard (e.g. Pagan and Ullah (1999), page 21).

Assumption 22 *The parameter space Θ is a nonempty compact subset of $\mathbb{R}^{p+1}$, and the sample analog $\hat{Q}(\theta)$ of Q is defined on an enlargement Θ' around Θ in $\mathbb{R}^{p+1}$.*

Assumption 22 is needed because of the possibility that Λ_0 contains an element on the boundary of Θ.

Assumption 23 *$\hat{W}$ converges to W uniformly in $\theta \in \Theta$ and is continuous for all $\theta \in \Theta'$.*

Let $d_H(A, B)$ be the Hausdorff distance between two sets A, B, i.e.

$$d_H(A, B) = \max\left\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right\}$$

Theorem 11 *If Assumptions 19 – 23 hold, then $d_H(\hat{\Lambda}, \Lambda_0) = o_p(1)$. Moreover, for any $q > 0$, $\mathbb{P}(\Lambda_0 \subseteq \hat{\Lambda}) = 1 - O(n^{-q})$.*

Note that q can be arbitrarily large due to the fact that the probability of Λ_0 being contained in $\hat{\Lambda}$ converges to one at an exponential rate. $\hat{\Lambda}$ is a set estimator which shrinks to Λ_0 with probability one. Next, the consistency of point estimator $\tilde{\theta}$ is established below.

Assumption 24 $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\log p_\theta\left(Y|X\right)\right|\right] < \infty$.

Assumption 24 is a standard assumption in ML estimation, e.g. Newey and McFadden (1986).

Theorem 12 *Under Assumptions 16 – 24, $\tilde{\theta} \xrightarrow{p} \theta_0$.*

Let $X_j^{[k]}$ be the k–th variable in regressors X_j. Similar notation for $\beta_j^{[k]}$.

Assumption 25 *For $j = 1, 2$, $X_j^{[1]}$ is a continuous argument and $\beta_j^{[1]} \neq 0$. Let $\tilde{X}_j$ be all the X variables without X_j^1, i.e., $\tilde{X}_j = (X_j^{[2]}, ..., X_j^{[k_j]}; X_{3-j})$. Assume further $\mathbb{E}\left[\sup_t f_{X_j^{[1]}|\tilde{X}_j}(t|\tilde{X}_j) \cdot \|\tilde{X}_j\|\right] < \infty$, where $f_{X_j^{[1]}|\tilde{X}_j}$ is the conditional density of $X_j^{[1]}$ given $\tilde{X}_j$.*

The first half of Assumption 25 is also used in Manski (1985). Assumption 25 guarantees $\mathbb{E}\left|1\left[X_i \in \Pi_3(\theta)\right] - 1\left[X_i \in \Pi_3(\theta_0)\right]\right| = O\left(\|\theta - \theta_0\|\right)$ for θ in a small neighborhood of θ_0.

Because $P_\theta(\cdot|x)$ is twice continuously differentiable at θ_0 when $x \in \text{interior1}\left(\Pi_3\left(\theta_0\right)\right)$, I can define a conditional score given $x \in \text{interior1}\left(\Pi_3\left(\theta_0\right)\right)$ by

$$s(y, x; \theta_0) = \frac{\partial \log P_\theta(y|x)}{\partial \theta}\Big|_{\theta=\theta_0}$$

For x in the boundary of $\Pi_3(\theta_0)$, $P_\theta(\cdot|x)$ may not be differentiable at $\theta = \theta_0$. So for all y, I let $s(y, x; \theta_0) = \bar{s}$ for those x's, where $|\bar{s}| < \infty$. Note that, under Assumption 18 the set of elements of x on the boundary of $\Pi_3(\theta_0)$ has measure zero.

Theorem 13 *Under Assumptions 16 – 25, there is*

$$\sqrt{n}\left(\bar{\theta} - \theta_0\right) \xrightarrow{d} N(0, V_\epsilon^{-1}),$$

where $V_\epsilon = \mathbb{E}\{1\left[X \in \Psi(B_\epsilon(\theta_0))\right] \times s(Y, X; \theta_0)s'(Y, X; \theta_0)\}$.

Based on the $\sqrt{n}$–consistent estimator $\bar{\theta}$, I estimate $\Pi_3(\theta_0)$ by $\Psi\left(B_{\epsilon_n}(\bar{\theta})\right)$ to obtain a more efficient estimator $\hat{\theta}$. Note that $\Pi_3(\bar{\theta})$ is one estimator of $\Pi_3(\theta_0)$. The objective function, however, may not be continuous in any neighborhood of θ_0 if I replace $\Psi\left(B_{\epsilon_n}(\bar{\theta})\right)$ with $\Pi_3(\bar{\theta})$ in equation (2.6).

Assumption 26 *$\epsilon_n \to 0$ and $n^{1/2}\epsilon_n \to \infty$.*

Under Assumption 26, $\Psi\left(B_{\epsilon_n}(\bar{\theta})\right)$ estimates $\Pi_3(\theta_0)$ from inside.

Theorem 14 *Under Assumptions 16 – 26,*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, V^{-1}),$$

where $V = \mathbb{E}\{1\left[X \in \Pi_3(\theta_0)\right] \times s(Y, X; \theta_0)s'(Y, X; \theta_0)\}$. Furthermore, matrix $V - V_\epsilon$ is positive semidefinite.

From Theorem 14, $\hat{\theta}$ is a more efficient $\sqrt{n}$–consistent estimator than $\bar{\theta}$ and exploits all the information contained in the subset $\Pi_3(\theta_0)$, so I use it as the final estimator. Note that, because $\Pi_3(\theta_0)$ is always a proper subset of $\Pi_2(\theta_0)$, $\hat{\theta}$ doesnot achieve full efficiency even when $\Pi_2(\theta_0)$ is the whole support of X. After $\hat{\theta}$ is obtained, it is possible to test the hypothesis that $\Pi_2(\theta_0) = \mathcal{X}$. If the null hypothesis is accepted, a standard ML estimator using all the data could achieve the efficiency bound. I do not pursue this possibility here.

2.3.5 Extensions

In the above discussion, the estimation procedure only exploits the subset of the data where a unique MSBE exists. Now I describe two extensions of my estimation procedure, which use the information outside of $\Pi_3(\theta_0)$ and could improve the performance of the estimator.

The last step estimator $\hat{\theta}$ can be modified to incorporate the inequality moment constraints using all data by defining a new estimator $\hat{\hat{\theta}}$ as

$$\hat{\hat{\theta}} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \left\{1\left[X_i \in \Psi\left(B_{\epsilon_n}(\bar{\theta})\right)\right] \log P_\theta(Y_i|X_i)\right\} - \frac{c_n}{n}\sum_{i=1}^{n}\left\{\left[\hat{t}(X_i,\theta)\right]_-' \hat{W}^{1/2}(\theta)\right\}^2$$

where c_n is a non–decreasing positive sequence. There is no asymptotic efficiency improvement for this modified estimator because none of the inequality constraints is binding asymptotically, but the finite sample performance of the estimator may be better.

Another extension is to impose an equilibrium selection mechanism to a larger subset, $\Pi_1(\theta_0)$. For example, assume the equilibrium selection mechanism depends on x and varies smoothly with x, which means that the equilibrium played in all markets with the same x is assumed to be identical and will not suddenly switch when x is slightly changed. Then I can construct a more efficient estimator, similar to Tamer (2003), which uses all the information in $\Pi_1(\theta_0)$.

2.4 Appendix A

Proof of Theorem 8

First, given $x \in \Pi_1(x)$, define a mapping $T(u) : \kappa_x \to \kappa_x$ as follows

$$\begin{aligned} T_1(u) + \alpha_{01}\mathbb{P}\left[U_2 \le u_2 | U_1 = T_1(u)\right] = x_1'\beta_{01} \\ T_2(u) + \alpha_{02}\mathbb{P}\left[U_1 \le u_1 | U_2 = T_2(u)\right] = x_2'\beta_{02} \end{aligned} \tag{2.11}$$

The mapping $T(u)$ defined above also depends on x. Without abuse of notations, I drop the the index x in $T(u)$ to make notations simple. $T(u)$ is well–defined for $x \in \Pi_1(x)$ and is a continuous mapping under Assumption 15. Hence a fixed point $u^* = (u_1^*, u_2^*) \in \kappa_x$ exists by Brouwer fixed point theorem. Hence

$$x_1'\beta_{01} - \alpha_{01}\mathbb{P}\left(U_2 \leq u_2^*|U_1 = u_1^*\right) - u_1^* = 0; \quad x_2'\beta_{02} - \alpha_{02}\mathbb{P}\left(U_1 \leq u_1^*|U_2 = u_2^*\right) - u_2^* = 0$$

When $x \in \Pi_1(\theta_0)$, because $h_1(u_1, u_2^*, \theta_0) \leq 0$ for $u_1 \in [x_1'\beta_{01} - \alpha_{01}, x_1'\beta_{01}]$, then it can be verified that $S_1 = 1[u_1 \leq u_1^*]$ is a best response of player 1. Parallel analysis holds for player 2. Hence the following strategy profile $\{1[u_1 \leq u_1^*], 1[u_2 \leq u_2^*]\}$ can construct a BNE. □

proof of Theorem 9

Prove by contradiction. Given $x \in \Pi_2(\theta_0)$, assume $u = (u_1^*, u_2^*)$ and $v = (v_1^*, v_2^*)$ are the cutoff values that define two MSBEs. Denote $T(\cdot)$ as the mapping defined in (2.11). Hence, $T(u) = u, T(v) = v$.

Form a continuously differentiable function $\phi(t)$ by

$$\phi(t) = \frac{< T(u) - T(v), T[v + t(u-v)] >}{\|T(u) - T(v)\|}$$

Note that $\phi(1) - \phi(0) = \|T(u) - T(v)\| = \|u - v\|$, and also $\phi(1) - \phi(0) = \int_0^1 \phi'(t)dt$.

For any $t \in (0,1)$, we have

$$\begin{aligned}\phi'(t) &= \frac{< T(u) - T(v), T'[v + t(u-v)](u-v) >}{\|T(u) - T(v)\|} = \frac{< u - v, T'[v + t(u-v)](u-v) >}{\|u - v\|} \\ &\leq \frac{\|u - v\| \times \|T'[v + t(u-v)](u-v)\|}{\|u - v\|} = \|T'[v + t(u-v)]\| \times \|u - v\| < \|u - v\| \quad a.e.\end{aligned}$$

The first inequality comes from the Cauchy Schwartz inequality and the last inequality is based on the definition of $\Pi_2(\theta_0)$, which implies $\|T'[v + t(u-v)]\| < 1$ a.e. Hence $\phi(1) - \phi(0) = \int_0^1 \phi'(t)dt < \|u - v\|$, contradiction. □

Proof of lemma 3

Firstly,

$$Y_1 = 1\left[X_1'\beta_{01} - \alpha_{01}\mathbb{E}(Y_2|X, U_1) - U_1 \geq 0\right].$$

Because $0 \leq \mathbb{E}(Y_2|X, U_1) \leq 1$, it follows that

$$1\left[X_1'\beta_{01} - \alpha_{01} - U_1 \geq 0\right] \leq Y_1 \leq 1\left[X_1'\beta_{01} - U_1 \geq 0\right],$$

which gives us $\mathbb{P}(U_1 \leq X_1'\beta_{01} - \alpha_{01}|X, U_2) \leq \mathbb{E}(Y_1|X, U_2) \leq \mathbb{P}(U_1 \leq X_1'\beta_{01}|X, U_2)$. Similarly, $\mathbb{P}(U_2 \leq X_2'\beta_{02} - \alpha_{02}|X, U_1) \leq \mathbb{E}(Y_2|X, U_1) \leq \mathbb{P}(U_2 \leq X_2'\beta_{02}|X, U_1)$. Then

$$1\left[X_1'\beta_{01} - \alpha_{01}\mathbb{P}(U_2 \leq X_2'\beta_{02}|X, U_1) - U_1 \geq 0\right] \leq Y_1$$

$$\leq 1\left[X_1'\beta_{01} - \alpha_{01}\mathbb{P}(U_2 \leq X_2'\beta_{02} - \alpha_{02}|X, U_1) - U_1 \geq 0\right]$$

Because (U_1, U_2) is jointly normal distributed, then for any $x \in \mathcal{X}$,

$$1\left[x_1'\beta_{01} - \alpha_{01}\Phi\left(\frac{x_2'\beta_{02} - \rho_0 U_1}{\sqrt{1-\rho_0^2}}\right) - U_1 \geq 0\right] \leq \mathbb{E}(Y_1|x)$$
$$\leq 1\left[x_1'\beta_{01} - \alpha_{01}\Phi\left(\frac{x_2'\beta_{02} - \alpha_{02} - \rho_0 U_1}{\sqrt{1-\rho_0^2}}\right) - U_1 \geq 0\right]$$

Note that

$$1\left[x_1'\beta_{01} - \alpha_{01}\Phi\left(\frac{x_2'\beta_{02} - \rho_0 U_1}{\sqrt{1-\rho_0^2}}\right) - U_1 \geq 0\right] \geq \Phi\left[x_1'\beta_{01} - \alpha_{01}\Phi(c_{01} + \rho_0\alpha_{01}/\sqrt{1-\rho_0^2})\right]$$

holds for both $U_1 \geq x_1'\beta_{01} - \alpha_{01}$ and $U_1 \leq x_1'\beta_{01} - \alpha_{01}$. Thus $\mathbb{E}(Y_1|x) \geq g_1^L(x, \theta_0)$. Similarly, $\mathbb{E}(Y_1|x) \leq g_1^U(x, \theta_0)$ and parallel analysis holds for $\mathbb{E}(Y_2|x)$. Thus $\theta_0 \in \Lambda_0$. □

Proof of Lemma 4

Under Assumption 16, $h_1(u, \theta_0) = u_1 + \alpha_{01}\Phi(\frac{u_2 - \rho_0 u_1}{\sqrt{1-\rho_0^2}})$ and $h_2(u, \theta_0) = u_2 + \alpha_{02}\Phi(\frac{u_1 - \rho_0 u_2}{\sqrt{1-\rho_0^2}})$. First consider the constraint for h_1 in $\Pi_2(\theta_0)$.

Because $\frac{\partial}{\partial u_1}h_1(u, \theta_0) > \frac{\partial}{\partial u_2}h_1(u, \theta_0)$ is equivalent to

$$2(1-\rho_0^2)\ln \iota_{01} < (u_2 - \rho_0 u_1)^2.$$

When $\iota_{01} \geq 1$, above constraint holding for all $(u_1, u_2) \in \kappa_x$ is equivalent to

$$x_2'\beta_{02} - \rho_0 x_1'\beta_{01} + \rho_0\alpha_{01} < -\sqrt{2(1-\rho_0^2)\ln \iota_{01}} \quad a.e.$$

or

$$x_2'\beta_{02} - \alpha_{02} - \rho_0 x_1'\beta_{01} > \sqrt{2(1-\rho_0^2)\ln \iota_{01}} \quad a.e.$$

Consider the first constraint in the definition of $\Pi_3(\theta_0)$ for $\iota_{01} \geq 1$,

$$\left(x_2'\beta_{02} - \rho_0 x_1'\beta_{01} - \frac{\alpha_{02} - \rho_0\alpha_{01}}{2}\right)^2 \geq \left(\frac{\alpha_{02} + \rho_0\alpha_{01}}{2} + \nu_1(\theta_0)\right)^2$$

which is equivalent to the constraint for h_1 in $\Pi_2(\theta_0)$. A similar argument holds for h_2. Hence, $\Pi_2(\theta_0) = \Pi_3(\theta_0)$ when $\iota_{01}, \iota_{02} \geq 1$.

If $\iota_{0j} < 1$, then the j-th constraint in the definition of $\Pi_2(\theta_0)$ is not binding, which means $\Pi_3(\theta_0) \subseteq \Pi_2(\theta_0)$. □

Proof of Theorem 10

If $\theta \in \Lambda_0$, then the following inequalities hold for any $x \in \mathcal{X}$,

$$\Phi\left(x_1'\beta_1 - \alpha_1\right) \leq \mathbb{E}\left(Y_1|x\right) \leq \Phi\left(x_1'\beta_1\right)$$
$$\Phi\left(x_2'\beta_2 - \alpha_2\right) \leq \mathbb{E}\left(Y_2|x\right) \leq \Phi\left(x_2'\beta_2\right)$$

Note that θ_0 also satisfies above constraints, then

$$x_1'\beta_1 - \alpha_1 \leq x_1'\beta_{01};\;\; x_1'\beta_{01} - \alpha_{01} \leq x_1'\beta_1$$

Which implies that

$$x_1'\beta_{01} - \alpha_{01} \leq x_1'\beta_1 \leq x_1'\beta_{01} + \alpha_1$$

And a similar argument holds for $x_2'\beta_2$. Because of Assumption 17, the support of $(X_1'\beta_1, X_2'\beta_2)$ is unbounded given the value of θ. Hence, by definition,

$$\Phi\left[x_1'\beta_1 - \alpha_1\Phi\left(c_1 + \frac{\rho\alpha_1}{\sqrt{1-\rho^2}}\right)\right] \leq E\left(Y_1|x\right) \leq \Phi\left[x_1'\beta_1 - \alpha_1\Phi\left(c_1 - \frac{\alpha_2}{\sqrt{1-\rho^2}}\right)\right]$$

where $c_1 = (x_2'\beta_2 - \rho x_1'\beta_1)/\sqrt{1-\rho^2}$ can be arbitrary large (or small). Hence β_1 and α_1 are point identified in the limit, i.e.,

$$\lim_{x_2'\beta_2 \to \infty} \Phi\left(x_1'\beta_1 - \alpha_1\right) - E\left(Y_1|x\right) = 0;\;\; \lim_{x_2'\beta_2 \to -\infty} \Phi\left(x_1'\beta_1\right) - E\left(Y_1|x\right) = 0$$

Similarly, β_2 and α_2 are also point identified. Then for any $\theta \in \Lambda_0$,

$$(\alpha, \beta) = (\alpha_0, \beta_0)$$

Consider a subset of $\Psi(\Lambda_0)$,

$$\bigcap_{\theta \in \Lambda} \left[x_2'\beta_2 - \alpha_2 - \rho x_1'\beta_1 \geq \sqrt{2(1-\rho^2)\ln\iota_1}\,\right] \supset \bigcap_{\rho \in [0,1]} \left[x_2'\beta_{02} - \alpha_{02} - \rho x_1'\beta_{01} \geq \sqrt{2(1-\rho^2)\ln\iota_1}\,\right]$$

Since $\sqrt{2(1-\rho^2)\ln\iota_1}$ is bounded above, then $\mathbb{P}\left[X \in \Psi(\Lambda_0)\right] > 0$ under Assumption 17. □

Proof of Theorem 11

By definition,

$$\Lambda_0 = \{\theta \in \Theta : \mathbb{P}[m(X,\theta) \geq 0] = 1\} = \{\theta \in \Theta : f(x)m(x,\theta) \geq 0, \text{ for all } x \in \mathscr{X}\},$$

Let $\tilde{\Lambda} = \{\theta \in \Theta : f(X_i)m(X_i,\theta) \geq 0, \text{ for } i = 1, ..., n\}$. It follows that $d(\Lambda_0, \tilde{\Lambda}) \xrightarrow{p} 0$. Let $t(x,\theta) = f(x)m(x,\theta)$. By Assumptions 19 – 21, $\sup_x |\mathbb{E}\hat{t}(x,\theta) - t(x,\theta)| = O(h^2)$ and $\sup_x |\hat{t}(x,\theta) - \mathbb{E}\hat{t}(x,\theta)| =$

$O_p(\sqrt{nh^p})$. Hence, with Assumption 19,

$$\sup_x |\hat{t}(x,\theta) - t(x,\theta)| = o_p(n^{-\gamma}).$$

By definition, $\hat{\Lambda} = \{\theta \in \Theta : f(X_i)m(X_i,\theta) + \hat{t}(X_i,\theta) - t(X_i,\theta) + n^{-\gamma} \geq 0, \text{ for } i = 1, ..., n\}$.

Given arbitrary $\epsilon > 0$,

$$\mathbb{P}\left[d(\tilde{\Lambda}, \hat{\Lambda}) > \epsilon\right] \leq \mathbb{P}\left[\exists\theta \in \tilde{\Lambda} : d(\theta, \hat{\Lambda}) > \epsilon\right] + \mathbb{P}\left[\exists\theta \in \hat{\Lambda} : d(\theta, \tilde{\Lambda}) > \epsilon\right]$$

When $\hat{t}(X_i,\theta) - t(X_i,\theta) + n^{-\gamma} \geq 0$ for all i, there is $\tilde{\Lambda} \subseteq \hat{\Lambda}$. Then

$$\mathbb{P}\left[\exists\theta \in \tilde{\Lambda} : d(\theta, \hat{\Lambda}) > \epsilon\right] \leq \mathbb{P}\left[\exists i : \hat{t}(X_i,\theta) - t(X_i,\theta) + n^{-\gamma} < 0\right] \to 0.$$

For any $\delta > 0$, let $\tilde{\Lambda}_\delta = \{\theta \in \Theta : f(X_i)m(X_i,\theta) + \delta \geq 0, \text{ for } i = 1, ..., n\}$. Then

$$\mathbb{P}\left[\exists\theta \in \hat{\Lambda} : d(\theta, \tilde{\Lambda}) > \epsilon\right] = \lim_{\delta\downarrow 0} \mathbb{P}\left[\exists\theta \in \hat{\Lambda} : d(\theta, \tilde{\Lambda}_\delta) > \epsilon\right] \leq \lim_{\delta\downarrow 0} \mathbb{P}\left[\exists i : \hat{t}(X_i,\theta) - t(X_i,\theta) + n^{-\gamma} < \delta\right]$$

Hence

$$\lim_{n\to\infty} \mathbb{P}\left[\exists\theta \in \hat{\Lambda} : d(\theta, \tilde{\Lambda}) > \epsilon\right] = 0,$$

which implies that $d(\hat{\Lambda}, \tilde{\Lambda}) \xrightarrow{p} 0$, therefore $d(\hat{\Lambda}, \Lambda_0) \xrightarrow{p} 0$.

For the second part of this theorem,

$$\begin{aligned}
\mathbb{P}(\Lambda_0 \not\subseteq \hat{\Lambda}) &\leq \mathbb{P}\left[\exists\theta \in \Lambda_0 : \max_i \|[\hat{t}(X_i,\theta) + n^{-\gamma}e]_-\| > 0\right] \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left[\exists\theta \in \Lambda_0 : \|[\hat{t}(X_i,\theta) + n^{-\gamma}e]_-\| > 0\right] = n\mathbb{P}\left[\exists\theta \in \Lambda_0 : \|[\hat{t}(X_1,\theta) + n^{-\gamma}e]_-\| > 0\right] \\
&\leq n\sum_{j=1}^{2} \mathbb{P}\left\{\exists\theta \in \Lambda_0 : \frac{1}{nh^p}\sum_{l=1}^{n} K\left(\frac{X_l - X_1}{h}\right)\left[Y_{jl} - g_j^L(X_1,\theta)\right] + n^{-\gamma} < 0\right\} \\
&\quad + n\sum_{j=1}^{2} \mathbb{P}\left\{\exists\theta \in \Lambda_0 : \frac{1}{nh^p}\sum_{l=1}^{n} K\left(\frac{X_l - X_1}{h}\right)\left[g_j^U(X_1,\theta) - Y_{jl}\right] + n^{-\gamma} < 0\right\}. \quad (2.12)
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbb{P}\left\{\exists\theta \in \Lambda_0 : \frac{1}{nh^p}\sum_{l=1}^{n} K\left(\frac{X_l - X_1}{h}\right)\left[Y_{1l} - g_1^L(X_1,\theta)\right] + n^{-\gamma} < 0\right\} \\
&\leq \mathbb{P}\left\{\exists\theta \in \Lambda_0 : f(X_1)\left[\mathbb{E}(Y_{11}|X_1) - g_1^L(X_1,\theta)\right] - \frac{1}{nh^p}\sum_{l=1}^{n} K\left(\frac{X_l - X_1}{h}\right)\left[Y_{1l} - g_1^L(X_1,\theta)\right] > n^{-\gamma}\right\} \\
&\leq \mathbb{P}\left[\left|f(X_1)\mathbb{E}(Y_{11}|X_1) - \frac{1}{nh^p}\sum_{l=1}^{n} K\left(\frac{X_l - X_1}{h}\right) Y_{1l}\right| \geq 0.5n^{-\gamma}\right]
\end{aligned}$$

$$+\mathbb{P}\left[\left|f(X_1)-\frac{1}{nh^p}\sum_{l=1}^{n}K\left(\frac{X_l-X_1}{h}\right)\right|\geq 0.5n^{-\gamma}\right].$$

The last inequality uses the fact that $0\leq g_1^L(x,\theta)\leq 1$ for any value of x and θ. From Lemma B1 and B2,

$$n^{1+q}\mathbb{P}\left[\left|f(X_1)\mathbb{E}(Y_{11}|X_1)-\frac{1}{nh^p}\sum_{l=1}^{n}K\left(\frac{X_l-X_1}{h}\right)Y_{1l}\right|\geq 0.5n^{-\gamma}\right]\to 0.$$

Similarly,

$$n^{1+q}\mathbb{P}\left[\left|f(X_1)-\frac{1}{nh^p}\sum_{l=1}^{n}K\left(\frac{X_l-X_1}{h}\right)\right|\geq 0.5n^{-\gamma}\right]\to 0.$$

Then

$$n^{1+q}\mathbb{P}\left\{\exists\theta\in\Lambda_0:\frac{1}{nh^p}\sum_{l=1}^{n}K\left(\frac{X_l-X_1}{h}\right)\left[Y_{1l}-g_1^L(X_1,\theta)\right]+n^{-\gamma}<0\right\}\to 0.$$

Similar results can be obtained for the other three terms in the RHS of equation (2.12). Hence $n^q\mathbb{P}(\Lambda_0\not\in\hat{\Lambda})\to 0$. □

Proof of Theorem 12

Let $L_n^{[1]}(\theta)=\frac{1}{n}\sum_{i=1}^{n}\{1\left[x_i\in\Psi(\hat{\Lambda})\right]\log P_\theta(Y_i|X_i)\}$ and $L^{[1]}(\theta)=\mathbb{E}\{1\left[X\in\Psi(\Lambda_0)^o\right]\log P_\theta(Y|X)\}$, where $\Psi(\Lambda_0)^o$ means the interior of $\Psi(\Lambda_0)$.

For $x\in\mathscr{X}$, let $\mathscr{S}_x=\{\theta\in\Theta:(x,\theta)\text{ satisfies (2.8)}\}$. By definition, $\mathscr{S}_x$ is a close set in Θ. When $\theta\in\mathscr{S}_x$, $P_\theta(\cdot|x)$ is defined by (2.9); otherwise by (2.10). Note that $P_\theta(\cdot|x)$ is upper semi–continuous in θ. Hence θ_0 is identified by $L^1(\theta)$, i.e., $\sup_{\|\theta-\theta_0\|\geq\epsilon}L^{[1]}(\theta)<L^{[1]}(\theta_0)$ for some $\epsilon>0$. So it's sufficient to show $\sup_\theta|L_n^{[1]}(\theta)-L^{[1]}(\theta)|=o_p(1)$.

$$\begin{aligned}\sup_\theta|L_n^{[1]}(\theta)-L^{[1]}(\theta)|&\leq\sup_\theta\left|\frac{1}{n}\sum_{i=1}^{n}\{1\left[X_i\in\Psi(\Lambda_0)^o\right]\log P_\theta(Y_i|X_i)\}-L^{[1]}(\theta)\right|\\&+\sup_\theta\left|\frac{1}{n}\sum_{i=1}^{n}\{1\left[X_i\in\Psi(\hat{\Lambda})\right]\log P_\theta(Y_i|X_i)\}-\frac{1}{n}\sum_{i=1}^{n}\{1\left[X_i\in\Psi(\Lambda_0)^o\right]\log P_\theta(Y_i|X_i)\}\right|=o_p(1)\end{aligned}$$

The last inequality comes from the Uniform Law of Large Number [5] and Lemma 7 in appendix B. □

Proof of Theorem 13

The proof of consistency similarly follows the proof in Theorem 12 by replacing Λ_0 with $B_\epsilon(\theta_0)$ and using Assumption 25.

[5] Note that $P_\theta(y|x)$ can be written as $p_1(y,x,\theta)\times 1\left[\theta\in\mathscr{S}_x\right]+p_2(y,x,\theta)\times 1\left[\theta\in\mathscr{S}_x^c\right]$, where $p_1(y,x,\theta)$ equals to the RHS of equation (2.9) and $p_2(y,x,\theta)$ is defined as the RHS of equation (2.10). Because both p_1 and p_2 are the classes of continuous functions indexed by θ and by definition $\mathscr{S}_x$ is the class of ellipsoids indexed by θ, then $P_\theta(\cdot|x)$ is the VC class of functions indexed by θ. Then Uniform LLN applies here.

Remember that $A_\epsilon(\tilde\theta) = \{\theta \in \Theta : \|\theta - \tilde\theta\|_\infty \le \epsilon\}$ and $B_\epsilon(\tilde\theta)$ is the collection of all extreme points of $A_\epsilon(\tilde\theta)$, then $\Psi(B_\epsilon(\tilde\theta)) = \Psi(A_\epsilon(\tilde\theta))$. Let $L_n^{[2]}(\theta) = \frac{1}{n}\sum_{i=1}^n \{1\left[X_i \in B_\epsilon(\bar\theta)\right] \log P_\theta(Y_i|X_i)\}$. So $L_n^2(\theta)$ is twice continuously differentiable at $\theta \in A_\epsilon^o(\tilde\theta)$, where $A_\epsilon^o(\tilde\theta)$ means the interior of $A_\epsilon(\tilde\theta)$. Because $\tilde\theta \xrightarrow{p} \theta_0$ and $\bar\theta \xrightarrow{p} \theta_0$, then

$$\mathbb{P}\left[\bar\theta \in A_\epsilon^o(\tilde\theta)\right] \to 1.$$

When $\bar\theta \in A_\epsilon^o(\tilde\theta)$, $L_n^2(\theta)$ is twice continuously differentiable at $\theta = \bar\theta$, so

$$\frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \Psi\left(B_\epsilon(\tilde\theta)\right)\right] s(Y_i, X_i; \bar\theta) = 0.$$

Where $s(Y_i, X_i; \theta)$ is the derivative of $\log P_\theta(Y_i|X_i)$ with respect to θ. By Taylor expansion

$$\frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \Psi\left(B_\epsilon(\tilde\theta)\right)\right] s(Y_i, X_i; \theta_0) + \frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \Psi\left(B_\epsilon(\tilde\theta)\right)\right] (\partial s/\partial\theta')(Y_i, X_i; \theta^\dagger)\left(\bar\theta - \theta_0\right) = 0$$

where $\theta^\dagger$ is between $\tilde\theta$ and θ_0. Hence

$$\begin{aligned}
&\sqrt{n}(\bar\theta - \theta_0) \\
&= -\left\{\frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \Psi\left(B_\epsilon(\tilde\theta)\right)\right](\partial s/\partial\theta')(Y_i, X_i; \theta^\dagger)\right\}^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n 1\left[X_i \in \Psi\left(B_\epsilon(\tilde\theta)\right)\right] s(Y_i, X_i; \theta_0) \\
&= -\left\{\mathbb{E}1\left[X \in \Psi\left(B_\epsilon(\theta_0)\right)\right](\partial s/\partial\theta')(Y, X; \theta_0) + o_p(1)\right\}^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n 1\left[X_i \in \Psi\left(B_\epsilon(\tilde\theta)\right)\right] s(Y_i, X_i; \theta_0)
\end{aligned}$$

The last step comes from ULLN and the fact that both $1\left[X_i \in \Psi\left(B_\epsilon(\theta)\right)\right]$ and $(\partial s/\partial\theta')(Y_i, X_i; \theta)$ are VC class of functions indexed by $\theta \in \mathcal{N}_\epsilon(\theta_0)$. Further, by Lemma 8, it follows that

$$\sqrt{n}(\bar\theta - \theta_0) \xrightarrow{d} \mathbb{N}(0, V_\epsilon^{-1})$$

□

Proof of Theorem 14

Let $L^{[3]}(\theta) = \mathbb{E}\left\{1\left[X \in \Pi_3(\theta_0)\right] \log P_\theta(Y|X)\right\}$ and $L_n^{[3]}(\theta) = \frac{1}{n}\sum_{i=1}^n 1\left[X \in \Psi(B_{\epsilon_n}(\tilde\theta))\right] \log P_\theta(Y|X)$.

Note that $P_\theta(\cdot|x)$ is not a continuous in θ. For any $x \in \mathscr{X}$, however, I can define a twice continuously differentiable probability measure $\tilde P_\theta(\cdot|x)$, such that (1) $\tilde P_\theta(\cdot|x) = P_\theta(\cdot|x)$ for $\theta \in \mathscr{S}_x$; (2) $\tilde P_\theta(\cdot|x) \ne \tilde P_{\theta_0}(\cdot|x)$ for $\theta \ne \theta_0$; (3) $\tilde P_\theta(\cdot|x) \ge P_\theta(\cdot|x)$ for all $\theta \in \Theta$.[6]

Let $\tilde L_n^{[3]}(\theta) = \frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \Psi(B_{\epsilon_n}(\bar\theta))\right] \log \tilde P_\theta(Y_i|X_i)$ and $\tilde L^{[3]}(\theta) = \mathbb{E}\left\{1\left[X \in \Pi_3(\theta_0)\right] \log \tilde P_\theta(Y|X)\right\}$.

[6]One example of $\tilde P_\theta(\cdot|x)$ satisfying all the requirements is: when there are multiple solutions in equation (2.3), select one among those solutions such that $u^*(x;\theta)$ is twice continuously differentiable in θ; then let $\tilde P_\theta(\cdot|x)$ is defined by equation (2.9).

Let further

$$\breve{\theta} = \arg\max_{\theta \in \Theta} \tilde{L}_n^{[3]}(\theta).$$

Note that $\tilde{L}_n^{[3]}(\theta) = L_n^{[3]}(\theta)$ for $\theta \in A_{\epsilon_n}(\bar{\theta})$.[7] Then if $\breve{\theta} \in A_{\epsilon_n}(\bar{\theta})$,

$$L_n^{[3]}(\breve{\theta}) = \tilde{L}_n^{[3]}(\breve{\theta}) \geq \tilde{L}_n^{[3]}(\theta) \geq L_n^{[3]}(\theta)$$

for all $\theta \in \Theta$, which implies that $\hat{\theta} = \breve{\theta}$. By $\sqrt{n}$–consistency of $\bar{\theta}$ and Assumption 26, it is sufficient to prove

$$\sqrt{n}\left(\breve{\theta} - \theta_0\right) \xrightarrow{d} \mathbb{N}(0, V^{-1}).$$

The consistency of $\breve{\theta}$ similarly follows the proof in Theorem 12 by replacing Λ_0 with $\Pi_3(\theta_0)$ and using Assumption 25. For the asymptotic normality of $\breve{\theta}$, similarly as the proof in Theorem 13, using Taylor expansion and Uniform LLN I have

$$\sqrt{n}(\breve{\theta} - \theta_0) = -\left\{\mathbb{E}1\left[X \in \Pi_3\left(\theta_0\right)\right](\partial s/\partial\theta')(Y, X; \theta_0) + o_p(1)\right\}^{-1} \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n} 1\left[X_i \in \Psi\left(B_{\epsilon_n}(\bar{\theta})\right)\right] s(Y_i, X_i; \theta_0)$$

By Lemma 9, there is

$$\sqrt{n}(\breve{\theta} - \theta_0) \xrightarrow{d} \mathbb{N}(0, V^{-1}).$$

□

[7] By definition, $A_{\epsilon_n}(\tilde{\theta}) = \{\theta \in \Theta : \|\theta - \tilde{\theta} \leq \epsilon_n\|\}$, see Section 3.1. If $\theta \in A_{\epsilon_n}(\tilde{\theta})$, $\theta \in \mathscr{S}_x$ for all $x \in \Psi(B_{\epsilon_n}(\tilde{\theta}))$.

2.5 Appendix B

Lemma 5 *Under Assumption 19 to 21, for any $q > 0$,*

$$n^{1+q}\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=1}^{n}\left[K\left(\frac{X_l - X_1}{h}\right)Y_{1l} - \mathbb{E}\left(K\left(\frac{X_l - X_1}{h}\right)Y_{1l}\Big|X_1\right)\right]\right| \geq 0.5n^{-\gamma}\right\} \to 0$$

proof: Let $Z_l(x) = K\left(\frac{X_l - x}{h}\right)Y_{1l}$ for $l = 2, ..., n$. Consider

$$\begin{aligned}
&\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=1}^{n}\left[K\left(\frac{X_l - X_1}{h}\right)Y_{1l} - \mathbb{E}\left(K\left(\frac{X_l - X_1}{h}\right)Y_{1l}\Big|X_1\right)\right]\right| \geq 0.5n^{-\gamma}\Big|X_1 = x\right\} \\
&\quad \leq \mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=2}^{n}[Z_l(x) - \mathbb{E}Z_l(x)]\right| \geq 0.5n^{-\gamma} - \left|\frac{K(0)}{nh^p}[Y_{11} - \mathbb{E}(Y_{11}|X_1 = x)]\right|\Big|X_1 = x\right\} \\
&\qquad \leq \mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=2}^{n}[Z_l(x) - \mathbb{E}Z_l(x)]\right| \geq 0.5n^{-\gamma} - \frac{2\overline{K}}{nh^p}\right\}
\end{aligned}$$

where $\overline{K}$ is an upper bound of bounded kernel function $K(\cdot)$ and $\overline{K} > 0$. Because

$$\begin{aligned}
&\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=2}^{n}[Z_l(x) - \mathbb{E}Z_l(x)]\right| \geq 0.5n^{-\gamma} - \frac{2\overline{K}}{nh^p}\right\} = \mathbb{P}\left\{\left|\sum_{l=2}^{n}[Z_l(x) - \mathbb{E}Z_l(x)]\right| \geq 0.5n^{1-\gamma}h^p - 2\overline{K}\right\} \\
&\leq 2\exp\left\{-\frac{\left[0.5n^{1-\gamma}h^p - 2\overline{K}\right]^2}{2\sum_{l=2}^{n}Var[Z_l(x)] + 2\overline{K}n^{1-\gamma}h^p/3}\right\} = 2\exp\left\{-\frac{\left[0.5n^{1-\gamma}h^p - 2\overline{K}\right]^2}{2(n-1)Var[Z_2(x)] + 2\overline{K}n^{1-\gamma}h^p/3}\right\}
\end{aligned}$$

where the inequality comes from Bernstein's Inequality and the fact $\mathbb{P}\left[|Z_l(x) - \mathbb{E}Z_l(x)| \leq 2\overline{K}\right] = 1$ for all n. Then

$$\begin{aligned}
Var[Z_2(x)] &= \mathbb{E}\left[K^2\left(\frac{X-x}{h}\right)Y_1^2\right] - \left\{\mathbb{E}\left[K\left(\frac{X-x}{h}\right)Y_1\right]\right\}^2 \\
&= \mathbb{E}\left[K^2\left(\frac{X-x}{h}\right)\mathbb{E}(Y_1|X)\right] - \left\{\mathbb{E}\left[K\left(\frac{X-x}{h}\right)\mathbb{E}(Y_1|X)\right]\right\}^2
\end{aligned}$$

Let $\mathbb{E}(Y_1|x) = g(x)$ and $\varphi(x) = g(x)f(x)$, then

$$\begin{aligned}
Var(Z_2) &= h^p\int_{\mathbb{R}^p}K^2(U)\varphi(x - hU)dU - \left[h^p\int_{\mathbb{R}^p}K(U)\varphi(x - hU)dU\right]^2 \\
&\leq h^p\int_{\mathbb{R}^p}K^2(U)\varphi(x - hU)dU \leq \overline{\varphi}h^p\int_{\mathbb{R}^p}K^2(U)dU
\end{aligned}$$

where $\overline{\varphi}$ is the upper bound of $\varphi(\cdot)$ and $\overline{\varphi} > 0$. Let $M_1 = \overline{\varphi}\int K^2(U)dU < \infty$. Note the the left side does not depend on x. Hence

$$n^{1+q}\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=1}^{n}\left[K\left(\frac{X_l-X_1}{h}\right)Y_{1l}-\mathbb{E}\left(K\left(\frac{X_l-X_1}{h}\right)Y_{1l}\Big|X_1\right)\right]\right|\geq 0.5n^{-\gamma}\right\}$$
$$\leq 2n^{1+q}\int_{\mathbb{R}^p}\exp\left\{-\frac{\left[0.5n^{1-\gamma}h^p-2\overline{K}\right]^2}{2(n-1)Var[Z_2(x)]+2\overline{K}n^{1-\gamma}h^p/3}\right\}dF_X$$
$$\leq 2n^{1+q}\exp\left\{-\frac{\left[0.5n^{1-\gamma}h^p-2\overline{K}\right]^2}{2(n-1)M_1h^p+2\overline{K}n^{1-\gamma}h^p/3}\right\}$$

When n is large enough, the RHS is bounded by $n^{1+q}\exp\left(-Mn^{1-2\gamma}h^p\right)$ for some $M>0$, which converges to zero due to the Assumption 19. □

Lemma 6 *Under Assumption 19 to 21, for any $q>0$,*

$$n^{1+q}\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=1}^{n}\mathbb{E}\left[K\left(\frac{X_l-X_1}{h}\right)Y_{1l}\Big|X_1\right]-f(X_1)\mathbb{E}(Y_{11}|X_1)\right|\geq 0.5n^{-\gamma}\right\}\to 0$$

proof: For $x\in\mathscr{X}$, consider

$$\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=1}^{n}\mathbb{E}\left[K\left(\frac{X_l-X_1}{h}\right)Y_{1l}\Big|X_1\right]-f(X_1)\mathbb{E}(Y_{11}|X_1)\right|\geq 0.5n^{-\gamma}\Big|X_1=x\right\}$$
$$=\mathbb{P}\left[\left|\frac{n-1}{nh^p}\mathbb{E}\left[K\left(\frac{X_2-x}{h}\right)Y_{12}\right]+\frac{K(0)}{nh^p}\mathbb{E}(Y_{11}|X_1=x)-f(x)\mathbb{E}(Y_{11}|X_1=x)\right|\geq 0.5n^{-\gamma}\right]$$
$$=\mathbb{P}\left\{\left|\frac{1}{h^p}\mathbb{E}\left[K\left(\frac{X_2-x}{h}\right)Y_{12}\right]-\frac{1}{nh^p}\mathbb{E}\left[K\left(\frac{X_2-x}{h}\right)Y_{12}-K(0)g(x)\right]-f(x)g(x)\right|\geq 0.5n^{-\gamma}\right\}$$

Because $|K(\cdot)|\leq\overline{K}$ and $0\leq g(x)\leq 1$ and $0.5n^{-\gamma}-\overline{K}/nh^p>0$ when n is large enough, the LHS will be less than

$$\mathbb{P}\left[\left|\int_{\mathbb{R}^p}K(U)\,\varphi(x-hU)dU-\varphi(x)\right|\geq 0.5n^{-\gamma}-\frac{2\overline{K}}{nh^p}\right].\tag{2.13}$$

By Assumption 21, $\varphi(x)$ is twice continuously differentiable function and its second derivative is bounded, then

$$\int_{\mathbb{R}^p}K(U)\,\varphi(x-hU)dU-\varphi(x)=\frac{1}{2}h^2\int_{\mathbb{R}^p}\sum_{k=1}^{p}\sum_{k'=1}^{p}\frac{\partial^2\varphi(x^\dagger)}{\partial x_k\partial x_{k'}}U_kU_{k'}K(U)dU.$$

where $x^\dagger$ depends on the value of x and hU. By Assumption 20 (2), there exists some $M_2>0$ which does not depend on x such that

$$\left|\int_{\mathbb{R}^p}K(U)\,\varphi(x-hU)dU-\varphi(x)\right|\leq M_2h^2.$$

Then, equation (2.13) is less than

$$\mathbb{P}\left(M_2 h^2 \geq 0.5 n^{-\gamma} - \frac{\overline{K}}{nh^p}\right)$$

Hence,

$$\mathbb{P}\left\{\left|\frac{1}{nh^p}\sum_{l=1}^{n}\mathbb{E}\left[K\left(\frac{X_l - X_1}{h}\right)Y_{1l}\Big|X_1\right] - f(X_1)\mathbb{E}(Y_{11}|X_1)\right| \geq 0.5n^{-\gamma}\right\}$$
$$\leq \int_{\mathbb{R}^p}\mathbb{P}\left(M_2h^2 \geq 0.5n^{-\gamma} - \frac{\overline{K}}{nh^p}\right)dF_X = \mathbb{P}\left(M_2h^2 \geq 0.5n^{-\gamma} - \frac{\overline{K}}{nh^p}\right)$$

By Assumption 19, then the RHS of above term equals to zero when n is large enough. □

Lemma 7 *Under Assumption 24, if* $\|\hat{\Lambda} - \Lambda\| = o_p(1)$ *and* $\mathbb{P}(\Lambda \not\subseteq \hat{\Lambda}) = o(1)$*, then*

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\left[1\left(X_i \in \Psi(\hat{\Lambda})\right) - 1\left(X_i \in \Psi(\Lambda_0)^o\right)\right]\log P_\theta(Y_i|X_i)\right\}\right| = o_p(1)$$

proof: Let $\Gamma_n = \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\left[1\left(X_i \in \Psi(\hat{\Lambda})\right) - 1\left(X_i \in \Psi(\Lambda_0)^o\right)\right]\log P_\theta(Y_i|X_i)\right\}\right|$. Then for any $\epsilon > 0$ and $\xi > 0$,

$$\begin{aligned}
\mathbb{P}(|\Gamma_n| > \epsilon) &= \mathbb{P}\left(|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}\right) + \mathbb{P}\left(|\Gamma_n| > \epsilon; \Lambda_0 \not\subseteq \hat{\Lambda}\right)\\
&\leq \mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right] + \mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) > \xi\right] + \mathbb{P}\left(\Lambda_0 \not\subseteq \hat{\Lambda}\right)\\
&\leq \mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right] + \mathbb{P}\left[d(\Lambda_0, \hat{\Lambda}) > \xi\right] + \mathbb{P}\left(\Lambda_0 \not\subseteq \hat{\Lambda}\right)
\end{aligned}$$

Hence

$$\lim_{n\to\infty}\mathbb{P}(|\Gamma_n| > \epsilon) \leq \lim_{n\to\infty}\mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right] \tag{2.14}$$

Let $Y_n = \frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \Psi(\Lambda_0); X_i \notin \Psi(\hat{\Lambda})\right] \times \sup_{\theta\in\Theta}|\log P_\theta(Y_i|X_i)|$. Because $|\Gamma_n| \leq |Y_n|$, then

$$\begin{aligned}
\mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right] &\leq \mathbb{P}\left[|Y_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right]\\
&\leq \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \Psi(\Lambda_0)^o; X_i \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta\in\Theta}|\log P_\theta(Y_i|X_i)| > \epsilon\right\},
\end{aligned}$$

where $\Lambda_\xi = \{\theta \in \Theta : d(\theta, \Lambda_0) \leq \xi\}$. By Chebyshev inequality, it follows that

$$\begin{aligned}
&\mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right]\\
&\leq \frac{\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \Psi(\Lambda_0)^o; X_i \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta\in\Theta}|\log P_\theta(Y_i|X_i)|\right\}}{\epsilon}\\
&= \frac{\mathbb{E}\left\{1\left[X \in \Psi(\Lambda_0)^o; X \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta\in\Theta}|\log P_\theta(Y|X)|\right\}}{\epsilon}
\end{aligned}$$

Because $\lim_{\xi \downarrow 0} 1\left[X \in \Psi(\Lambda_0)^o; X \notin \Psi(\Lambda_\xi)\right] = 0$ and with Assumption 24, there is

$$\mathbb{E}\left\{1\left[X \in \Psi(\Lambda_0)^o; X \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta \in \Theta} |\log P_\theta(Y|X)|\right\} \leq \mathbb{E}\left\{\sup_{\theta \in \Theta} |\log P_\theta(Y|X)|\right\} < \infty.$$

Then

$$\lim_{\xi \downarrow 0} \mathbb{E}\left\{1\left[X \in \Psi(\Lambda_0)^o; X \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta \in \Theta} |\log P_\theta(Y|X)|\right\} = 0$$

Because equation (2.14) holds for all $\xi > 0$, then

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(|\Gamma_n| > \epsilon\right) &\leq \lim_{\xi \downarrow 0} \lim_{n\to\infty} \mathbb{P}\left[|\Gamma_n| > \epsilon; \Lambda_0 \subseteq \hat{\Lambda}; d(\Lambda_0, \hat{\Lambda}) \leq \xi\right] \\
&\leq \lim_{\xi \downarrow 0} \lim_{n\to\infty} \frac{\mathbb{E}\left\{1\left[X \in \Psi(\Lambda_0)^o; X \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta \in \Theta} |\log P_\theta(Y|X)|\right\}}{\epsilon} \\
&= \lim_{\xi \downarrow 0} \frac{\mathbb{E}\left\{1\left[X \in \Psi(\Lambda_0)^o; X \notin \Psi(\Lambda_\xi)\right] \times \sup_{\theta \in \Theta} |\log P_\theta(Y|X)|\right\}}{\epsilon} = 0.
\end{aligned}$$

□

Lemma 8 *Assume $\tilde{\theta} \xrightarrow{p} \theta_0$, then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1\left[X_i \in \Psi\left(B_\epsilon(\tilde{\theta})\right)\right] s(Y_i, X_i; \theta_0) \xrightarrow{d} \mathbb{N}(0, V_\epsilon)$$

proof: This proof is expressed in standard empirical process notation. Let $Z = (X, Y)$ and $\mathscr{Z}$ be the support of Z. For a measurable function $h : \mathscr{Z} \times \Theta \to \mathbb{R}^d$, let $\mathbb{P}_n h(\cdot, \theta) = n^{-1} \sum_{i=1}^{n} h(Z_i, \theta)$ and $\mathbb{P}h(\cdot, \theta) = \int_{\mathscr{Z}} h(z, \theta) P(dz)$.

Let $h(z, \theta) = 1\left[x \in \Psi\left(B_\epsilon(\theta)\right)\right] s(y, x; \theta_0)$, $\mathbb{G}_n(\theta) = \mathbb{P}_n h(\cdot, \theta) - \mathbb{P}h(\cdot, \theta)$. Because $1\left[x \in \Psi(B_\epsilon(\theta))\right]$ indexed by θ is a VC class of functions, then by empirical processes method (see Pollard (1989)), for every sequence of positive numbers $\{\delta_n\}$ converging to zero that

$$\sup\left\{n^{1/2}\left|\mathbb{G}_n(\theta) - \mathbb{G}_n(\theta_0)\right| : \|\theta - \theta_0\| \leq \delta_n\right\} = o_p(1).$$

which implies that

$$\begin{aligned}
n^{1/2}\mathbb{P}_n h(\cdot, \tilde{\theta}) &= n^{1/2}\mathbb{G}_n(\tilde{\theta}) + n^{1/2}\mathbb{P}h(\cdot, \tilde{\theta}) \\
&= n^{1/2}\left[\mathbb{G}_n(\tilde{\theta}) - \mathbb{G}_n(\theta_0)\right] + n^{1/2}\mathbb{G}_n(\theta_0) + n^{1/2}\mathbb{P}h(\cdot, \tilde{\theta}) \\
&= o_p(1) + n^{1/2}\mathbb{P}_n h(\cdot, \theta_0) + n^{1/2}\left[\mathbb{P}h(\cdot, \tilde{\theta}) - \mathbb{P}h(\cdot, \theta_0)\right] \xrightarrow{d} \mathbb{N}(0, V_\epsilon)
\end{aligned}$$

The last step comes from the fact that (1) $n^{1/2}\mathbb{P}_n h(\cdot, \theta_0) \xrightarrow{d} \mathbb{N}(0, V_\epsilon)$; (2)$\mathbb{P}h(\cdot, \theta_0) = 0$; (3) when $\|\tilde{\theta} - \theta_0\| \leq \epsilon$, $\Psi\left(B_\epsilon(\tilde{\theta})\right) \subseteq \Pi_3(\theta_0)$. Thus $\mathbb{P}h(\cdot, \tilde{\theta}) = 0$. □

Lemma 9 *Assume* $\sqrt{n}(\bar{\theta} - \theta_0) = O_p(1)$*, then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1\left[X_i \in \Psi\left(B_{\epsilon_n}(\bar{\theta})\right)\right] s(Y_i, X_i; \theta_0) \xrightarrow{d} \mathbb{N}(0, V)$$

proof: Similarly as the proof of 8, let $\tilde{h}(\cdot, \theta, \epsilon) = 1\left[x \in \Psi\left(B_\epsilon(\theta)\right)\right] s(y, x; \theta_0)$. Let further $\mathbb{P}_n h(\cdot, \theta, \epsilon) = n^{-1} \sum_{i=1}^{n} h(Z_i, \theta, \epsilon)$, $\mathbb{P} h(\cdot, \theta, \epsilon) = \int_{\mathscr{Z}} h(z, \theta, \epsilon) P(dz)$ and $\mathbb{G}_n(\theta, \epsilon) = \mathbb{P}_n h(\cdot, \theta, \epsilon) - \mathbb{P} h(\cdot, \theta, \epsilon)$. Then

$$\begin{aligned}
n^{1/2} \mathbb{P}_n h(\cdot, \bar{\theta}, \epsilon_n) &= n^{1/2} \mathbb{G}_n(\bar{\theta}, \epsilon_n) + n^{1/2} \mathbb{P} h(\cdot, \bar{\theta}, \epsilon_n) \\
&= n^{1/2}\left[\mathbb{G}_n(\bar{\theta}, \epsilon_n) - \mathbb{G}_n(\theta_0, 0)\right] + n^{1/2} \mathbb{G}_n(\theta_0, 0) + n^{1/2} \mathbb{P} h(\cdot, \bar{\theta}, \epsilon_n) \\
&= o_p(1) + n^{1/2} \mathbb{P}_n h(\cdot, \theta_0, 0) + n^{1/2}\left[\mathbb{P} h(\cdot, \bar{\theta}, \epsilon_n) - \mathbb{P} h(\cdot, \theta_0, 0)\right] \xrightarrow{d} \mathbb{N}(0, V).
\end{aligned}$$

□

Chapter 3

Semiparametric Estimation of Binary Decision Games of Incomplete Information with Correlated Private Signals

3.1 Introduction

This paper studies the identification and estimation of a static binary decision game of incomplete information. We make no parametric assumptions on the joint distribution of private signals and allow them to be correlated. We show that the parameters of interest can be point–identified subject to a scale normalization under mild support requirements for the regressors (publicly observed states) and errors (private signals). Following Manski and Tamer (2002b), we use the modified maximum score estimator with estimated bounds and show its consistency in our setting.

Static binary decision game models have many applications. Bjorn and Vuong (1984), for example, study labor force participation. Binary decision game models are more widely adopted in the empirical industrial organization literature to study firms' entry behavior (e.g.Berry (1992), Bresnahan and Reiss (1991a,b), Ciliberto and Tamer (2010), Jia (2008)). It is also applied to study the peer effects in individual decisions (Krauth (2007)).

Our paper contributes to the literature in several respects. First, we do not require (conditional) independence of private signals across players, as Aguirregabiria and Mira

(2007b), Bajari et al. (2009), de Paula and Tang (2010),[1] Grieco (2010), Pesendorfer and Schmidt-Dengler (2003b) and Tang (2009) do. Instead, we assume that the private signals are positively regression dependent, conditional on the publicly observed states, which includes (conditional) independence as a special case. Positive regression dependence is a weaker notion of positive dependence than alternatives such as positive affiliation or decreasing inverse hazard rates, but stronger than positive correlation. We provide a numerical example showing that ignoring the correlation between the private signals results in inconsistent estimates.

Second, we do not make any parametric assumption on the joint distribution of private signals, which distinguishes our paper from Xu (2009), which assumes that the private signals are jointly normally distributed. Aradillas-Lopez (2008) proposes an estimation procedure for incomplete information games with linear payoff functions without parametric assumptions on private signals. In his paper, the belief of a player is the conditional expectation of the opponent's action given publicly observed states and her own action. In contrast, we employ the Bayesian Nash Equilibrium (BNE) concept,[2] where the conditioning variables are not only the publicly observed states but also the player's private signal.

Third, to accommodate heteroskedasticity of unknown form, instead of conditional mean independence, we assume conditional quantile independence of the private signals given the publicly observed states.[3]

We achieve point identification at infinity (of regressor values). We show that when players play a Monotone Strategy Bayesian Equilibrium (MSBE), the equilibrium strategies can be represented by a semiparametric binary response model with an unobserved regressor. We derive (estimable) bounds for the unobserved regressor. With further mild assumptions on the support of regressors (similar to those in Manski (1975)) and the tail behavior of the errors,[4] the bounds on the unobserved regressor can be shown to converge to each other at the infinity of regressor values. Point identification then follows in the same way as with the maximum score estimator. The strategy of identification at infinity is first adopted by Chamberlain (1986), Heckman (1990) and Manski (1975, 1985, 1988). In the

[1] de Paula and Tang also study the case in which private signals are dependent. For this case, they assume that game observations are clustered into (non-singleton) groups such that the equilibrium played within each cluster is the same (their assumption 3).

[2] See Fudenberg and Tirole (1991), Myerson (1997) and Osborne and Rubinstein (1999).

[3] We assume that conditional on the publicly observed states, the medians of private signals equal to 0 without loss of generality.

[4] It is trivially satisfied when the publicly observed states are exogenous.

empirical game literature, it can also be found in Bajari, Hong, and Ryan (2004) and Tamer (2003) in their studies of complete information games with multiple equilibria.

Our estimation approach is different from, *e.g.*, Aguirregabiria and Mira (2007b), Bajari et al. (2009), Pesendorfer and Schmidt-Dengler (2003b), Tamer (2003) and Xu (2009), where (pseudo) likelihood based approaches are used. Aradillas-Lopez (2007) studied a pairwise-differencing estimation procedure for games of incomplete information.[5] Tang (2009) takes a single index approach. Our estimator follows the principles of the maximum score estimator (see Manski (1975, 1985, 1988)). Manski and Tamer (2002b) propose the modified maximum score estimator for regressions with interval data on regressors or outcomes. We apply their approach to our model. We first estimate the bounds on the unobserved regressor and then show that the two–step modified maximum score estimator is consistent.

This paper is organized as follows. In section 2 we introduce the model. In section 3 we discuss point identification. In section 4, we propose a two-step modified maximum score estimator and show its consistency. In section 5, we discuss identification and estimation under weaker conditions. We provide a simple simulation example to illustrate the performance of our estimator in the finite samples in Section 6.

3.2 Model

We study the following 2–by–2 static game of incomplete information:

	$Y_2 = 1$	$Y_2 = 0$
$Y_1 = 1$	$X_1'\beta_1 - \alpha_1 - U_1, X_2'\beta_2 - \alpha_2 - U_2$	$X_1'\beta_1 - U_1, 0$
$Y_1 = 0$	$0, X_2'\beta_2 - U_2$	$0, 0$

In this game, two risk–neutral players, $j = 1, 2$, each simultaneously make choice $Y_j \in \{0, 1\}$. The first number in each cell is the payoff of player 1 under the corresponding choice profile. A state of the game is (X, U), where $X = (X_1', X_2')'$ and $U = (U_1, U_2)'$. $X_j \in \mathscr{X}_j \subseteq \mathbb{R}^{d_j}$, $d_j \in \mathbb{N}^+$, is a vector of publicly observed variables. $U_j \in \mathbb{R}$ is the private signal observed only by player j. Let F_{XU} be the joint distribution function of (X, U). We assume that F_{XU} is common knowledge.

In this structure, $\beta_j \in \mathbb{R}^{d_j}$ and $\alpha_j \in \mathbb{R}$ are the parameters of interest. α_j measures the

[5]Aradillas-Lopez (2008) discusses briefly the possibility of applying Khan and Tamer (2009) methods to the incomplete information game in which belief is based on unobservables.

strategic effect: how the action of the other player ($-j$) affects the payoff of player j choosing $Y_j = 1$. Let $\beta = (\beta_1', \beta_2')'$, $\alpha = (\alpha_1, \alpha_2)'$. We assume $\alpha_j \geq 0$ for $j = 1, 2$.[6]

Player j chooses action 1 if and only if her expected payoff is greater than 0,

$$Y_j = \mathbf{1}\left[X_j'\beta_j - \alpha_j \mathbb{P}(Y_{-j} = 1|X, U_j) - U_j \geq 0\right], \tag{3.1}$$

where $\mathbf{1}[\cdot]$ is the indicator function. The term $\mathbb{P}(Y_{-j} = 1|X, U_j)$ is player j's expectation about the other player's action (based on player j's information).

A special type of BNE that we are interested in is the MSBE. With an MSBE, the equilibrium strategies are weakly monotone functions, *i.e.*, there exists a cutoff–value function profile $u^* = (u_1^*, u_2^*) : \mathscr{X} \to \mathbb{R}^2$ such that for each j,

$$Y_j = \mathbf{1}\left[U_j \leq u_j^*(X)\right]. \tag{3.2}$$

Athey (2001), among others, establishes a sufficient condition for the existence of an MSBE in finite actions games, namely a *Single Crossing Condition (SCC)*. In the model we are considering here, the *SCC* is satisfied if assumption 27 holds.

Assumption 27 *The conditional density $f_{U|X}(\cdot|\cdot)$ of (U_1, U_2) given X exists. For $j \in \{1, 2\}$, for any $t \in \mathbb{R}$ and $x \in \mathscr{X}$, $u_j + \alpha_j \mathbb{P}\left(U_{-j} \leq t | X = x, U_j = u_j\right)$ is non–decreasing in u_j.*

Assumption 27 implies that given her opponent plays a monotone strategy, player i's best response is nondecreasing in her private signal. It imposes a restriction on the conditional copula function $C(\cdot, \cdot, X)$ of U_1 and U_2 given X. [7] Assumption 27 is trivially satisfied when U_1 and U_2 are independent, conditional on X. In a special case where U

[6] Other cases can be analyzed in a similar way.

[7] We use a simple case to illustrate this. Suppose that U and X are independent, then

$$\begin{aligned} u_1 + \alpha_1 \mathbb{P}\left(U_2 \leq u_2 | X = x, U_1 = u_1\right) &= u_1 + \alpha_1 F_{U_2|U_1}(u_2|u_1) \\ &= u_1 + \alpha_1 \frac{\partial F(u_1, u_2)/\partial u_1}{f_{U_1}(u_1)} = u_1 + \frac{\alpha_1}{f_{U_1}(u_1)} \frac{\partial C\left(F_{U_1}(u_1), F_{U_2}(u_2)\right)}{\partial u_1}. \end{aligned}$$

Suppose that $\alpha_1 > 0$ and the copula function is twice differentiable, then the condition stated in the assumption says that

$$\frac{\partial^2 C(s_1, s_2)}{\partial s_1^2} \geq -\frac{1}{\alpha_1 f_{U_1}(F_{U_1}^{-1}(s_1))}.$$

Similarly, we have

$$\frac{\partial^2 C(s_1, s_2)}{\partial s_2^2} \geq -\frac{1}{\alpha_2 f_{U_2}(F_{U_2}^{-1}(s_2))}.$$

is parametrized to be bivariately normal distributed with mean zero and unit variance, assumption 27 implies that the correlation between U_1 and U_2 is bounded above (below one), vis–a–vis the magnitude of α.

Lemma 10 *Suppose that assumption 27 holds. Then for any $X = x \in \mathscr{X}$, there exists an MSBE with equilibrium strategies characterized by equation (3.2). proof: See Appendix 3.7.*

Note that assumption 27 does not rule out the possibility of the existence of Non–Monotone Strategy Bayesian Equilibria. Under assumption A, however, we implicitly assume in the following analysis that only one MSBE is played in equilibrium. (Weak) Monotonicity property of the equilibrium strategies is widely assumed, or implicitly made in the empirical game models of incomplete information, for instance in Guerre, Perrigne, and Vuong (2000). The conditional independence assumption in the literature also implies that all BNEs in the models are MSBE.

Furthermore, assumption 27 is a sufficient condition for the existence of MSBE. Our analysis stays valid as long as an MSBE exists. In addition, if there is an identifiable subset of $\mathscr{X}$ on which an MSBE is played, our analysis can still be applied by using data belonging to the subset.

3.3 Identification

In this section we study the identification of (α, β). Econometricians observe X and Y, but not the private signals U. Our identification strategy is as follows. We first derive (estimable) upper and lower bounds for the equilibrium strategies of the players under the assumption that U_1 and U_2 are positively regression dependent conditional on X (assumption 28 below). Then we show that the bounds can be arbitrarily close to each other when some argument of the regressors goes to infinity (Lemma 11). These results delivers the point–identification of (α, β).

Assumption 28 ***(positive regression dependence)*** *For $j \in \{1, 2\}$, for any $x \in \mathscr{X}$ and any $t \in \mathbb{R}$, $\mathbb{P}(U_{-j} \leq t | X = x, U_j = u_j)$ is non–increasing in u_j.*

Positive regression dependence is a weaker notion of positive dependence than other alternatives such as positive affiliation and decreasing inverse hazard rate.[8]

[8] de Castro (2007) provides an example. Let the density of $U = (u_1, u_2) \in [0, 1]^2$ be

$$f(u_1, u_2) = k/[1 + (u_1 - u_2)^2]$$

Under assumption 28, we can represent the equilibrium strategies by a semiparametric binary regression model with interval data (see Manski and Tamer (2002b)). To motivate, suppose that $\mathbb{P}(U_{-j} \leq t | X = x, U_j = u_j)$ is continuous in u_j for all $t \in \mathbb{R}$ and $x \in \mathscr{X}$ (we actually do not assume this in the result of Theorem 15). Given the equilibrium is MSBE, for any $x \in \mathscr{X}$, when player j receives a private signal with value $u_j^*(x)$, her expected payoff equals to zero,

$$x_j' \beta_j - \alpha_j \mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right) - u_j^*(x) = 0. \tag{3.3}$$

Then it follows from the definition of MSBE,

$$Y_j = 1[U_j \leq u_j^*(x)] = 1\left[U_j \leq x_j' \beta_j - \alpha_j \mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right)\right]. \tag{3.4}$$

Note that the term $\mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right)$ is an unobservable regressor of the binary response model characterized by equation (3.4). Further, let $\nu_j^0(x) = \mathbb{P}\left(Y_{-j} = 1 | X = x, Y_j = 0\right)$ and $\nu_j^1(x) = \mathbb{P}\left(Y_{-j} = 1 | X = x, Y_j = 1\right)$. Assumption 28 implies that,

$$\begin{aligned} \nu_j^0(x) = \mathbb{P}&\left[U_{-j} \leq u_{-j}^*(x) | X = x, U_j > u_j^*(x)\right] \\ &\leq \mathbb{P}\left[U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right] \\ &\leq \mathbb{P}\left[U_{-j} \leq u_{-j}^*(x) | X = x, U_j \leq u_j^*(x)\right] = \nu_j^1(x). \end{aligned}$$

Hence the unobserved regressor, as a function of x, is bounded by a pair of estimable functions. The result is formally summarized by Theorem 15.

Theorem 15 *Suppose that assumption 27 and 28 are satisfied. Then for player $j = 1, 2$, the equilibrium strategy of player j is represented by the structure*

$$Y_j = \mathbf{1}\left[U_j \leq X_j' \beta_j - \alpha_j \nu_j(X)\right]; \quad \mathbb{P}\left(\nu_j^0(X) \leq \nu_j(X) \leq \nu_j^1(X)\right) = 1, \tag{3.5}$$

where

$$\nu_j(x) = \mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right). \tag{3.6}$$

proof: See Appendix 3.7.

---

for some $k > 0$. Then U_1 and U_2 is positive regression dependent, but the other two properties don't hold.

We have a few comments on Theorem 15. First, the bounds ν_j^0 and ν_j^1 are estimable. We study the identification taking ν_j^0 and ν_j^1 as known. Second, if U_1 and U_2 are independent conditional on X, then $\nu_j^1(x) = \nu_j^0(x)$ for all $x \in \mathscr{X}$. If the inequality in assumption 28 holds strictly on a subset $\tilde{\mathscr{X}_j}$ of $\mathscr{X}_j$ (a violation to conditional independence), then $\nu_j^1(x) > \nu_j^0(x)$ for all $x \in \tilde{\mathscr{X}_j}$. Theorem 2 thus provides a testable implication for conditional independence provided players play one MSBE.[9] Further, we can relax Assumption 28 by requiring $\mathbb{P}(U_{-j} \leq t | X = x, U_j = u_j)$ be monotonic in u_j for all t and x, i.e., our model allows for both positive and negative regression dependence. Similar result as stated in Theorem 15 follows by redefining ν_j^1 and ν_j^0. We maintain assumption 28 for the ease of notation.

Assumption 29 Med $\left(U_j | X = x\right) = 0$ *for* $j \in \{1, 2\}$ *and all* $x \in \mathscr{X}$.

Assumption 30 *For* $j = 1, 2$, *there exists no proper linear subspace of* $\mathbb{R}^{d_j}$ *having probability 1 under* F_{X_j}.

Assumptions 29 and 30 are also made in Manski (1985) for a single agent binary response model. Assumption 29 imposes a conditional median independence restriction on the private signals and normalizes the median to be 0. This assumption allows for heteroskedasticity of unknown form. Assumption 30 excludes multicollinearity.

Let $X_{j,1}$ be the first regressor for player j. Let $\tilde{X}_j = (X_{j,2}, \cdots, X_{j,d_j})'$. We define $\beta_{j,1}$ and $\tilde{\beta}_j$ in the same way.

Assumption 31 $\beta_{j,1} \neq 0$. *The distribution of* $X_{j,1}$ *conditional on* $(\tilde{X}_j, X_{-j})$ *has everywhere positive density with respect to the Lebesgue measure.*

Assumption 31 requires that for each player there exists a special regressor which is continuously distributed and has unbounded support conditional on the rest of regressors. Manski (1985) makes a assumption similar to assumption 31. We require that the conditioning variables include not only the rest of the regressors of player j, but also all of player $-j$'s regressors, which implies an exclusive restriction to the model. This requirement excludes the possibility of using a state variable that is common to both players, *e.g.*, a macroeconomic variable, as the special regressor. Similar exclusion restriction can also be found in Bajari, Chernozhukov, Hong, and Nekipelov (2009) and de Paula and Tang (2010).

[9] We thank Aureo de Paula for his comments on this. In this paper, we focus on identification and estimation and leave rigorous investigation on this issue as future research.

Assumption 32 *For all $u \in \mathbb{R}$ and all $x_j \in \mathscr{X}_j$,*

$$\lim_{t\to+\infty} \mathbb{P}(U_{-j} \leq t - \alpha_{-j} | X_j = x_j, X'_{-j}\beta_{-j} = t, U_j \geq u) = 1,$$
$$\lim_{t\to-\infty} \mathbb{P}(U_{-j} \leq t | X_j = x_j, X'_{-j}\beta_{-j} = t, U_j \leq u) = 0.$$

Assumption 32 requires that the conditional tail probability of U_{-j} is arbitrarily small when the conditioning variable $X'_{-j}\beta_{-j}$ approaches $\pm\infty$. Assumption 32 is trivially satisfied when U and X are independent, or when the support of the distribution of U is bounded.

Lemma 11 *Suppose that assumptions 27 through 32 hold. Then for all $\epsilon > 0$ and all $x_j \in \mathscr{X}_j$,*

$$\lim_{t\to+\infty} \mathbb{P}\left(\nu_j^0(X) \leq 1 - \epsilon | X_j = x_j, X'_{-j}\beta_{-j} = t\right) = 0,$$
$$\lim_{t\to-\infty} \mathbb{P}\left(\nu_j^1(X) \leq \epsilon | X_j = x_j, X'_{-j}\beta_{-j} = t\right) = 1.$$

proof: See Appendix 3.7.

Note that since $\nu_j^0(x) \leq \nu_j^1(x) \leq 1$ for all x, it follows from Lemma 11 that both ν_j^0 and ν_j^1 converge to 1 in probability as $x'_{-j}\beta_{-j}$ goes to $+\infty$. Meanwhile, both bounds converge to 0 in probability as $x'_{-j}\beta_{-j}$ goes to $-\infty$.

In Manski (1975, 1985) and Manski and Tamer (2002b), the parameters of interest can only be identified up to scale. For the same reason, we can only identify (α_j, β_j) up to scale for each j in our model. We normalize $|\beta_{j,1}| = 1$.

Theorem 16 *Suppose that assumptions 27 through 32 hold. Then (α, β) is point identified.*
proof: See Appendix 3.7.

Note that, the special regressor in our model plays an extra role under assumption 32 other than in Manski (1985): when $x'_1\beta_1$ goes to $+\infty$ $(-\infty)$, both ν_2^1 and ν_2^0 converge to 1 (0). Thus we take advantage of both infinities, at which the unobserved regressor behaves as an observed 0–1 variable. We can then achieve point identification of (α, β) in the same way as the maximum score estimator does.

3.4 Estimation

We follow the estimation approach proposed by Manski and Tamer (2002b). The difference is that we non–parametrically estimate the unobserved bounds ν_j^0 and ν_j^1 in the first step.

Assumption 33 *Let* $Z_i = (X_i', Y_i')' \in \mathbb{R}^{d_1+d_2+2}$ *for* $i = 1, 2, \cdots, n$ *be an i.i.d sample, where* $Y_i = (Y_{1i}, Y_{2i})'$ *and* $X_i = (X_{1i}', X_{2i}')'$.

Let $p_j(x) = \mathbb{P}(Y_j = 1|X = x)$ and $\delta_j(x) = 1\left[p_j(x) \geq \frac{1}{2}\right]$. To keep notations simple, we write h_j for $\left(p_j, \nu_j^0, \nu_j^1\right)$ and h for (h_1, h_2). Let

$$\Theta_j = \{(a_j, b_j) \in \mathbb{R} \times \mathbb{R}^{d_j} : |b_{j,1}| = 1;\ \|(a_j, b_j)\| \leq \mathbb{M}\|\}.$$

and $\Theta = \Theta_1 \times \Theta_2$, where $\mathbb{M}$ is a large positive number to ensure Θ is compact and $(\alpha_j, \beta_j) \in \Theta_j$. Let further $a = (a_1, a_2)$, $b = (b_1, b_2)$ and $\operatorname{sgn}(\cdot)$ be the sign function.[10]

Lemma 12 *Suppose that assumption 27 through 32 are satisfied, then*

$$(\alpha, \beta) = \underset{(a,b)\in\Theta}{\operatorname{argmax}}\, L(a, b, h),$$

where $L(a, b, h) = \sum_{j=1}^{2} \int g_j\left(Z, a_j, b_j, h_j\right) dF_z$ *and*

$$g_j\left(Z, a_j, b_j, h_j\right) = \left(2Y_j - 1\right) \times \left\{\delta_j(X) \operatorname{sgn}\left[X_j' b_j - a_j \nu_j^0(X)\right] + \left(1 - \delta_j(X)\right) \operatorname{sgn}\left[X_j' b_j - a_j \nu_j^1(X)\right]\right\}.$$

proof: See Appendix 3.7.

Let $\hat{h} = (\hat{h}_1, \hat{h}_2)'$ be a nonparametric estimator for h. Then we define the estimator for (α, β) as:

$$(\hat{\alpha}, \hat{\beta}) = \underset{(a,b)\in\Theta}{\operatorname{argmax}}\, L_n\left(a, b, \hat{h}\right), \tag{3.7}$$

where the sample criterion function L_n is

$$L_n(a, b, h) = \frac{1}{2n} \sum_{j=1}^{2} \sum_{i=1}^{n} g_j\left(Z_i, a_j, b_j, h_j\right). \tag{3.8}$$

We impose the following assumptions to establish the consistency of our estimator.

Assumption 34 (α, β) *is in the interior of the compact parameter space* Θ.

[10] $\operatorname{sgn}(x) = 1, 0$ and -1 respectively when $x > 0$, $x = 0$ and $x < 0$.

Assumption 35 *For $j \in \{1,2\}$, assume $\mathbb{P}\left[p_j(X) = \frac{1}{2}\right] = 0$; Moreover, for any $(a,b) \in \Theta$, assume that $\mathbb{P}\left[X_j' b_j - a_j \nu_j^0(X) = 0\right] = 0$ and $\mathbb{P}\left[X_j' b_j - a_j \nu_j^1(X) = 0\right] = 0$.*

The first equation of assumption 35 is also assumed in Manski and Tamer (2002b).[11] The second and third equations are rank conditions imposed on the augmented random vector $(X, \nu_j^0(X))$ and $(X, \nu_j^1(X))$ respectively. In assumption 31, we assume that $X_{j,1}$ has positive density everywhere with respect to Lebesgue measure, given all the other regressors. Given we normalize $|\beta_{j,1}| = 1$, the second and third equation will be satisfied as long as ν_j^0 and ν_j^1 are continuous function and not linear in $x_{j,1}$.

Assumption 36 *For $j = 1,2$, for any $\delta \in (0, \sup_x f(x)]$, $\sup_{\{x: f(x) \geq \delta\}} |\hat{p}_j(x) - p_j(x)| \xrightarrow{p} 0$, $\sup_{\{x: f(x) \geq \delta\}} |\hat{\nu}_j^0(x) - \nu_j^0(x)| \xrightarrow{p} 0$ and $\sup_{\{x: f(x) \geq \delta\}} |\hat{\nu}_j^1(x) - \nu_j^1(x)| \xrightarrow{p} 0$.*

Assumption 36 requires that the first step estimates be uniformly consistent on any subset of the support where the density of X is bounded away from 0. Bierens (1996b)(Theorem 10.3.1) provides conditions under which assumption 36 is satisfied for kernel estimators.

Theorem 17 *Suppose that assumptions 27 through 36 are satisfied, then $\hat{\alpha} \xrightarrow{p} \alpha$ and $\hat{\beta} \xrightarrow{p} \beta$. proof: See Appendix 3.7.*

3.5 Partial Identification under Weaker Conditions

Assumptions 30, 31 and 32 are support restrictions for the regressors and private signals. Without them, (α, β) is not point–identified. The data however defines a subset of Θ that contains true parameter value. In particular, for $j = 1,2$, let Θ_j^I be the collection of (a_j, b_j) such that for all $x \in \mathscr{X}$,

$$\mathbf{1}[x_j' b_j - a_j \nu_j^1(x) \geq 0] \leq \text{Med}(Y_j | X = x) \leq \mathbf{1}[x_j' b_j - a_j \nu_j^0(x) \geq 0].$$

Let $\Theta^I = \Theta_1^I \times \Theta_2^I$. We call Θ^I the identified set. Note that by Theorem 15, $(\alpha, \beta) \in \Theta^I$ under assumptions 27 and 28 when players play an MSBE.

Following the modified maximum score estimator proposed by Manski and Tamer (2002b), we define the set estimator $\widehat{\Theta}^I$ for Θ^I as

[11] In Manski and Tamer (2002b), it is assumed that $\mathbb{P}\{(x, \nu_1, \nu_0) : \mathbb{P}(y = 1 | x, \nu_1, \nu_0) = 1 - \alpha\} = 0$, where ν_0 and ν_1 are bounds of the unobserved regressor.

$$\widehat{\Theta}^I = \left\{ (a^*, b^*) \in \Theta : L_n\left(a^*, b^*, \hat{h}\right) \geq \sup_{(a,b)\in\Theta} L_n\left(a, b, \hat{h}\right) - \kappa_n \right\} \tag{3.9}$$

for some $\kappa_n \to 0$.

To establish consistency, we adopt a one direction Hausdorff distance measure $\rho(A, B)$,

$$\rho(A, B) = \sup_{a\in A} \inf_{b\in B} \|a - b\|,$$

where $\|\cdot\|$ is the Euclidean norm. When either A or B is empty set, the distance is $+\infty$.

Theorem 18 *Suppose that assumptions 27 though 29, and 33 through 36 are satisfied, then*

$$\rho\left(\widehat{\Theta}^I, \Theta^I\right) \xrightarrow{p} 0.$$

Suppose in addition that $\sup_{(a,b)\in\Theta} \left|L_n\left(a, b, \hat{h}\right) - L(a, b, h)\right| = o_p(\kappa_n)$, *then* $\rho\left(\Theta^I, \widehat{\Theta}^I\right) \xrightarrow{p} 0.$
proof: See Appendix 3.7.

3.6 Experiment

3.6.1 Finite Sample Performance

In this section we use a numerical experiment to illustrate the performance of our estimator in a finite–size sample. Let $d_1 = d_2 = 2$, $X_1 = (X_{1,1}, 1)'$ and $X_2 = (X_{2,1}, 1)'$, where $(X_{1,1}, X_{2,1})$ has a mean zero bivariate normal distribution with identity covariance matrix. Let U_1 and U_2 have a mean zero bivariate normal distribution with variance σ_j^2 and correlation coefficient r. We let U be independent of X.

We let $\sigma_1 = \sigma_2 = 1$, $r = 0.5$, $\beta_{1,2} = \beta_{2,2} = 0$, $\beta_{1,1} = \beta_{2,1} = 1$, $\alpha_1 = \alpha_2 = 1$. It can be shown that an MSBE exists under this design, *i.e.*, for each x, there exist cutoff values $u_1^*(x)$ and $u_2^*(x)$, such that player j chooses 1 whenever her private signal $U_j \leq u_j^*(X)$. We compute $u_j^*(x)$ by solving the following equations for each X in the sample:

$$u_1^* = \beta_{1,1}x_{1,1} + \beta_{1,2} - \alpha_1\Phi\left(\frac{\sigma_2 u_2^* - \rho\sigma_1 u_1^*}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)$$

$$u_2^* = \beta_{2,1}x_{2,1} + \beta_{2,2} - \alpha_2\Phi\left(\frac{\sigma_1 u_1^* - \rho\sigma_2 u_2^*}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\right),$$

where $\Phi(\cdot)$ is the c.d.f of standard normal distribution.

Since U_1 and U_2 are joint normal with positive correlation, they are positively regression dependent. Moreover, assumption 32 is satisfied since U and X are independent. The design also satisfies the support assumptions 30 and 31. Theorem 15 implies that the problem can be transformed into a semiparametric binary regression model:

$$Y_j = \mathbf{1}[\beta_{j,1}X_{j,1} + \beta_{j,2} - \alpha_j \nu_j(X) - U_j \geq 0],$$

with $\mathbb{P}(\nu_j^0(X) \leq \nu_j(X) \leq \nu_j^1(X)) = 1$.

Table 3.1 shows the composition of one random sample with size $n = 3000$.

Table 3.1. Sample composition

Choice profile	percentage
$Y = (1,1)$	13.93%
$Y = (1,0)$	27.17%
$Y = (0,1)$	25.47%
$Y = (0,0)$	33.43%

Figure 3.1 plots ν_1^0, ν_1^1 and p_1 and their kernel estimates. For presentation purpose, we fix $x_{1,1} = 0$, but a similar pattern holds for other values of $x_{1,1}$. The upper-left panel shows p_1 and its estimate $\hat{p}_1$, as functions of $x_{2,1}$. In this design, $\beta_{2,1} > 0$, large $x_{2,1}$ implies the public observed states is in favor of player 2. p_1 is a decreasing function of $x_{2,1}$ for each given $x_{1,1}$ reflects the negative strategy effect ($\alpha_j > 0$). The estimate $\hat{p}$ captures this pattern.

The upper-right panel and lower-left panel show functions ν_1^1 and ν_1^0 and their estimates. When $x_{2,1}$ is large, both lower bounds and upper bound of unobserved belief $\nu(x)$ are close to zero; when $x_{2,1}$ is small, both bounds are close to 0. The lower-right panel shows that the estimates of both bounds contain the true belief $\nu(x)$ well.

We estimate the parameters with kernel estimators. We use standard normal density as the kernel function (second order kernel); we choose bandwidth equals to $1.06n^{-\frac{1+\tau}{6}}$ for small $\tau > 0$ to eliminate the bias.[12] For each sample size, we compute three different estimates: two step modified MSE proposed by our paper, infeasible modified MSE and infeasible MSE. For infeasible modified MSE, we assume that econometricians know the true bounds ν_j^1 and ν_j^0. It is the estimate in Manski and Tamer (2002b), where the bounds are observed by econometricians; for infeasible MSE, we assume that econometricians know the true belief ν_j. Since the latter two estimators use information which is typically not

[12]It should be noted that we could remove bias further by using the technics developed by Bierens (1987).

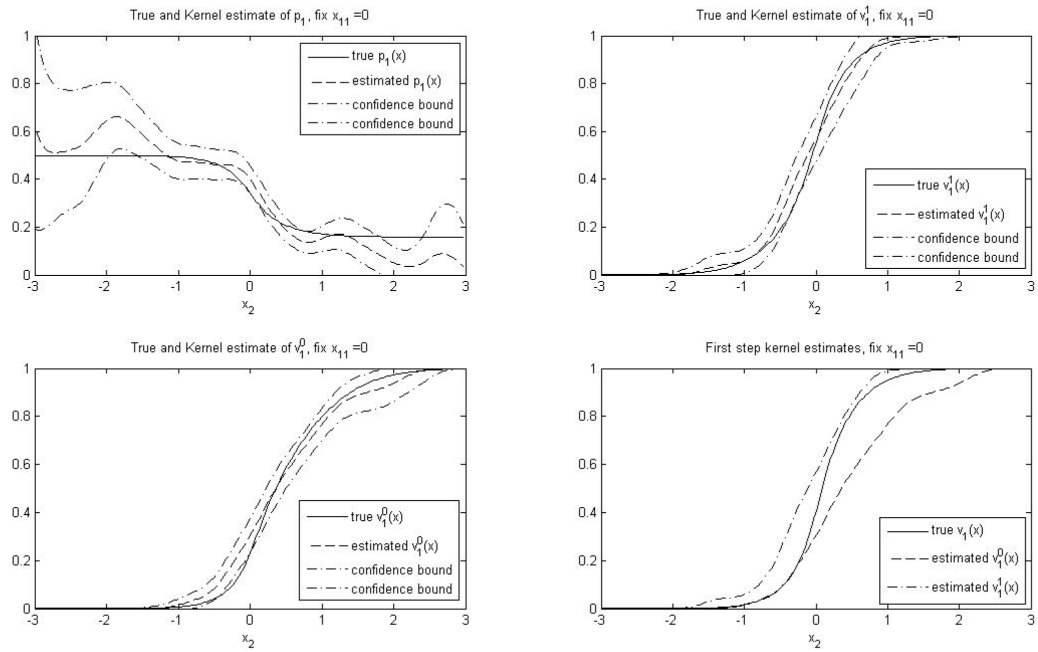

Figure 3.1. Kernel estimates of ν_1^0, ν_1^1, and p_1, fix $x_{1,1} = 0$

available in the applications that we are interested in (*e.g.*, entry games), their performance sets an upper bound for the performance of our estimator.

Table 3.2. Finite sample mean and mean squared error for $\hat{\alpha}_1$

Sample size	Two step modified MSE	Infeasible modified MSE	Infeasible MSE
$n = 1000$	1.115(0.478)	1.074(0.484)	1.010(0.464)
$n = 3000$	1.102(0.356)	1.070(0.336)	1.009(0.295)
$n = 10000$	1.085(0.242)	1.061(0.244)	1.006(0.187)

Root mean squared errors reported in the parentheses.

Table 3.2 reports the finite sample mean of the estimates for α_1, based on 1000 replications (the results for α_2 are similar). The root mean squared errors (RMSE) are reported in parentheses. As we expect, the infeasible modified MSE and infeasible MSE have less finite sample bias and a smaller mean square error than the two step modified MSE. In relatively large samples, the performance of our estimator is very close to the performance of the infeasible modified MSE, suggesting that based on the choices for the kernel and bandwidth, the first step estimation error is negligible. Please note that we achieve point identification when the bounds converge at infinity, where only a small portion of data is available. As a result, the performance of the infeasible MSE is better than the infeasible modified MSE and our estimator.

3.6.2 Two step Modified MSE and misspecified MLE

In this subsection, we provide a simulation example to show that ignoring the correlation between the private signals results in inconsistent estimates. In particular, we investigate the performance of the Maximum Likelihood Estimator (MLE) of which the Conditional Choice Probability (CCP) is estimated in the first step. We consider a data generating process in which U_1 and U_2 have a mean zero bivariate normal distribution with correlation coefficient $r = 0.7$. $(X_{1,1}, X_{2,1})$ is drew from a mixture of two normal distribution: with probability 0.75, $(X_{1,1}, X_{2,1})$ is drew from a zero mean joint normal with covariance equals to 0.16 times identity matrix; with probability 0.25, $(X_{1,1}, X_{2,1})$ is drew from a zero mean joint normal with covariance equals to 100 times identity matrix. Other parameters are set to be the same as in subsection 3.6.1. It can be shown that an MSBE exists under this design.

Consider the following model,

$$Y_j = \mathbf{1}[\beta_{j,1}X_{j,1} + \beta_{j,2} - \alpha_j \mathbb{P}(Y_{-j} = 1|X) - U_j \geq 0]. \tag{3.10}$$

We estimate the term $\mathbb{P}(Y_{-j} = 1|X)$ nonparametricly in the first step, then estimate (3.10) by ML and correctly specified marginal distribution of U_j. Note that when $r \neq 0$, equation (3.10) may still be misspecified since $\mathbb{P}(Y_{-j} = 1|X) \neq \mathbb{P}(Y_{-j} = 1|X, U_j)$ in general.

Table 3.3. Finite sample performance of modified MSE and misspecified MLE

$r = 0.7$	two step modified MSE	Mis–specified MLE
$n = 3,000$	1.199(0.481)	1.683(0.700)
$n = 10,000$	1.089(0.287)	1.628(0.633)

Root mean squared errors reported in the parentheses.

Figure 3.2 plots the kernel smoothed finite sample distribution of both estimates, under sample size $3,000$ and $10,000$ respectively. The true value of the parameter is 1. It can be seen from the figure that in large sample, the inference based on a misspecified MLE may be distorted because of the existence of the bias. Table 3.6.2 reports the mean of the finite sample distribution and the RMSE. Under this design, the RMSE of the misspecified MLE is larger than the two step modified MSE at both sample size. When sample size increases

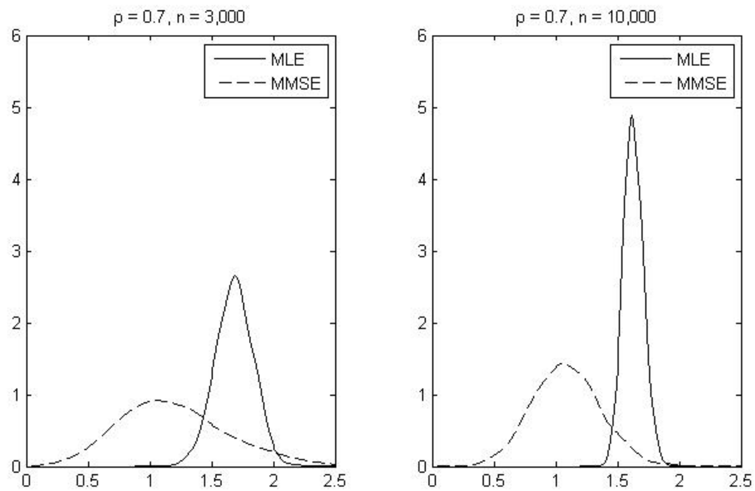

Figure 3.2. Finite sample distribution of modified MSE and misspecified MLE

from $3,000$ to $10,000$, the RMSE of the two step modified MSE decreases more than 40%, while the pseduo MLE RMSE only decreases by 10%.

3.7 Proofs

Proof of Lemma 10

For any given x, considering the following "truncation" on the payoffs of the players when choosing action 1:

$$\begin{aligned}\pi_1^*(x_1, U_1) = \left(x_1'\beta_1 - \alpha_1 \mathbf{1}[Y_2 = 1] - U_1\right) \mathbf{1}[0 < x_1'\beta_1 - \alpha_1 - U_1 < \alpha_1] \\ + \mathbf{1}[x_1'\beta_1 - \alpha_1 - U_1 \geq \alpha_1] - \mathbf{1}[x_1'\beta_1 - \alpha_1 - U_1 \leq 0\alpha_1]. \quad (3.11)\end{aligned}$$

$\pi_2^*(x_2, U_2)$ is similarly defined. We apply Theorem 1 in Athey (2001) to show an MSBE exists in the "truncated game". Note that by construction both $\pi_1^*(x_1, U_1)$ and $\pi_2^*(x_2, U_2)$ are bounded for every given $x \in \mathscr{X}$, which implies that the assumption A1 in Athey (2001) holds. Also note that assumption 27 ensures that the truncated game satisfies the SCC condition. Applying Theorem 1 in Athey (2001), it follows that there exists one MSBE for the "truncated game". It is straightforward to see that this MSBE is also an equilibrium in the original game.

Proof of Theorem 15

When $\mathbb{P}(U_{-j} \leq t|X = x, U_j = u_j)$ is continuous in u_j for any $t \in \mathbb{R}$ and $x \in \mathscr{X}$, the result of Theorem 15 holds by the arguments in section 3. Without this continuity assumption, let,

$$\nu_j^+(x) = \lim_{u_j \downarrow u_j^*(x)} \mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j\right),\ \nu_j^-(x) = \lim_{u_j \uparrow u_j^*(x)} \mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j\right).$$

Under assumption 28, $\mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j\right)$ is a non–increasing function in u_j. So $\nu_j^+(x)$ and $\nu_j^-(x)$ are well defined and $\nu_j^+(x) \leq \nu_j^-(x)$.

From the definition of MSBE it follows that, for all $u_j > u_j^*(x)$

$$x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j\right) - u_i \leq 0.$$

Hence

$$\lim_{u_j \downarrow u_j^*(x)} \left\{x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j\right) - u_j\right\} = x_j'\beta_j - \alpha_j \nu_j^+(x) - u_j^*(x) \leq 0.$$

Similarly, $x_j'\beta_j - \alpha_j \nu_j^-(x) - u_j^*(x) \geq 0$. It implies that $\nu_j^+(x) \geq \nu_j^-(x)$. So we have $\nu_j^+(x) \geq \nu_j^-(x)$. Hence

$$x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j^*(x)\right) - u_j^*(x) = 0.$$

Thus

$$Y_j = 1[U_j \leq u_j^*(x)] = 1\left[U_j \leq x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1|X = x, U_j = u_j^*(x)\right)\right]. \square$$

Proof of Lemma 11

Suppose the first condition in assumption 32 is satisfied, then,

$$\begin{aligned}\nu_1^0(x) &= \mathbb{P}\left(U_2 \leq u_2^*|X = x, Y_1 = 0\right) = \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2\nu_2|X = x, Y_1 = 0\right) \\ &\geq \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2|X = x, Y_1 = 0\right) = \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2|X = x, U_1 \geq u_1^*(x)\right) \\ &\geq \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2|X = x, U_1 \geq x_1'\beta_1\right).\end{aligned}$$

The first inequality is from the fact that $\nu_2 \leq 1$; the second inequality is from the fact $u_1^* \leq x_1'\beta_1$ and assumption 28. Hence, for any $\epsilon \in (0,1)$,

$$\mathbb{P}\left(U_2 \leq t - \alpha_2|X_1 = x_1, X_2'\beta_2 = t, U_1 \geq x_1'\beta_1\right) \leq \mathbb{E}\left[\nu_1^0(X)|X_1 = x_1, X_2'\beta_2 = t\right]$$

$$\leq (1-\epsilon)\mathbb{P}\left(\nu_1^0(X) \leq 1-\epsilon | X_j = x_j, X_2'\beta_2 = t\right) + \mathbb{P}\left(\nu_1^0(X) > 1-\epsilon | X_j = x_j, X_2'\beta_2 = t\right)$$
$$= -\epsilon\mathbb{P}\left(\nu_1^0(X) \leq 1-\epsilon | X_j = x_j, X_2'\beta_2 = t\right) + 1.$$

Let $t \to +\infty$, under assumption 32 we have

$$\lim_{t\to+\infty} \mathbb{P}\left(\nu_1^0(X) \leq 1-\epsilon | X_j = x_j, X_2'\beta_2 = t\right) = 0.$$

Likewise, the second equation of Lemma 11 also holds.

Proof of Theorem 16

This proof follows from the proof of Lemma 2 in Manski (1985).

Fix $\tilde{X}_1 = \tilde{x}_1$ and $\tilde{X}_2 = \tilde{x}_2$ be arbitrary values. It is enough to show that for any $(\tilde{b}_1, a_1) \neq (\tilde{\beta}_1, \alpha_1)$, either $\mathbb{P}(X_{1,1} + \tilde{x}_1'\tilde{b}_1 - a_1\nu_1^1 \geq 0 > X_{1,1} + \tilde{x}_1'\tilde{\beta}_1 - \alpha_1\nu_1^0 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0$ or $\mathbb{P}(X_{1,1} + \tilde{x}_1'\tilde{\beta}_1 - \alpha_1\nu_1^0 \geq 0 > X_{1,1} + \tilde{x}_1'\tilde{b}_1 - a_1\nu_1^1 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0$. Or equivalently, to show,

$$\mathbb{P}(\alpha_1\nu_1^0 - \tilde{x}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1\nu_1^1 - \tilde{x}_1'\tilde{b}_1 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0,$$

or

$$\mathbb{P}(a_1\nu_1^0 - \tilde{x}_1'\tilde{b}_1 > X_{1,1} \geq \alpha_1\nu_1^1 - \tilde{x}_1'\tilde{\beta}_1 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0.$$

We show the first one holds as long as $(\tilde{b}_1, a_1) \neq (\tilde{\beta}_1, \alpha_1)$. Note that if there were no terms associate with ν_1^1 and ν_1^0, the identification is essentially the same as the identification in Lemma 2 in Manski (1985).

We consider the following two cases.

Case 1. $\tilde{b}_1 \neq \tilde{\beta}_1$. By assumption 30, $\tilde{x}_1'\tilde{b}_1 \neq \tilde{x}_1\tilde{\beta}_1$. Without loss of generality, assume $\tilde{x}_1'\tilde{b}_1 > \tilde{x}_1\tilde{\beta}_1$. Let $\epsilon \in (0, \frac{1}{4}(\tilde{x}_1'\tilde{b}_1 - \tilde{x}_1'\tilde{\beta}_1))$. Then $\tilde{x}_1'\tilde{\beta}_1 + \epsilon < \tilde{x}_1'\tilde{b}_1 - \epsilon$. By Lemma 11, there exists a $t_1 > 0$ such that for any $x_{1,1} \in (-\tilde{x}_1'\tilde{b}_1 + \epsilon, -\tilde{x}_1'\tilde{\beta}_1 - \epsilon)$,

$$\mathbb{P}(\max\left\{\alpha_1\nu_1^0, a_1\nu_1^1\right\} < \epsilon | X_1 = x_1, \tilde{X}_2 = \tilde{x}_2, X_{2,1} \leq -t_1) = 1.$$

For notation simplicity, we drop the conditioning variable $\tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2$ in the each probability term in the next equation. Then

$$\mathbb{P}(\alpha_1\nu_1^0(X) - \tilde{X}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1\nu_1^1(X) - \tilde{X}_1'\tilde{b}_1)$$
$$\geq \mathbb{P}(\alpha_1\nu_1^0(X) - \tilde{X}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1\nu_1^1(X) - \tilde{X}_1'\tilde{b}_1 | X_{2,1} < -t_1) \times \mathbb{P}(X_{2,1} < -t_1)$$
$$\geq \mathbb{P}(-\epsilon - \tilde{x}_1'\tilde{\beta}_1 > X_{1,1} \geq \epsilon - \tilde{x}_1'\tilde{b}_1 | X_{2,1} < -t_1) \times \mathbb{P}(X_{21} < -t_1) > 0.$$

The last inequality holds by assumption 31.

Case 2. $\tilde{b}_1 = \tilde{\beta}_1$. Then $\alpha_1 \neq a_1$. Without loss of generality, assume that $\alpha_1 > a_1$. Let η be a positive number satisfying $\alpha_1 - \eta > a_1 + \eta$. By Lemma 11, there exists some $t_2 > 0$ and such that

$$\mathbb{P}(\nu_1^0 \alpha_1 > \alpha_1 - \eta > a_1 + \eta > \nu_1^1 a_1 | X_1 = x_1, \tilde{X}_2 = \tilde{x}_2, X_{2,1} \geq t_2) = 1.$$

Still, we drop the conditioning variable $\tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2$ in the each probability term in the next equation. Then it follows,

$$\begin{aligned}
&\mathbb{P}(\alpha_1 \nu_1^0 (X) - \tilde{x}_1' \tilde{\beta}_1 > X_{1,1} \geq a_1 \nu_1^1 (X) - \tilde{x}_1' \tilde{b}_1) \\
&\qquad \geq \mathbb{P}(\alpha_1 \nu_1^0 (X) - \tilde{x}_1' \tilde{\beta}_1 > X_{1,1} \geq a_1 \nu_1^1 (X) - \tilde{x}_1' \tilde{b}_1 | X_{2,1} > t_2) \mathbb{P}(X_{2,1} > t_2) \\
&\qquad\qquad \geq \mathbb{P}(\alpha_1 - \eta - \tilde{x}_1' \tilde{\beta}_1 > X_{1,1} \geq a_1 + \eta - \tilde{x}_1' \tilde{b}_1 | X_{2,1} > t_2,) \mathbb{P}(X_{2,1} > t_2) > 0.
\end{aligned}$$

The last inequality again follows from assumption 31. The statement in the Lemma thus follows by combining case 1 and 2.

Proof of Lemma 12

We need to show that $\int g_1 (Z, \alpha_1, \beta_1, h) \, dF_Z (Z) > \int g_1 (Z, a_1, b_1, h) \, dF_Z (Z)$ for all $(a_1, b_1) \neq (\alpha_1, \beta_1)$.

Notice that $\mathscr{X}$ can be divided into three regions, A, B and C, as follows:

$$\begin{aligned}
A &= \left\{ x : x_1' \beta_1 - \alpha_1 \nu_1^0 (x) < 0 \right\}, \\
B &= \left\{ x : x_1' \beta_1 - \alpha_1 \nu_1^1 (x) \geq 0 \right\}, \\
C &= \left\{ x : x_1' \beta_1 - \alpha_1 \nu_1^0 (x) \geq 0 > x_1' \beta_1 - \nu_1^1 (x) \right\}.
\end{aligned}$$

For any $x \in A$, $p_1 (x) = \mathbb{P} \left(U_1 \leq X_1' \beta_1 - \nu_1 (X) | X = x \right) \leq \mathbb{P} \left(U_1 \leq 0 | X = x \right) = \frac{1}{2}$, Hence

$$\begin{aligned}
\int \mathbf{1}[X \in A] g_1 (Z, a_1, b_1, h) \, dF_Z (Z) &= \int \mathbf{1}[X \in A] \left(1 - 2 p_1 (\xi) \right) dF_X (\xi) \\
&= \int \mathbf{1}[X \in A] \left| 2 p_1 (\xi) - 1 \right| dF_X (\xi).
\end{aligned}$$

Similarly, we have

$$\int \mathbf{1}[X \in B] g_1 (Z, a_1, b_1, h) \, dF_Z (Z) = \int \mathbf{1}[X \in B] \left| 2 p_1 (\xi) - 1 \right| dF_X (\xi),$$

and

$$\begin{aligned}
\int \mathbf{1}[X \in C] g_1 (Z, a_1, b_1, h) \, dF_Z (Z) &= \int \mathbf{1}[X \in C] \left(2 p_1 (\xi) - 1 \right) sgn \left(p_1 (\xi) > \frac{1}{2} \right) dF_X (\xi) \\
&= \int \mathbf{1}[X \in C] \left| 2 p_1 (\xi) - 1 \right| dF_X (\xi).
\end{aligned}$$

It follows that $\int g_1(Z,\alpha_1,\beta_1,h)\,dF_Z(Z) = \int |2p_1(\xi)-1|\,dF_X(\xi)$.

Now consider any $(a_1,b_1) \neq (\alpha_1,\beta_1)$. Let $\tilde{x} = (\tilde{x}_1,\tilde{x}_2)$ be arbitrary. From Lemma 16 we know that one of the following is true:

$$\mathbb{P}(\alpha_1\nu_1^0 - \tilde{X}_1'\tilde{\beta}_1 > X_{11} \geq a_1\nu_1^1 - \tilde{X}_1'\tilde{b}_1|\tilde{X}_1 = \tilde{X}_1, \tilde{X}_2 = \tilde{X}_2) > 0$$

or

$$\mathbb{P}(a_1\nu_1^0 - \tilde{X}_1'\tilde{b}_1 > X_{11} \geq \alpha_1\nu_1^1 - \tilde{X}_1'\tilde{\beta}_1|\tilde{X}_1 = \tilde{X}_1, \tilde{X}_2 = \tilde{X}_2) > 0.$$

Call the set of (x_{11},x_{21}) which satisfy the above two conditions as $D(\tilde{x})$. No matter which is true, we have

$$\int_{D(\tilde{x})} g_1(Z,a_1,b_1,h)\,dF_{Z|\tilde{X}}\left(Z|\tilde{X}=\tilde{x}\right) = -\int_{D(\tilde{x})} |2p_1(\xi)-1|\,dF_{X|\tilde{X}}\left(\xi|\tilde{X}=\tilde{x}\right) < 0,$$

since $p_1(X) = 1/2$ with probability 0. So it follows:

$$\int g_1(Z,a_1,b_1,h)\,dF_Z(Z) - \int g_1(Z,a_1,b_1,h)\,dF_Z(Z)$$
$$= 2\int\int_{D(\tilde{x})} |2p_1(\xi)-1|\,dF_{X|\tilde{X}}\left(\xi|\tilde{X}=\tilde{x}\right)dF_{\tilde{X}}(\tilde{x}) > 0.$$

Proof of Theorem 17

This part of the proof consists two main parts. Lemma 13 shows that if the first step estimator is consistent, the effect of first stage estimation error on objective function will only be the order of $o_p(1)$. Lemma 15 verifies the uniform weak law of large number (under the true h).

Lemma 13 *Suppose assumptions 33 through 36 hold. Then*

$$\sup_{(a,b)\Theta}\left|L_n\left(a,b,\hat{h}\right) - L_n(a,b,h)\right| = o_p(1).$$

proof: Let $\boldsymbol{v} = \left(\nu_1^1,\nu_1^0,\nu_2^1,\nu_2^0\right)$ and $\hat{\boldsymbol{v}}$ be the corresponding estimates. Similarly define $p = (p_1,p_2)$ and $\hat{p} = (\hat{p}_1,\hat{p}_2)$. To show the stated expression, it is enough to show that

$$\sup_{(a,b)\in\Theta}|L_n(a,b,\hat{p},\hat{\boldsymbol{v}}) - L_n(a,b,p,\hat{\boldsymbol{v}})| = o_p(1), \tag{3.12}$$
$$\sup_{(a,b)\in\Theta}|L_n(a,b,p,\hat{\boldsymbol{v}}) - L_n(a,b,p,\boldsymbol{v})| = o_p(1). \tag{3.13}$$

Consider equation (3.12) first. Due to the additive form of the objective function, it is sufficient to show the the right hand side of (3.14) is $o(1)$; the other terms can be dealt with using similar arguments.

$$\mathbb{E}\left\{\sup_{(a,b)\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_{1i}\left[\boldsymbol{1}\left(\hat{p}_1\left(X_i\right)<1/2\right)-\boldsymbol{1}\left(p_1\left(X_i\right)<1/2\right)\right]\boldsymbol{1}\left(X_{1i}'b_1-a_1\hat{v}_1^1\left(X_i\right)\geq 0\right)\right|\right\}$$

$$\leq \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left|\boldsymbol{1}\left(\hat{p}_1\left(X_i\right)<1/2\right)-\boldsymbol{1}\left(p_1\left(X_i\right)<1/2\right)\right|\right\}$$

$$=\mathbb{E}\left|\boldsymbol{1}\left(\hat{p}_1\left(X_i\right)<1/2\right)-\boldsymbol{1}\left(p_1\left(X_i\right)<1/2\right)\right|$$

$$\leq \mathbb{P}\left[\hat{p}_1\left(X_i\right]\leq 1/2<p_1(X_i)\right)+\mathbb{P}\left[p_1\left(X_i\right)<1/2\leq\hat{p}_1(X_i)\right]. \quad (3.14)$$

where $\widetilde{Y}_{1i}=2Y_{1i}-1$.

We show that $\mathbb{P}\left[\hat{p}_1\left(X_i\right)<1/2\leq p_1(X_i)\right]=o(1)$ *only. Let* $\kappa>0$ *and* $\delta>0$ *be some small positive numbers,*

$$\mathbb{P}\left[\hat{p}_1\left(X_i\right)<1/2\leq p_1(X_i)\right]$$

$$\leq \mathbb{P}\left[\hat{p}_1\left(X_i\right)<1/2, 1/2\leq p_1(X_i)<1/2+\kappa\right]+\mathbb{P}\left[\hat{p}_1\left(X_i\right)<1/2, p_1(X_i)\geq 1/2+\kappa\right]$$

$$\leq \mathbb{P}\left[1/2\leq p_1(X_i)<1/2+\kappa\right]+\mathbb{P}\left[p_1(X_i)-\hat{p}_1(X_i)\geq\kappa\right]$$

$$\leq \mathbb{P}\left[1/2\leq p_1(X_i)<1/2+\kappa\right]+\mathbb{P}\left[p_1(X_i)-\hat{p}_1(X_i)\geq\kappa, f(X_i)>\delta\right]+\mathbb{P}\left[f(X_i)\leq\delta\right]. \quad (3.15)$$

The first term on the right hand side of (3.15) can be made arbitrarily small by taking $\kappa\downarrow 0$ *under assumption 35; the second term disappears because of assumption 36; the third term goes to zero as* $\delta\downarrow 0$.

Now we show $\sup_{(a,b)\Theta}\left|L_n\left(a,b,p,\hat{\boldsymbol{v}}\right)-L_n\left(a,b,p,\boldsymbol{v}\right)\right|=o_p\left(1\right)$. *Again, we only consider the term for player* 1 *associate with* v_1^1 *(the left hand side of (3.16)). The other terms can be dealt with under similar arguments.*

$$\sup_{(a,b)\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(2Y_{1i}-1\right)\boldsymbol{1}\left(p_1\left(X_i\right)<\frac{1}{2}\right)\right.$$

$$\left.\times\left[\boldsymbol{1}\left(X_{1i}'b_1-a_1\hat{v}_1^1\left(X_i\right)\geq 0\right)-\boldsymbol{1}\left(X_{1i}'b_1-a_1v_1^1\left(X_i\right)\geq 0\right)\right]\right|$$

$$\leq \sup_{(a,b)\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{1}\left(p_1\left(X_i\right)<1/2\right)\boldsymbol{1}\left(a_1v_1^1\left(X_i\right)\leq X_{1i}'b_1<a_1\hat{v}_1^1\left(X_i\right)\right)$$

$$+\sup_{(a,b)\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{1}\left(p_1\left(X_i\right)<1/2\right)\boldsymbol{1}\left(a_1\hat{v}_1^1\left(X_i\right)\leq X_{1i}'b_1<a_1v_1^1\left(X_i\right)\right) \quad (3.16)$$

We only show that the first term on the right hand side of (3.16) is $o_p(1)$. *For some given* $\delta>0$,

$$\sup_{(a,b)\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{1}\left(p_1\left(X_i\right)<1/2\right)\boldsymbol{1}\left(a_1v_1^1\left(X_i\right)\leq X_{1i}'b_1<a_1\hat{v}_1^1\left(X_i\right)\right)$$

$$\leq \sup_{(a,b)\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{1}\left(a_1v_1^1\left(X_i\right)\leq X_{1i}'b_1<a_1\hat{v}_1^1\left(X_i\right)\right)\boldsymbol{1}(f(X_i)\geq\delta)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{1}(f(X_i)<\delta)$$

$$\leq \sup_{(a,b)\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(a_1 \nu_1^1 (X_i) \leq X_{1i}' b_1 < a_1 \nu_1^1 (X_i) + a_1 \eta_n\right) + \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(f(X_i) < \delta)$$

$$\leq \sup_{(a,b)\in\Theta} \frac{1}{n}\sum_{i=1}^{n} m(X_i, a, b, \eta_n) - Pm(X_i, a, b, \eta_n)$$

$$+ \sup_{(a,b)\in\Theta} Pm(X_i, a, b, \eta_n) + \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(f(X_i) < \delta), \quad (3.17)$$

where $m(X_i, a, b, \eta_n) = \mathbf{1}\left(0 \leq X_{1i}' b_1 - a_1 \nu_1^1 (X_i) < a_1 \eta_n\right)$ *and* Pm *is the expectation of* h *with respect to the first argument. The second inequality holds for some* $\eta_n \xrightarrow{p} 0$ *by assumption 36.*

Let $\mathscr{M} = \{m(\cdot, a, b, c) : (a,b) \in \Theta, c \in \mathbb{C}\}$ *be a collection of* h *functions indexed by* (a,b,c)*, where* $\mathbb{C}$ *is a compact subset of* $\mathbb{R}$ *and contains* 0*. It is easy to verify that* $\mathscr{M}$ *is VC class. The first term in on the right hand side of (3.17) is* $o_p(1)$ *since* $\frac{1}{\sqrt{n}}(\sum_{i=1}^{n} m(X_i, a, b, c) - Pm(X_i, a, b, c))$ *weakly converges to a stochastic process whose index ranges over a compact set.*

The third term on the right hand side of (3.17) can be made arbitrarily small by taking $\delta \downarrow 0$*. As for the second term, let* $W_{(a,b)} = X_{1i}' b_1 - a_1 \nu_1^1 (X_i)$ *and* $f_{W_{(a,b)}}$ *be the density of* $W_{(a,b)}$*.*[13]

$$\sup_{(a,b)\in\Theta} Pm(X_i, a, b, \eta_n) = \sup_{(a,b)\in\Theta} \mathbb{P}\left(0 \leq X_{1i}' b_1 - a_1 \nu_1^1 (X_i) < a_1 \eta_n\right) = \eta_n O\left(\sup_{(a,b)\in\Theta} a_1 f_{W_{(a,b)}}(0)\right).$$

The right hand side is $o_p(1)$ *because* $f_{W_{(a,b)}}(0)$ *is well–defined by assumption 35 and is continuous in* $(a,b) \in \Theta$ *by its definition, thus* $\sup_{(a,b)\in\Theta} a_1 f_{W_{(a,b)}}(0)$ *is finite by the compactness of* Θ*.*

Combine both parts, we finish the proof of Lemma 13.

Lemma 14 *Suppose assumptions 33 through 36 are satisfied, then:*

$$\sup_{(a,b)\in\Theta} \left|L_n(a,b,h) - L(a,b,h)\right| = o_p(1).$$

proof: This follows directly from Lemma 2.4 of Newey and McFadden (1986).

Lemma 15 *Suppose assumptions 33 through 36 are satisfied, then:*

$$\sup_{(a,b)\in\Theta} \left|L_n\left(a,b,\hat{h}\right) - L(a,b,h)\right| = o_p(1).$$

proof: Notice that

$$\sup_{(a,b)\in\Theta} \left|L_n\left(a,b,\hat{h}\right) - L(a,b,h)\right|$$

$$\leq \sup_{(a,b)\in\Theta} \left|L_n\left(a,b,\hat{h}\right) - L_n(a,b,h)\right| + \sup_{(a,b)\in\Theta} \left|L_n(a,b,h) - L(a,b,h)\right|.$$

---

[13]The density exists for every (a,b) because of assumption 31.

The first term is $o_p(1)$ by Lemma 13, and the second term is $o_p(1)$ by 14.

Now we show that $(\hat{\alpha}, \hat{\beta}) \xrightarrow{p} (\alpha, \beta)$. Let $\Theta^\epsilon = \{(a, b) : \|(a, b) - (\alpha, \beta)\| < \epsilon\}$. Then Θ/Θ^ϵ is compact. Since (α, β) is the unique maximizer of the population objective function, we have for some $\eta > 0$

$$\mathbb{P}\left((\hat{\alpha}, \hat{\beta}) \in \Theta/\Theta^\epsilon\right) \leq \mathbb{P}\left(L\left(\hat{\alpha}, \hat{\beta}, h\right) - L\left(\alpha, \beta, h\right) < -\eta\right).$$

However,

$$\begin{aligned} &L\left(\hat{\alpha}, \hat{\beta}, h\right) - L\left(\alpha, \beta, h\right) \\ &\quad = \left[L\left(\hat{\alpha}, \hat{\beta}, h\right) - L_n\left((\hat{\alpha}, \hat{\beta}), \hat{h}\right)\right] + \left[L_n\left(\hat{\alpha}, \hat{\beta}, \hat{h}\right) - L_n\left(\alpha, \beta, \hat{h}\right)\right] + \left[L_n\left(\alpha, \beta, \hat{h}\right) - L\left(\alpha, \beta, h\right)\right] \\ &\qquad \geq o_p(1) + L_n\left(\hat{\alpha}, \hat{\beta}, \hat{h}\right) - \sup_{(a,b)} L_n\left(a, b, \hat{h}\right) + o_p(1) = o_p(1). \end{aligned}$$

The first and the third term come is $o_p(1)$ because of Lemma 15. The middle term is greater than 0 by definition of the estimator. So we can conclude that $\mathbb{P}\left((\hat{\alpha}, \hat{\beta}) \in \Theta/\Theta^\epsilon)\right)$ converges to 0, which implies $\|(\hat{\alpha}, \hat{\beta} - (\alpha, \beta))\| \xrightarrow{p} 0$.

Proof of Theorem 18

The proof of Theorem 17 carries through for each $(\hat{\alpha}, \hat{\beta}) \in \hat{\Theta}$, which implies that

$$\sup_{\hat{\Theta}} \rho\left((a, b), \Theta^I\right) \xrightarrow{p} 0.$$

On the other hand, following the proof of Manski and Tamer (2002b) Proposition 3(b), we have $\sup_{\Theta^I} \rho\left((a, b), \hat{\Theta}\right) \xrightarrow{p} 0$. The conclusion follows.

Chapter 4

Tighter Bounds in Triangular Systems

4.1 Introduction

The primary objective of our paper is to obtain identification results — that are stronger than are currently available in the literature under alternative conditions — for the nonparametric triangular model

$$\begin{cases} Y = g(X, U), \\ X = h(Z, V), \end{cases} \tag{4.1}$$

where $Y \in \mathscr{S}_y \subset \mathbb{R}, X \in \mathscr{S}_x \subset \mathbb{R}^d, Z \in \mathscr{S}_z \subset \mathbb{R}^{d_z}$ are observables, g, h are unknown functions, and $U \in \mathscr{U} = (0,1], V \in \mathscr{V} \subseteq \mathscr{U}^d$ are errors. We refer to X as endogenous regressors and Z as instruments and use bold face symbols to denote random variables and regular face symbols for (nonrandom) values the corresponding random variable can take. Similar to Chesher (2005), the regressors need not be continuous and the objective is identification of the object

$$\psi^* = \psi(x^*, \tau^*, v^*) = g\big(x^*, Q_{U|V}(\tau^*|v^*)\big) \tag{4.2}$$

for given values of $(\tau^*, x^*, v^*) \in \mathscr{U} \times \mathscr{S}_x \times \mathscr{V}$, where $Q_{U|V}(\tau|v) = \inf\{u : \mathbb{P}[U \leq u|V = v] \geq \tau\}$.

If, for the sake of intuition, one attaches the labels 'earnings' to Y, 'education' to X, 'demographics' to Z, 'talent' to V, and '(market) success' to U, then $\psi(x^*, 0.5, 0.5)$ can be interpreted as the (counterfactual) earnings of someone with median success and median talent if she were given education x^*.[1] Identification of marginal effects such as $\psi(x^*, 0.5, 0.5) - \psi(x^{**}, 0.5, 0.5)$ naturally follows from the identification of $\psi(x^*, 0.5, 0.5)$ and $\psi(x^{**}, 0.5, 0.5)$.

A model similar to 4.1 was studied in Chesher (2003) and Chesher (2005). Chesher (2003) used a strict monotonicity assumption, excluding discrete–valued X, to identify the partial derivatives of g with respect to X. Ma and Koenker (2006) and Jun (2009) proposed a parametric and a semiparametric estimator of Chesher's 2003 model, respectively. Chesher (2005) is more closely related to our paper in the sense that X is allowed to be discrete and that the object of interest is also ψ^*. The object of estimation in Newey, Powell, and Vella (1999) and Pinkse (2000) is $g\big(x^*, \mathbb{E}(U|V = v^*)\big)$, which is similar to ψ^*, but in those papers the errors are assumed additively separable in both equations in 4.1.

In Chesher (2003) regressors are assumed to be continuous in which case point identification of the partial derivatives of g can be achieved by using strict monotonicity conditions on the second argument of g and h. However, when X is discrete, as in the example of the years of schooling, strict monotonicity cannot hold. In Chesher (2005) (identified) bounds are obtained for ψ^* under weak monotonicity, a dependence condition on U and V, and 'local exclusion' and 'local exogeneity' conditions on the instrument Z.[2] We present our results under 'global' rather than local conditions, i.e. we impose a global exclusion restriction (Z does not enter g) and assume that Z is independent of U, V. Global conditions are stronger than local ones, but we note that those conditions are not testable and that global conditions are more common in multi–equations models.[3] Further, our global conditions allow us to replace the rank condition in Chesher (2005) ($\mathscr{R}$) with an alternative, weaker, rank condition ($\mathscr{R}^*$) which allows for the construction of tighter bounds on ψ^* than those obtained in Chesher (2005).[4] Moreover, in the case of binary regressors $\mathscr{R}$ is never satisfied, but $\mathscr{R}^*$ developed in this paper usually holds and in some cases leads to point identification of ψ^*; the example we provide exploits continuous variation in Z. A more precise and detailed discussion follows in the next section. Section 4.3 contains an empirical example illustrating the difference between $\mathscr{R}$ and $\mathscr{R}^*$.

Results similar to those developed in this paper can in principle be established under

[1] We use the term 'success' to emphasize the potential dependence between U, V.

[2] Starting from $Y = \tilde{g}(X, Z, U)$, Chesher (2005) assumed that there exist $z_1, z_2 \in \mathscr{S}_z$ such that for some r^*, $Q_{U|V,Z}(\tau^*|v^*, z_1) = Q_{U|V,Z}(\tau^*|v^*, z_2) = r^*$ (local exogeneity) and $\tilde{g}(x^*, z_1, r^*) = \tilde{g}(x^*, z_2, r^*)$ (local exclusion).

[3] This is true for traditional linear simultaneous equations models, as well as for the bulk of the modern literature, e.g. Imbens and Newey (2009).

[4] Cases exist in which the bounds are the same.

local conditions, also. However, obtaining much tighter bounds under conditions that are meaningfully different from the global ones results in conditions that are exceedingly difficult to interpret; see Jun, Pinkse, and Xu (2009), which is available on our website. Chesher (2005) establishes that his bounds are tight in a point identification example; our paper does not provide insights as to whether or not Chesher's bounds under his conditions are tight more generally.

Alternatively, one can conduct the analysis conditional on a subset $\mathscr{S}_z^*$ of instrument values. Our results go through without modification provided that all conditions and results are interpreted conditional on $Z \in \mathscr{S}_z^*$. Since conditioning on $\mathscr{S}_z^*$ amounts to throwing away information, doing so generally yields wider bounds than if global exclusion/exogeneity can be assumed to hold without such conditioning. But it is weaker than global exclusion/exogeneity and it does allow for instruments to enter into the g–function directly, albeit subject to the strong condition that the g–function value is the same for all $z \in \mathscr{S}_z^*$.

The methodology developed in this paper can be applied in other settings. For instance, Jun, Pinkse, and Xu (2010) provide an extension of the identification method proposed in this paper, which is applied to the model of Vytlacil and Yildiz (2007); the Vytlacil–Yıldız results for binary endogenous regressors are extended to cover discrete endogenous regressors that can take more than two values and their support restrictions are relaxed in the case of binary regressors.

The proof of our theorem relies on an inversion of the conditional distribution function $\Pi(y|x^*, v^*) = \mathbb{P}[Y \leq y | X = x^*, V = v^*]$. Therefore, our methodology can be used to derive bounds on other functionals of Π such as the mean $\delta(x^*, v^*) = \mathbb{E}[Y|X = x^*, V = v^*]$, which is in fact the quantity of interest in Manski and Tamer (2002a).

There are certain similarities between Manski and Tamer (2002a) and Chesher (2005), and indeed our paper. There are however several differences besides the difference in object of interest (mean versus quantile) noted earlier. First, the primary objective in Manski and Tamer (2002a) is estimation, whereas in Chesher (2005) and here it is identification. Second, in Manski and Tamer (2002a) upper and lower bounds (v_0^{MT}, v_1^{MT}) on V are assumed to be available while Chesher (2005) provides conditions (involving instrumental variables) under which such bounds are available and can be used. Finally, if the Manski and Tamer (2002a) bounds were used in the quantile context, the bounds that obtain after inversion of Π would be the ones obtained by Chesher (2005), not the ones provided in this paper; see appendix 4.6.2 for details.

Although we only provide identification results in this paper, the identification approach here can be implemented in practice. We are currently developing an estimator for ψ^* in a separate paper. This estimator assumes the existence of continuous instruments, which we do not assume for our identification result in the present paper. Developing an estimator which takes full advantage of the weakest set of identification results contained in this paper could be challenging.

Our paper is organized as follows. Section 4.2 contains the main results established in this paper. In section 4.3 we illustrate our proposal using the Angrist and Krueger (1991) data set.

4.2 Main Results

4.2.1 Assumptions

Consider again the model in 4.1. The objective remains to find identifiable bounds on ψ^* defined in 4.2 for given values of τ^*, x^*, v^*. We make the following assumptions.

Assumption 37 *$U, V_1, \ldots, V_d$ have (marginal) $U(0,1]$–distributions.*

Assumption 38 *g is non–decreasing in u for all x and $h(z,v) = \left[h_1(z,v_1), \ldots, h_d(z,v_d)\right]^\top$, where h_j is nondecreasing and left–continuous in v_j for all values of z for $j = 1, \ldots, d$.*

Assumption 39 *U, V are independent of Z.*

Assumption 40 *U is positive regression dependent on V, i.e. $Q_{U|V}(\tau|v)$ is nondecreasing in v for all values of τ.*

Assumption 41 *$\mathscr{X}(x^*, v^*) = \{z \in \mathscr{S}_z : \; h(z, v^*) = x^*\}$ is nonempty.*

Given that g, h are unknown, the distributional conditions in assumption 37 plus the weak monotonicity and left–continuity conditions in assumption 38 by themselves amount to normalizations; see appendix 4.6.1. The assumption that only one error enters each h_j–equation is general since no dependence conditions are imposed between the V_j's. Assumptions 37 and 38 do become restrictive, however, when paired with the positive regression dependence condition in assumption 40. Assumption 39 is restrictive, as was discussed in the introduction. In the context of (4.1) and imposing assumption 39, the only addition in assumptions 37, 38 and 40 over what is assumed in Chesher (2005) is that the direction of monotonicity of $Q_{U|V}(\tau|v)$ is specified. This is innocuous, because the same analysis can be repeated under the assumption of the other direction of monotonicity after which one can compare the resulting bounds with the bounds based on assumption 40.

Assumption 41 requires that the type of individual for which bounds are desired exists. If there are no demographic characteristics z that yield an education level x^* for someone of talent v^*, then our procedure does not yield meaningful bounds for the earnings of someone with education level x^* and talent v^* for any level of success τ. Assumption 41 could be restrictive if one conditions on a subset $\mathscr{S}_z^*$ of demographic profiles, as discussed in the introduction.

4.2.2 Basics

We start by stating a lemma, which shows that if V were observable then ψ^* would be directly estimable from the data; V then plays the role of a control variable. The assumptions made above are presumed to hold for all lemmas.

Lemma 16 *For all $\tau \in \mathscr{U}$, $\psi(x^*, \tau, v^*) = Q_{Y|X,V}(\tau|x^*, v^*)$. proof:*

Lemma 16 implies that ψ^* can alternatively be interpreted as the τ^*–quantile of the earnings distribution of individuals with education x^* and talent v^*. Note that

$$h(z,v) = \begin{bmatrix} h_1(z,v_1) \\ \vdots \\ h_d(z,v_d) \end{bmatrix} = \begin{bmatrix} h_1\big(z, Q_{V_1}(v_1)\big) \\ \vdots \\ h_d\big(z, Q_{V_d}(v_d)\big) \end{bmatrix} = \begin{bmatrix} h_1\big(z, Q_{V_1|Z}(v_1|z)\big) \\ \vdots \\ h_d\big(z, Q_{V_d|Z}(v_d|z)\big) \end{bmatrix} = \begin{bmatrix} Q_{X_1|Z}(v_1|z) \\ \vdots \\ Q_{X_d|Z}(v_d|z) \end{bmatrix}, \tag{4.3}$$

where the second to fourth equalities follow from assumptions 37, 39 and 38, respectively. Hence, if the conditional distribution of X_j given $Z = z$ is continuous for all z then lemma 16 implies that h_j is invertible in its second argument and that $V_j = F_{X_j|Z}(X_j|Z)$, where $F_{X_j|Z}$ is the conditional distribution function of X_j given Z. Therefore, the V_j's that correspond to continuous X_j's can be recovered from the data. For this reason we only discuss the case in which the elements of X are all discrete from hereon.

Let

$$V_j(x_j, z) = \big(\mathbb{P}[X_j < x_j|Z = z],\ \mathbb{P}[X_j \le x_j|Z = z]\big] \tag{4.4}$$

for $j = 1, 2, \cdots, d$. Then, for any $v \in V(x,z) \equiv V_1(x_1, z) \times \cdots \times V_d(x_d, z)$ we have

$$h(z,v) \overset{4.3}{=} \left[Q_{X_1|Z}(v_1|z), \; \cdots \; , Q_{X_d|Z}(v_d|z)\right]^{\mathrm{tr}} = x, \tag{4.5}$$

where the last equality follows from the definition of $V_j(x_j, z)$ in equation (4.4). Therefore, $V(x,z)$ is a set of talent levels for which individuals with demographics z achieve education level x,

$$V(x,z) = \{v \in \mathscr{U}^d : \; h(z,v) = x\}. \tag{4.6}$$

Please note that since $V(x,z)$ depends only on (a conditional distribution function of) observables, it is identified for all $(x,z) \in \mathscr{S}_x \times \mathscr{S}_z$.

4.2.3 Basic Rank Condition

Let

$$\begin{cases} \mathscr{G}^+(x,v) = \left\{V \in \mathscr{U}^d : \exists z \in \mathscr{S}_z : V = V(x,z) \geq v\right\}, \\ \mathscr{G}^-(x,v) = \left\{V \in \mathscr{U}^d : \exists z \in \mathscr{S}_z : V = V(x,z) \leq v\right\}, \end{cases} \tag{4.7}$$

where $V \geq v$ ($V \leq v$) means that no vectors in V have elements that are strictly less (greater) than the corresponding element of v. Intuitively, for $V(x,z)$ to belong to $\mathscr{G}^+(x,v)$, demographics z must be so unfavorable as to ensure that anyone with demographics z but talent less than v would not be able to achieve education x.

We now turn to the first of the two rank conditions mentioned in the introduction, namely $\mathscr{R}$.

Condition 1 ($\mathscr{R}$) *Neither $\mathscr{G}^+(x^*, v^*)$ nor $\mathscr{G}^-(x^*, v^*)$ is empty.* □

$\mathscr{R}$ is due to Chesher (2005) as, under local conditions, is lemma 18 below. $\mathscr{R}$ requires the instrument to be strong enough to ensure that

$$\mathbb{P}\left[X_j \leq x_j^* | Z = z\right] \leq v^* \leq \mathbb{P}\left[X_j < x_j^* | Z = \tilde{z}\right], \quad j = 1, \ldots, d, \tag{4.8}$$

for some $z, \tilde{z} \in \mathscr{S}_z$. If $\mathscr{R}$ is satisfied, then the result of lemma 18 below follows almost immediately, using lemma 17 along the way. Let $Q_{U|V}(\tau|V)$ denote the τ quantile of the conditional distribution of U given that $V \in V$.

Lemma 17 *For all $\tau \in \mathscr{U}$ and all $(x,z) \in \mathscr{S}_x \times \mathscr{S}_z$, if $V(x,z) \neq \emptyset$, then $Q_{Y|X,Z}(\tau|x,z) = g\{x, Q_{U|V}(\tau|V(x,z))\}$. proof:*

Lemma 18 *Under $\mathscr{R}$ (condition 1),*

$$\sup_{\{z\in\mathscr{S}_z:V(x^*,z)\leq v^*\}} Q_{Y|X,Z}(\tau^*|x^*,z) \leq \psi^* \leq \inf_{\{z\in\mathscr{S}_z:V(x^*,z)\geq v^*\}} Q_{Y|X,Z}(\tau^*|x^*,z), \tag{4.9}$$

or equivalently

$$\sup_{V\in\mathscr{G}^-(x^*,v^*)} g\big(x^*,Q_{U|V}(\tau^*|V)\big) \leq \psi^* \leq \inf_{V\in\mathscr{G}^+(x^*,v^*)} g\big(x^*,Q_{U|V}(\tau^*|V)\big). \tag{4.10}$$

proof:

Lemma 18 is a sensible result, which has the following intuition. Find a demographic profile z such that individuals must have talent no less (greater) than v^* to achieve education x^*. Individuals with demographics z and education x^* then have a success distribution no less (more) favorable than those of individuals with talent equal to v^* by assumptions 39 and 40 and the same level of education. Hence, by assumption 38 ψ^* must be no less (greater) than the τ^* quantile of the earnings distribution of individuals with demographics z and education x^*. Out of all such profiles z, select the one resulting in the tightest upper (lower) bound.

Two problems with $\mathscr{R}$ (condition 1) are that (i) it may not hold and (ii) the classes $\mathscr{G}^+(x^*,v^*)$, $\mathscr{G}^-(x^*,v^*)$, even when nonempty, may not be large. In fact, $\mathscr{R}$ cannot be satisfied if x^* is a scalar and equals the highest or lowest value possible. For instance, if X is binary (college–educated or not) then $V(0,z) = \big(0,\mathbb{P}[X=0|Z=z]\big]$ and $V(1,z) = \big(\mathbb{P}[X=0|Z=z],1\big]$ for all z; $V(1,z)$ has upper limit equal to one since there is no nontrivial upper bound to the talent of individuals with a college education. For vector–valued x^*, the problem is still more severe.

Further, note that each value of z generates at most one element in either $\mathscr{G}^+(x^*,v^*)$ or $\mathscr{G}^-(x^*,v^*)$. This fact, together with the global exogeneity of Z, suggests that $\mathscr{G}^+(x^*,v^*)$ and $\mathscr{G}^-(x^*,v^*)$ may be too small; a new rank condition is needed. A more detailed discussion follows in the next subsection.

4.2.4 New Rank Condition

We now develop our new, weaker, rank condition $\mathscr{R}^*$. It is based on the idea that the collection $\{V(x^*,z) : z\in\mathscr{S}_z\}$ can in fact generate larger classes of sets that are useful for bounding ψ^* than $\mathscr{G}^-(x^*,v^*)$ and $\mathscr{G}^+(x^*,v^*)$. To be more specific, consider the example of binary X (college–educated or not) again. In this example $V(1,z_1) - V(1,z_2)$ is the set of talent–levels for which individuals would attend college with demographics z_1 but not with z_2. If $V(1,z_1) - V(1,z_2) \leq v^*$ then the success distribution of college–educated individuals whose talent is in the range $V(1,z_1) - V(1,z_2)$ is no more favorable than those of college–educated individuals with talent v^*. This can be the case even when neither $V(1,z_1) \leq v^*$ nor $V(1,z_2) \leq v^*$.

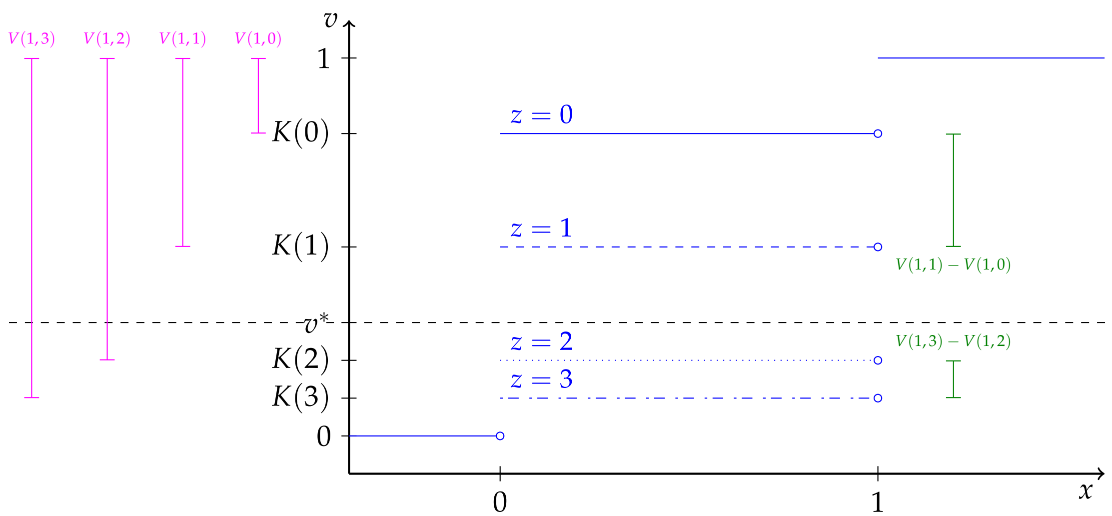

Figure 4.1. Example of how sets are combined.

The situation is illustrated in 4.1 in which X is binary, $\mathscr{S}_z = \{0,1,2,3\}$ and $K(z) = \mathbb{P}[X = 0|Z = z]$. In the graphed example, demographic profiles 0 and 1 ensure that all those with a college education must have talent no less than v^*, but there is no demographic value that makes individuals with talent no better than v^* attend college. So $\mathscr{R}$ is not satisfied; $\mathscr{G}^+(1,v^*) = \{V(1,1), V(1,0)\}$ and $\mathscr{G}^-(1,v^*) = \emptyset$. Therefore, the method used in lemma 18 provides an upper bound for ψ^*, but it does not provide a meaningful lower bound. As will be shown below, there is information available for the construction of an upper bound that is not contained in $\mathscr{G}^+(1,v^*)$. Indeed, we can also construct an upper bound by looking at the group of individuals who would attend college with $z = 1$ but not with $z = 0$. The bound provided by $V(1,1) - V(1,0)$ may well be tighter than the bound provided by either $V(1,1)$ or $V(1,0)$. Likewise, $V(1,3) - V(1,2)$, the group of talent levels that would result in a college education with $z = 3$ but not with $z = 2$, can be used to construct a lower bound. A more complicated example, involving vector–valued X, can be found at the end of this subsection.

Lemma 19 is our starting point. Let $\phi^*(V) = \phi(\tau^*, V) = g\big(x^*, Q_{U|V}(\tau^*|V)\big)$. Our main result, theorem 19, below is based on the fact that

$$\phi^*(V) \geq \psi^* \tag{4.11}$$

whenever $V \geq v^*$; this is a direct implication of assumption 40. Lemma 19 shows how sets can be combined. Let $\mu(V) = \mathbb{P}[V \in V]$ and let $\mathscr{K} = \{V \subset \mathscr{U}^d : V \neq \emptyset,\ \mu(V)$ and $\phi(\tau, V)$ are identified for all $\tau \in \mathscr{U}\}$.

Lemma 19 *For any* $V_1, V_2 \in \mathscr{K}$,

1. *If* $V_1 \subset V_2, \mu(V_2 - V_1) > 0$, *then* $V_2 - V_1 \in \mathscr{K}$.
2. *If* $V_1 \cap V_2 = \emptyset, \mu(V_1 \cup V_2) > 0$, *then* $V_1 \cup V_2 \in \mathscr{K}$.

proof:

Lemma 19 can be applied to $V(x,z)$–sets because $\phi\big(\tau, V(x,z)\big)$ is identified for all τ by lemma 17 and because $\mu\big(V(x,z)\big) = \mathbb{P}[X = x|Z = z]$ is identified.[5] If one applies either operation described in lemma 19 to $V_1 = V(x, z_1)$ and $V_2 = V(x, z_2)$ then the resulting set V_3 belongs to $\mathscr{K}$, and the procedure can be iterated. Doing so ultimately leads to a *Dynkin system* or *λ system* (Billingsley, 1995, p.41) of measurable sets.

Let $\mathscr{V}(x) = \{V : V \neq \emptyset,\ \exists z \in \mathscr{S}_z : V(x,z) = V\}$. In 1 below one can take $\mathscr{D}_0 = \mathscr{A} = \mathscr{V}(x^*)$, $\mathscr{D}_1$ to be the collection of sets that contains all sets in $\mathscr{A}$ plus all sets that arise when one applies lemma 19 to all combinations of elements in $\mathscr{A}$, $\mathscr{D}_2$ to be the collection of all sets in $\mathscr{D}_1$ plus all sets that arise when one applies lemma 19 to all combinations of elements in $\mathscr{D}_1$, and so forth. Ultimately, one ends up with $\mathscr{D} = \mathscr{D}_\infty$.

Definition 1 *Let $\mathscr{A}$ be a collection of measurable subsets of $\mathscr{U}^d$. Then $\mathscr{D} = \mathscr{D}(\mathscr{A})$ is the collection $\mathscr{D}_\infty$ in the following iterative scheme. Let $\mathscr{D}_0 = \mathscr{A}$. Then for all $t \geq 0$, $\mathscr{D}_{t+1}$ consists of all sets A^* such that at least one of the following three conditions is satisfied.*

1. $A^* \in \mathscr{D}_t$,
2. $\exists A_1, A_2 \in \mathscr{D}_t : A_1 \subset A_2,\ \mu(A_2 - A_1) > 0,\ A^* = A_2 - A_1$,
3. $\exists A_1, A_2 \in \mathscr{D}_t : A_1 \cap A_2 = \emptyset,\ \mu(A_1 \cup A_2) > 0,\ A^* = A_1 \cup A_2$. □

[5] Since $V(x,z)$ depends only on the conditional distribution of X given Z, $V(x,z)$ itself is also identified.

We will use $\mathscr{D}(x)$ in lieu of $\mathscr{D}\big(\mathscr{V}(x)\big)$ to emphasize its dependence on x. We are now in a position to state our rank condition. Let

$$\begin{cases} \mathscr{J}^-(x,v) = \{V \in \mathscr{D}(x) : V \leq v\}, \\ \mathscr{J}^+(x,v) = \{V \in \mathscr{D}(x) : V \geq v\}. \end{cases} \tag{4.12}$$

Condition 2 ($\mathscr{R}^*$) *Neither $\mathscr{J}^-(x^*,v^*)$ nor $\mathscr{J}^+(x^*,v^*)$ is empty.* □

If one compares $\mathscr{R}^*$ to $\mathscr{R}$ (condition 2 to 1), $\mathscr{R}^*$ is weaker since $\mathscr{J}^-$, $\mathscr{J}^+$ contain all elements of $\mathscr{G}^-,\mathscr{G}^+$, respectively. Only in rare circumstances are $\mathscr{R}$ and $\mathscr{R}^*$ the same.

Theorem 19 *Under assumptions 37–41, if $\mathscr{R}^*$ is satisfied then*

$$\sup_{V\in\mathscr{J}^-(x^*,v^*)} g\big(x^*, Q_{U|V}(\tau^*|V)\big) \leq \psi^* \leq \inf_{V\in\mathscr{J}^+(x^*,v^*)} g\big(x^*, Q_{U|V}(\tau^*|V)\big), \tag{4.13}$$

where the bounds are identified. proof:

It is instructive to compare the bounds in 4.13 to those resulting from $\mathscr{R}$ in 4.10. Because $\mathscr{J}^-$, $\mathscr{J}^+$ are larger classes than $\mathscr{G}^-,\mathscr{G}^+$, the bounds in 4.13 are generally tighter than those in 4.10, and hence also than those in 4.9.

An example of the difference between the bounds in equations (4.13) and (4.10) arises when we consider 4.1; note for instance that $\mathscr{G}^-(1,v^*) = \varnothing$, whereas $\mathscr{J}^-(1,v^*) = \{V(1,3) - V(1,2)\}$. The difference between the bounds arising from $\mathscr{R}$ and $\mathscr{R}^*$ becomes extreme when the instrument has continuous variation. Consider again 4.1, but with $\mathscr{S}_z = \mathbb{R}$ and for the special case that $h(z,v) = I\big(v > H(z)\big)$ for all v,z, where H is a continuous distribution function. $\mathscr{R}$ is as before not satisfied, so it produces no bounds. Now $\mathscr{R}^*$. Note that $V(0,z) = \big(0,H(z)\big]$ and $V(1,z) = \big(H(z),1\big]$. Take $z^* = H^{-1}(v^*)$. If $Q_{U|V}(\tau^*|v)$ is continuous at $v = v^*$, then

$$\begin{aligned} &\lim_{t\to\infty} Q_{U|V}\big(\tau^*|V(1,z^*-1/t) - V(1,z^*)\big) \\ &\qquad = \lim_{t\to\infty} Q_{U|V}\big(\tau^*|V(1,z^*) - V(1,z^*+1/t)\big) \\ &\qquad\qquad = Q_{U|V}(\tau^*|v^*), \end{aligned}$$

and using $\mathscr{R}^*$ ψ^* is then point–identified for $x^* = 1$ (and similarly for $x^* = 0$) since

$V(1, z^* - 1/t) - V(1, z^*) \in \mathscr{J}^-(1, v^*)$ and $V(1, z^*) - V(1, z^* + 1/t) \in \mathscr{J}^+(1, v^*)$ for all $t \geq 1$.

The bounds in theorem 19 are sharp under the stated conditions.[6] To see this, note that the inequalities $\phi^*(V_0) \leq \psi^* \leq \phi^*(V_1)$ for all $V_0, V_1 \subset \mathscr{U}$ such that $V_0 \leq v^* \leq V_1$ cannot be improved on and that $\phi^*(V)$ is not identified unless it is expressed as a mapping from one of $\mathscr{P}, \mathscr{P}^2, \cdots, \mathscr{P}^\infty$ to $\mathbb{R}$, where $\mathscr{P} = \{p : \mathbb{R} \to [0,1] : p(y) = \mathbb{P}[Y \leq y | X = x^*, Z = z]$ for some $z \in \mathscr{S}_z\}$. Because of the limitations of local identification conditions, Chesher's (2005) analysis is restricted to using $\mathscr{P}$ only but assumption 39 enables us to use all of $\mathscr{P}, \mathscr{P}^2, \ldots, \mathscr{P}^\infty$.

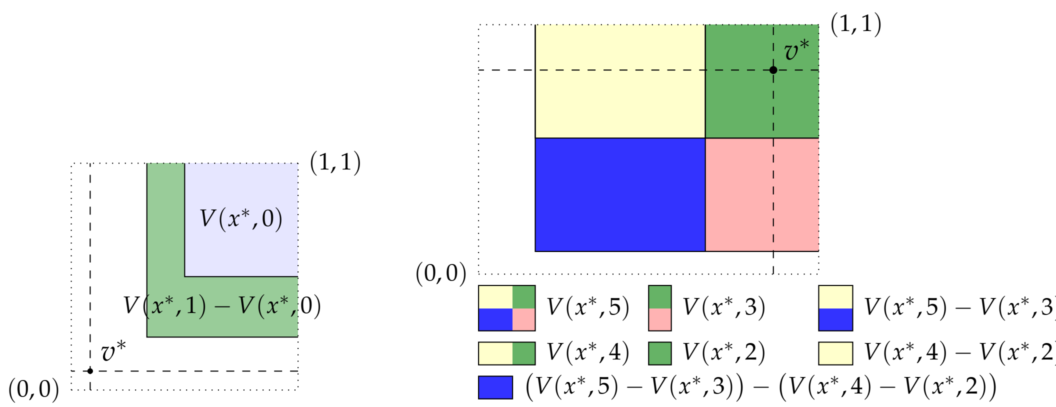

Figure 4.2. How to obtain upper and lower bounds when $x^* = (1, 1)$.

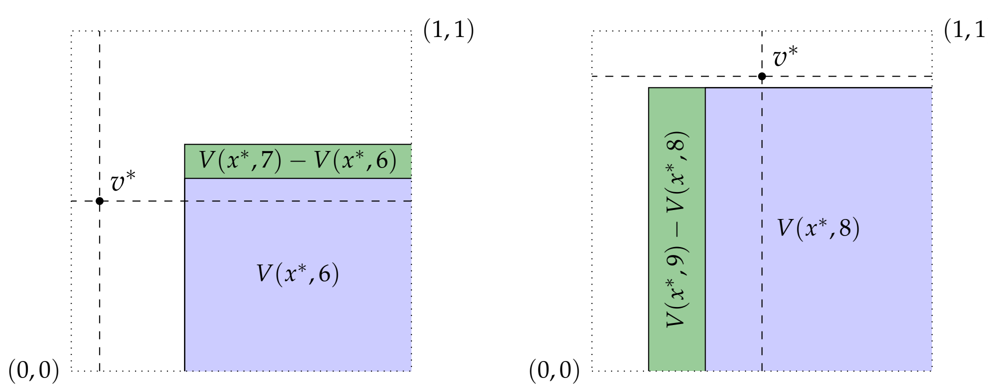

Figure 4.3. How to obtain upper and lower bounds when $x^* = (1, 0)$.

As mentioned earlier, we conclude with an example for vector–valued X. In our examples X contains two binary variables: a college education dummy and a field of

[6] I.e. the bounds cannot be improved without further restrictions.

specialization dummy (marketable or not). See figures 4.2 and 4.3. For $x^* = (1,1)$, the $V(x^*, z)$–sets are rectangles in $\mathscr{U}^2$ including $(1,1)$ and $V(x^*, 1) - V(x^*, 0)$ is simply the difference between two such rectangles. $V(x^*, 0), V(x^*, 1), V(x^*, 1) - V(x^*, 0)$ can all be used to construct upper bounds. But to obtain a lower bound at all, one must use at least four $V(x^*, z)$–rectangles, as indicated in the second graph in 4.2. The rectangle below v^* is the collection of $v = (v_1, v_2)$–points for which v_1 is a level of (college–)talent sufficient to obtain a college degree with demographics $z = 4$ or $z = 5$ but not with $z = 2$ or $z = 3$ and for which v_2 is a level of (marketable field–)talent sufficient to have a marketable field of specialization with demographics $z = 3$ or $z = 5$ but not with $z = 2$ or $z = 4$. As 4.3 illustrates, finding upper and lower bounds for $x^* = (1,0)$ is easier than finding a lower bound for $x^* = (1,1)$.

4.3 Revisiting Angrist and Krueger (1991)

We now illustrate the difference between $\mathscr{R}$ and $\mathscr{R}^*$ by using the Angrist and Krueger (1991) data. Angrist and Krueger (1991) estimated a wage equation with years of schooling as a(n endogenous) regressor. They used quarter of birth dummies as instruments. Chesher (2005) concluded that the Angrist and Krueger (1991) instruments do not satisfy $\mathscr{R}$ for any value of years of schooling and for any level of talent. In this section we determine whether $\mathscr{R}^*$ is satisfied.

Before we proceed, we comment on the plausibility of global exclusion, exogeneity and rank conditions. Exclusion restrictions and invariance restrictions of the distribution of latent variables given instruments are at best partially testable[7] so they are usually justified by economic reasoning. For instance, the wage equation in Angrist and Krueger (1991) does not include the birth–quarter variables, the exogeneity of which the authors justified by arguing the independence of ability and birth–quarters. Rank conditions, on the other hand, are restrictions on the joint distribution of observables. So they are generally testable once exogeneity of instruments is assumed. Potential failure of rank conditions has received more attention than failure of exclusion restrictions, for instance in the weak instrument literature.

To simplify our discussion we define X to be limited to the values $\{0, 1, 2\}$; no more than 6 years, 7–12 years, and more than 12 years of education, respectively. The instrument

[7] E.g. by means of a test of overidentifying restrictions

Z equals the quarter of birth (1–4). 4.1 summarizes the effect of the instruments, where we pretend that the estimated probabilities equal the true probabilities.

	z			
$x \downarrow$	1	2	3	4
0	0.0317	0.0315	0.0280	0.0270
1	0.6119	0.6038	0.5977	0.5946
2	1.0000	1.0000	1.0000	1.0000
n_z	81,671	80,138	86,856	80,844

Table 4.1. The effect of the birth quarter instrument on the education

Because $\mathbb{P}[X < x|Z = z] < \mathbb{P}[X \leq x|Z = \tilde{z}]$ for all $x, z, \tilde{z}$, $\mathscr{R}$ (or equivalently 4.8) is not satisfied, irrespective of the value of x^* and v^*. The birth quarter instruments are hence too weak for $\mathscr{R}$ to be satisfied.

$\mathscr{R}^*$ is different, however. For instance, consider $x^* = 2$ and $v^* = 0.6$. Since $V(2,3) \subset V(2,4)$ and $V(2,4) - V(2,3) = \big(0.5946, 0.5977\big] \leq v^* \leq V(2,1) = \big(0.6119, 1.0000\big]$, $\mathscr{R}^*$ is satisfied for $x^* = 2$ and $v^* = 0.6$.

The above example is flawed in two respects. First, we used estimated rather than true probabilities. Note however that the sample size is large and that this is just an example to illustrate the possibility of using $\mathscr{R}^*$ when $\mathscr{R}$ is not satisfied. Second, in the above example, Z can only take a small number of different values. With more variation in the instrument, the difference in identifying potential of $\mathscr{R}$ and $\mathscr{R}^*$ increases exponentially.

4.4 Proofs of lemmas

Proof of lemma 16

By assumption 38, $Q_{Y|X,V}(\tau|x^*, v^*) = g\big(x, Q_{U|X,V}(\tau|x^*, v^*)\big)$. Recall that by assumption 41, $\mathscr{Z}(x^*, v^*)$ is nonempty. Thus, $X = x^*, V = v^* \Leftrightarrow Z \in \mathscr{Z}(x^*, v^*), V = v^*$, such that $Q_{U|X,V}(\tau|x^*, v^*) = Q_{U|V,Z}\big(\tau|v^*, \mathscr{Z}(x^*, v^*)\big)$, which equals $Q_{U|V}(\tau|v^*)$ by assumption 39. □

Proof of lemma 17

By assumption 38, $Q_{Y|X,Z}(\tau|x, z) = g\big(x, Q_{U|X,Z}(\tau|x, z)\big)$. Since $X = x, Z = z \Leftrightarrow V \in V(x, z), Z = z$ and by assumption 39,

$$Q_{U|X,Z}(\tau|x, z) = Q_{U|V,Z}\big(\tau|V(x, z), z\big) = Q_{U|V}\big(\tau|V(x, z)\big). \square$$

Proof of lemma 18

We establish the upper bound; the argument for the lower bound is virtually identical. Let z be such that $V(x^*, z) \geq v^*$. Then by lemma 17,

$$Q_{Y|X,Z}(\tau^*|x^*, z) = g\{x^*, Q_{U|V}(\tau^*|V(x^*, z))\} \geq g\{x^*, Q_{U|V}(\tau^*|v^*)\} = \psi^*,$$

where the weak inequality follows from assumption 38 and the fact that for any $u \in \mathscr{U}$

$$\begin{aligned}\mathbb{P}\big[U \leq u|V \in V(x^*, z)\big]\mathbb{P}\big[V \in V(x^*, z)\big] &= \int_{V(x^*,z)} \mathbb{P}\big[U \leq u|V = v\big]dv \\ &\leq \int_{V(x^*,z)} \mathbb{P}\big[U \leq u|V = v^*\big]dv = \mathbb{P}\big[U \leq u|V = v^*\big]\mathbb{P}\big[V \in V(x^*, z)\big],\end{aligned}$$

which implies that $Q_{U|V}(\tau^*|V(x^*, z)) \geq Q_{U|V}(\tau^*|v^*)$. □

Proof of lemma 19

We show (i) where (ii) follows similarly. Note that for any y by the conditions on V_1, V_2,

$$\mathbb{P}\big[g(x^*, U) \leq y|V \in V_2 - V_1\big] = \frac{\mathbb{P}\big[g(x^*, U) \leq y|V \in V_2\big]\mu(V_2) - \mathbb{P}\big[g(x^*, U) \leq y|V \in V_1\big]\mu(V_1)}{\mu(V_2 - V_1)}. \tag{4.14}$$

Now $\mathbb{P}\big[g(x^*, U) \leq y|V \in V_j\big]$ is identified for $j = 1, 2$ and all y because $\phi(\tau, V_j)$ is identified for $j = 1, 2$ and all $\tau \in \mathscr{U}$. Further, since V_1, V_2 are disjoint and $\mu(V_1), \mu(V_2)$ are identified by assumption, so is $\mu(V_2 - V_1) = \mu(V_2) - \mu(V_1)$. So the left hand side in 4.14 is identified for all y. Invert the conditional distribution function to obtain the conditional quantile. □

4.5 Proof of theorem

Proof of Theorem 19

Recall that $\mathscr{V}(x^*) \subset \mathscr{K}$ by the discussion following lemma 19. Therefore, $\mathscr{D}(x^*) \subset \mathscr{K}$ by lemma 19 and by construction of $\mathscr{D}(x^*)$. Combining this with equation 4.11 (and its converse when $V \leq v^*$) concludes the proof. □

4.6 Miscellaneous

4.6.1 Normalization

Consider an arbitrary function $g^*(x, u^*)$ with arbitrarily distributed u^*, where g^* is weakly increasing in its second argument. Letting ω be the quantile function of u^*, there exists a uniform random variable U such that $u^* = \omega(U)$. Therefore, $g(x,u) = g^*\big(x, \omega(u)\big)$, is still weakly increasing in u. Since h is weakly increasing in its second argument and V is uniform, we have $h(z,v) = Q_{X|Z}(v|z)$. Left–continuity holds, because $\mathbb{P}[X \leq x|Z = z]$ is a CADLAG function of x and $Q_{X|Z}(v|z) = \inf\{x : \mathbb{P}[X \leq x|Z = z] \geq v\}$.

4.6.2 Manski and Tamer

Below is a somewhat more detailed discussion of the relationship between Manski and Tamer (2002a) and the present paper.

Manski and Tamer (2002a) assume that V is scalar–valued and that upper and lower bounds (V_0^{MT}, V_1^{MT}) on V are observed. They further assume monotonicity of $\delta(x^*, v)$ in v and that $\mathbb{E}[Y|X, V, V_0^{MT}, V_1^{MT}] = \mathbb{E}[Y|X, V]$ a.s..

Replacing Y, V_0^{MT}, V_1^{MT} with $\mathbb{1}(Y \leq y), \mathbb{P}[X < x^*|Z], \mathbb{P}[X \leq x^*|Z]$, respectively,[8] facilitates a comparison of Manski and Tamer (2002a) with Chesher (2005) and the current paper. The availability of instruments in the triangular model provides us with more structure to be exploited. To simplify the discussion, suppose that both $\mathbb{P}[X < x^*|Z = z]$ and $\mathbb{P}[X \leq x^*|Z = z]$ are invertible in z, such that $(V_0^{MT}, V_1^{MT}) = \big(\mathbb{P}[X < x^*|Z = z], \mathbb{P}[X \leq x^*|Z = z]\big)$ if and only if $Z = z$. Then, equation (A3) in the proof of proposition 1 in Manski and Tamer (2002a) would become

$$\begin{aligned} \mathbb{P}\big[Y \leq y|X = x^*, V = \mathbb{P}[X < x^*|Z = z]\big] &\leq \mathbb{P}[Y \leq y|X = x^*, Z = z] \\ &\leq \mathbb{P}\big[Y \leq y|X = x^*, V = \mathbb{P}[X \leq x^*|Z = z]\big]. \end{aligned} \tag{4.15}$$

Note here that the monotonicity assumption of Manski and Tamer (2002a)[9] in combination with the inversion of the distribution functions in 4.15 leads to the quantile version of proposition 1 of Manski and Tamer (2002a), which coincides with the bounds in 4.9 that our main theorem, theorem 19, improves upon.

[8] $\mathbb{1}$ is the indicator function.

[9] I.e. $\mathbb{E}[\mathbb{1}\{Y \leq y\}|X = x^*, V = v]$ is weakly decreasing in v.

Bibliography

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

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

AGUIRREGABIRIA, V., AND P. MIRA (2007b): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 31–53.

ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does compulsory school attendance affect schooling and earnings?," *Quarterly Journal of Economics*, 106, 979–1014.

ARADILLAS-LOPEZ, A. (2007): "Pairwise Difference Estimation of Incomplete Information Games," Working papers, Princeton University.

——— (2008): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Working papers, Princeton University.

ARADILLAS-LOPEZ, A., AND E. TAMER (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business & Economic Statistics*, 26, 261–310.

ATHEY, S. (2001): "Single crossing properties and the existence of pure strategy equilibria in games of incomplete information," *Econometrica*, pp. 861–889.

AUMANN, R. J. (1964): "Mixed vs. behavior strategies in infinite extensive games," *Annals of Mathematics Studies*, 52, 627–630.

BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2009): "Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game," *Unpublished manuscript, Dep. Econ., Univ. Minn., Minneapolis*.

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

BAJARI, P., H. HONG, AND S. RYAN (2004): "Identification and estimation of discrete games of complete information," *NBER Working paper*.

BERRY, S. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, pp. 889–917.

BERRY, S., AND E. TAMER (2006): "Identification in Models of Oligopoly Entry," in *Advances in economics and econometrics: theory and applications, ninth World Congress*, p. 46. Cambridge Univ Pr.

BIERENS, H. (1987): "Kernel estimators of regression functions," in *Advances in Econometrics: Fifth World Congress*, vol. 1, pp. 99–144.

——— (1996a): "Topics in advanced econometrics," *Cambridge Books*.

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

BIERENS, H. J. (1996b): *Topics in Advanced Econometrics*, no. 9780521565110 in Cambridge Books. Cambridge University Press.

BILLINGSLEY, P. (1995): *Probability and measure*. Wiley, New York, third edn.

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

BRESNAHAN, T. F., AND P. C. REISS (1990): "Entry in Monopoly Markets," *The Review of Economic Studies*, 57(4), 531–553.

——— (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48(1-2), 57–81.

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

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

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

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

CHAMBERLAIN, G. (1986): "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32(2), 189–218.

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

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

CHESHER, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

——— (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525–1550.

CILIBERTO, F., AND E. TAMER (2010): "Market structure and multiple equilibria in airline markets," *Econometrica, formthcoming*.

DE CASTRO, L. I. (2007): "Affiliation, equilibrium existence and the revenue ranking of auctions," Economics working papers, Universidad Carlos III, Departamento de Economï£¡a.

DE PAULA, A., AND X. TANG (2010): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," Working papers.

ELLICKSON, P., AND S. MISRA (2008): "Supermarket pricing strategies," *Marketing Science*, 27(5), 811–828.

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

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. MIT Press.

GRIECO, P. (2010): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," Working papers.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.

HARSANYI, J. C. (1967–68): "Games with Incomplete Information Played by "Bayesian" Players," *Management Science*, 14, 159–182 and 320–334 and 486–502.

HECKMAN, J. (1990): "Varieties of selection bias," *The American Economic Review*, pp. 313–318.

HOTELLING, H. (1929): "Stability in competition," *The economic journal*, 39(153), 41–57.

HOTZ, V. J., AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60(3), 497–529.

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

IMBENS, G., AND W. NEWEY (2009): "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77(5), 1481–1512.

JIA, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

JUN, S. J. (2009): "Local structural quantile effects in a model with a nonseparable control variable," *Journal of Econometrics*, 151, 82–97.

JUN, S. J., J. PINKSE, AND H. XU (2009): "Tighter bounds in triangular models: supplementary material," Discussion paper, The Pennsylvania State University.

——— (2010): "Discrete endogenous variables in weakly separable models," Discussion paper, Penn State.

KHAN, S., AND E. TAMER (2009): "Irregular Identification, Support Conditions and InverseWeight Estimation," Working papers.

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

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

KRAUTH, B. (2007): "Peer and selection effects on youth smoking in California," *Journal of Business and Economic Statistics*, 25(3), 288–298.

LAFFONT, J.-J., H. OSSARD, AND Q. VUONG (1995): "Econometrics of First-Price Auctions," *Econometrica*, 63(4), 953–980.

LEE, L. (2007): "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models," *Journal of Econometrics*, 137(2), 489–514.

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

MA, L., AND R. KOENKER (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471–506.

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

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

MANSKI, C., AND E. TAMER (2002a): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519–546.

MANSKI, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3(3), 205–228.

——— (1985): "Semiparametric analysis of discrete response : Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27(3), 313 – 333.

——— (1988): *Analog estimation methods in econometrics*. Chapman and Hall New York.

MANSKI, C. F., AND E. TAMER (2002b): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

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

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

MYERSON, R. (1997): *Game theory: analysis of conflict*. Harvard Univ Pr.

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

NEWEY, W. K., AND D. MCFADDEN (1986): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier.

NEWEY, W. K., J. L. POWELL, AND F. VELLA (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67(3), 565–603.

OSBORNE, M., AND A. RUBINSTEIN (1999): *A course in game theory*. MIT press.

PAGAN, A., AND A. ULLAH (1999): *Nonparametric econometrics*. Cambridge University Press, New York, NY, USA.

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

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2003b): "Identification and Estimation of Dynamic Games," Working Paper 9726, National Bureau of Economic Research.

PINKSE, J. (2000): "Nonparametric two–step regression estimation when regressors and error are dependent," *Canadian Journal of Statistics*, 28(2), 289–300.

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

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

——— (2010): "The Future of Spatial Econometrics," Forthcoming.

POLLARD, D. (1989): "Asymptotics via Empirical Processes," *Statistical Science*, 4(4), 341–354.

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

POWELL, J. L. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54(6), 1435–1460.

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

ROBINSON, P. (2008): "Developments In the Analysis of Spatial Data," *the Journal of The Japan Statistical Society*, 38(1), 87–96.

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

SEIM, K. (2006a): "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 37(3), 619–640.

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

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

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

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

TANG, X. (2009): "Estimating Simultaneous Games with Incomplete Information under Median Restrictions," Working papers.

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

VYTLACIL, E., AND N. YILDIZ (2007): "Dummy endogenous variables in weakly separable models," *Econometrica*, 75(1), 757–779.

XU, H. (2009): "Estimation of Bayesian Nash Equilibria In Two-player Games With Correlated Private Signals," Working papers, Pennsylvania State University.

Haiqing Xu

Curriculum Vitae

EDUCATION:

- Ph.D., Economics, Penn State University, expected August 2011
- M.A., Economics, Peking University, June 2005
- B.A., Management, Peking University, June 2000

Ph.D. THESIS:

- Discrete Choice Models with Endogeneity

  Thesis Advisors: Professors Joris Pinkse and Sung Jae Jun

FIELDS:

- Econometrics, and Applied Econometrics

PAPERS:

- "Discrete Choice Models With Local Interactions: A Game Theoretical Approach," **Job market paper**, November 2010
- "Tighter Bounds in Triangular Systems," joint with Sung Jae Jun and Joris Pinkse. *the Journal of Econometrics, 161 (2011), 122-128*
- "Discrete Endogenous Variables in Weakly Separable Models," joint with Sung Jae Jun and Joris Pinkse. Revised and Submitted to *the Econometrics Journal*, October 2010
- "Semiparametric Identification and Estimation of Binary Decision Games of Incomplete Information with Correlated Private Signals," joint with Yuanyuan Wan, submitted, May 2010
- "Estimation of Bayesian Nash Equilibria in Static Discrete Games with Correlated Private Signals", May 2009

PAPERS IN PROGRESS:

- "Identification of Treatment Effects In a Triangular System of equations," joint with Sung Jae Jun, Joris Pinkse and Nese Yildiz
- "Estimating a Nonparametric Triangular Model with Binary Endogenous Regressors," joint with Sung Jae Jun, and Joris Pinkse
- "Causal Discourse in a Game of Incomplete Information," joint with Karim Chalak, and Halbert White