Estimation of Discrete Games with Correlated Types[*]

Haiqing Xu,[†]
Department of Economics
The University of Texas at Austin

November 2012

This paper focuses on the identification and estimation of static games of incomplete information with correlated types. Instead of making the assumption of (conditional) independence of players' types to simplify the equilibria set, I establish a method that allows to identify subsets of the space of covariates (i.e. publicly observed state variables in payoff functions), for which there exists a unique *Bayesian Nash Equilibrium* (BNE) and the equilibrium strategies are monotone functions. The unique monotone pure strategy BNE can be characterized in a simple manner, based on which I propose an estimation procedure exploiting the information contained in the subset of the covariate space, and establish the consistency and the limiting distribution of the estimator.

Key Words: Incomplete Information Game, Monotone Pure Strategy BNE, Maximum Likelihood Estimation

JEL Classification Codes: C35; C62; C72.

[*]This paper is a revision of Chapter 2 of my thesis. I am indebted to Joris Pinkse and Sung Jae Jun for their guidance and support. I am grateful to Quang Vuong, a co–editor and three anonymous referees, whose comments and suggestions have helped generate a substantially improved paper. I also thank Herman Bierens, Frank Erickson, Edward Green, Han Hong, Hiroyuki Kasahara, Nianqing Liu, Xun Tang, Yuanyuan Wan and seminar participants at 2009 Far East and South Asia Meeting of the Econometric Society for helpful comments. Any and all errors are my own.
[†]Department of Economics, The University of Texas at Austin, BRB 3.160, Mailcode C3100, Austin, TX, 78712, h.xu@austin.utexas.edu

1. Introduction

This paper focuses on the identification and estimation of static games of incomplete information with correlated types. Instead of making the assumption of (conditional) independence of players' types to simplify the equilibria set, I establish a method that allows to identify subsets of the space of covariates (i.e. publicly observed state variables in payoff functions), for which there exists a unique *Bayesian Nash Equilibrium* (BNE) and the equilibrium strategies are monotone functions. The unique monotone pure strategy BNE can be characterized in a simple manner, based on which I propose an estimation procedure exploiting the information contained in the subset of the covariate space, and establish the consistency and the limiting distribution of the estimator.

Static discrete games, like the one I study, are of interest because of their empirical applications. Bjorn and Vuong (1984), for example, studies labor force participation. Recently, this class of games are more widely adopted in the empirical industrial organization to study firms' entry behavior (e.g. Berry, 1992; Bresnahan and Reiss, 1990, 1991a,b; Ciliberto and Tamer, 2009; Jia, 2008). In much of this literature, an agent's payoff often depends on not only her covariate variables, but also other agents' choices. Therefore, the strategic effects are embedded in the equilibrium solution to the simultaneous equations (i.e., best responses of the game).

In this paper, I study a full parametric binary game of incomplete information, which might have multiple equilibria.[1] The proposed methodology contributes to the literature in two aspects. First, I allow players' types to be correlated, which is motivated by empirical concerns. In much of the incomplete information game literature, e.g., Aguirregabiria and Mira (2002), Bajari, Hong, Krainer, and Nekipelov (2010) and Pesendorfer and Schmidt-Dengler (2003), the identification strategy relies heavily on the fact that a player's equilibrium beliefs about her rivals' choices depend on observed state variables only and can be nonparametrically estimated thereof, which is mainly a consequence of the (conditional) independent types condition. In contrast, I assume that players' private payoff shocks (types) conform to joint normal distribution and are positively correlated with each other. The correlation coefficient is also a parameter of interest in my structural model.

The quest for correlated types in discrete games is motivated by several considerations. The (conditional) independence of types implies that players' actions should also be conditionally

[1] Aradillas-Lopez (2010) estimates the same game structure without making parametric restrictions on types, by assuming that players do not have exact knowledge about the distributions involved and then using an equilibrium concept defined in Aumann (1987).

independent given the covariates, which may not happen in the data.[2] Moreover, this restriction implies that all equilibrium solutions must be monotone pure strategy BNEs, which is convenient but rules out the possibility of non–monotone strategy BNE. Second, allowing correlation is also important for reasons of the model specification. For example, consider two firms entering a local market: one would expect the private payoff shocks on the profitability of entry to be positively correlated with each another, if the shocks depend on some common factors of the local market and each player only observes the integrated value of the shock, but can not decompose it into the idiosyncratic noise and the other part from the common factors.

Second, the proposed approach does not make any assumption on equilibrium selection mechanism. In the literature, the multiple equilibria issue invokes ad–hoc equilibrium selection assumptions in data-generating process, i.e., when there are multiple equilibria, only one equilibrium is being played in data (see, e.g., Bajari, Hong, Krainer, and Nekipelov, 2010; Tang, 2010; Aradillas-Lopez, 2010; Wan and Xu, 2009; Sweeting, 2009). Dropping the independence assumption even complicates the multiple equilibria issue. First, it is difficult to characterize all the equilibria, especially for the ones with non–monotone strategies. Second, the number of equilibria is unknown.[3] Hence, even if one imposes an equilibrium selection rule, it is difficult to implement in practice.

This paper extends a novel approach called "*level–k rationality*" in Aradillas-Lopez and Tamer (2008) for the identification of a structural model and show that, in a subset of the space of covariates (i.e. publicly observed state variables), there is a unique BNE, in which equilibrium strategies are monotone functions. Moreover, this subset can be identified in a straightforward manner, and therefore is estimable.[4]

The (unique) monotone pure strategy BNE can be characterized in a simple manner. In the presence of correlation, it is costly to obtain a closed–form solution for the equilibrium in general. In the binary decision game considered in this paper, an important insight is that a monotone pure strategy is fully characterized by a cutoff value in the support of type. Therefore, a numerical solution of BNE can be solved as a fixed point in a vector space, if the equilibrium is a monotone pure strategy BNE.

[2]There are other two possible sources for the observed correlation between players' actions: unobserved heterogeneity (see Grieco, 2010) and multiple equilibria (see De Paula and Tang, 2012).

[3]Suppose U conforms a binormal distribution and U_1 and U_2 are conditionally independent given X. Then it is known that the number of equilibria ranges from 1 to 3, see, e.g., (see Grieco, 2010). However, such kind of results does not obtain if U_1 and U_2 are positively correlated.

[4]De Paula and Tang (2012) propose a semiparametric test procedure for multiplicity of equilibria in the data under the assumption that private information are conditionally independent.

This paper is organized as follows. Section 2 describes the game model. Section 3 provides characterization of BNEs and monotone pure strategy BNEs. I show that there is a unique BNE, which has monotone pure strategies, given regressors belonging to a subset. In section 4 and 5, I establish the identification and estimation of the structural parameters, respectively. Section 6 provides Monte Carlo experiment studies to illustrate the performance of the proposed estimator in finite samples and Section 7 concludes. All proofs are in the appendix.

2. The Model

Consider the following 2–by–2 static game of incomplete information:

		PLAYER 2 $Y_2 = 1$	$Y_2 = 0$
PLAYER 1	$Y_1 = 1$	$X_1'\beta_1 - \alpha_1 - U_1,\ X_2'\beta_2 - \alpha_2 - U_2$	$X_1'\beta_1 - U_1,\ 0$
	$Y_1 = 0$	$0,\ X_2'\beta_2 - U_2$	$0,\ 0$

TABLE 1: Two–player simultaneous move game of incomplete information

where $X = (X_1, X_2) \in \mathscr{S}_X \subseteq \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ represents public information to both players. The payoff shock $U_j \in \mathbb{R}$ $(j = 1, 2)$ is player j's private information, which is only observed by j, not his rival. Y_j is the choice of player j. Let $U = (U_1, U_2)$ be independent of X,[5] and conforms to a joint normal distribution with unit variances and correlation parameter $\rho_0 \in [0, 1)$[6], which is assumed to be common knowledge of both players. $\beta_j \in \mathbb{R}^{k_j}$ and $\alpha_j \in \mathbb{R}_+$ are coefficients in the payoff function and α_j measures the size of the strategic effect.[7] Let $\theta_0 = \left(\alpha_1, \alpha_2, \beta_1', \beta_2', \rho_0\right)' \in \Theta$ be the parameters of interest. Throughout this paper, I use $\theta = \left(a_1, a_2, b_1', b_2', \rho\right)'$ to denote a generic parameter value in the parameter space $\Theta \subseteq \mathbb{R}^{2+k_1+k_2} \times [0, 1]$.

A game and the according equilibria with the similar setup can also be found in Pesendorfer and Schmidt-Dengler (2003) and references therein. In this incomplete information game, I adopt the standard pure strategy BNE solution concept (see, e.g., Aumann, 1964; Harsanyi, 1967–68). In equilibrium, player j's strategy is a function $s_j^*(X, U_j)$, where $s_j^* : \mathbb{R}^{k_1+k_2} \times \mathbb{R} \to \{0, 1\}$ maps all j's information to a binary decision. Player j chooses s_j^* in a way such that it maximizes her expected payoff: choosing $s_j^* = 1$ if and only if $X_j'\beta_j - \alpha_j \mathbb{E}\left[s_{-j}^*(X, U_{-j})|X, U_j\right] - U_j \geq 0$, where $\mathbb{E}(s_{-j}^*|X, U_j)$ is the beliefs of her rival's move in equilibrium. In other words, fix $X = x \in$

[5] The independence between X and U is not essential for the equilibrium analysis in Section 3.

[6] Aradillas-Lopez (2010) developed a semiparametric approach without a parametric specification of the distribution of U.

[7] α_j is restricted to be nonnegative only for the brevity of notation.

$\mathscr{S}_X$, the equilibrium strategy profile $s^* = \{s_1^*(x,\cdot), s_2^*(x,\cdot)\}$ is a fixed point solving the following simultaneous equations system

$$s_j(x,u_j) = \mathbf{1}\left\{x_j'\beta_j - \alpha_j \mathbb{E}\left[s_{-j}(x,U_{-j})|U_j = u_j\right] - u_j \geq 0\right\}, \quad \text{for } j = 1,2, \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function. Note that I drop the conditioning variable $X = x$ in j's belief term $\mathbb{E}\left[s_{-j}(x,U_{-j})|U_j = u_j\right]$ due to the independence between X and U. In equation (1), player j's equilibrium strategy depends on not only j's observed state variables x_j, but also rival's state variables x_{-j}. This is because x_{-j} affects player j's expectation on her rival's choice.

This binary game of incomplete information can be interpreted as an entry model, where two firms simultaneously decide whether to enter a local market or not (see, e.g., Ciliberto and Tamer, 2009). Before they make their decisions, information X is disclosed publicly, and each player observes a private payoff shock for entry profit. There are interactions between the players' strategies: α_j measures the magnitude of strategic impacts. Moreover, each player's entry profit is parametrized by a linear sum of the publicly observed term, the private shock and the strategic effect. Note that asymmetry in this game arises as long as $X_1'\beta_1 \neq X_2'\beta_2$, which reflects the fact that one player may have a commonly known advantage or disadvantage for entering this local market.

From the above discussion, for a given realization $x = (x_1, x_2)$ of public states X, a BNE is a fixed point in the functional space. To obtain such a solution, a convenient assumption is widely used in the literature that U_1 and U_2 are conditionally independent given X, which means that an individual's private payoff shock does not contain any additional information for the rival's choice. When U_1 and U_2 are allowed to be positively correlated, difficulty arises to characterize each equilibrium, especially non–monotone strategy BNEs. Therefore, it is costly to know the whole equilibria set.

3. Characterization of Equilibria Set

As a special class of pure strategy BNE, a monotone pure strategy BNE can be characterized in a much simpler manner. Hence, it is feasible to characterize the equilibria set for some values of public states if these realizations of public state variables permit only monotone pure strategy BNEs, especially when the equilibrium is unique.

Given public information $X = x$, a monotone pure strategy BNE can be characterized by a vector $u^*(x) = (u_1^*(x), u_2^*(x)) \in \mathbb{R}^2$, such that for $j = 1, 2$,

$$s_j^*(x, u_j) = \mathbb{1}\left[u_j \le u_j^*(x)\right], \tag{2}$$

where the $u^*(x)$ satisfies the following mutual consistency conditions

$$x_j'\beta_j - \alpha_j \mathbb{P}(U_{-j} \le u_{-j}^*(x) | U_j = u_j) - u_j \ge 0 \Longleftrightarrow u_j \le u_j^*(x), \tag{3}$$

for $j = 1, 2$. Hence, given $X = x$, a monotone strategy BNE obtains by solving a fixed point $u^*(x)$ in the vector space $\mathbb{R}^2$.

I now define a subset $\mathcal{M}(\theta_0)$ of the covariate space, which depends on the underlying parameter value θ_0, such that for any $x \in \mathcal{M}(\theta_0)$ all the equilibria in the game are monotone strategy BNEs. Later, I will refine $\mathcal{M}(\theta_0)$ to a smaller subset $\mathcal{U}(\theta_0)$, in which each value of public states will admit a unique monotone pure strategy BNE.

For $j = 1, 2$, let the function $h_j(\cdot; \theta_0) : \mathbb{R}^2 \to \mathbb{R}$ be defined as follows:

$$h_j(u; \theta_0) = u_j + \alpha_j \mathbb{P}(U_{-j} \le u_{-j} | U_j = u_j).$$

In Definition 1 below I will define a rectangular $\mathcal{I}(x; \theta_0)$ on the support $\mathscr{S}_U$ through a recursion scheme, which corresponds to Aradillas-Lopez and Tamer (2008)'s "level–k rationality" — a notion weaker than the BNE solution concept.

Definition 1. *For any $x \in \mathscr{S}_X$, let $\mathscr{V}_{j,1}^-(x; \theta_0) = x_j'\beta_j - \alpha_j$ and $\mathscr{V}_{j,1}^+(x; \theta_0) = x_j'\beta_j$. Let further*

$$\mathscr{V}_{j,k}^-(x; \theta_0) = x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \le \mathscr{V}_{-j,k-1}^+(x; \theta_0) | U_j = \mathscr{V}_{j,k-1}^-(x; \theta_0)\right),$$
$$\mathscr{V}_{j,k}^+(x; \theta_0) = x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \le \mathscr{V}_{-j,k-1}^-(x; \theta_0) | U_j = \mathscr{V}_{j,k-1}^+(x; \theta_0)\right).$$

Let $\mathscr{V}_j^-(x; \theta_0) = \lim_{k\to\infty} \mathscr{V}_{j,k}^-(x; \theta_0)$ and $\mathscr{V}_j^+(x; \theta_0) = \lim_{k\to\infty} \mathscr{V}_{j,k}^+(x; \theta_0)$. Moreover, define $\mathcal{I}_{j,k}(x; \theta_0) = \left[\mathscr{V}_{j,k}^-(x; \theta_0), \mathscr{V}_{j,k}^+(x; \theta_0)\right]$, $\mathcal{I}_j(x; \theta_0) = \left[\mathscr{V}_j^-(x; \theta_0), \mathscr{V}_j^+(x; \theta_0)\right]$, and $\mathcal{I}(x; \theta_0) = \mathcal{I}_1(x; \theta_0) \times \mathcal{I}_2(x; \theta_0)$.

Throughout the following analysis, I will use $\mathscr{V}_{j,k}^-(x)$, $\mathscr{V}_{j,k}^+(x)$, $\mathscr{V}_j^-(x)$ and $\mathscr{V}_j^+(x)$ in lieu of $\mathscr{V}_{j,k}^-(x; \theta_0)$, $\mathscr{V}_{j,k}^+(x; \theta_0)$, $\mathscr{V}_j^-(x; \theta_0)$ and $\mathscr{V}_j^+(x; \theta_0)$, respectively, to simplify my notation and emphasize their dependence on x. Noted that $\mathscr{V}_j^-(x)$ and $\mathscr{V}_j^-(x; \theta_0)$ are well–defined as the limits of sequences, because one can verify that both $\{\mathscr{V}_{j,k}^-(x)\}_{k=1}^\infty$ and $\{\mathscr{V}_{j,k}^+(x)\}_{k=1}^\infty$ are monotone sequences.

It should also be noted that $\mathscr{V}_j^-(x)$ and $\mathscr{V}_j^+(x)$ satisfy the following conditions:

$$\mathscr{V}_j^-(x) = x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \le \mathscr{V}_{-j}^+(x) | U_j = \mathscr{V}_j^-(x)\right),$$
$$\mathscr{V}_j^+(x) = x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \le \mathscr{V}_{-j}^-(x) | U_j = \mathscr{V}_j^+(x)\right).$$

By definition, there is $\mathcal{I}_{j,1}(x;\theta_0) \supseteq \cdots \supseteq \mathcal{I}_{j,k}(x;\theta_0) \supseteq \mathcal{I}_j(x;\theta_0)$ for all $k \in \mathbb{N}$.

For each $k \in \mathbb{N}$, let

$$\mathcal{M}_k(\theta_0) = \left\{x \in \mathscr{S}_X : \partial h_j(u;\theta_0)/\partial u_j \ge 0 \text{ for all } u \in \mathcal{I}_{1,k}(x;\theta_0) \times \mathcal{I}_{2,k}(x;\theta_0), j = 1,2\right\}$$

and

$$\mathcal{M}(\theta_0) \equiv \mathcal{M}_\infty(\theta_0) = \left\{x \in \mathscr{S}_X : \partial h_j(u;\theta_0)/\partial u_j \ge 0 \text{ for all } u \in \mathcal{I}(x;\theta_0), j = 1,2\right\}.$$

By definition, $\{\mathcal{M}_k(\theta_0)\}_{k=1}^\infty$ is a monotone increasing sequence of subsets on the support $\mathscr{S}_X$ and $\mathcal{M}(\theta_0)$ is the limit of the sequence.

The definition of $\mathcal{M}_k(\theta_0)$ is guided by Reny (2011), Theorem 4.1: h_j is required to be a non–decreasing function of u_j only on the support $\mathcal{I}_k(x;\theta_0)$, instead of the whole support $\mathbb{R}^2$. This condition is weaker than the *single crossing condition (SCC, see Athey, 2001)*, a sufficient condition for the existence of monotone pure strategy BNE. To see this, for instance, let $k = 1$. When $u_j \le x_j'\beta_j - \alpha_j$ (or $u_j \ge x_j'\beta_j$), player j's optimal decision is to choose $s_j^*(x, u_j) = 1$ (or $s_j^*(x, u_j) = 0$), which is irrelevant of the rival's strategy. Hence, the fact that h_j is non–decreasing within the interval $\mathcal{I}_{j,1}(x;\theta_0)$ guarantees a monotone best response to any rival's strategy. This argument can be generalized to $k = 2, 3, \cdots$ using "level–k rationality" in Aradillas-Lopez and Tamer (2008): in any equilibrium solution of BNE, it is for sure that player j's equilibrium response is: for any $k \in \mathbb{N}$, $s_j^*(x, u_j) = 1$ if $u_j < \mathscr{V}_{j,k}^-(x)$; $s_j^*(x, u_j) = 0$ if $u_j > \mathscr{V}_{j,k}^+(x)$. Theorem 1 summarizes the discussion above.

Theorem 1. *Suppose $X = x \in \mathcal{M}(\theta_0)$. All pure strategy BNEs in the game with $X = x$ are monotone pure strategy BNEs. Moreover, for any monotone pure strategy BNE, w.l.o.g., characterized by $u^*(x) \in \mathbb{R}^2$, there is $u^*(x) \in \mathcal{I}(x;\theta_0)$.*

Proof. See Appendix A.1 □

Note that, if $x \in \mathcal{M}(\theta_0)$, the expression $x_j'\beta_j - \alpha_j \mathbb{P}(U_{-j} \le u_{-j}^* | U_j = u_j) - u_j$ is a continuously decreasing function of u_j on the support $\mathcal{I}(x;\theta_0)$. Thus, that condition (3) is equivalent to

$$x_j'\beta_j - \alpha_j \mathbb{P}\big(U_{-j} \le u_{-j}^*(x) | U_j = u_j^*(x)\big) - u_j^*(x) = 0. \tag{4}$$

Hence, for $X = x \in \mathcal{M}(\theta_0)$, the set of equilibria obtains by solving equations (4). However, there can be multiple monotone pure strategy BNE's here, like in Bresnahan and Reiss (1990, 1991a), and Tamer (2003). Instead of imposing some equilibrium selection mechanism, I characterize a subset $\mathcal{U}(\theta_0)$ of $\mathcal{M}(\theta_0)$, which admits a unique monotone pure strategy BNE.[8]

For each $k \in \mathbb{N}$, let

$$\mathcal{U}_k(\theta_0) = \{x \in \mathscr{S}_X : \ \partial h_j(u;\theta_0)/\partial u_j > \partial h_j(u;\theta_0)/\partial u_{-j}, \ a.e. \ \forall \ u \in \mathcal{I}_{1,k}(x;\theta_0) \times \mathcal{I}_{2,k}(x;\theta_0), j = 1,2\}$$

and

$$\mathcal{U}(\theta_0) \equiv \mathcal{U}_\infty(\theta_0) = \{x \in \mathscr{S}_X : \ \partial h_j(u;\theta_0)/\partial u_j > \partial h_j(u;\theta_0)/\partial u_{-j} \ a.e. \ \forall \ u \in \mathcal{I}(x;\theta_0), j = 1,2\}.$$

Similar to $\{\mathcal{M}_k(\theta_0)\}_{k=1}^{\infty}$, the sequence of subsets $\{\mathcal{U}_k(\theta_0)\}_{k=1}^{\infty}$ is monotone increasing on the support $\mathscr{S}_X$ and $\mathcal{U}(\theta_0)$ is the limit of the sequence. It should also be noted that there is $\mathcal{U}_k(\theta_0) \subseteq \mathcal{M}_k(\theta_0)$ due to the fact $\partial h_j(u;\theta_0)/\partial u_{-j} \geq 0$ a.s..

Theorem 2. *Suppose $X = x \in \mathcal{U}(\theta_0)$. The game with $X = x$ has a unique BNE, which is a monotone pure strategy BNE.*

Proof. See Appendix A.2. □

By the assumption on the distribution of U, the conditions to define $\mathcal{U}(\theta_0)$ can expressed explicitly, i.e., $x \in \mathcal{U}(\theta_0)$ if and only if

$$1 - \frac{(1+\rho_0)\alpha_j}{\sqrt{2\pi(1-\rho_0^2)}} \exp\left\{-\frac{t^2}{2(1-\rho_0^2)}\right\} \geq 0, \tag{5}$$

holds for all $\mathscr{V}_{-j}^{-}(x) - \rho_0\mathscr{V}_j^{+}(x) \leq t \leq \mathscr{V}_{-j}^{+}(x) - \rho_0\mathscr{V}_j^{-}(x)$ and $j = 1,2$. Thus, if the model parameters satisfy $\frac{(1+\rho_0)\alpha_j}{\sqrt{2\pi(1-\rho_0^2)}} \leq 1$ for $j = 1,2$, then equation (5) always holds, i.e., $\mathcal{U}(\theta_0) = \mathscr{S}_X$. Moreover, one can also show that a sufficient condition for $x \in \mathcal{U}(\theta_0)$ is: either the inequalities

$$\mathscr{V}_1^{-}(x) \geq \Delta(\theta_0), \quad \text{and } \mathscr{V}_2^{+}(x) \leq -\Delta(\theta_0)$$

or

$$\mathscr{V}_1^{-}(x) \leq -\Delta(\theta_0), \quad \text{and } \mathscr{V}_2^{+}(x) \geq \Delta(\theta_0)$$

[8] It should be noted that the multiple equilibria issue exists even if U_1 and U_2 are assumed to be independent. Here is a simple example: $\alpha_1 = \alpha_2 = 4$, $x_1'\beta_1 = x_2'\beta_2 = 2$, and $\rho_0 = 0$. In this setup, three monotone strategy BNEs can be found.

holds, where $\Delta(\theta_0) = \sqrt{2\frac{1-\rho_0}{1+\rho_0}\ln\left\{\frac{(1+\rho_0)\alpha_{\max}}{\sqrt{2\pi(1-\rho_0^2)}},1\right\}}$ and $\alpha_{\max} = \max\{\alpha_1,\alpha_2\}$. To see this, one can show that for $j=1,2$ there is either $\mathscr{V}_{-j}^{-}(x)-\rho_0\mathscr{V}_j^{+}(x) \geq \sqrt{2(1-\rho_0^2)\ln\left\{\frac{(1+\rho_0)\alpha_j}{\sqrt{2\pi(1-\rho_0^2)}},1\right\}}$ or $\mathscr{V}_{-j}^{+}(x)-\rho_0\mathscr{V}_j^{-}(x) \leq -\sqrt{2(1-\rho_0^2)\ln\left\{\frac{(1+\rho_0)\alpha_j}{\sqrt{2\pi(1-\rho_0^2)}},1\right\}}$; thus equation (5) holds.

4. Identification

In the following analysis, I discuss the identification of the structural parameter θ_0 in the sense of Hurwicz (1950); Koopmans and Reiersol (1950), i.e. whether there is a unique structural parameter $\theta_0 \in \Theta$ to rationalize the conditional distribution of Y given X. Let $\Theta = \mathbb{B}\times[0,\bar{\alpha}]^2\times[0,\bar{\rho}]$ be a compact space where $B \subseteq \mathbb{R}^{k_1+k_2}$. Suppose that the subset $\mathcal{U}(\theta_0)$ is known and has a strictly positive probability measure.[9] Then, θ_0 is identified. To see this, let one first condition on $X = x \in \mathcal{U}(\theta_0)$. Then, there is

$$\mathbb{E}(Y_j|X=x) = \Phi(u_j^*(x)),$$

where Φ is the c.d.f. of the standard normal distribution. Therefore $u_j^*(x) = \Phi^{-1}\left(\mathbb{E}(Y_j|X=x)\right)$. Further, arbitrarily pick $(p_1,p_2) \in \mathscr{S}_{\mathbb{E}(Y_1|X),\mathbb{E}(Y_2|X)|X\in\mathcal{U}(\theta_0)}$. It follows that

$$\mathbb{E}\left[Y_1Y_2|\mathbb{E}(Y_1|X)=p_1,\mathbb{E}(Y_2|X)=p_2,X\in\mathcal{U}(\theta_0)\right] = \mathbb{P}\left[U_1\leq\Phi^{-1}(p_1);U_2\leq\Phi^{-1}(p_2)\right],$$

from which ρ_0 is identified. This is because

$$\begin{aligned}\frac{\partial\mathbb{E}\left[Y_1Y_2|\mathbb{E}(Y_1|X)=p_1,\mathbb{E}(Y_2|X)=p_2,X\in\mathcal{U}(\theta_0)\right]}{\partial p_1} &= \frac{\partial\mathbb{P}\left[\Phi(U_1)\leq p_1;\Phi(U_2)\leq p_2\right]}{\partial p_1}\\ &= \mathbb{P}\left[\Phi(U_2)\leq p_2|\Phi(U_1)=p_1\right] = \Phi\left(\frac{\Phi^{-1}(p_2)-\rho_0\Phi^{-1}(p_1)}{\sqrt{1-\rho_0^2}}\right), \quad (6)\end{aligned}$$

in which the second equality follows Darsow, Nguyen, and Olsen (1992). Therefore,

$$\frac{\Phi^{-1}(p_2)-\rho_0\Phi^{-1}(p_1)}{\sqrt{1-\rho_0^2}} = \Phi^{-1}\left(\frac{\partial\mathbb{E}\left[Y_1Y_2|\mathbb{E}(Y_1|X)=p_1,\mathbb{E}(Y_2|X)=p_2,X\in\mathcal{U}(\theta_0)\right]}{\partial p_1}\right). \quad (7)$$

Note that the RHS of equation (7) is known from the conditional distribution of Y given X.[10]

[9] For example, when $\frac{1+\bar{\rho}}{1-\bar{\rho}}\times\frac{\bar{\alpha}^2}{2\pi}\leq 1$, $\mathcal{U}(\theta_0) = \mathbb{R}^{k_1+k_2}$.

[10] It should be noted that the differentiability of $\mathbb{E}\left[Y_1Y_2|\mathbb{E}(Y_1|X)=p_1,\mathbb{E}(Y_2|X)=p_2,X\in\mathcal{U}(\theta_0)\right]$ involves a full rank condition on the support of $\mathscr{S}_{\mathbb{E}(Y_1|X),\mathbb{E}(Y_2|X)|X\in\mathcal{U}(\theta_0)}$, which is testable.

It is straightforward that the term $\rho_0/\sqrt{1-\rho_0^2}$ is identified from equation (7) by taking further derivative with respect to p_1 on both sides of the equation. Since ρ_0 and $\rho_0/\sqrt{1-\rho_0^2}$ are one to one map, then ρ_0 is also identified. It should also be noted that the identification of ρ_0 allows a nonparametric setup for the payoff functions as long as $\mathcal{U}(\theta_0)$ is known.

Moreover, given the knowledge of ρ_0 and $u_j^*(X)$, (α_j, β_j) can be identified by equation (4), i.e.

$$X_j'\beta_j - \alpha_j\Phi\left(\frac{u_{-j}^*(X) - \rho_0 u_j^*(X)}{\sqrt{1-\rho_0^2}}\right) - u_j^*(X) = 0$$

under an additional rank condition, i.e. the matrix $\mathbb{E}(Z_j'Z_j)$ has a full rank for which $Z_j = \left[X_j', \Phi\left(\frac{u_{-j}^*(X) - \rho_0 u_j^*(X)}{\sqrt{1-\rho_0^2}}\right)\right]'$. It should be noted that the full rank condition is a testable restriction given the identification of ρ_0 and $u_j^*(\cdot)$.

An alternative identification strategy for θ_0 is to use information criteria, which is less constructive: conditioning on $X \in \mathcal{U}(\theta_0)$, suppose that the information matrix is invertible;[11] then

$$\theta_0 = \text{argmax}_{\theta\in\Theta}\ \mathbb{E}\left[\mathbf{1}\left\{X \in \mathcal{U}(\theta_0)\right\} \times \ln \mathbb{P}_\theta(Y|X)\right],$$

where

$$\mathbb{P}_\theta(Y = y|X = x) = \begin{cases} \mathbb{P}_\theta(U_1 \le u_1^*(x,\theta), U_2 \le u_2^*(x,\theta)) & \text{if } y = (1,1), \\ \mathbb{P}_\theta(U_1 > u_1^*(x,\theta), U_2 \le u_2^*(x,\theta)) & \text{if } y = (0,1), \\ \mathbb{P}_\theta(U_1 \le u_1^*(x,\theta), U_2 > u_2^*(x,\theta)) & \text{if } y = (1,0), \\ \mathbb{P}_\theta(U_1 > u_1^*(x,\theta), U_2 > u_2^*(x,\theta)) & \text{if } y = (0,0). \end{cases}$$

For a given θ, $u^*(x,\theta) = (u_1^*(x,\theta), u_2^*(x,\theta))$ obtains by the following simultaneous equations: for $j = 1, 2$,[12]

$$x_j'b_j - a_j\Phi\left(\frac{u_{-j}^*(X) - \rho u_j^*(X)}{\sqrt{1-\rho^2}}\right) - u_j^* = 0.$$

4.1. **Unknown $\mathcal{U}(\theta_0)$.** The difficulty arises when $\mathcal{U}(\theta_0)$ is unknown, which is because of the dependence of $\mathcal{U}(\theta_0)$ on the underlying parameter θ_0. As a consequence, the identification of θ_0 hinges on a fixed point argument: Let $\psi : \mathscr{B} \to \Theta$, where $\mathscr{B} \equiv \{\mathcal{U}(\theta) : \theta \in \Theta\}$, be the mapping which corresponds to the identification approach discussed above. Thus, $\theta_0 = \psi(\mathcal{U}(\theta_0))$, from which θ_0 is identified under conditions ensuring that it is the unique fixed point of the equation.

[11]The invertibility of the information matrix could be satisfied if (i) $\mathbb{P}\left[u^*(X,\theta) \neq u^*(X,\theta_0) | X \in \mathcal{U}(\theta_0)\right] > 0$ for all $\theta \neq \theta_0$; and (ii) $\mathbb{P}\left[X \in \mathcal{U}(\theta_0)\right] > 0$.

[12]When $\theta \neq \theta_0$, there could be multiple solutions to equations (4) even for $x \in \mathcal{U}(\theta_0)$. In this case, I can choose $u_j^*(x,\theta) = x_j'b_j$ as a convention.

In this paper, however, I propose an alternative identification strategy which is constructive when $\mathcal{U}(\theta_0)$ is unknown. The procedure takes two steps: First, I identify a subset $\Theta_I \subseteq \Theta$, which contains θ_0 and is small enough such that $\mathcal{C}(\Theta_I) \equiv \bigcap_{\theta \in \Theta_I} \mathcal{U}(\theta)$ has a strictly positive probability measure. Because $\mathcal{C}(\Theta_I)$ is a subset of $\mathcal{U}(\theta_0)$, thus θ_0 is identified by replacing $\mathcal{U}(\theta_0)$ with $\mathcal{C}(\Theta_I)$.[13] The above discussion is summarized in the next theorem.

Theorem 3. *Suppose $\theta_0 \in \Theta_I \subseteq \Theta$ and $\mathbb{P}\left[X \in \mathcal{C}(\Theta_I)\right] > 0$. Moreover, if (i) conditional on $X \in \mathcal{C}(\Theta_I)$, $\mathbb{E}(Y|X)$ has a non–degenerated continuous support in $[0,1]^2$; and (ii) $\mathbb{E}\left[\mathbf{1}\left\{X \in \mathcal{C}(\Theta_I)\right\} Z_j' Z_j\right]$ has a full rank for $j = 1, 2$, then θ_0 is identified.*

By a similar argument to the identification analysis using $\mathcal{U}(\theta_0)$ in the beginning of this section, the proof of Theorem 3 is straightforward and therefore omitted.

4.2. **Finding Θ_I and "level–∞ rationality".** It is crucial to construct the subset Θ_I by which θ_0 is partially identified. Aradillas-Lopez and Tamer (2008) proposed a novel approach to identify a set containing θ_0 by using the restrictions called "level–k rationality" ($k \to \infty$), which are implied by the BNE solution concept.

Under the current setup, the constraints of "level–1 rationality" can be derived as follows: consider the equilibrium response for player $j = 1, 2$,

$$Y_j = \mathbf{1}\left[X_j'\beta_j - \alpha_j \mathbb{E}(Y_{-j}|X, U_j) - U_j \geq 0\right]. \tag{8}$$

Because the belief term $0 \leq \mathbb{E}(Y_{-j}|X, U_j) \leq 1$, thus, no matter how his rival behaves, player j's equilibrium response can always be bounded in the following way:

$$\mathbf{1}\left[\mathscr{V}_{j,1}^-(X) - U_j \geq 0\right] \leq Y_j \leq \mathbf{1}\left[\mathscr{V}_{j,1}^+(X) - U_j \geq 0\right]. \tag{9}$$

Therefore, $Y_j = 1$ if $U_j < \mathscr{V}_{j,1}^-(X)$, and $Y_j = 0$ if $U_j > \mathscr{V}_{j,1}^+(X)$, which are the restrictions derived from "level–1 rationality". Note that "level–1 rationality" is silent about the rational response of Y_j when $\mathscr{V}_{j,1}^-(X) \leq U_j \leq \mathscr{V}_{j,1}^+(X)$.

The restrictions of the "level–2 rationality" can be derived similarly: from equation (9) we have

$$\mathbb{P}\left(\mathscr{V}_{-j,1}^-(X) - U_{-j} \geq 0 | X, U_j\right) \leq \mathbb{E}(Y_{-j}|X, U_j) \leq \mathbb{P}\left(\mathscr{V}_{-j,1}^+(X) - U_{-j} \geq 0 | X, U_j\right).$$

[13] Aradillas-Lopez (2010) also suggests to focus a subset of observables where BNE is likely to be unique.

Thus for $\mathscr{V}_{j,1}^{-}(X) \leq U_j \leq \mathscr{V}_{j,1}^{+}(X)$, it follows that

$$\mathbb{P}\left(\mathscr{V}_{-j,1}^{-}(X) - U_{-j} \geq 0 | X, U_j = \mathscr{V}_{j,1}^{+}(X)\right) \leq \mathbb{E}(Y_{-j}|X, U_j) \leq \mathbb{P}\left(\mathscr{V}_{-j,1}^{+}(X) - U_{-j} \geq 0 | X, U_j = \mathscr{V}_{j,1}^{-}(X)\right).$$

Then, by equation (8) and the fact that $\alpha_j \geq 0$, it follows that

$$\mathbf{1}\left[\mathscr{V}_{j,2}^{-}(X) - U_j \geq 0\right] \leq Y_j \leq \mathbf{1}\left[\mathscr{V}_{j,2}^{+}(X) - U_j \geq 0\right]. \tag{10}$$

Therefore, $Y_j = 1$ if $U_j < \mathscr{V}_{j,2}^{-}(X)$, and $Y_j = 0$ if $U_j > \mathscr{V}_{j,2}^{+}(X)$. Note that $\mathscr{V}_{j,1}^{-}(X) \leq \mathscr{V}_{j,2}^{-}(X) \leq \mathscr{V}_{j,2}^{+}(X) \leq \mathscr{V}_{j,1}^{+}(X)$, which means that higher level of rationality provides additional restrictions.

Moreover, applying "level–k rationality" for $k \in \mathbb{N} \cup \{\infty\}$ recursively, there is

$$\mathbf{1}\left[\mathscr{V}_{j,k}^{-}(X) - U_j \geq 0\right] \leq Y_j \leq \mathbf{1}\left[\mathscr{V}_{j,k}^{+}(X) - U_j \geq 0\right]. \tag{11}$$

Now I am ready to define Θ_I: let $\theta = (a_1, a_2, b_1', b_2', \rho)'$ be a generic parameter value in Θ and

$$\Theta_I = \{\theta \in \Theta : \Phi\big(\mathscr{V}_j^{-}(x;\theta)\big) \leq \mathbb{E}(Y_j|X = x) \leq \Phi\big(\mathscr{V}_j^{+}(x;\theta)\big), \ \ \forall x \in \mathscr{S}_X, \ j = 1, 2\}.$$

By definition, $\theta_0 \in \Theta_I$. Replacing $\mathscr{V}_j^{-}(x;\theta)$ and $\mathscr{V}_j^{+}(x;\theta)$ respectively with $\mathscr{V}_{j,k}^{-}(x;\theta)$ and $\mathscr{V}_{j,k}^{+}(x;\theta)$ in the definition of Θ_I, one can define $\Theta_{I,k}$ in a similar manner.

4.3. **Rank condition and the support of covariates.** Essentially, $\mathbb{P}[X \in \mathcal{C}(\Theta_I)] > 0$ is a rank condition, which requires the support of X to be rich enough. To characterize the subset $\mathcal{C}(\Theta_I)$, however, the difficulties arises as follows: The distribution of Y given X might not be well defined due to the issue multiple equilibria (see the discussion of "incompleteness" in Tamer, 2003). In another word, the subset Θ_I can not be characterized without the knowledge of the equilibrium selection mechanism for the multiple equilibria.

To answer the important question that how large the set $\mathcal{C}(\Theta_I)$ is, I derive a subset of it, which can be characterized simply. Let Θ be compact, and $\bar{\alpha}$ and $\bar{\rho}$ be the upper bounds for α_j and ρ, respectively. Let further $\Delta^*(\rho) = \sqrt{2\frac{1-\rho}{1+\rho}\ln\left\{\frac{(1+\rho)\bar{\alpha}}{\sqrt{2\pi(1-\rho^2)}}, 1\right\}}$ and $\gamma^*(\rho) = -\Delta^*(\rho) + \bar{\alpha} \times \Phi\left(\sqrt{\frac{1+\rho}{1-\rho}}\Delta^*(\rho)\right)$. Moreover, I define $\gamma_0^* = \sup_{\rho \in [0,\bar{\rho}]} \gamma^*(\rho)$ and

$$\Pi = \left\{x \in \mathscr{S}_X : \mathbb{E}(Y_1|X = x) \geq \Phi(\gamma_0^*); \ \mathbb{E}(Y_2|X = x) \leq \Phi(-\gamma_0^*)\right\} \bigcup \left\{x \in \mathscr{S}_X : \mathbb{E}(Y_1|X = x) \leq \Phi(-\gamma_0^*); \ \mathbb{E}(Y_2|X = x) \geq \Phi(\gamma_0^*)\right\}.$$

Theorem 4. *By definition,* $\Pi \subseteq \mathcal{C}(\Theta_I)$.

Proof. See Appendix A.3. □

By Theorem 4, the rank condition for $\mathcal{C}(\Theta_I)$ will be satisfied if one has $\mathbb{P}(X \in \Pi) > 0$. Because $x_j'\beta_j - \alpha_j \leq \mathscr{V}_j^-(x) \leq \mathscr{V}_j^+(x) \leq x_j'\beta_j$, it could be shown using equation (11) that if $x_1'\beta_1 - \alpha_1 \geq \gamma_0^*; x_2'\beta_2 \leq -\gamma_0^*$ (or $x_1'\beta_1 \leq -\gamma_0^*; x_2'\beta_2 - \alpha_2 \geq \gamma_0^*$), then $\mathbb{E}(Y_1|X = x) \geq \Phi(\gamma_0^*)$ and $\mathbb{E}(Y_2|X = x) \leq \Phi(-\gamma_0^*)$, which provides $x \in \Pi$. This means that a large support of $(X_1'\beta_1, X_2'\beta_2)$ is sufficient for $\mathbb{P}(X \in \Pi) > 0$. Figure 1 provides a numerical example in which Π is described by the shadow areas.

It should also be emphasized on that the subset Π depends on the value of $\overline{\alpha}$ through γ_0^*. For the compactness of Θ, $\bar{\alpha}$ need to be *ad hoc* chosen reasonably large such that $\alpha_j \in [0, \bar{\alpha}]$. Larger $\bar{\alpha}$, more stringent support conditions are required for the covariates X to achieve identification. If $\overline{\alpha}$ is set to be arbitrarily large, however, one has to assume a full support of $(X_1'\beta_1, X_2'\beta_2)$ on $\mathbb{R}^2$ to ensure the rank condition, by which the proposed identification strategy becomes an identification–at–infinity argument, see, e.g., Tamer (2003); Bajari, Hong, Krainer, and Nekipelov (2010) for discrete games of complete information.

Now I give a numerical discussion of choosing γ_0^* for some given $\overline{\alpha}$. By definition,

$$\gamma_0^* = \sup_{\rho \in [0, \overline{\rho}]} -\sqrt{\frac{1-\rho}{1+\rho} \ln \max\left\{\frac{1+\rho}{1-\rho} \times \frac{\overline{\alpha}^2}{2\pi}, 1\right\}} + \overline{\alpha} \times \Phi\left(\sqrt{\ln \max\left\{\frac{1+\rho}{1-\rho} \times \frac{\overline{\alpha}^2}{2\pi}, 1\right\}}\right).$$

Since it can be shown that the function $g(t) \equiv -\sqrt{t \ln \max\left\{\frac{\overline{\alpha}^2}{2\pi t}, 1\right\}} + \overline{\alpha} \times \Phi\left(\sqrt{\ln \max\left\{\frac{\overline{\alpha}^2}{2\pi t}, 1\right\}}\right)$ is (weakly) monotone decreasing in $t \in [0, 1]$. Therefore, given $\rho \in [0, \bar{\rho}]$ in the parameter space Θ, it follows that $\gamma_0^* = \gamma^*(\overline{\rho})$. Further, one can show that $\overline{\alpha}/2 \leq \gamma_0^* \leq \overline{\alpha}$.[14] It is also understood that $\frac{1+\overline{\rho}}{1-\overline{\rho}} \times \frac{\overline{\alpha}^2}{2\pi} > 1$, otherwise $\mathcal{U}(\theta_0)$ is known as the full support.

Table 1 provides γ_0^* for different combinations of $\overline{\alpha}$ and $\overline{\rho}$. It should be noted that the standard deviation of U_i has been normalize to be 1. Hence, the value of $\overline{\alpha}$ imposes an upper bound for the strategic component at the scale of the error's standard deviation.

Figure 1 illustrates the size of Π in the space of covariates in a simple setup for $\overline{\alpha} = 1.5$ and 2, respectively, and $\overline{\rho} = 0.6$. The payoff functions for both players are identical: $X_j\beta - \alpha Y_{-j} - U_j$ in which $\beta = 1$ and $\alpha = 1.5$ are fixed and $\mathscr{S}_X \subseteq \mathbb{R}^2$; moreover, the correlations coefficient parameter ρ_0 is 0.3 and 0.5, respectively. The subsets Π are represented by the shadow areas in figure 1.

[14] Note that $\gamma^*(\rho)$ is approaching to $\overline{\alpha}$ as $\rho \to 1$.

TABLE 1. γ_0^* for different values of $\overline{\alpha}$

	$\overline{\alpha}=1$	1.5	2	2.5	3	4	$\cdots$
$\overline{\rho}=0$	—	—	—	—	1.5772	2.3659	$\cdots$
0.4	—	—	1.0589	1.4500	1.8729	2.7623	$\cdots$
0.5	—	0.7537	1.1144	1.5266	1.9620	2.8690	$\cdots$
0.6	—	0.7886	1.1830	1.6125	2.0597	2.9830	$\cdots$
0.7	—	0.8464	1.2668	1.7120	2.1703	3.1091	$\cdots$
0.8	0.5257	0.9299	1.3732	1.8331	2.3023	3.2565	$\cdots$
0.9	0.6123	1.0577	1.5234	1.9986	2.4793	3.4504	$\cdots$
$\simeq 1$	1.0000	1.5000	2.0000	2.5000	3.0000	4.0000	$\cdots$

"— " refers to the degenerated case: $\frac{1+\overline{\rho}}{1-\overline{\rho}} \times \frac{\overline{\alpha}^2}{2\pi} \leq 1$

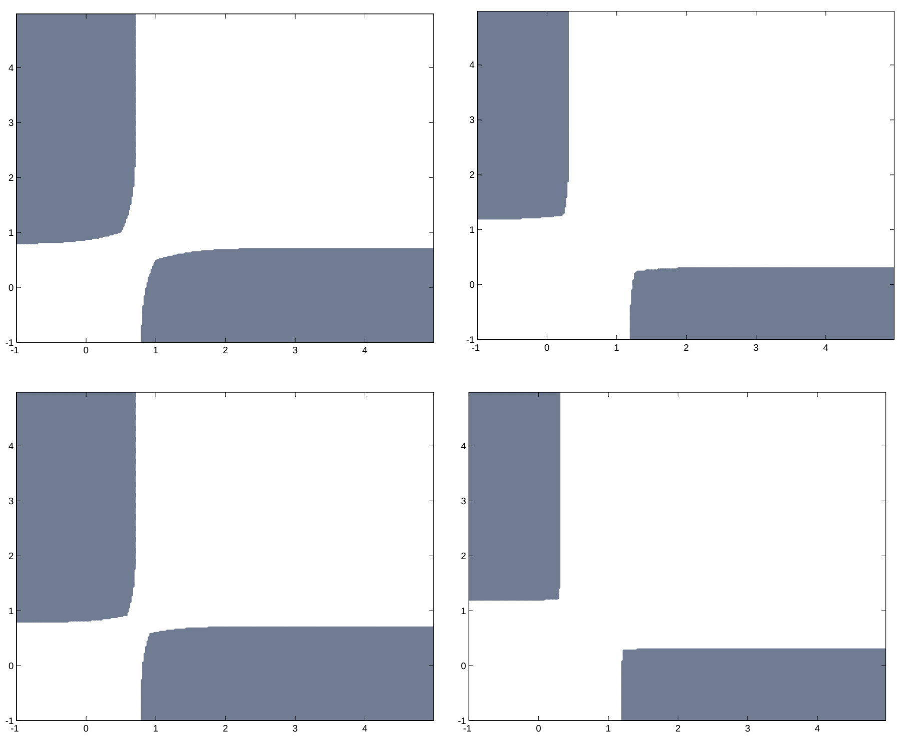

FIGURE 1. Examples of Π with $\overline{\alpha}=1.5$ (left) and 2 (right); $\rho_0 = 0.3$ (upper) and 0.5 (down)

5. OUTLINE OF ESTIMATION STRATEGY

The estimation approach is naturally suggested by the identification strategy in Section 4. 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 $Y_i = (Y_{1i}, Y_{2i})'$.

The estimation takes two steps. I now proceeds with introducing my first–step estimator:

$$\widetilde{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \widetilde{\Pi})\right] \log \mathbb{P}_\theta(Y_i|X_i), \tag{12}$$

where $\mathbb{P}_\theta$ is the conditional probability of Y given X defined in section 4, and $\widetilde{\Pi}$ is a consistent estimator of Π such that $\mathbf{1}(X \in \widetilde{\Pi}) - \mathbf{1}(X \in \Pi) \xrightarrow{p} 0$. Note that a uniformly consistent estimator of $\mathbb{E}(Y_j|X)$ is sufficient to define $\mathbf{1}(X \in \widetilde{\Pi})$, i.e. if X is continuously distributed,

$$\begin{aligned}\mathbf{1}(X_i \in \widetilde{\Pi}) \equiv \mathbf{1}&\left\{\sum_{\ell\neq i}[Y_{1\ell} - \Phi(\gamma_0^*)]\, K\left(\frac{X_\ell - X_i}{h}\right) \geq 0\right\} \times \mathbf{1}\left\{\sum_{\ell\neq i}[Y_{2\ell} - \Phi(-\gamma_0^*)]\, K\left(\frac{X_\ell - X_i}{h}\right) \leq 0\right\} \\ &+ \mathbf{1}\left\{\sum_{\ell\neq i}[Y_{1\ell} - \Phi(-\gamma_0^*)]\, K\left(\frac{X_\ell - X_i}{h}\right) \leq 0\right\} \times \mathbf{1}\left\{\sum_{\ell\neq i}[Y_{2\ell} - \Phi(\gamma_0^*)]\, K\left(\frac{X_\ell - X_i}{h}\right) \geq 0\right\},\end{aligned}$$

where K and h are the kernel function and the smoothing bandwidth, respectively. Under additional conditions, which are standard in the literature, it could be shown that $\mathbf{1}(X \in \widetilde{\Pi}) - \mathbf{1}(X \in \Pi) \xrightarrow{p} 0$. If X is discrete, $\mathbf{1}(X \in \widetilde{\Pi})$ can also be defined similarly by plugging into a nonparametric estimator of $\mathbb{E}(Y_j|X)$, but one needs to rule out the case that the distribution of X has a mass point on the boundary of Π.

Assumption A. *Let* $\mathbf{1}(X \in \widetilde{\Pi}) - \mathbf{1}(X \in \Pi) \xrightarrow{p} 0$.

Assumption B. *Let* $\mathscr{S}_X$ *be compact and* $\mathbb{P}(X \in \Pi) > 0$.

Assumption C. *Let* Θ *be compact and* $\mathbb{E}\left\{\sup_{\theta\in\Theta}|\ln \mathbb{P}_\theta\,(Y|X)|^{1+\epsilon}\right\} < \infty$ *for some* $\epsilon > 0$.

Assumption A is a high level condition but has been well studied in the nonparametric estimation literature and only for the brevity of presentation. The first half of assumption B is standard in the literature and the second half constitutes a rank condition as discussed in Section 4.3. Assumption C is slightly stronger than the condition $\mathbb{E}\left\{\sup_{\theta\in\Theta}|\ln \mathbb{P}_\theta\,(Y|X)|\right\} < \infty$, which is a standard assumption in MLE literature, e.g. Newey and McFadden (1986).

Theorem 5. *Suppose assumptions A through C hold. Then* $\widetilde{\theta} \xrightarrow{p} \theta_0$.

Proof. See Appendix A.4. □

The consistent estimator $\widetilde{\theta}$ allows me to exploit information further in a different subset of the data, i.e., $\mathcal{V}(\theta, \delta)$, which is a subset of $\mathcal{U}(\widetilde{\theta})$ and satisfies regularity conditions. For fixed $\delta > 0$, let

$\gamma(\theta) = -\Delta(\theta) + a_{\max} \times \Phi\left(\sqrt{\frac{1+\rho}{1-\rho}}\Delta(\theta)\right)$[15] and

$$\mathcal{V}(\theta,\delta) = \left\{x \in \mathscr{S}_X : x_1'b_1 \geq \gamma(\theta) + \delta(1+\|x\|); x_2'b_2 - a_2 \leq -\gamma(\theta) - \delta(1+\|x\|)\right\}$$
$$\bigcup \left\{x \in \mathscr{S}_X : x_1'b_1 - a_1 \leq -\gamma(\theta) - \delta(1+\|x\|); x_2'b_2 \geq \gamma(\theta) + \delta(1+\|x\|)\right\}.$$

It can be shown that $\mathcal{V}(\theta,\delta) \subseteq \mathcal{U}(\theta)$ for all $\theta \in \Theta$ and for any fixed $\delta \in \mathbb{R}^+$, $\{\mathcal{V}(\theta,\delta) : \theta \in \Theta\}$ is a VC class of sets.

Lemma 1. *For fixed $\delta > 0$, $\mathcal{V}(\theta,\delta) \subseteq \mathcal{U}(\theta)$.*

Proof. See Appendix A.5. □

Lemma 2. *Fix $\delta \in \mathbb{R}^+$. The collection $\{\mathcal{V}(\theta,\delta) : \theta \in \Theta\}$ is a VC class of sets.*

Proof. See Appendix A.6. □

By definition, there exists $\epsilon_\delta > 0$ such that for any $\|\theta - \theta_0\| \leq \epsilon_\delta$, there is $\mathcal{V}(\theta,\delta) \subseteq \mathcal{V}(\theta_0,0) \subseteq \mathcal{U}(\theta_0)$. Thus, by consistency of $\widetilde{\theta}$, $\mathbb{P}\left[\mathcal{V}(\widetilde{\theta},\delta) \subseteq \mathcal{U}(\theta_0)\right] \to 1$ as n goes to infinity. Thus, my second–step estimator is defined as

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right] \log \mathbb{P}_\theta(Y_i|X_i). \tag{13}$$

Assumption D. *Let θ_0 be an interior point of Θ.*

Assumption D is standard in the literature for MLE, see, e.g. Newey and McFadden (1986).

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

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

The first half of Assumption E is also used in Manski (1985). Assumption E guarantees $\mathbb{E}\left|1\left[X_i \in \mathcal{V}(\theta,0)\right] - 1\left[X_i \in \mathcal{V}(\theta_0,0)\right]\right| = O\left(\|\theta - \theta_0\|\right)$ for θ in a small neighborhood of θ_0.

Further, let $s(y,x;\theta)$ be the score function, i.e., $s(y,x,\theta) = \partial \log \mathbb{P}_\theta(y|x)/\partial\theta$.

Theorem 6. *Suppose assumptions A through E hold and $\widetilde{\theta} \xrightarrow{p} \theta_0$. Then $\hat{\theta} \xrightarrow{p} \theta_0$. Moreover,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_\delta^{-1}),$$

[15] Note that $\gamma(\theta) \to a_{\max}$ as $\rho \to 1$. Similar to the discussion of $\gamma^*(\rho)$ in Section 4.3, one can show that $\gamma(\theta)$ is an increasing function in ρ and $a_{\max}/2 \leq \gamma(\theta) \leq a_{\max}$.

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

Proof. See Appendix A.7 □

Similar to Chernozhukov and Hong (2002), one can repeat above second–step estimation procedure one or more times, using sample $\mathcal{V}(\hat{\theta}, \delta_n)$ in place of $\mathcal{V}(\tilde{\theta}, \delta)$, where δ_n is a deterministic sequence with $\delta_n \downarrow 0$ (slower than $n^{-1/2}$). The updated estimator will achieve greater efficiency.[16]

6. Monte Carlo Studies

In this section, I use numerical experiments to illustrate the performance of the proposed estimator and also that ignoring the correlation between the private information results in inconsistent estimates and possibly misleading inference. In particular, I investigate the performance of the pseudo–MLE when the players' types are misspecified to be independent.

If U_1 and U_2 are independent, a two–step MLE would be based on the following model,

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

in which $\mathbb{P}(Y_{-j} = 1|X)$ can be nonparametrically estimated in the first stage.[17] It is a misspecified model, since $\mathbb{P}(Y_{-j}|X) \neq \mathbb{P}(Y_{-j}|X, U_j)$ in general.

I evaluate the performance of my proposed estimator and the two–step pseudo–MLE in the following examples. I now specify the distribution of $X = (X_1, X_2) \in \mathbb{R}^2$ on a compact support as follows: let Z_1 and Z_2 be two independent random variables with uniform distribution on $[0, 2.5]$; let further $X_1 = Z_1 - Z_2$ and $X_2 = 2 - Z_2$. Note that the rank condition in Theorem 3 is satisfied under such a specification. Let $\beta_1 = \beta_2 = 1$, $\alpha_1 = \alpha_2 = 1.5$. Let further $\rho_0 = 0.3$ and $\rho_0 = 0.5$ in two experiments, respectively. Moreover, I choose sample size $n = 1000, 3000, 5000$.

To generate observables $\{(X_i, Y_i) : i = 1, \cdots, n\}$, I need to solve equilibrium for each observation. If there are multiple monotone pure strategy BNEs, or no pure monotone pure strategy BNE exists, then the following equation system of (u_1^*, u_2^*) would have multiple solutions:

$$X_1\beta_1 - \alpha_1 \Phi\left(\frac{u_2^* - \rho_0 u_1^*}{\sqrt{1-\rho_0^2}}\right) = u_1^*$$

$$X_2\beta_2 - \alpha_2 \Phi\left(\frac{u_1^* - \rho_0 u_2^*}{\sqrt{1-\rho_0^2}}\right) = u_2^*.$$

[16] Such a result and other details for the asymptotic properties are available upon request to the author.
[17] In my experiments, I actually compute the term $\mathbb{P}(Y_{-j} = 1|X)$ in the first stage, instead of estimating it.

Denote $K(x)$ to be the number of solutions and $(u^*_{1,k}(x,\theta_0), u^*_{1,k}(x,\theta_0))$ to be the k–th solution. Then I use $Y_j = \mathbf{1}(U_j \leq \bar{u}^*_j(x;\theta_0))$, where $\bar{u}^*_j(x,\theta_0) = \sum_{k=1}^{K(x)} u^*_{j,k}(x,\theta_0)/K(x)$, to mimic the data generated from multiple equilibria or non–monotone–pure–strategy BNE.

Table 2 shows the composition of one random sample with $\rho_0 = 0.5$ and $N = 1000$. In the

TABLE 2. Sample composition

Choice profile	Percentage
$Y = (1,1)$	6.2%
$Y = (1,0)$	25.5%
$Y = (0,1)$	50.8%
$Y = (0,0)$	17.5%

estimation, I choose a compact parameter space: $\Theta = [-5,5]^2 \times [0,2]^2 \times [0,0.6]$, for which $\bar{\alpha} = 2$ and $\bar{\rho} = 0.6$. From Table 1, $\gamma^*_0 = 1.1830$. For each design, I simulate $R = 100$ samples and calculate summary statistics from empirical distributions of estimators from these simulations, including mean (MEAN), median (MEDIAN), standard deviation (SD), and root of mean squared error (RMSE). Note that RMSE is estimated using the empirical distribution of estimators and the knowledge of the true parameters in the designs.

In the first stage estimator, $\mathbb{E}(Y_i|X)$ is estimated using kernel method in which I employ a standard second–order normal kernel with bandwidth $h = 1.06 \times N^{-1/6}$. Table 3 reports summary statistics for the first–stage estimator $\tilde{\beta}_1$ and $\tilde{\alpha}_1$ in the setting $\rho_0 = 0.5$.

TABLE 3. Finite sample behavior of $\tilde{\beta}_1$ and $\tilde{\alpha}_1$ in the setting $\rho_0 = 0.5$

	$\tilde{\beta}_1$					$\tilde{\alpha}_1$				
N	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
1000	1.00	1.0899	1.0382	0.4108	0.4205	1.50	1.5367	1.5027	0.1777	0.1814
3000	1.00	1.0023	1.0264	0.0669	0.0707	1.50	1.5053	1.5097	0.1160	0.1161
5000	1.00	1.0118	1.0056	0.0527	0.0540	1.50	1.5120	1.5077	0.0857	0.0865

Tables 4 and 5 make a comparison the performance of the proposed estimator and the misspecified MLE using summary statistics in the setting $\rho_0 = 0.5$. In a misspecified model, the correlation between private information is falsely assumed away. Instead of using the usual two–stage approach, in which the first step is a nonparametric estimation of the equilibrium belief $\mathbb{E}(Y_{-j}|X)$, I adopt the true value of the equilibrium belief $\mathbb{E}(Y_{-j}|X)$ for the second–stage Probit to avoid the finite sample bias from the nonparametric estimation, which will conceivably improve the performance of the final estimator of $(\alpha_j.\beta_j)$. The summary statistics suggest that misspecified MLE are inconsistent estimators for both α_1 and β_1. In contrast, the proposed estimator converges in both bias and variance as the sample size increases.

TABLE 4. Proposed estimator $\hat{\beta}_1$ and misspecified MLE for β_1 in the setting $\rho_0 = 0.5$

	Proposed estimator $\hat{\beta}_1$					Misspecified MLE				
N	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
1000	1.00	1.0019	0.9939	0.0837	0.0838	1.00	1.1276	1.1327	0.0792	0.1502
3000	1.00	1.0032	0.9976	0.0531	0.0532	1.00	1.1156	1.1119	0.0459	0.1243
5000	1.00	1.0041	1.0059	0.0361	0.0363	1.00	1.1164	1.1135	0.0349	0.1215

TABLE 5. Proposed estimator $\hat{\alpha}_1$ and misspecified MLE for α_1 in the setting $\rho_0 = 0.5$

N	Proposed estimator $\hat{\alpha}_1$					Misspecified MLE				
	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
1000	1.50	1.5444	1.5214	0.1559	0.1621	1.50	1.7509	1.7339	0.1151	0.2760
3000	1.50	1.5099	1.4998	0.0848	0.0854	1.50	1.7453	1.7514	0.0731	0.2559
5000	1.50	1.5074	1.4990	0.0613	0.0618	1.50	1.7472	1.7466	0.0544	0.2531

The proposed method also estimates the correlation coefficient parameter ρ_0. Table 6 reports summary statistics for $\hat{\rho}$ in both settings $\rho_0 = 0.3$ and $\rho_0 = 0.5$. There is also evidence for improvement of the estimator in terms of each of the summary statistics as the sample size increases.

TABLE 6. Finite sample behavior of $\hat{\rho}$

N	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
1000	0.30	0.3644	0.3500	0.1797	0.1916	0.50	0.5101	0.6000	0.1396	0.1399
3000	0.30	0.3122	0.3050	0.1002	0.1009	0.50	0.5063	0.5100	0.0906	0.0908
5000	0.30	0.2975	0.3000	0.0700	0.0701	0.50	0.5048	0.5000	0.0711	0.0713

7. CONCLUSION

It is worth emphasizing that the approach established in this paper hinges crucially on two features of the game model: first, there is no unobserved complete information structural term in the payoff functions.[18] When there are payoff variables (V) that are observed by both players but not by researchers, the proposed approach does not work. Additional model restrictions would be necessary such that one could obtain $\mathbb{E}(Y|X, V)$ from inverting $\mathbb{E}(Y|X)$.

Second, the proposed approach does not naturally extend to binary games with more than two players. This is due to the issue of multiple equilibria, which generally exist in a large subset

[18]A model featured with unobserved heterogeneity and independent private information also generates dependence among players' choices conditional on covariates, see Grieco (2010).

of the covariate space when the number of players $I \geq 3$. Moreover, the way I construct Π is to choose a small choice probability for one player and a large one for the other. When there are more than two players, it is impossible to choose covariates in such a way that each player's choice's probability belongs to different categories ("small" or "large" probability).

It should also be noted that the proposed method could be generalized to a discrete game with ordered multiple choices, but not multinomial games (for an illustration of multinomial game, see, e.g. Bajari, Hong, Krainer, and Nekipelov, 2010). When the error term is a multidimensional random vector rather than a scale, difficulties arise to characterize monotone pure strategy BNEs.

Finally, the joint normal distribution of private information is not essential to the proposed method, especially for the marginal normal distribution. See Appendix B for a detailed discussion.

References

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.

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

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. (1987): "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica: Journal of the Econometric Society*, pp. 1–18.

Aumann, R. J. (1964): "Mixed vs. behavior strategies in infinite extensive games," *Annals of Mathematics Studies*, 52, 627–630.

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.

Berry, S. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, pp. 889–917.

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.

Chernozhukov, V., and H. Hong (2002): "Three-step censored quantile regression and extramarital affairs," *Journal of the American Statistical Association*, 97(459), 872–882.

Ciliberto, F., and E. Tamer (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1928.

Darsow, W., B. Nguyen, and E. Olsen (1992): "Copulas and Markov processes," *Illinois Journal of Mathematics*, 36(4), 600–642.

de Castro, L. I. (2007): "Affiliation, equilibrium existence and the revenue ranking of auctions," Economics working papers, Universidad Carlos III, Departamento de Economa.

De Paula, A., and X. Tang (2012): "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, 80(1), 143–172.

Grieco, P. (2010): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," Discussion paper, Working papers.

Harsanyi, J. C. (1967–68): "Games with Incomplete Information Played by "Bayesian" Players," *Management Science*, 14, 159–182 and 320–334 and 486–502.

Hurwicz, L. (1950): "Generalization of the concept of identification," *Statistical Inference in Dynamic Economic Models (T. Koopmans, ed.). Cowles Commission, Monograph*, 10, 245–257.

Jia, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

Koopmans, T., and O. Reiersol (1950): "The identification of structural characteristics," *The Annals of Mathematical Statistics*, 21(2), 165–181.

Kosorok, M. (2008): *Introduction to empirical processes and semiparametric inference*. Springer Verlag.

Manski, C. F. (1985): "Semiparametric analysis of discrete response : Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27(3), 313 – 333.

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.

Pesendorfer, M., and P. Schmidt-Dengler (2003): "Identification and Estimation of Dynamic Games," Working Paper 9726, National Bureau of Economic Research.

Pollard, D. (1989): "Asymptotics via Empirical Processes," *Statistical Science*, 4(4), 341–354.

Reny, P. (2011): "On the Existence of Monotone Pure-Strategy Equilibria in Bayesian Games," *Econometrica*, 79(2), 499–553.

Sweeting, A. (2009): "The strategic timing of radio commercials: An empirical analysis using multiple equilibria," *RAND Journal of Economics*, 40(4), 710–742.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70(1), 147–165.

Tang, X. (2010): "Estimating simultaneous games with incomplete information under median restrictions," *Economics Letters*, 108(3), 273–276.

Wan, Y., and H. Xu (2009): "Semiparametric estimation of Binary decision of games of incomplete information with correlated private signals," Working papers, Department of Economics, The Pennsylvania State University.

Appendix A.

Let $\mathcal{B}$ be the collection of Boreal subsets in $\mathbb{R}$. For any $x \in \mathscr{S}_X$, let further

$$\mathcal{K}_j(x) = \{B \in \mathcal{B} : (-\infty, \mathscr{V}_j^-(x)] \subseteq B \text{ and } [\mathscr{V}_j^+(x), +\infty) \cap B = \emptyset\}.$$

Note that by *level–k rationality* with $k = \infty$, player j's equilibrium response must satisfy: $Y_j = 1$ for $U_j \leq \mathscr{V}_j^-(x)$ and $Y_j = 0$ for $U_j \geq \mathscr{V}_j^+(x)$ (for a detailed argument, see the discussion in Section 4.2.) Hence, I can restrict my attention to the strategy profiles which is defined as

$$s_1(x, u_1) = \mathbb{1}(u_1 \in \mathcal{A}_1), \quad s_2(x, u_2) = \mathbb{1}(u_2 \in \mathcal{A}_2)$$

where $(\mathcal{A}_1, \mathcal{A}_2) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$.

Lemma 3. *Suppose $X = x$. Suppose for any given $(\mathcal{A}_1, \mathcal{A}_2) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$ and for $j = 1, 2$, the function $u_j + \alpha_j \mathbb{P}(U_{-j} \in \mathcal{A}_{-j} | U_j = u_j)$ is an increasing function of $u_j \in \mathcal{I}_j(x_j; \theta_0)$. Then conditional on $X = x$, all pure strategy BNEs in this game are monotone strategy BNEs.*

Proof. Fix x. Suppose a strategy profile $\{s_1^*(x, \cdot), s_2^*(x, \cdot)\}$ is a pure strategy BNE. Then there exists $(\mathcal{A}_1^*, \mathcal{A}_2^*) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$, such that $s_j^*(x, u_j) = \mathbb{1}(u_j \in \mathcal{A}_j^*)$ and $\{s_1^*(x, \cdot), s_2^*(x, \cdot)\}$ satisfies the best response equations (1). Because

$$x_j'\beta_j - \alpha_j \mathbb{P}\left[s_{-j}^*(x, U_{-j}) = 1 | U_j = u_j\right] - u_j = x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \in \mathcal{A}_{-j}^* | U_j = u_j\right) - u_j,$$

which is a decreasing function of u_j. Then there exists a $u_j^*(x)$ such that equations (1) can be represented as $s_j^*(x, u_j) = \mathbb{1}(u_j \leq u_j^*(x))$, which implies that the equilibrium strategies have to be monotone functions. □

A.1. Proof of Theorem 1.

Proof. By Lemma 3, it suffices to show that for any $(\mathcal{A}_1, \mathcal{A}_2) \in \mathcal{K}_1(x) \times \mathcal{K}_2(x)$, $u_j + \alpha_j \mathbb{P}(U_{-j} \in \mathcal{A}_{-j} | U_j = u_j)$ is an increasing function of u_j in $\mathcal{I}_j(x_j; \theta_0)$. W.L.O.G. I take $j = 1$. Let ϕ be the p.d.f. of the standard normal distribution. Because

$$u_1 + \alpha_1 \mathbb{P}(U_2 \in \mathcal{A}_2 | U_1 = u_1) = u_1 + \frac{\alpha_1}{\sqrt{1 - \rho_0^2}} \int_{\mathcal{A}_2} \phi\left(\frac{t - \rho_0 u_1}{\sqrt{1 - \rho_0^2}}\right) dt,$$

which is differentiable in u_1, then it is equivalent to show that for all $u_1 \in \mathcal{I}_1(x_1;\theta_0)$

$$1 - \frac{\rho_0\alpha_1}{1-\rho_0^2}\int_{\mathcal{A}_2}\phi'\left(\frac{t-\rho_0 u_1}{\sqrt{1-\rho_0^2}}\right)dt \geq 0.$$

Since $\phi'(t) = -t\phi(t)$ for any $t \in \mathbb{R}$, then

$$1 - \frac{\rho_0\alpha_1}{1-\rho_0^2}\int_{\mathcal{A}_2}\phi'\left(\frac{t-\rho_0 u_1}{\sqrt{1-\rho_0^2}}\right)dt = 1 + \frac{\rho_0\alpha_1}{\sqrt{1-\rho_0^2}}\int_{\overline{\mathcal{A}}_2(u_1)} s\phi\left(s\right)ds$$

where $\overline{\mathcal{A}}_2(u_1)$ is a linear transformation of the set $\mathcal{A}_2$, i.e., $\overline{\mathcal{A}}_2(u_1) = \frac{\mathcal{A}_2-\rho_0 u_1}{\sqrt{1-\rho_0^2}} \equiv \left\{\frac{t-\rho_0 u_1}{\sqrt{1-\rho_0^2}} : t \in \mathcal{A}_2\right\}$.

Therefore, I need to show, for all $u_1 \in \mathcal{I}_1(x)$

$$1 + \frac{\rho_0\alpha_1}{\sqrt{2\pi(1-\rho_0^2)}}\int_{\overline{\mathcal{A}}_2(u_1)} s\exp\left(-s^2/2\right)ds \geq 0.$$

Note that the LHS is minimized by choosing $\mathcal{A}_2$ in $\mathcal{K}_2(x)$ such that $\overline{\mathcal{A}}_2(u_1)$ contains all possible negative elements, i.e. $\mathcal{A}_2^*(u_1) = (-\infty, \mathscr{V}_2^-(x)] \cup \{t \in [\mathscr{V}_2^-(x), \mathscr{V}_2^+(x)] : t - \rho_0 u_1 \leq 0\}$. It is straightforward to see that there exists $\overline{u}_2(u_1) \in \mathcal{I}_2(x;\theta_0)$ such that $\mathcal{A}_2^*(u_1) = (-\infty, \overline{u}_2(u_1)]$. Hence, it suffices to show for all $(u_1, \overline{u}_2) \in \mathcal{I}(x;\theta_0)$, there is

$$1 + \frac{\rho_0\alpha_1}{\sqrt{2\pi}\sqrt{1-\rho_0^2}}\int_{-\infty}^{\frac{\overline{u}_2-\rho_0 u_1}{\sqrt{1-\rho_0^2}}} s\exp\left(-s^2/2\right)ds \geq 0. \tag{14}$$

By the definition of $\mathcal{M}(\theta_0)$, equation (14) is satisfied. □

A.2. **Proof of Theorem 2.**

Proof. Prove by contradiction. Fix $x \in \mathcal{U}(\theta_0)$. Suppose $u^*(x) = (u_1^*(x), u_2^*(x))$ and $v^*(x) = (v_1^*(x), v_2^*(x))$ are the cutoff values that define two different monotone strategy BNEs. For notational brevity, here I suppress the dependence on x of u^* and v^*. By the *"level–k rationality"* argument, both u^* and v^* belong to $\mathcal{I}(x,\theta_0)$. Define $T(\cdot) : \mathcal{I}(x;\theta_0) \to \mathcal{I}(x;\theta_0)$ as follows

$$\begin{aligned} x_1'\beta_1 - \alpha_1\mathbb{P}\left[U_2 \leq u_2 | U_1 = T_1(u)\right] - T_1(u) = 0, \\ x_2'\beta_2 - \alpha_2\mathbb{P}\left[U_1 \leq u_1 | U_2 = T_2(u)\right] - T_2(u) = 0. \end{aligned} \tag{15}$$

Note that $T(\cdot)$ is well–defined, i.e., for any fixed $u \in \mathcal{I}(x;\theta_0)$, there exists a unique $T(u)$ satisfying equations (15), due to the monotonicity of $\alpha_j\mathbb{P}(U_{-j} \leq u_{-j}|U_j = u_j) + u_j$ in u_j on $\mathcal{I}(x;\theta_0)$. Hence, $T(u^*) = u^*, T(v^*) = v^*$.

Define a continuously differentiable function $\varphi(t)$ by

$$\varphi(t) = \frac{\langle T(u^*) - T(v^*), T[v^* + t(u^* - v^*)]\rangle}{\|T(u^*) - T(v^*)\|}$$

Note that $\varphi(1) - \varphi(0) = \|T(u^*) - T(v^*)\| = \|u^* - v^*\|$, and also $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt$. Moreover, $\forall t \in (0,1)$, we have

$$\varphi'(t) = \frac{\langle u^* - v^*, T'[v^* + t(u^* - v^*)](u^* - v^*)\rangle}{\|u^* - v^*\|} \\ \leq \frac{\|u^* - v^*\| \times \|T'[v^* + t(u^* - v^*)](u^* - v^*)\|}{\|u^* - v^*\|} < \|u^* - v^*\| \quad a.e.$$

The first inequality comes from the Cauchy Schwartz inequality and the last inequality is based on the the fact that $T'_{jj} = 0$ and the conditions for $x \in \mathcal{U}(\theta_0)$ implies that $|T'_{12}|, |T'_{21}| < 1$ for all $t \in (0,1)$. Hence $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt < \|u^* - v^*\|$, contradiction. □

A.3. **Proof of Theorem 4.**

Proof. W.L.O.G., let $X = x$ satisfy that $\mathbb{E}(Y_1|X = x) \geq \Phi(\gamma_0^*)$ and $\mathbb{E}(Y_2|X = x) \leq \Phi(-\gamma_0^*)$. It suffices to show that for any $\theta \in \Theta_I$, $x \in \mathcal{U}(\theta)$.

Fix $\theta \in \Theta_I$. W.L.O.G., let $\frac{1+\rho}{1-\rho} \times \frac{\bar{\alpha}^2}{2\pi} > 1$. By the definition of Θ_I,

$$\Phi\big(\mathscr{V}_1^+(x;\theta)\big) \geq \mathbb{E}(Y_1|X = x) \geq \Phi(\gamma_0^*), \quad \Phi\big(\mathscr{V}_2^-(x;\theta)\big) \leq \mathbb{E}(Y_2|X = x) \leq \Phi(-\gamma_0^*).$$

Since $\gamma_0^* \geq \gamma^*(\rho)$, it follows that

$$\mathscr{V}_1^+(x;\theta) \geq \gamma^*(\rho), \quad \mathscr{V}_2^-(x;\theta) \leq -\gamma^*(\rho). \tag{16}$$

Moreover, because

$$\mathscr{V}_1^-(x;\theta) = x_1'b_1 - a_1\Phi\left(\frac{\mathscr{V}_2^+(x;\theta) - \rho\mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}}\right), \quad \mathscr{V}_1^+(x;\theta) = x_1'b_1 - a_1\Phi\left(\frac{\mathscr{V}_2^-(x;\theta) - \rho\mathscr{V}_1^+(x;\theta)}{\sqrt{1-\rho^2}}\right),$$
$$\mathscr{V}_2^-(x;\theta) = x_2'b_2 - a_2\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\right), \quad \mathscr{V}_2^+(x;\theta) = x_2'b_2 - a_2\Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\right).$$

It follows that

$$\mathscr{V}_1^+(x;\theta) - \mathscr{V}_1^-(x;\theta) = a_1 \left[\Phi\left(\frac{\mathscr{V}_2^+(x;\theta) - \rho \mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}} \right) - \Phi\left(\frac{\mathscr{V}_2^-(x;\theta) - \rho \mathscr{V}_1^+(x;\theta)}{\sqrt{1-\rho^2}} \right) \right],$$

$$\mathscr{V}_2^+(x;\theta) - \mathscr{V}_2^-(x;\theta) = a_2 \left[\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho \mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}} \right) - \Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho \mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}} \right) \right].$$

Therefore, by equation (16),

$$\mathscr{V}_1^-(x;\theta) + a_1 \left[\Phi\left(\frac{\mathscr{V}_2^+(x;\theta) - \rho \mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}} \right) - \Phi\left(\frac{\mathscr{V}_2^-(x;\theta) - \rho \mathscr{V}_1^+(x;\theta)}{\sqrt{1-\rho^2}} \right) \right] \geq \gamma^*(\rho),$$

$$\mathscr{V}_2^+(x;\theta) - a_2 \left[\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho \mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}} \right) - \Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho \mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}} \right) \right] \leq -\gamma^*(\rho);$$

which implies that

$$\mathscr{V}_1^-(x;\theta) + \overline{\alpha} \times \Phi\left(\frac{\mathscr{V}_2^+(x;\theta) - \rho \mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}} \right) > \gamma^*(\rho), \tag{17}$$

$$\mathscr{V}_2^+(x;\theta) - \overline{\alpha} \times \left[1 - \Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho \mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}} \right) \right] < -\gamma^*(\rho). \tag{18}$$

Thus, there exists some $\epsilon > 0$ such that for $\gamma_\epsilon^*(\rho) = \gamma^*(\theta) + \epsilon$,

$$\mathscr{V}_1^-(x;\theta) + \overline{\alpha} \times \Phi\left(\frac{\mathscr{V}_2^+(x;\theta) - \rho \mathscr{V}_1^-(x;\theta)}{\sqrt{1-\rho^2}} \right) \geq \gamma_\epsilon^*(\rho),$$

$$\mathscr{V}_2^+(x;\theta) - \overline{\alpha} \times \left[1 - \Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho \mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}} \right) \right] \leq -\gamma_\epsilon^*(\rho).$$

Moreover, I use a recursive approach to obtain bounds for $\mathscr{V}_1^-(x;\theta)$ and $\mathscr{V}_2^+(x;\theta)$. Define $\ell_{1,1}^-(\theta) \equiv \gamma_\epsilon^*(\rho) - \overline{\alpha}$, $\ell_{2,1}^+(\theta) \equiv -\gamma_\epsilon^*(\rho) + \overline{\alpha}$ and $\ell_{1,k}^-(\theta) = \gamma_\epsilon^*(\rho) - \overline{\alpha} \times \Phi\left(\frac{\ell_{2,k-1}^+(\theta) - \rho \ell_{1,k-1}^-(\theta)}{\sqrt{1-\rho^2}} \right)$ and $\ell_{2,k}^+(\theta) = -\gamma_\epsilon^*(\rho) + \overline{\alpha} \times \left[1 - \Phi\left(\frac{\ell_{1,k-1}^-(\theta) - \rho \ell_{2,k-1}^+(\theta)}{\sqrt{1-\rho^2}} \right) \right]$. Note that $\{\ell_{1,k}^-(\theta)\}_{k\geq 1}$ is a decreasing sequence and $\{\ell_{2,k}^+(\theta)\}_{k\geq 1}$ is increasing. Define $\ell_1^-(\theta) = \lim_k \ell_{1,k}^-(\theta)$ and $\ell_2^+(\theta) = \lim_k \ell_{2,k}^+(\theta)$. By equation (17), $\mathscr{V}_1^-(x;\theta) \geq \ell_{1,1}^-(\theta)$ and $\mathscr{V}_2^+(x;\theta) \leq \ell_{2,1}^+(\theta)$, which further imply that $\mathscr{V}_1^-(x;\theta) \geq \ell_{1,2}^-(\theta)$ and $\mathscr{V}_2^+(x;\theta) \leq \ell_{2,2}^+(\theta)$, and so on and so forth. Thus $\mathscr{V}_1^-(x;\theta) \geq \ell_{1,k}^-(\theta)$ and $\mathscr{V}_2^+(x;\theta) \leq \ell_{2,k}^+(\theta)$ for all $k \in \mathbb{N}$. In the limit, there is $\mathscr{V}_1^-(x;\theta) \geq \ell_1^-(\theta)$ and $\mathscr{V}_2^+(x;\theta) \leq \ell_2^+(\theta)$.

Next I will solve bounds for $\ell_1^-(\theta)$ and $\ell_2^+(\theta)$. Note that $\ell_{1,1}^-(\theta) = -\ell_{2,1}^+(\theta)$, which implies $\ell_{1,2}^-(\theta) = -\ell_{2,2}^+(\theta)$, and so on and so forth. Thus $\ell_1^-(\theta) = -\ell_2^+(\theta)$. Therefore, $\ell_1^-(\theta)$ solves the following equation: $t + \overline{\alpha}\Phi\left(-\sqrt{\frac{1+\rho}{1-\rho}} \times t\right) = \gamma_\epsilon^*(\rho)$. It is the smallest solution if there are multiple of them. Because $\frac{1+\rho}{1-\rho} \times \frac{\bar{\alpha}^2}{2\pi} > 1$, the function $g(t) \equiv t + \bar{\alpha}\Phi\left(-\sqrt{\frac{1+\rho}{1-\rho}} \times t\right)$ is (locally) maximized and minimized at $t = \Delta^*(\rho)$ and $t = -\Delta^*(\rho)$, respectively, where $\Delta^*(\rho) = \sqrt{\frac{1-\rho}{1+\rho} \ln\left(\frac{1+\rho}{1-\rho} \times \frac{\bar{\alpha}^2}{2\pi}\right)}$. By the shape of $g(\cdot)$, the equation $t + \bar{\alpha}\Phi\left(-\sqrt{\frac{1+\rho}{1-\rho}} \times t\right) = \gamma^*_\epsilon(\rho)$ has a unique solution which is larger than $\Delta^*(\rho)$, i.e. $\ell_1^-(\theta) \geq \Delta^*(\rho)$.

Therefore, we obtain bounds for $\mathscr{V}_1^-(x;\theta)$ and $\mathscr{V}_2^+(x;\theta)$

$$\mathscr{V}_1^-(x;\theta) \geq \Delta^*(\rho) \geq \Delta(\theta), \qquad \mathscr{V}_2^+(x;\theta) \leq -\Delta^*(\rho) \leq -\Delta(\theta).$$

By Theorem 2, $x \in \mathcal{U}(\theta)$. Because θ is arbitrarily chosen, then $x \in \mathcal{C}(\Theta_I)$. □

A.4. **Proof of Theorem 5.**

Proof. Let $L_n(\theta) = \frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \Pi)\right] \log \mathbb{P}_\theta(Y_i|X_i)$ and $G_n(\theta) = \frac{1}{n}\sum_{i=1}^n 1\left[X_i \in \widetilde{\Pi})\right] \log \mathbb{P}_\theta(Y_i|X_i)$. By Newey and McFadden (1986) (Theorem 2.5), it suffices to show $L_n(\widetilde{\theta}) = \sup_{\theta\in\Theta} L_n(\theta) + o_p(1)$. By the definition of $\widetilde{\theta}$, it suffices to show

$$\sup_{\theta\in\Theta} |L_n(\theta) - G_n(\theta)| = o_p(1)$$

Note that

$$\sup_{\theta\in\Theta} |L_n(\theta) - G_n(\theta)| \leq \frac{1}{n}\sum_{i=1}^n \left|\mathbf{1}(X_i \in \Pi) - \mathbf{1}(X_i \in \widehat{\Pi})\right| \times \sup_{\theta\in\Theta} |\ln \mathbb{P}_\theta(Y_i|X_i)|$$

Then, it suffices to show

$$\mathbb{E}\left[\left|\mathbf{1}(X_i \in \Pi) - \mathbf{1}(X_i \in \widehat{\Pi})\right| \times \sup_{\theta\in\Theta} |\ln \mathbb{P}_\theta(Y_i|X_i)|\right] \to 0. \tag{19}$$

Moreover, by assumptions A and C and Holder's Inequality, condition (19) holds. □

A.5. **Proof of Lemma 1.**

Proof. Fix θ and δ. W.L.O.G., let $x \in \mathcal{V}(\theta,\delta)$ satisfy $x_1'b_1 \geq \gamma(\theta) + \delta(1 + \|x\|)$; $x_2'b_2 - a_2 \leq -\gamma(\theta) - \delta(1 + \|x\|)$.

Note that

$$\mathscr{V}_1^-(x;\theta) = x_1'b_1 - \alpha_1\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\right),$$

$$\mathscr{V}_1^+(x;\theta) = x_2'b_2 - \alpha_2\Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\right).$$

Therefore,

$$\mathscr{V}_1^-(x;\theta) + \alpha_{\max}\Phi\left(\frac{\mathscr{V}_1^+(x;\theta) - \rho\mathscr{V}_2^-(x;\theta)}{\sqrt{1-\rho^2}}\right) \geq x_1'b_1 > \gamma(\theta),$$

$$\mathscr{V}_1^+(x;\theta) - \alpha_{\max}\left[1 - \Phi\left(\frac{\mathscr{V}_1^-(x;\theta) - \rho\mathscr{V}_2^+(x;\theta)}{\sqrt{1-\rho^2}}\right)\right] \leq x_2'b_2 - \alpha_2 < -\gamma(\theta)$$

where $\gamma(\theta) = -\Delta(\theta) + \alpha_{\max} \times \Phi\left(\sqrt{\frac{1+\rho}{1-\rho}}\Delta(\theta)\right)$.

Thus, by a similar argument as that in the proof for Theorem 4, it follows that

$$\mathscr{V}_1^-(x;\theta) \geq \Delta(\theta), \quad \mathscr{V}_2^+(x;\theta) \leq -\Delta(\theta),$$

which implies that $x \in \mathcal{U}(\theta)$. □

A.6. **Proof of Lemma 2.**

Proof. By Lemma 9.12 in Kosorok (2008), the class $\mathcal{G}_0$ of functions with the form $x_1'c_1 + c_0$ with (c_0, c_1) ranging over $\mathbb{R} \times \mathbb{R}^{k_1}$ is a VC class of functions. The class $\mathcal{G}_1$ of functions $x_1'b_1 - \gamma(\theta)$ with b_1 ranging over $\mathbb{R}^{k_1}$ and $\gamma(\theta) \in \mathbb{R}$ is also a VC class of functions. This is because for any $\theta \in \Theta$, $x_1'b_1 - \gamma(\theta)$ can be written as $x_1'c_1 + c_0$ for some (c_0, c_1). Then $\mathcal{G}_1$ is a sub–class of $\mathcal{G}_0$, therefore $\mathcal{G}_1$ is also a VC class of functions with no greater index. Moreover, by Part (v) in Lemma 9.9, Kosorok (2008), the class of functions with the form $x_1'b_1 - \gamma(\theta) - \delta(1 + \|x\|)$ is a VC class of functions for fixed $\delta \in \mathbb{R}^+$. Therefore, the class of sets $\{x \in \mathscr{S}_X : x_1'b_1 \geq \gamma(\theta) + \delta(1 + \|x\|)\}$ is a VC class of subsets. By Lemma 9.7 (ii) in Kosorok (2008), $\{\mathcal{V}(\theta, \delta) : \theta \in \Theta\}$ is a VC class of subsets. □

A.7. **Proof of Theorem 6.**

Proof. For the consistency of $\hat{\theta}$, all the proofs simply follow that for theorem 5. For the second part of this theorem, by definition of $\hat{\theta}$, there is

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{1}\left[X_i \in \mathcal{V}(\widetilde{\theta}, \delta)\right] s(Y_i, X_i; \hat{\theta}) = 0.$$

By Taylor expansion

$$\frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right] s(Y_i, X_i;\theta_0) + \frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right]\left(\frac{\partial}{\partial\theta}s(Y_i, X_i;\theta^{\dagger})\right)'(\hat{\theta}-\theta_0) = 0$$

where $\theta^{\dagger}$ is between $\hat{\theta}$ and θ_0. Hence

$$\begin{aligned}&\sqrt{n}(\hat{\theta}-\theta_0)\\ &\quad = -\left\{\frac{1}{n}\sum_{i=1}^{n} 1\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right]\left(\frac{\partial}{\partial\theta}s(Y_i, X_i;\theta^{\dagger})\right)'\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} 1\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right] s(Y_i, X_i;\theta_0)\end{aligned}$$

By the ULLN, assumption E and the fact that $1\left[X_i \in \mathcal{V}(\theta,\delta)\right]$ belongs to VC class of functions indexed by $\theta \in \mathcal{N}_{\epsilon}(\theta_0)$, there is

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right]\left(\frac{\partial}{\partial\theta}s(Y_i, X_i;\theta^{\dagger})\right)' \xrightarrow{p} \mathbb{E}\left\{\mathbf{1}\left[X_i \in \mathcal{V}(\theta_0,\delta)\right]\left(\frac{\partial}{\partial\theta}s(Y_i, X_i;\theta_0)\right)'\right\}.$$

Hence, it suffices to show

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{1}\left[X_i \in \mathcal{V}(\widetilde{\theta},\delta)\right] s(Y_i, X_i;\theta_0) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{1}\left[X_i \in \mathcal{V}(\theta_0,\delta)\right] s(Y_i, X_i,\theta_0) = o_p(1).$$

Let $h(Y,X;\theta,\delta) = 1\left[X \in \mathcal{V}(\theta,\delta)\right] s(Y,X;\theta_0)$, $\mathbb{G}_n(\theta) = n^{-1}\sum_{i=1}^{n} h(Y_i, X_i;\theta,\delta) - \mathbb{E}h(Y,X;\theta,\delta)$. Because $1\left[x \in \mathcal{V}(\theta,\delta)\right]$ indexed by θ is a VC class of functions, then by empirical processes method (see Pollard, 1989), for every sequence of positive numbers $\{\epsilon_n\}$ converging to zero that

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

which implies that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i, X_i;\widetilde{\theta},\delta) &= n^{1/2}\mathbb{G}_n(\widetilde{\theta}) + n^{1/2}\mathbb{E}h(Y,X;\widetilde{\theta},\delta)\\ &= n^{1/2}\left[\mathbb{G}_n(\widetilde{\theta})-\mathbb{G}_n(\theta_0)\right] + n^{1/2}\mathbb{G}_n(\theta_0) + n^{1/2}\mathbb{E}h(Y,X;\widetilde{\theta},\delta)\\ &= o_p(1) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i, X_i;\theta_0,\delta) + n^{1/2}\left[\mathbb{E}h(Y,X;\widetilde{\theta},\delta) - \mathbb{E}h(Y,X;\theta_0,\delta)\right]\end{aligned}$$

Because (1)$\mathbb{E}h(Y,X;\theta_0,\delta) = 0$; (2) $\widetilde{\theta} \xrightarrow{p} \theta_0$, then $\mathbb{P}\{\mathcal{V}(\widetilde{\theta},\delta) \subseteq \mathcal{V}(\theta_0,0)\} \to 1$. Thus $\mathbb{E}h(Y,X;\widetilde{\theta},\delta) = 0$ with probability approaching to one. Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i, X_i;\widetilde{\theta},\delta) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(Y_i, X_i;\theta_0,\delta) = o_p(1).$$

□

Appendix B. A general result using copula functions

Let $C(v_1, v_2; \rho_0)$ be the copula function of the joint distribution of (U_1, U_2), i.e. $C(v_1, v_2; \rho_0) = F_U\left(F_1^{-1}(v_1), F_2^{-1}(v_2); \rho_0\right)$. Let further F and f be the marginal c.d.f. and p.d.f. of U_j ($j = 1, 2$), respectively. Then the conditions to define $\mathcal{U}(\theta_0)$ can be written as: $x \in \mathcal{U}(\theta_0)$ if and only if

$$1 + \alpha_j \times \frac{\partial^2 C(F(u_1), F(u_2); \rho_0)}{\partial v_j^2} \times f(u_j) \geq \alpha_j \times \frac{\partial^2 C(F(u_1), F(u_2); \rho_0)}{\partial v_1 \partial v_2} \times f(u_{-j})$$

for all $u \in \mathcal{I}(x; \theta_0)$ and $j = 1, 2$.

Assumption F. *The p.d.f. f satisfies: $f(u) = f(-u)$ for all $u \in \mathbb{R}$ and $f\left(F^{-1}(\tau)\right)$ is increasing in $\tau \in (0, 1/2]$.*

Assumption G. *The copula function C satisfies: (i) $\partial C^2(v_1, v_2; \rho_0)/\partial v_j^2 \leq 0$; (ii) $\partial C^2(v_1, v_2; \rho_0)/\partial v_1 \partial v_2$ is monotone increasing in v_j and monotone decreasing in v_{-j} on the support $(v_j, v_{-j}) \in (0, 1/2] \times [1/2, 1)$; (iii) $\partial C^2(v_1, v_2; \rho_0)/\partial v_j^2$ is monotone decreasing in v_j and monotone increasing in v_{-j} on the support $(v_j, v_{-j}) \in (0, 1/2] \times [1/2, 1)$.*

Assumption F imposes weak restrictions on the shape of the c.d.f. of U_j, which can be satisfied by, e.g., standard normal or standard logistic distribution. Assumption F implies that $F^{-1}(\tau) = -F^{-1}(1 - \tau)$. Assumption G essentially restricts the dependence structure between private information. Assumption G–(i) is equivalent to the positive regression dependence condition (see, e.g., de Castro, 2007, for a definition and examples of positive regression dependence). Note that

$$\frac{\partial C^2(v_1, v_2; \rho_0)}{\partial v_1 \partial v_2} = \frac{f_U(F^{-1}(v_1), F^{-1}(v_2))}{f(F^{-1}(v_1)) \times f(F^{-1}(v_2))}.$$

Therefore, $\partial C^2(v_1, v_2; \rho_0)/\partial v_1 \partial v_2$ is always positive. Assumption G can be satisfied by, e.g., an FGM copula $\mathcal{C}(\nu_1, \nu_2; \rho_0) = \nu_1 \nu_2 \left[1 + \rho_0 (1 - \nu_1)(1 - \nu_2)\right]$ with $0 \leq \rho_0 \leq 1$. It is straightforward to verify assumption G–(i), (ii) and (iii), since

$$\frac{\partial C^2(v_1, v_2; \rho_0)}{\partial v_1 \partial v_2} = 1 + \rho_0 \left[-v_1 - v_2 + 2\nu_1 \nu_2\right], \quad \frac{\partial C^2(v_1, v_2; \rho_0)}{\partial v_j^2} = 2\rho_0 v_{-j}(v_{-j} - 1).$$

Lemma 4. *Suppose assumptions F and G hold. Let $\tau = \tau(\theta_0) \in (0, 1/2]$ solve*[19]

$$1 + \alpha_{\max} \times \frac{\partial^2 C(\tau, 1-\tau; \rho_0)}{\partial v_j^2} \times f\left(F^{-1}(\tau)\right) \geq \alpha_{\max} \times \frac{\partial^2 C(\tau, 1-\tau; \rho_0)}{\partial v_1 \partial v_2} \times f\left(F^{-1}(1-\tau)\right).$$

Then a sufficient condition for $x \in \mathcal{U}(\theta_0)$ is: either $\mathscr{V}_1^-(x) \geq F^{-1}(1-\tau(\theta_0))$; $\mathscr{V}_2^+(x) \leq F^{-1}(\tau(\theta_0))$, or $\mathscr{V}_1^-(x) \leq F^{-1}(\tau(\theta_0))$; $\mathscr{V}_2^-(x) \geq F^{-1}(1-\tau(\theta_0))$.

Proof. It directly follows from assumptions F and G.

Further, I define Π as follows: let $\Pi \equiv \{x \in \mathscr{S}_X : \mathbb{E}(Y_1|X) \geq F(\tilde{\gamma}_0^*), \mathbb{E}(Y_2|X) \leq 1 - F(\tilde{\gamma}_0^*)\} \cup \{x \in \mathscr{S}_X : \mathbb{E}(Y_1|X) \leq 1 - F(\tilde{\gamma}_0^*), \mathbb{E}(Y_2|X) \geq F(\tilde{\gamma}_0^*)\}$, where $\tilde{\gamma}^*(\theta) \equiv F^{-1}(\tau(\theta)) + \overline{\alpha} \times \frac{\partial C(\tau(\theta), 1-\tau(\theta); \rho)}{\partial v_1}$ and $\tilde{\gamma}_0^* = \sup_{\theta \in \Theta} \tilde{\gamma}^*(\theta)$. By a similar argument as that in the proof of Theorem 4, one can show that $\Pi \subseteq \mathcal{C}(\Theta_I)$.

[19]By assumptions F and G, there are at most one solution. It is understood that if there no such a solution, it corresponds to the degenerated case, i.e., $\mathcal{U}(\theta_0)$ is the whole support in the covariate space. For notational brevity, let $\tau(\theta_0) = 1/2$ when there is no solution.