Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 09-023

"Estimating Simultaneous Games with Incomplete Information under Median Restrictions"

by

Xun Tang

http://ssrn.com/abstract=1431493

Estimating Simultaneous Games with Incomplete Information under Median Restrictions

Xun TANG[1]

University of Pennsylvania

April 30, 2009

Abstract

I estimate a simultaneous discrete game with incomplete information where players' private information are only required to be median independent of observed states and can be correlated with observable states. This median restriction is weaker than other assumptions on players' private information in the literature (e.g. perfect knowledge of its distribution or its independence of the observable states). I show index coefficients in players' utility functions are point-identified under an exclusion restriction and fairly weak conditions on the support of states. This identification strategy is fundamentally different from that in a single-agent binary response models with median restrictions, and does not involve any parametric assumption on equilibrium selection in the presence of multiple Bayesian Nash equilibria. I then propose a two-step extreme estimator for the linear coefficients, and prove its consistency.

KEYWORDS: Games with incomplete information, semiparametric identification, median restrictions, consistent estimation

JEL CODES: C14, C35, C51

[1]Corresponding author: Xun Tang, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. Tel: 215-898-7409. Email: xuntang@econ.upenn.edu. Fax: 215-573-2057.

This paper evolves from a chapter of my Ph.D. dissertation at Northwestern University. I am grateful to Elie Tamer for detailed discussions and comments. I also thank Andrés Aradillas-López, Hanming Fang, Stefan Hoderlein, Bo Honoré, Joel Horowitz, Shakeeb Khan, Charles Manski, Rosa Matzkin, Ulrich Müller, Aureo De Paula, Robert Porter, Kyungchul Song, Petra Todd and the econometrics seminar participants at Northwestern, Princeton, and University of Pennsylvania for helpful feedbacks. Financial support from the Center of Studies in Industrial Organizations at Northwestern University is gratefully acknowledged.

1 Introduction

Consider a simple 2-by-2 simultaneous discrete game with incomplete information where the space of pure strategies $\{1, 0\}$ is the same for both players $i = 1, 2$. The payoff structure is :

	0	1
0	0,0	$0, X'\beta_2 - \epsilon_2$
1	$X'\beta_1 - \epsilon_1, 0$	$X'\beta_1 + \delta_1 - \epsilon_1, X'\beta_2 + \delta_2 - \epsilon_2$

where $X \in \mathbb{R}^K$ are states observed by players and econometricians, and $\epsilon_i \in \mathbb{R}^1$ is private information (PI) observed by player i only but not the rival or econometricians.[2] The rows and the columns correspond to the strategies of players 1 and 2 respectively. The first expressions in each of the four cells correspond to the payoffs for player 1, while the second expressions correspond to those for player 2. The joint distribution of PI given X (denoted by $F_{\epsilon|X}$ where $\epsilon \equiv (\epsilon_1, \epsilon_2)$) and parameters $\theta \equiv \{\beta_i, \delta_i\}_{i=1,2}$ are common knowledge among players. Let Ω_R denote the support of a generic random variable R. A *pure strategy* for player i is a mapping $g_i : \Omega_X \otimes \Omega_{\epsilon_i} \to \{0, 1\}$.[3] A *pure-strategy Bayesian Nash equilibrium* (BNE) is summarized by a pair of mappings $A_i : \Omega_X \to \Omega_{\epsilon_i}$ that maps from a $x \in \Omega_X$ to a subset of Ω_{ϵ_i} (so that the pure strategies are $g_i(x, \varepsilon_i) = 1(\varepsilon_i \in A_i(x))$ where $1(.)$ is the indicator function) and

$$
\begin{aligned}
A_1^*(x) &= \{\varepsilon_1 : \varepsilon_1 \leq x'\beta_1 + \delta_1 P(\epsilon_2 \in A_2^*(x)|\varepsilon_1, x)\} \\
A_2^*(x) &= \{\varepsilon_2 : \varepsilon_2 \leq x'\beta_2 + \delta_2 P(\epsilon_1 \in A_1^*(x)|\varepsilon_2, x)\}
\end{aligned}
$$

Obviously $A_i^*(x)$ depends on $\{\delta_i, \beta_i\}_{i=1,2}$ and $F_{\epsilon|X=x}$, and is independent of realized values of $(\varepsilon_1, \varepsilon_2)$.[4] The existence of BNE follows from an application of Brouwer's Fixed Point Theorem but uniqueness is not guaranteed. Econometricians only know $\delta_i < 0$ for $i = 1, 2$, and the following restrictions are satisfied by the data-generating process (DGP).

CMI (Conditional and median independence) ϵ_1 is independent of ϵ_2 conditional on X with $Med(\epsilon_i|X = x) = 0$ for $i = 1, 2$ almost everywhere on Ω_X.

SE (Single equilibrium) If multiple BNE exist, only a single equilibrium is played in the DGP.

[2] I use upper case letters to denote random variables or vectors (such as ϵ), and lower case letters for their realizations (such as ε).

[3] Our approach allows for the dependence of support of ϵ on states X. We use Ω_ϵ to denote the support of private information only for the sake of simplicity in exposition.

[4] More generally, pure-strategies should take the form $g_i(x, \varepsilon_i) = 1(\varepsilon_i \in A_i(\varepsilon_i, x))$, but this can be easily represented as $g_i(x, \varepsilon_i) = 1(\varepsilon_i \in A_i^*(x))$ with $A_i^*(x) \equiv \{\varepsilon_i : \varepsilon_i \in A_i(\varepsilon_i, x)\}$.

Under CMI and SE, the observed choice probabilities $p(x) \equiv [p_1(x)\ p_2(x)]$ in the DGP (where $p_i(x)$ denotes the probability that i chooses 1 under state x) must solve,

$$\begin{bmatrix} p_1(x) \\ p_2(x) \end{bmatrix} = \begin{bmatrix} F_{\epsilon_1|X=x}(x'\beta_1 + p_2(x)\delta_1) \\ F_{\epsilon_2|X=x}(x'\beta_2 + p_1(x)\delta_2) \end{bmatrix} \tag{1}$$

We shall argue below that under CMI and SE, a pair $(\theta, F_{\epsilon_1,\epsilon_2|X})$ generates $p(x)$ if and only if it generates $p_i(x)$ in two simultaneous "binary response" models $Y_i = 1(X'\beta_i + p_{-i}(X)\delta_i - \epsilon_i \geq 0)$ for $i = 1, 2$. The interaction term δ_i needs to be normalized to -1 for $i = 1, 2$ to identify β_i under median restrictions alone.

This paper defines and characterizes the identification region of β_i, gives sufficient conditions for point identification of $\{\beta_i\}_{i=1,2}$ and proposes a semiparametric two-step extreme estimator that is consistent for the identified set (under the Hausdorff metric). The identification strategy in this game theoretic context is fundamentally different from arguments used in a single-agent, binary response model with median restrictions (such as Manski (1985), Manski and Tamer (2002)). This is because in a Bayesian Nash equilibrium, opponents' choice probabilities enter players' payoffs as an additional state variable, thus invalidating the large support assumption on regressors that has been used in the identification in single-agent binary response models under median independence. Our solutions is to exploit an exclusion restriction in linear index utilities and introduce a fairly weak condition that only requires regressors' support to be closed under scalar contractions (i.e. multiplications with some constant $|\alpha| < 1$). Then point-identification can be achieved with bounded support of regressors.

Several recent works have estimated simultaneous discrete games with incomplete information under different assumptions. Aradillas-Lopez (2005) studied a case where (ϵ_1, ϵ_2) are jointly independent from observable states X. He extended the semiparametric likelihood estimator in Klein and Spady (1993) to this game-theoretic framework. He gave sufficient and necessary conditions for the uniqueness of BNE, which is necessary for a well-defined likelihood. Bajari, Hong, Krainer and Nekipelov (2007) showed a general function $u(X)$ can be identified nonparametrically provided private information (PI) are independently and identically distributed across players given X and that $F_{\varepsilon_1,\varepsilon_2|X}$ is perfectly known to the econometrician. In comparison, I formulate the BNE as a system of two binary regressions (where rivals' equilibrium choice probabilities enter as an additional regressor) under much weaker restrictions of median independence. Point identification of index coefficients is attained under the weak median restriction at a moderate cost of some additional (yet fairly general) conditions on the support of regressors.

2 Identification of Index Coefficients

2.1 Partial identification

Let $\mathcal{F}_{CMI}$ denote the set of distributions of private information $F_{\epsilon|X}$ that satisfy CMI. Let $(\theta^0, F^0_{\epsilon|X})$ denote the true index coefficients and the distribution of private information (PI) in the DGP, and let $p^* \equiv (p_1^*, p_2^*)$ where $p_i^*(x; \theta^0, F^0_{\epsilon|X})$ is the probability that " i chooses 1 in state x" actually observed in the DGP with true parameters $(\theta^0, F^0_{\epsilon|X})$. For any generic pair of coefficients $\theta \equiv \{\beta_i, \delta_i\}_{i=1,2}$ and PI distribution $G_{\epsilon|X}$, let $\psi(x; \theta, G_{\epsilon|X})$ denote the set of choice probabilities under state x implied by a model with $\theta, G_{\epsilon|X}$. That is,

$$\psi(x; \theta, G_{\epsilon|X}) \equiv \{(p_1, p_2) \in [0,1]^2 : (p_1, p_2) \text{ solves (1) given } x, \theta, G_{\epsilon|X=x}\} \tag{2}$$

The mapping ψ is non-empty by the Brouwer's fixed point theorem but might be a correspondence in general as uniqueness is not guaranteed. Define

$$\chi(p^*, b, G_{\varepsilon|X}) \equiv \{x : (p_1^*(x), p_2^*(x)) \in \psi(x; b, G_{\epsilon|X})\} \tag{3}$$

Let Θ denote the parameter space for θ.

Definition 1 *Given a p^* observed in the DGP with $\theta^0 \in \Theta$ and $F^0_{\epsilon|X} \in \mathcal{F}_{CMI}$, a parameter θ is observationally equivalent (denoted $\overset{o.e.}{\sim}$) to θ^0 under $\mathcal{F}_{CMI}$ if $\exists F_{\epsilon|X} \in \mathcal{F}_{CMI}$ such that $\Pr\{X \in \chi(p^*, \theta, F_{\epsilon|X})\} = 1$. The identification region of θ^0 under $\mathcal{F}_{CMI}$ is the subset of Θ such that $\theta \overset{o.e.}{\sim} \theta^0$ under $\mathcal{F}_{CMI}$ for all θ in the subset. We say θ^0 is point-identified under $\mathcal{F}_{CMI}$ if this identification region is the singleton θ^0.*

Two remarks are necessary. First, identification is relative to the BNE outcome p^* actually observed in the DGP (which, under SE, solves (1)). Second, the definition of " $\overset{o.e.}{\sim}$ " only requires *marginal* choice probabilities of the players to be rationalizable by the implied BNE under $(\theta, F_{\epsilon|X})$, even though econometricians get to observe their *joint* choice probabilities. This is because under the CMI and SE, the joint choice probability equals a product of two marginal probabilities, and our point of departure is the choice probabilities observed in DGP satisfy this testable implication so that the identification region will not be vacuously empty.

Let $\theta_i \equiv (\beta_i, \delta_i)$ and Θ_i denote the corresponding parameter space. Suppose the model is correctly specified for some $\theta^0, F^0_{\epsilon_1,\epsilon_2|X} \in \Theta \otimes \mathcal{F}_{CMI}$. Let $\mathcal{F}^i_{MI}$ denote the set of marginal distributions of ϵ_i that correspond to a joint distribution $F_{\epsilon|X}$ in $\mathcal{F}_{CMI}$. For any $x \in \Omega_X$ and a pair of generic parameters $\theta \in \Theta$, $G_{\epsilon|X} \in \mathcal{F}_{CMI}$, define the analogs to (2) and (3) in the context of single-agent decisions:

$$\psi_i(x;\theta_i, G_{\epsilon_i|X}, p^*_{-i}) \equiv G_{\epsilon_i|x}(x'\beta_i + p^*_{-i}(x)\delta_i)$$

and

$$\chi_i(p^*, \theta_i, G_{\epsilon_i|X}) \equiv \{x : p^*_i(x) = \psi_i(x;\theta_i, G_{\epsilon_i|X}, p^*_{-i})$$

Definition 2 *Given an BNE outcome p^* observed in a DGP with $(\theta^0, F^0_{\epsilon|X})$, θ_i is unilaterally observationally equivalent to θ^0_i (denoted $\overset{u.o.e.}{\sim}$) under $\mathcal{F}^i_{MI}$ if $\exists F_{\epsilon_i|\mathbf{X}} \in \mathcal{F}^i_{MI}$ such that $\Pr(X \in \chi_i(p^*, \theta_i, F_{\epsilon_i|X})) = 1$. Then θ^0_i is unilaterally point-identified in Θ_i under $\mathcal{F}^i_{MI}$ (given p^*) if $\forall F_{\epsilon_i|X} \in \mathcal{F}^i_{MI}$, $\Pr(X \in \chi_i(\theta_i, F_{\epsilon_i|\mathbf{X}}, p^*)) < 1$ for all $\theta_i \neq \theta^0_i$ in Θ_i.*

Lemma 1 *Suppose CMI and SE hold. Given an BNE outcome p^* observed in the DGP, $\theta \overset{o.e.}{\sim} \theta^0$ under $\mathcal{F}_{CMI}$ if and only if $\theta_i \overset{u.o.e.}{\sim} \theta^0_i$ under $\mathcal{F}^i_{MI}$ for both $i = 1, 2$.*

Let $\tilde{x}_i$ denote the vector of "augmented regressors" for player i (i.e. $\tilde{x}_i \equiv [x, p^*_{-i}(x)]$). Note $\tilde{x}_i$ depends on true parameters $\theta^0, F^0_{\epsilon|X}$ through the rival choice probabilities in BNE $p^*_{-i}(x)$. Let

$$\xi_{i,\theta_i} \equiv \{x : \text{"}\tilde{x}'_i\theta_i \leq 0 \textit{ and } p^*_i(x) > \tfrac{1}{2}\text{" or "}\tilde{x}'_i\theta_i \geq 0 \textit{ and } p^*_i(x) < \tfrac{1}{2}\text{"}\} \tag{4}$$

We have suppressed the dependence of ξ_{i,θ_i} on p^*_i (and therefore θ^0, $F^0_{\epsilon|X}$) in the definition for notational ease.

Lemma 2 *Suppose CMI and SE hold. Then $\theta \overset{o.e.}{\sim} \theta^0$ under $\mathcal{F}_{CMI}$ if and only if $\Pr(X \in \cup_{i=1,2}\xi_{i,\theta_i}) = 0$. Furthermore, the identification region of θ^0 under $\mathcal{F}_{CMI}$ is convex.*

2.2 Point identification

The proof of point-identification of θ^0 (up to scale) in the game is fundamentally different from that in a binary regression with median independence of the errors, even though Lemma 1 suggests an intuitive link between the identification of θ^0 in simultaneous games and that in a system of two binary response decisions. This is because in both individual binary responses in (1), the opponent's choice probabilities $p^*_{-i}(x)$ enter as an additional regressor in $\tilde{x}_i \equiv [x, p^*_{-i}(x)]$. And the conditions that would point-identify θ^0 up to scale in a single-agent binary regression with median independence (such as the richness of support of $\tilde{x}_i$ in Manski (1985) and Manski and Tamer (2002))) are not satisfied by this "additional regressor" whose support is confined in $[0,1]$ and depends on more primitive conditions on true parameters in the DGP $\theta^0, F^0_{\epsilon|X}$. We shall establish our point-identification result using an exclusion restriction on the index utilities and some qualitatively different, novel conditions on the support Ω_X. Let Θ^B denote the parameter space of $\beta \equiv (\beta_1, \beta_2)$, Θ^B_i denote the parameter space of β^0_i, and $\beta^0 \equiv (\beta^0_1, \beta^0_2)$ and $\delta^0 \equiv (\delta^0_1, \delta^0_2)$ denote the actual index in the DGP.

ER (Exclusion Restriction) For $i = 1, 2$, there exists an index $h_i \in \{1, 2, ..., K\}$ such that $\beta^0_{i,h_i} = 0$, $\beta^0_{-i,h_i} \neq 0$.

PS (Parameter Space) For $i = 1, 2$, β^0_i is in the interior of Θ^B_i, where Θ^B_i is a convex, compact subset of $\mathbb{R}^K$.

SN (Scale normalization) For $i = 1, 2$, $|\delta^0_i| = 1$.

The exclusion restriction requires that for each player i at least one of the regressors (β^0_{i,h_i}) do not enter the utility for i, but enter the utility of the opponent $-i$. Thus, β^0_{i,h_i} (and therefore x_{h_i}) only affects $p^*_i(x)$ indirectly through $p^*_{-i}(x)$. This restriction is crucial for our identification arguments as it ensures that from the perspective of player i the opponent choice probabilities p^*_{-i} can vary while the index utilities $X'\beta^0_i$ is fixed, thus allowing any index coefficient β_i such that $X'\beta_i \neq X'\beta^0_i$ with positive probability to be distinguishable from β^0_i provided certain general conditions on Ω_X are satisfied. Such exclusion restrictions arise naturally in lots of empirical applications in industrial organizations. For example, consider static entry/exit games between firms located in different geographical regions. The indices $X'\beta_i$ are interpreted as conditional medians of monopoly profits. There can be commonly observed geographical features (such as local demographics of the workforce with in a region, etc) that only affect the profitability of the local firms but not that of others. Finally, note the scale normalization is necessary for identifying (β^0_1, β^0_2), as is obvious from representation of the simultaneous game as parallel binary responses.

Let $\Omega_{W|Z=z}$ denote the conditional support of a generic variable W given a realized value of another generic variable Z at z.

PID-1 (Private information distributions) For i=1,2, (i) for all $x \in \Omega_X$ where Ω_X is a compact support, $F^0_{\epsilon_i|x}$ are Lipschitz continuous with positive densities on the support of ϵ_i with an unknown positive constant C_{F_i}; (ii) there exists an unknown constant $K^i_{F_j} > 0$ such *that for* $j = 1, 2$,

$$\sup_{t \in \mathbb{R}^1} |F^0_{\epsilon_j|\bar{x}_{-h_i},\tilde{x}_{h_i}}(t) - F^0_{\epsilon_j|\bar{x}_{-h_i},x_{h_i}}(t)| \leq K^i_{F_j}|\tilde{x}_{h_i} - x_{h_i}|$$

for all $\bar{x}_{-h_i} \in \Omega_{X_{-h_i}}$ *and* $x_{h_i}, \tilde{x}_{h_i}$ *on a compact support* $\Omega_{X_{h_i}|\bar{x}_{-h_i}}$.

PID-(i) requires the marginal distributions of ϵ_1, ϵ_2 conditional on any x respectively not to increase too fast, and rules out discontinuities (jumps) in the distributions. *PID-(ii)* requires the marginal distributions of ϵ_1 and ϵ_2 given any $\bar{x}_{-h_i}$ "not to perturb too much" as x_{h_i} changes. It is satisfied if ϵ_j is independent of X_{h_i} conditional on X_{-h_i}. These two restrictions enable an application of a version of the fixed point theorems to show the actual choice probabilities p_i^* observed in the DGP, as a solution to (1) in BNE, are continuous in the regressors X_{h_i} (which is excluded from the index $X_i'\beta_i$) conditional on all other regressors.

Lemma 3 *Suppose CMI, SE, ER, PS, SN and PID-1 are satisfied with* $|C_{F_1}C_{F_2}| < 1$*. Then* $p_i^*(x_{h_i}, \bar{x}_{-h_i})$ *is continuous in* x_{h_i} *for any* $\bar{x}_{-h_i} \in \Omega_{X_{-h_i}}$.

The next lemma ensures for $i = 1, 2$, the actual opponent choice probabilities p_{-i}^* observed is dense given any x_{-h_i}, in the sense that there is a positive probability that p_{-i}^* falls within any open interval on a certain closed interval in $[0, 1]$. We prove this by using the continuity of p_i^* in x_{h_i} and the following conditions on the distribution and support of X_{h_i} given any $\bar{x}_{-h_i}$.

PID-2 (P.I. distributions) For $i = 1, 2$, *there exists a constant interval* $[a^i, b^i] \in (0, 1)$ *and constants* $c^i, d^i \in \mathbb{R}^1$ *(with* $c^i < d^i$*) such that* $\Pr(\epsilon_i < c^i|x) < a_i$ *and* $\Pr(\epsilon_i > d^i|x) < 1 - b^i$ *for all* $x \in \Omega_X$.

RSX-1 (Rich support of X) For $i = 1, 2$, *and for all* $\bar{x}_{-h_i} \in \Omega_{X_{-h_i}}$, X_{h_i} *is continuously distributed with a compact support on* $\mathbb{R}^1$ *and* $\Pr(X'\beta^0_{-i} \in I|\bar{x}_{-h_i}) > 0$ *for any open interval* I *in* $[c^i - 1, d^i + 1]$ *as defined above.*

Assumption *PID-2* implies for both $i = 1, 2$, there exists an interval on the real line such that the probability for ϵ_i to lie in this interval is uniformly bounded below by $b^i - a^i$ for all $x \in \Omega_X$. Among other things, this restriction can be satisfied if the support of ϵ_i is bounded for all $x \in \Omega_X$, or if ϵ_i are independent of X for $i = 1, 2$. The assumption *RSX-1* is plausible because $\beta^0_{2,h_1} \neq 0$ and it can be satisfied if x_{h_i} has sufficiently large support conditional on any x_{-h_i}. Note this assumption is compatible with the compactness of the support Ω_X.

Lemma 4 *Suppose CMI, SE, ER, PS, SN, PID-1,2, RSX-1 are satisfied. Then for any open interval* $I \subset [a^{-i}, b^{-i}]$, $\Pr\{p^*_{-i}(X) \in I|\ X_{-h_i} = \bar{x}_{-h_i}\} > 0$ *for all* $\bar{x}_{-h_i} \in \Omega_{X_{-h_i}}$ *and* $i = 1, 2$.

Now we are ready to prove the point-identification of β^0 under median independence.

RSX-2 (Rich support of X) For $i = 1, 2$, *(i) for all nonzero vector* $\lambda \in \mathbb{R}^{K-1}$, $\Pr(X'_{-h_i}\lambda \neq 0) > 0$*; (ii) there exists an unknown constant* $\bar{C} < \infty$ *such that*

$$\Pr(\text{"}a^{-i} < \min\{|X'b_i|, |X'b_i'|\} \leq \bar{C}\text{"} \ \wedge \ \text{"}X'b_i \neq X'b_i'\text{"}\ ) > 0$$

for all $b_i, b_i' \in \Theta_i^B$*; (iii) For all* $S \subseteq \Omega_{X_{-h_i}}$ *such that* $P(X_{-h_i} \in S) > 0$, $\Pr(X_{-h_i} \in \alpha S) > 0$ $\forall \alpha \in (-1, 1)$ *where* $\alpha S \equiv \{\tilde{x}_{-h_i} : \tilde{x}_{-h_i} = \alpha x_{-h_i}\ for\ some\ x_{-h_i} \in S\}$.

The condition (i) in RSX-2 ensures there is a positive probability that $X'\beta_i \neq X'\beta_i'$ for any pair $\beta_i \neq \beta_i'$ in Θ_i^B, which is necessary for condition (ii) in RSX-2. The condition (iii) in *RSX-2* requires support of X_{-h_i} to be closed under scalar multiplications with α where $|\alpha| < 1$. Together with condition (ii), this ensures the random interval between $X'\beta_i$ and $X'\beta_i'$ must intersect with $[a^{-i}, b^{-i}]$ with positive probability for all pairs in Θ_i^B. Then Lemma 4 can be used for showing the identification of β^0 (remarkably with the support of states being compact).

Proposition 1 *Suppose CMI, SE, ER, PS, SN, PID-1,2 and RSX-1,2 are satisfied and* $\delta_i < 0$ *for* $i = 1, 2$. *Then* $\beta^0 = (\beta_1^0, \beta_2^0)$ *is point-identified under* $\mathcal{F}_{CMI}$.

Proof of Proposition 1. We begin with the case of player 1. Note x_{h_1} affects both $p_2^*(X)$ and $p_1^*(X)$ (through $p_2^*(X)$) but not $X'\beta_1^0$ since $\beta_{1,h_1}^0 = 0$ by ER. By condition *(i)* in *RSX-2*, $\Pr\{X'_{-h_1}(\beta_{1,-h_1} - \beta_{1,-h_1}^0) \neq 0\} > 0$ for all $\beta_1 \neq \beta_1^0$ in Θ_1^B. Let I_{β_1,β_1^0} denote the random interval between $X'_{-h_1}\beta_{1,-h_1}$ and $X'_{-h_1}\beta_{1,-h_1}^0$. We shall prove the claim that "under the conditions of the proposition, there is positive probability that $I_{\beta_1,\beta_1^0} \cap [a^1, b^1]$ is an non-degenerate interval with an interior". (We shall refer to this claim as the "non-degenerate random intersection" (*NDRI*) claim hereafter.) By condition *(ii)* in *RSX-2*, we have

$$\Pr(\text{"}a^2 < \min\{|X'\beta_1|, |X'\beta_1^0|\} \leq \bar{C}\text{"} \wedge \text{"}X'\beta_1 \neq X'\beta_1^0\text{"}) > 0$$

for any $\beta_1 \neq \beta_1^0$. Denote the intersection of these three events in the display above by event "A" and denote the event that "$sign(X'\beta_1) \neq sign(X'\beta_1^0)$" by event "$B$". Let B^c denote the complement of event B. Then it must be true that either $\Pr(A \wedge B) > 0$ or

$\Pr(A \wedge B^c) > 0$. Suppose the former is true. Then it follows immediately that the *NDRI* claim is true. Suppose the latter is true. Then it must be the case either

$$\Pr\left(\text{"}a^2 < \min\{X'\beta_1, X'\beta_1^0\} \leq \bar{C}\text{"} \wedge \text{"}\min\{X'\beta_1, X'\beta_1^0\} > 0\text{"} \wedge \text{"}X'\beta_1 \neq X'\beta_1^0\text{"}\right) > 0$$

or

$$\Pr\left(\text{"}a^2 < \min\{-X'\beta_1, -X'\beta_1^0\} \leq \bar{C}\text{"} \wedge \text{"}\max\{X'\beta_1, X'\beta_1^0\} < 0\text{"} \wedge \text{"}X'\beta_1 \neq X'\beta_1^0\text{"}\right) > 0$$

In either case, condition *(iii)* in *RSX-2* (support of X closed under scalar multiplication with $-1 < \alpha < 1$) implies that the *NDRI* claim is true (even when $\min\{|X'\beta_1|, |X'\beta_1^0|\} > b^2$). Besides, by Lemma 4, $\Pr\{\bar{p}_2^*(X_{h_1}, \bar{x}_{-h_1}) \in I|\bar{x}_{-h_1}\} > 0$ for all open interval $I \subset [a^2, b^2]$ and all $\bar{x}_{-h_i} \in \Omega_{X_{-h_i}}$ under the conditions of the proposition, and note the random interval I_{β_1,β_1^0} only depends on X_{-h_1} but not X_{h_1} due to the exclusion restriction. It then follows $\Pr(p_2^*(X) \in I_{\beta_1,\beta_1^0}) > 0$ for all $\beta_1 \neq \beta_1^0$, and therefore

$$\Pr(sign(X'\beta_1 - p_2^*(X)) \neq sign(X'\beta_1^0 - p_2^*(X))) > 0$$

Hence no $\beta_1 \neq \beta_1^0$ is observationally equivalent to β_1^0 under $\mathcal{F}_{CMI}$. Proof for the case with player 2 follows from similar arguments. Q.E.D.

3 Estimation

We define an extreme estimator for the identification region of β^0 under $\mathcal{F}_{CMI}$ (denoted Θ_{ID}^B) by minimizing a non-negative, random function $\hat{Q}_n(b)$ constructed from empirical distributions of choices and states observed in data. The idea is that the limiting function of $\hat{Q}_n(.)$ as $n \to \infty$ (denoted $Q(.)$) is equal to zero if and only if $b \in \Theta_{ID}^B$. Thus the set of minimizers of $\hat{Q}_n(.)$ converge to Θ_{ID}^B in probability (denoted $\xrightarrow{p}$) under the Hausdorff metric between sets, provided $\hat{Q}_n \xrightarrow{p} Q$ uniformly over the parameter space Θ^B. We start by defining the limiting function $Q(.)$ first. Define the non-stochastic function for $i = 1, 2$,

$$Q_i(b_i) \equiv E[\Lambda(1/2 - p_i^*(X))(X'b_i - p_{-i}^*(X))_+^2 + \Lambda(p_i^*(X) - 1/2)(X'b_i - p_{-i}^*(X))_-^2]$$

where $1(.)$ is the indicator function, $a_+ \equiv \max(0, a)$, $a_- \equiv \max(0, -a)$, and where $\Lambda : [-\frac{1}{2}, \frac{1}{2}] \to [0, +\infty)$ is a smooth function such that $\Lambda(c) = 0$ for all $c \leq 0$ and $\Lambda(c) > 0$ for all $c > 0$. Let $Q(b) = Q_1(b_1) + Q_0(b_0)$ for all $b = (b_1, b_0)$ in Θ^B.

Lemma 5 *Suppose CMI and SE are satisfied and* $\Pr\{X'b_i = p_{-i}^*(X)\} = 0$ *for all* $b_i \in \Theta_i^B$ *and* $i = 1, 2$. *Then* $Q(b) \geq 0$ *for all* $b \in \Theta^B$ *and* $Q(b) = 0$ *if and only if* $b \in \Theta_{ID}^B$.

Let g and j be indices for cross-sectional units in the DGP (i.e. the games) and n denote the number of games in the data (i.e. the sample size). Let $Y_{i,g}$ be the choice of player i in game g. Define kernel estimates for $f_0(x_g)$ and $h_0(x_g) \equiv E(Y_{i,g}|X_g = x_g)f_0(x_g)$ as

$$\hat{f}(x_g) \equiv \left(n\sigma_n^k\right)^{-1} \sum_{j=1,j\neq g}^n K(\frac{x_j - x_g}{\sigma_n}) \ , \ \hat{h}(x_g) \equiv \left(n\sigma_n^k\right)^{-1} \sum_{j=1,j\neq g}^n y_{i,j} K(\frac{x_j - x_g}{\sigma_n})$$

where $K(.)$ is the kernel function and σ_n is the chosen smoothing parameter (the bandwidth). The nonparametric estimates for $p_i^*(x_g)$ is $\hat{p}(x_g) \equiv \hat{h}(x_g)/\hat{f}(x_g)$. Now construct the sample analog of $Q_i(b_i)$ for $i = 1, 2$:

$$\hat{Q}_{i,n}(b_i) = \frac{1}{n} \sum_{g=1}^n \Lambda(1/2 - \hat{p}_i(x_g))[x_g' b_i - \hat{p}_{-i}(x_g)]_+^2 + \Lambda(\hat{p}_i(x_g) - 1/2)[x_g' b_i - \hat{p}_{-i}(x_g)]_-^2$$

The two-step extreme estimator is defined as:

$$\hat{\Theta}_n = \arg\min_{\mathbf{b}\in\Theta^B} \hat{Q}_{1,n}(b_1) + \hat{Q}_{2,n}(b_2)$$

The estimator is set-valued in general, and the identification region Θ_{ID}^B is also set-valued in general. We shall show the estimator is consistent for the identification region Θ_{ID}^B under the Hausdorff metric. The metric between two sets A and B in $\mathbb{R}^K$ is defined as

$$d(A, B) \equiv \max\left\{\rho(A, B),\ \rho(B, A)\right\},\ where\ \rho(A, B) = \sup_{a\in A} \inf_{b\in B} ||a - b||$$

where $||.||$ is the Euclidean norm. Below we prove the two-step extreme estimator is consistent for the identification region Θ_{ID}^B under this metric. Regularity conditions for set consistency are collected below.

PAR (Parameter space) The identification region Θ_{ID}^B *is in the interior of a compact, convex parameter space* Θ^B.

RD (Regressors and disturbance) (i) the $(K + 1)$*-dimensional random vector* (X_g, ϵ_g) *is independently and identically distributed; (ii) the support of* X *(denoted* Ω_X*) is bounded, and its continuous coordinates have bounded joint density* $f_0(x_1, ., x_K)$*, and for both* $f_0(x)$ *and* $f_0(x)p_i^*(x)$ *are* m *times continuously differentiable on the interior of* Ω_X *with* $m > k$*; (iii)* $\Pr\{X'b_i = p_{-i}^*(X)\} = 0$ *for all* $b_i \in \Theta_i^B$ *and* $i = 1, 2$*; (iv)* $E\left[(X'b_i)^2\right] < \infty$ *for all* $b_i \in \Theta_i^B$.

KF (Kernel function) (i) $K(.)$ *is continuous and zero outside a bounded set; (ii)* $\int K(u)du = 1$ *and for all* $l_1 + .. + l_k < m$*,* $\int u_1^{l_1}...u_k^{l_k} K(u)du = 0$*; (iii)* $(\ln n)n^{-1/2}\sigma_n^{-K} \to 0$ *and* $\sqrt{n}(\ln n)\sigma_n^{2m} \to 0$.

SF (Smoothing functions) (i) $\Lambda : [-\frac{1}{2}, \frac{1}{2}] \to [0,1]$ *is such that* $\Lambda(c) = 0\ \forall c \leq 0$ *and* $\Lambda(c) > 0\ \forall c > 0$*; (ii)* Λ *is bounded with continuous, bounded first and second derivatives on the interior of the support.*

Condition (iii) in RD is a regularity condition for identification. Condition (i) in SF are also essential for the formulation of the identification region as the set of minimizers of Q. Conditions (i), (ii) in RD and the conditions in KF imply $\hat{p} \xrightarrow{p} p$ uniformly over Ω_X at a rate faster than $n^{-1/4}$, which, combined with smoothness property of Λ in condition (ii) of SF, contribute to the point-wise convergence of $\hat{Q}_n$ to Q in probability. The compactness of Θ and boundedness of Ω_X are technical conditions that make the integrand in the limiting function uniformly bounded over Θ.[5] These (weak) conditions ensure the sample analog $\hat{Q}_n$ converges in probability to Q pointwise. Given that $\hat{Q}_n$ is convex and continuous over the convex parameter space Θ^B, this point-wise convergence can be strengthened to uniform convergence over any compact subsets of Θ^B, which is the crucial condition for proving the consistency result below.

Proposition 2 *Suppose CMI, SE, PAR, RD, KF and SF are satisfied. Then (i)* $\hat{\Theta}_n$ *exists with probability approaching* 1 *and* $\Pr(\rho(\hat{\Theta}_n, \Theta^B_{ID}) > \varepsilon) \to 0$ *as* $n \to \infty$ *for all* $\varepsilon > 0$*; (ii) Suppose* $\sup_{b\in\Theta^B_{ID}} |\hat{Q}_n(b)| = O_p(a_n^{-1})$ *for some sequence of normalizing constants* $a_n \to \infty$ *and let* $\tilde{\Theta}_n = \{b \in \Theta^B : \hat{Q}_n(b) \leq \hat{c}/a_n\}$*, where* $\hat{c} \geq a_n\hat{Q}_n(b)$ *with probability approaching* 1 *and* $\hat{c}/a_n \xrightarrow{p} 0$*. Then* $\Pr(\rho(\Theta^B_{ID}, \tilde{\Theta}_n) > \varepsilon) \to 0$ *as* $n \to \infty$*.*

Proof of Proposition 2. First, we show $\sup_{b\in\Theta^B} |\hat{Q}_n(b) - Q(b)| \xrightarrow{p} 0$. To show this, it suffices to show $\sup_{b_i\in\Theta^B_i} |\hat{Q}_{i,n}(b_i) - Q_i(b_i)| = o_p(1)$. By Lemma 8.10 in Newey and McFadden (1994), under RD, TF and K, we have for $i = 1, 2$,

$$\sup_{x\in\Omega_X} |\hat{p}_i(x) - p^*_i(x)| = o_p(n^{-1/4}) \tag{5}$$

Apply a mean-value expansion of $\hat{Q}_{i,n}(b)$ around $p^*_{i,g} \equiv p^*_i(x_g)$:

$$\begin{aligned} \hat{Q}_{i,n}(b_i) \quad = \quad & \frac{1}{n}\sum\nolimits_{g=1}^{n} \left[\Lambda(1/2 - p^*_{i,g})(x'_g b_i - p^*_{-i,g})^2_+ + \Lambda(p^*_{i,g} - 1/2)(x'_g b_i - p^*_{-i,g})^2_-\right] + \\ & \frac{1}{n}\sum\nolimits_{g=1}^{n} \left\{ \begin{bmatrix} \tilde{\Lambda}^{(1)}_{i,g,-}(x'_g b_i - \tilde{p}_{-i,g})^2_- - \tilde{\Lambda}^{(1)}_{i,g,+}(x'_g b_i - \tilde{p}_{-i,g})^2_+ \\ -2\tilde{\Lambda}_{i,g,+}(x'_g b_i - \tilde{p}_{-i,g})_+ - 2\tilde{\Lambda}_{i,g,-}(x'_g b_i - \tilde{p}_{-i,g})_- \end{bmatrix}' \begin{bmatrix} \hat{p}_{i,g} - p^*_{i,g} \\ \hat{p}_{-i,g} - p^*_{-i,g} \end{bmatrix} \right\} \end{aligned} \tag{6}$$

where $\hat{p}_{i,g} \equiv \hat{p}_i(x_g)$ and $\tilde{p}_{i,g}$ is a shorthands for some point on the line segments between $(p^*_{i,g}, \hat{p}_{i,g})$, and $\tilde{\Lambda}_{i,g,+} \equiv \Lambda(1/2 - \tilde{p}_{i,g})$, $\tilde{\Lambda}_{i,g,-} \equiv \Lambda(\tilde{p}_{i,g} - 1/2)$, $\tilde{\Lambda}^{(1)}_{i,g,+} \equiv \Lambda'(1/2 - \tilde{p}_{i,g})$, $\tilde{\Lambda}^{(1)}_{i,g,-} \equiv$

[5] This may be stronger than necessary for consistency, as $\hat{Q}_n \xrightarrow{p} Q$ point-wise in Θ^B is needed.

$\Lambda'(\tilde{p}_{i,g} - 1/2)$. Define the first term in (6) by $\bar{Q}_{i,n}(b)$. Then by triangular inequality, for all $b_i \in \Theta_i^B$,

$$|\hat{Q}_{i,n}(b_i) - Q_i(b_i)| = |\hat{Q}_{i,n}(b_i) - \bar{Q}_{i,n}(b_i)| + |\bar{Q}_{i,n}(b_i) - Q_i(b_i)|$$

where the second term is $O_p(n^{-1/2})$ by the Central Limit Theorem. An application of the triangular inequality suggests the first term is bounded above by

$$\begin{bmatrix} \frac{1}{n}\sum_{g=1}^n |\tilde{\Lambda}^{(1)}_{i,g,-}(x_g'b_i - \tilde{p}_{-i,g})_-^2 - \tilde{\Lambda}^{(1)}_{i,g,+}(x_g'b_i - \tilde{p}_{-i,g})_+^2| \\ \frac{1}{n}\sum_{g=1}^n | - 2\tilde{\Lambda}_{i,g,+}(x_g'b_i - \tilde{p}_{-i,g})_+ - 2\tilde{\Lambda}_{i,g,-}(x_g'b_i - \tilde{p}_{-i,g})_-| \end{bmatrix}' \begin{bmatrix} \sup_{x\in\Omega_X} |\hat{p}_i - p_i^*| \\ \sup_{x\in\Omega_X} |\hat{p}_{-i} - p_{-i}^*| \end{bmatrix} \quad (7)$$

Note the absolute values of $\tilde{\Lambda}^{(1)}_{i,g,-}$, $\tilde{\Lambda}^{(1)}_{i,g,+}$, $\tilde{\Lambda}_{i,g,+}$ and $\tilde{\Lambda}_{i,g,-}$ are all bounded under SF, and

$$\frac{1}{n}\sum_{g=1}^n |(x_g'b_i - \tilde{p}_{-i,g})_-^2| \leq \frac{1}{n}\sum_{g=1}^n (x_g'b_i)^2 + 1 = O_p(n^{-1/2})$$

where the last equality follows from condition (iv) in RD and the Central Limit Theorem. Likewise we can show the first term in the product in (7) is $O_p(n^{-1/2})$. Since the second term of (7) is $o_p(n^{-1/4})$, the product in (7) is $o_p(n^{-3/4})$. Hence we have $|\hat{Q}_{i,n}(b_i) - Q_i(b_i)| = O_p(n^{-1/2})$ pointwise for all $b_i \in \Theta_i^B$. Note $\hat{Q}_{i,n}(b_i)$ is continuous and convex in b_i over Θ_i^B for all n. Convexity is preserved by pointwise limits, and hence Q_i is also convex and therefore continuous on the interior of Θ_i^B. Furthermore, by Andersen and Gill (1982) (and Theorem 2.7 in Newey and McFadden (1994)), the convergence in probability of $\hat{Q}_{i,n}(b_i)$ to $Q_i(b_i)$ must be uniform over Θ_i^B for $i = 1, 2$. It then follows $\sup_{b\in\Theta^B} |\hat{Q}_n(b) - Q(b)| \xrightarrow{p} 0$. The rest of the proof follows from arguments in Proposition 3 in Manski and Tamer (2002) and Theorem 3.1 in Chernozhukov, Hong and Tamer (2007), and is omitted for brevity. Q.E.D.

An obvious advantage of the estimator is that the objective function in the second step is a convex function in coefficients. The introduction of the sequence $\hat{c}/a_n$ in the definition of $\tilde{\Theta}_n$ in part *(ii)* is necessary for the more general case where Θ_{ID}^B is not a singleton. The perturbed estimator $\tilde{\Theta}_n$ is consistent for non-singleton Θ_I in Hausdorff metric. A possible choice of the pair $\hat{c}$ and a_n is $\log n$ and $n^{3/4}$ respectively.[6] A direction of future research will be to find regularity conditions on the joint distribution of (X, ϵ), and functions Λ, so that $\hat{Q}_n$ satisfies conditions for existence of polynomial minorant in Chernozhukov, Hong and Tamer (2007) and the rate of convergence can be derived.

[6] To see why a_n can be chosen to be $n^{3/4}$, note the first term in (6) is 0 for all $b \in \Theta_{ID}^B$ and the two terms in the product in (7) are $O_p(n^{-1/2})$ and $o_p(n^{-1/4})$ under the conditions stated.

4 Conclusion

In this paper, I estimate a simultaneous discrete game with incomplete information where players' private information is allowed to be correlated with observed states. Under the weak restriction of median independence of privation information, I characterize the identification region of the linear coefficients, and give sufficient conditions for the coefficient to be point identified. I propose a two-step extreme estimator and prove its consistency.

5 Appendix: Proofs of Lemmas

Proof of Lemma 1. *(Sufficiency)* Fix the BNE outcome p^* observed. Suppose $\theta_i \overset{u.o.e.}{\sim} \theta_i^0$ under $\mathcal{F}_{MI}^i$ for $i = 1, 2$. By definition $\exists \bar{F}_{\epsilon_i|X} \in \mathcal{F}_{MI}^i$ such that $\Pr\{p_i^*(X) = \bar{F}_{\epsilon_i|X}(X'\beta_i + p_{-i}^*(X)\delta_i)\} = 1$ for $i = 1, 2$. Hence $\Pr(p^*(X) \in \psi(X; \theta, \bar{F}_{\epsilon|X})) = 1$ where $\bar{F}_{\epsilon|X} \equiv \prod_{i=1,2} \bar{F}_{\epsilon_i|X} \in \mathcal{F}_{CMI}$, and $\theta \overset{o.e.}{\sim} \theta^0$ under $\mathcal{F}_{CMI}$. *(Necessity)* That $\theta \overset{o.e.}{\sim} \theta^0$ under $\mathcal{F}_{CMI}$ implies $\exists \bar{F}_{\epsilon|X} \in \mathcal{F}_{CMI}$ such that $\Pr\{p^*(X) \in \psi(X; \theta, \bar{F}_{\epsilon|X})\} = 1$. It follows that $\Pr\{p_i^*(X) = \bar{F}_{\epsilon_i|X}(X'\beta_i + p_{-i}^*(X)\delta_i)\}$ $= 1$ for $i = 1, 2$, where $\bar{F}_{\epsilon_i|X}$ are marginal distributions corresponding to $\bar{F}_{\epsilon|X}$. By definition, this means $\theta_i \overset{u.o.e.}{\sim} \theta_i^0$ under $\mathcal{F}_{MI}^i$ for both $i = 1, 2$. Q.E.D.

Proof of Lemma 2. (Necessity) Suppose θ is such that $\Pr(X \in \xi_{1,\theta_1}) > 0$. By definition for all $x \in \xi_{1,\theta_1}$ and all $G_{\epsilon_1|X} \in \mathcal{F}_{MI}^1$, either

$$\psi_1(x; \theta_1, G_{\epsilon_1|X}, p_2^*) \leq 1/2 < p_1^*(x)$$

or

$$\psi_1(x; \theta_1, G_{\epsilon_1|X}, p_2^*) \geq 1/2 > p_1^*(x)$$

Hence for all $x \in \xi_{1,\theta_1}$, $p^*(x) \notin \psi(x, \theta, G_{\epsilon|X})$ for any $G_{\epsilon|X} \in \mathcal{F}_{CMI}$. Therefore $\Pr(X \in \xi_{1,\theta_1}) > 0$ implies $\Pr\{X \in \chi(p^*, \theta, G_{\epsilon|X})\} < 1$ for any $G_{\epsilon|X} \in \mathcal{F}_{CMI}$ and θ is not observationally equivalent to θ^0 under $\mathcal{F}_{CMI}$. (Sufficiency) Suppose θ is such that $\Pr(X \in \cup_{i=1,2}\xi_{i,\theta_i}) = 0$. Let Σ^c denote the complement of a generic subset Σ of the support Ω_X. Then $\Pr(X \in \cap_{i=1,2}\xi_{i,\theta_i}^c) = 1$, where by definition $\xi_{i,\theta_i}^c = \{x \in \Omega_X$ such that "$\tilde{x}_i'\theta_i > 0$ and $p_i^*(x) \geq \frac{1}{2}$" or "$\tilde{x}_i'\theta_i < 0$ and $p_i^*(x) \leq \frac{1}{2}$" or "$\tilde{x}_i'\theta_i = 0$ and $p_i^*(x) = \frac{1}{2}$"$\}$. Then for all $x \in \Omega_X$, we can always construct $F_{\epsilon|x} \in \mathcal{F}_{CMI}$ such that $p_i^*(x) = F_{\epsilon_i|x}(x'\beta_i + p_{-i}^*(x)\delta_i)$ for both $i = 1, 2$. (Convexity) Suppose both θ and θ' are in the identification region of θ^0 under $\mathcal{F}_{CMI}$. That is,

$$\Pr(X \in \cup_{i=1,2}\xi_{i,\theta_i}) = \Pr(X \in \cup_{i=1,2}\xi_{i,\theta_i'}) = 0$$

Let $\theta^{\alpha} \equiv \alpha\theta + (1-\alpha)\theta'$ with $\theta_i^{\alpha} \equiv \alpha\theta_i + (1-\alpha)\theta_i'$. Define $\xi_{i,\theta_i^{\alpha}}$ to be a subset of Ω_X such that "$\tilde{x}_i'\theta_i^{\alpha} \leq 0$ *and* $p_i^*(x) > \frac{1}{2}$" or "$\tilde{x}_i'\theta_i^{\alpha} \geq 0$ *and* $p_i^*(x) < \frac{1}{2}$"}. Suppose $\exists x \in \xi_{i,\theta_i^{\alpha}}$ such that the first of the two events occur. This implies that either "$\tilde{x}_i'\theta_i \leq 0$ *and* $p_i^*(x) > \frac{1}{2}$" or "$\tilde{x}_i'\theta_i' \leq 0$ *and* $p_i^*(x) > \frac{1}{2}$" occurs. Hence either $x \in \xi_{i,\theta_i}$ or $x \in \xi_{i,\theta_i'}$. Symmetric arguments apply to show that $x \in \xi_{i,\theta_i} \cup \xi_{i,\theta_i'}$ if the other event happens. It follows $\xi_{i,\theta_i^{\alpha}} \subseteq \xi_{i,\theta_i} \cup \xi_{i,\theta_i'}$ for $i = 1, 2$. Therefore $\cup_{i=1,2}\xi_{i,\theta_i^{\alpha}} \subseteq (\cup_{i=1,2}\xi_{i,\theta_i}) \cup (\cup_{i=1,2}\xi_{i,\theta_i'})$ implies $\Pr(X \in \cup_{i=1,2}\xi_{i,\theta_i^{\alpha}}) = 0$, and θ^{α} is in the identification region of θ^0 under $\mathcal{F}_{CMI}$. Q.E.D.

Proof of Lemma 3. We prove the lemma from the perspective of player 1. The proof for the case of player 2 follows from the same argument. Fix $\bar{x}_{-h_1} \in \Omega_{X_{-h_1}}$. By the definition of a BNE and the assumption that p^* is rationalized by a single equilibrium only (i.e. assumption *SE*), we have

$$\begin{bmatrix} p_1^*(\bar{x}_{-h_1}, x_{h_1}) \\ p_2^*(\bar{x}_{-h_1}, x_{h_1}) \end{bmatrix} = \begin{bmatrix} F_{\epsilon_1|\bar{x}_{-h_1}, x_{h_1}}(\bar{x}_{-h_1}\beta_{1,-h_1}^0 + x_{h_1}\beta_{1,h_1}^0 + \delta_1^0 p_2^*(\bar{x}_{-h_1}, x_{h_1})) \\ F_{\epsilon_2|\bar{x}_{-h_1}, x_{h_1}}(\bar{x}_{-h_1}\beta_{2,-h_1}^0 + x_{h_1}\beta_{2,h_1}^0 + \delta_2^0 p_1^*(\bar{x}_{-h_1}, x_{h_1})) \end{bmatrix} \tag{8}$$

where $\beta_{1,h_1}^0 = 0$ under the exclusion restriction. Let $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$ denote the space of bounded, continuous functions on the compact support $\Omega_{X_{h_1}|\bar{x}_{-h_1}}$ under the sup-norm. By standard arguments, $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$ is a Banach Space. Define $C^{K_1}(\bar{x}_{-h_1})$ as a subset of functions in $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$ that map from $\Omega_{X_{h_1}|\bar{x}_{-h_1}}$ to $[0,1]$, and are Lipschitz continuous with some constant $k \leq K_1$. Then $C^{K_1}(\bar{x}_{-h_1})$ is bounded in the sup-norm and equi-continuous due to the Lipschitz continuity. Besides, $C^{K_1}(\bar{x}_{-h_1})$ is also closed in $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$. To see this, consider a sequence f_n in $C^{K_1}(\bar{x}_{-h_1})$ that converges in sup-norm to f_0. By the completeness of $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$, $f_0 \in C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$. Now suppose $f_0 \notin C^{K_1}(\bar{x}_{-h_1})$. Then $\exists x_{h_1}^a$, $x_{h_1}^b \in \Omega_{X_{h_1}|\bar{x}_{-h_1}}$ such that $|f_0(x_{h_1}^a) - f_0(x_{h_1}^b)| > K_1'|x_{h_1}^a - x_{h_1}^b|$ for some $K_1' > K_1$. By convergence of f_n, for all $\varepsilon > 0$, $|f_n(x_{h_1}^j) - f_0(x_{h_1}^j)| \leq \frac{\varepsilon}{2}|x_{h_1}^a - x_{h_1}^b|$ for $j = a, b$ for n big enough. Hence $\frac{|f_n(x_{h_1}^a) - f_n(x_{h_1}^b)|}{|x_{h_1}^a - x_{h_1}^b|} > K_1' - \varepsilon$ for n big enough. For any $\varepsilon < K_1' - K_1$, this implies for n big enough, f_n is not Lipschitz continuous with $k \leq K_1$. Contradiction. Hence $C^{K_1}(\bar{x}_{-h_1})$ is bounded, equi-continuous, and closed in $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$. By the Arzela-Ascoli Theorem, $C^{K_1}(\bar{x}_{-h_1})$ is a convex, compact subset of the normed linear space $C(\Omega_{X_{h_1}|\bar{x}_{-h_1}})$. Now substitute the second equation in (8) into the first one, and we have

$$\bar{p}_1(x_{h_1}) = \bar{F}_{\epsilon_1|x_{h_1}}\{\bar{x}_{-h_1}\beta_{1,-h_1}^0 + \delta_1^0 \bar{F}_{\epsilon_2|x_{h_1}}[\bar{x}_{-h_1}\beta_{2,-h_1}^0 + x_{h_1}\beta_{2,h_1}^0 + \delta_2^0 \bar{p}_1(x_{h_1})]\} \tag{9}$$

where $\bar{p}_1(x_{h_1})$ and $\bar{F}_{\epsilon_i|x_{h_1}}$ are shorthands for $p_1^*(x_{h_1}, \bar{x}_{-h_1})$ and $F_{\epsilon_i|x_{h_1}, \bar{x}_{-h_1}}$ conditional on $\bar{x}_{-h_1}$. Note the strategic interaction terms δ_i^0 are already normalized to have absolute value 1 under SN. Fix a x_{-h_1} and let $\bar{\tau}(x_{h_1})$ denote the right-hand side of (9). Suppose $\bar{p}_1(x_{h_1})$ is Lipschitz continuous with constant $k \leq K_1$ for some $K_1 > 0$. Then by the definition of the Lipschitz constants in *PID (i)-(ii)*, for all x_{h_1}, $\tilde{x}_{h1} \in \Omega_{X_{h_1}|\bar{x}_{-h_1}}$, $|\bar{\tau}(x_{h_1}) - \bar{\tau}(\tilde{x}_{h_1})| \leq D(K_1)|x_{h_1} - \tilde{x}_{h_1}|$, where

$$D(K_1) \equiv K^1_{F_1} + C_{F_1}[K^1_{F_2} + C_{F_2}(|\beta^0_{2,h_1}| + K_1)]$$

Since $\beta^0_{2,h_1} \neq 0$ and $|C_{F_1}C_{F_2}| < 1$, K_1 can be chosen such that $D(K_1) \leq K_1$. Therefore the right hand side of (9) is a continuous self-mapping from $C^{K_1}(\bar{x}_{-h_1})$ to $C^{K_1}(\bar{x}_{-h_1})$ for the K_1 chosen. It follows from Schauder's Fixed Point Theorem that there exists a solution $p^*_1(x_{h_1},\bar{x}_{-h_1})$ that is continuous in x_{h_1} for all $\bar{x}_{-h_1} \in \Omega_{X_{-h_1}}$.[7] Q.E.D.

Proof of Lemma 4. We prove the case for player $i = 1$. Fix $\bar{x}_{-h_1} \in \Omega_{X_{-h_1}}$. Then for all $x_{h_1} \in \Omega_{X_{h_1}|\bar{x}_{-h_1}}$,

$$\bar{p}_2(x_{h_1}) = \bar{F}_{\epsilon_2|x_{h_1}}[\bar{x}_{-h_1}\beta^0_{2,-h_1} + x_{h_1}\beta^0_{2,h_1} + \delta^0_2\bar{p}_1(x_{h_1})] \quad (10)$$

where $\bar{p}_i$ and $\bar{F}_{\epsilon_2|x_{h_1}}$ are shorthands for p^*_i and $F^0_{\epsilon_2|x_{h_1},\bar{x}_{-h_1}}$ for a fixed $\bar{x}_{-h_1}$, and δ^0_2 is normalized to have absolute value 1 under SN. Lemma 3 showed $\bar{p}_1(x_{h_1})$ is Lipschitz continuous in x_{h_1} given any $\bar{x}_{-h_1}$, and therefore the support of $\bar{p}_1(X_{h_1})$ given $\bar{x}_{-h_1}$ must be a connected interval contained in $[0, 1]$. By the conditions on $F_{\epsilon_i|X}$ in Lemma 3, $\bar{p}_2(x_{h_1})$ must also be Lipschitz continuous. Then *PID-2* and *RSX-1* imply the support of $\bar{p}_2(x_{h_1})$ is a subset of the open interval $(0, 1)$ that must cover $[a^2, b^2]$ for any $\bar{x}_{-h_1} \in \Omega_{X_{-h_1}}$, and $\Pr\{p^*_2(X) \in I|\bar{x}_{-h_1}\} > 0$ for all open interval $I \subset [a^2, b^2]$ and $\bar{x}_{-h_1} \in \Omega_{X_{-h_1}}$. Q.E.D.

Proof of Lemma 5. By construction, $Q(b)$ is non-negative $\forall b = (b_1, b_2) \in \Theta^B$. For $i = 1, 2$, by the law of total probability and the regularity condition that $\Pr(p^*_{-i}(X) = X'b_i) = 0$ for all $b_i \in \Theta^B_i$ (note under the normalization that when $\delta_i = -1$, the latter implies $\Pr(p^*_i(X) = 1/2) = 0$), we have

$$\begin{aligned} Q_i(b_i) &= E[\Lambda(1/2 - p^*_i(X))(X'b_i - p^*_{-i}(X))^2_+ + \Lambda(p^*_i(X) - 1/2)(X'b_i - p^*_{-i}(X))^2_-] \\ &= E[\Lambda(1/2 - p^*_i(X))(X'b_i - p^*_{-i}(X))^2_+ \mid p^*_i(X) < 1/2] \Pr(p^*_i(X) < 1/2) \\ + \quad & E[\Lambda(p^*_i(X) - 1/2)(X'b_i - p^*_{-i}(X))^2_- \mid p^*_i(X) > 1/2] \Pr(p^*_i(X) > 1/2) \end{aligned}$$

By definition for all $b = (b_1, b_2) \in \Theta^B_{ID}$, the following events must have zero probability for $i = 1, 2$,

$$\text{"}X'b_i \leq p^*_{-i}(X) \wedge p^*_i(X) > 1/2\text{" } \textit{or} \text{ "}X'b_i \geq p^*_{-i}(X) \wedge p^*_i(X) < 1/2\text{"} \quad (11)$$

Therefore $Q(b) = 0$ for all $b \in \Theta^B_{ID}$. On the other hand, for any $b \notin \Theta^B_{ID}$, at least one of the two events above must have positive probability for either $i = 1$ or 2. Without loss of

[7] Note here we have extended the SE assumption to restrict the single equilibrium played in the DGP to be from such Lipchitz-continuous Bayesian Nash equilibria.

generality, let the first event in (11) occur with positive probability for $i = 1$. Then that "$\Pr\{X'b_1 = p_2^*(X)\} = 0$ for all $b_1 \in \Theta_1^B$" implies $\Pr\{X'b_1 < p_2^*(X) \wedge p_1^*(X) > 1/2\} > 0$, which implies the second term in $Q_1(b_1)$ is strictly positive. Similar arguments can be applied to prove the first term in $Q_1(b_1)$ is strictly positive if the second event in (11) has positive probability for $i = 1$. The case with $i = 2$ follows from the same arguments. Hence $Q(b) > 0$ if and only if $b \notin \Theta_{ID}^B$. Q.E.D.

References

[1] Aradillas-López, A. (2005): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information", Working Paper, Princeton University

[2] Bajari, P., Hong, H., Krainer, J. and D. Nekipelov (2007): "Estimating Static Models of Strategic Interactions", Working Paper, University of Minnesota

[3] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and Inference in Partially Identified Models", Econometrica, Vol. 70, No. 5,1243-1284

[4] Klein, R. and R. Spady (1993): "An Efficient Semiparametric Estimator of Binary Response Models," Econometrica, 61, 387-421

[5] Manski, C. (1988): "Identification of Binary Response Models ", Journal of American Statistical Association, Vol 83, No. 403, p729-738

[6] Manski, C. and Tamer, E.(2002): "Inference on Regressions with Interval Data on a Regressor or Outcome", Econometrica, 70, 519-547

[7] Newey, W.K. and McFadden, D (1994): "Large Sample Estimation and Hypothesis Testing", Handbook of Econometrics, Volume IV, Chapter 36, Elsevier Science