SEMIPARAMETRIC ANALYSIS OF BINARY GAMES OF INCOMPLETE INFORMATION*

PAUL NIANQING LIU† AND HAIQING XU‡

[Preliminary]

ABSTRACT. This paper studies the identification and estimation in an I–player binary game of incomplete information. Our approach allows players' type to be correlated across players. By focusing on the monotone pure strategy Bayesian Nash Equilibrium (BNE), we show that the equilibrium strategies can be represented as a single–agent binary response model. Under weak restrictions, we show that the distribution of incomplete information can be nonparametrically identified. Further, we establish the identification of payoff functions in a linear–index setup. Following Klein and Spady (1993), we propose a three–stage estimation procedure and show that our estimator is $\sqrt{n}$–consistent, asymptotically normally distributed.

Keywords: Bayesian Nash Equilibrium, Discrete game, Incomplete information, Monotone strategy

Date: Thursday 29th November, 2012.

*We gratefully acknowledge Quang Vuong for his guidance and advice. We have greatly benefited from the discussions with Shakeeb Khan, Robin Sickles, Xun Tang, and Yuanyuan Wan. We also thank Jason Abrevaya, Bill Brown, Stephen Donald, Sukjin Han, Stefan Hoderlein, Brendan Kline, Natalia Sizova, and seminar participants at UT Austin and the Rice University for providing helpful comments.

† Department of Economics, The Pennsylvania State University, 408 Kern Graduate Building, University Park, PA, 16802, nul130@psu.edu.

‡Department of Economics, The University of Texas at Austin, BRB 3.160, Mailcode C3100, Austin, TX, 78712, h.xu@austin.utexas.edu.

1. Introduction

In this paper, we study the identification and estimation of static binary games of incomplete information with correlated private information (i.e. types). The range of applications of binary games includes, among others, models of entry (Bresnahan and Reiss, 1990, 1991; Berry, 1992; Jia, 2008; Seim, 2006), couples' retirement decisions (Banks, Blundell, and Casanova Rivas, 2010; Casanova, 2010), labor force participation (Bjorn and Vuong, 1984; Soetevent and Kooreman, 2007)), stock market analysts' recommendations (Bajari, Hong, Krainer, and Nekipelov, 2010), advertising (Sweeting, 2009), and social interactions (Brock and Durlauf, 2001a,b; Xu, 2011), etc.

To simply our exposition, we formally consider throughout this paper the equilibrium solution that can be represented by the following structural equations (i.e., best responses): for $i = 1, \cdots, I$,

$$Y_i = \mathbf{1}\Big\{ X_i'\beta_i + \sum_{j \neq i} \mathbb{P}(Y_j = 1 | X, U_i) - U_i \geq 0 \Big\}, \tag{1}$$

where subscript i is an index of players in the game; X_i is a vector of exogenous payoff relevant variables, while the error term U_i is i's private information, which is not observed by other players; We allow $U = (U_1, \cdots, U_I)$ to be correlated with each other under an unknown form. This model is a natural extension of Manski (1975, 1985)'s binary threshold crossing model in the single–agent setup to a structural model with strategic interactions.

This paper contributes to the existing discrete game literature in several respects. First, we do not require the (conditional) independence of private payoff shocks across players, which is widely adopted by most of the literature, e.g., Aguirregabiria and Mira (2007); Bajari, Hong, Krainer, and Nekipelov (2010); De Paula and Tang (2010); Grieco (2011); Pesendorfer and Schmidt-Dengler (2003) and Lewbel and Tang (2011) do; exceptions include Aradillas-Lopez (2010); Wan and Xu (2010) and Xu (2010).[1]

[1]The novel approach developed in Aradillas-Lopez (2010) assumes that players do not have exact knowledge about the distributions involved and then using an equilibrium concept defined in Aumann (1987).

Allowing correlated private signals is motivated primarily by empirical concerns. The (conditional) independence assumption of U is convenient but meanwhile imposes strong restrictions — players' choices must be conditionally independent, which could be invalidated by the data.[2] Moreover, in the social interaction framework, the correlation among players' private payoff shock represents the "homophily" effects in social behaviors, which is caused by the unobserved "similarity" in players' preference. In contrast, the peer effects is purely the strategic effects caused by interactions with other group members. Both effects accounting for the "herding" behavior in a society group can be identified and distinguished with each other in our model.

Second, we make no parametric assumptions on the joint distribution of private payoff shocks, which distinguish our paper from Xu (2010). We establish nonparametric identification results for the copula function of private payoff shocks, from which we can derive equilibrium belief function. In a similar semiparametric setup, Wan and Xu (2010) establish partial identification of payoff coefficients when types are positively regression dependent, and further achieve point identification under an additional support condition on regressors. The maximum score type estimator they suggested converges at $\sqrt[3]{n}$–rate. In this paper, we establish point identification of structural parameters under weak conditions. Moreover, the Klein–Spady type estimator we propose in this paper is $\sqrt{n}$–consistent.

The key in our semiparametric identification approach is to focus on the class of monotone pure strategy BNEs. Athey (2001) provided the seminar result that a monotone pure–strategy BNE exists whenever a Bayesian game obeys a Spence–Mirlees single–crossing restriction. McAdams (2003) and Reny (2011) extends Athey (2001)'s results. Applying Reny (2011) in our setup, we show that a monotone strategy BNE generally exists under weak conditions.

Third, we propose a Klein–Spady type pseudo maximum likelihood estimator for the structural parameter, which is shown to be $\sqrt{n}$–consistent. In the proposed estimation procedure, we estimate the belief component nonparmaetrically. Then, following Klein

[2]A model featured with unobserved heterogeneity and independent private signals also generates dependence among players' choices conditional on observed regressors (see Grieco, 2011).

and Spady (1993), we construct a pseudo loglikelihood function using the estimated beliefs as part of covariates. Monte–Carlo evidence indicates that there is only modest efficiency losses relative to the semiparametric estimation when the belief component is known to researchers.

The rest of the paper is arranged as follows. We introduce the setup of our game model in Section 2 and establish the existence for monotone pure strategy BNE in Section 3. Further, We discuss the semiparametric identification of the structural model in Section 4. In Section 5, we propose a Klein–Spady type estimator in a two–player setup. Section 6 provides Monte–Carlo simulations.

2. Model

We consider a static binary game of incomplete information, commonly referred to as a Discrete Bayesian game. There are a finite number of players, indexed by $i \in \mathcal{I} \equiv \{1, 2, \cdots, I\}$, and each player i simultaneously chooses an action $Y_i \in \{0, 1\}$.[3] Define $\mathcal{A} = \{0, 1\}^I$ as the action space of the game and let $y = (y_1, \cdots, y_I) \in \mathcal{A}$ be a generic element of $\mathcal{A}$. Following the convention, let $\mathcal{A}_{-i}$ and y_{-i} denote the action space and a profile of actions for all players but excluding player i, respectively.

For each player i, $X_i \in \mathbb{R}^{d_i}$ is a vector of payoff relevant random variables, which are publicly observed by all players. Define $X = (X_1, \cdots, X_I) \in \mathbb{R}^p$, where $p = \sum_{i=1}^{I} d_i$, as all the publicly observed information in the game. Player i's payoff shock U_i is i's private information, which is not observed by other players. Let $U = (U_1, \cdots, U_I)$ and F_{XU} be the c.d.f. of (X, U). The joint distribution F_{XU} is assumed to be common knowledge to all players.

The payoff for player i is described as follows,

$$\pi_i(y, x_i, u_i) = \begin{cases} x_i'\beta_i + \sum_{j \neq i} \alpha_{ij} y_j - u_i, & \text{if } y_i = 1, \\ 0, & \text{if } y_i = 0, \end{cases}$$

[3]For notational simplicity, we restrict players to make binary decisions and all of our results could be generalized to the case where the choice set for each player is finite, which is briefly discussed in the section 7.

where $\beta_i \in \mathbb{R}^{d_i}$ and $\alpha_{ij} \in \mathbb{R}$ $(i \neq j)$ are the parameters of interest. α_{ij} $(j \neq i)$ are strategic interaction parameters, which measures the *ceteris paribus* effects on i's payoff from j's choice. Our payoff function here is similar to the parametric case in Bajari, Hong, Krainer, and Nekipelov (2010).[4] The zero payoff for action $y_i = 0$ is a standard way of normalization.

Regarding to the payoff shock U, departing from the static discrete game literature (e.g., Bajari, Hong, Krainer, and Nekipelov, 2010), our analysis involves neither (conditional) independence restrictions between U_i and U_j nor parametric assumptions; only exceptions include Aradillas-Lopez (2010), Liu, Vuong, and Xu (2012), and Wan and Xu (2010).

Following the literature on Bayesian games, player i's decision rule is a function $Y_i = s_i(X, U_i)$, where $s_i : \mathscr{S}_X \times \mathbb{R} \to \{0,1\} \in \Delta_i$ maps all the information that i knows to a binary response and Δ_i is the strategy space of i. Note that X_{-i} also enters player i's decision rule s_i, since the opponents' decisions have effects on i's response through the strategic interactions.

Fix $x \in \mathscr{S}_X$. For any strategy profile $s = (s_1, \cdots, s_I) \in \bigtimes_i \Delta_i$ and $j \neq i$, we let $\sigma^s_{ij}(x, u_i)$ be the conditional probability $\mathbb{P}\left\{s_j(X, U_j) = 1 | X = x, U_i = u_i\right\}$, i.e.,

$$\sigma^s_{ij}(x, u_i) = \int_{\mathbb{R}} \mathbf{1}\left\{s_j(x, v) = 1\right\} f_{U_j|X,U_i}(v|x, u_i) dv$$

where $\mathbf{1}[\cdot]$ is the indicator function and $f_{U_j|X,U_i}$ is the conditional probability density function of U_j given X and U_i. Hence, the term $\sigma^s_{ij}(x, u_i)$ is player i's belief on the event $Y_j = 1$, given i's information (x, u_i) and the specified decision rule s.

The equilibrium concept we adopt is the pure strategy Bayesian Nash equilibrium (BNE). Similar to Bajari, Hong, Krainer, and Nekipelov (2010), the mixed strategy equilibrium is not considered hereafter, since with probability one, each player has a unique best response. Let $s^* = (s^*_1, \cdots, s^*_I)$ is the equilibrium strategy profile and $\sigma^*_{ij}(\cdot, \cdot)$ is a short notation for $\sigma^{s^*}_{ij}(\cdot, \cdot)$. In equilibrium, player i's equilibrium strategy satisfies a "mutual consistency"

[4] Aradillas-Lopez (2010), Lewbel and Tang (2011), and Wan and Xu (2010), among others, have also studied binary games with the same payoff structure but under a two–player framework.

requirement, i.e.

$$s_i^*(x, u_i) = \mathbf{1}\left[x_i'\beta_i + \sum_{j \neq i} \alpha_{ij}\sigma_{ij}^*(x, u_i) - u_i \geq 0\right]. \tag{2}$$

Equation (2) are indeed a simultaneous equation system, since player i's equilibrium beliefs σ_{ij}^* on the right hand depend on $s_j^*(x, \cdot)$, and vice versa. Therefore, s^* is defined as a fixed point to eq. (2). Although ensuring equilibrium existence in Bayesian games is a complex and deep subject in the literature, it is well known that a solution of such an equilibrium generally exists in a broad class of Bayesian games including the binary game under discussion (see, e.g., Vives, 1990).

3. MONOTONE PURE STRATEGY BNE

Monotone pure strategy BNEs, in which equilibrium strategies are monotone functions in private signals, are desirable in many applications in auction, entry, social interactions and global games for example. The seminar work on the existence of a monotone pure strategy BNE in games of incomplete information was provided Athey (2001) in both *supermodular* and *logsupermodular* games, and later extended by McAdams (2003) and Reny (2011).

To apply Theorem 4.1 in Reny (2011), we make the following assumption.

Assumption A. *Let the conditional distribution of U given X be absolutely continuous w.r.t. the Lebesgue measure and have positive and continuous conditional Radon–Nikodym densities $f_{U|X}$ a.e. over $\mathbb{R}^I$.*

Assumption A requires the conditional c.d.f. function $F_{U|X}$ to be twicely differentiable and have a full support on the Euclidean space.

Assumption B (Monotone Best Response Functions)**.** *For all $x \in \mathscr{S}_X$, $i \in \mathcal{I}$, and $v \in \mathbb{R}^I$, we have* $1 - \sum_{j \neq i}\left\{\alpha_{ij} \times \partial F_{U_j|X,U_i}(v_j|x, v_i)/\partial u_i\right\} \geq 0$.

Note that Assumption B is trivially satisfied if U are mutually independent. Assumption B also holds if $\alpha_{ij} \leq 0$ and U_i and U_j are positively regression dependent for all $i \neq j$.

Lemma 1. *Suppose that Assumptions A and B hold, then there exists at least one monotone pure strategy BNE in our binary discrete games.*

Proof. See Lemma 1 in Liu, Vuong, and Xu (2012). □

It should be noted that we are silent about the existence of non–monotone strategy BNEs under Assumption B in Lemma 1. Xu (2010) shows that non–monotone strategy BNEs can be ruled out under further restrictions on the correlation between private signals. Lemma 1 does not ensure either the uniqueness of monotone pure strategy BNE. Throughout our analysis, we assume that under Assumption B, only one monotone pure strategy BNE is played.

With a monotone pure strategy BNE, player i's equilibrium strategy is a weakly monotone functions of her private signal and can be characterized by a threshold function, i.e., fix $x \in \mathscr{S}_X$,

$$s_i^*(x, u_i) = \mathbf{1}\{u_i \leq u_i^*(x)\},$$

where $u_i^* : \mathscr{S}_X \to \mathbb{R}$. Further, the mutual consistency condition for BNEs requires that for all i

$$u_i \leq u_i^*(x) \Longleftrightarrow x_i'\beta_i + \sum_{j \neq i} \alpha_{ij} \times F_{U_j|X,U_i}\big(u_j^*(x)|x, u_i\big) - u_i \geq 0.$$

In a monotone pure strategy BNE, we can represent the equilibrium strategies as a semi–linear–index binary response model. For all $x \in \mathscr{S}_X$, let $\varphi_{ij}(x) = F_{U_j|X,U_i}\big(u_j^*(x)|x, u_i^*(x)\big)$ and $P_{ij} = \varphi_{ij}(X)$. Let further $P_i = [P_{ij}]_{j \neq i}$ and $\alpha_i = [\alpha_{ij}]_{j \neq i}$ be the $I-1$–dimensional random and deterministic vector, respectively.

Lemma 2. *Suppose that Assumptions A and B hold and that monotone pure strategy BNEs, $s^* = (s_1^*, \cdots, s_I^*)$, are played. Then the structural model can be represented as follows,*

$$Y_i = \mathbf{1}\left[U_i \leq X_i'\beta_i + P_i'\alpha_i\right], \tag{3}$$

Proof. See Lemma 2 in Liu, Vuong, and Xu (2012). □

4. Identification

In this section, we discuss the semiparametric identification of the structural parameters — α_i, β_i and $F_{U|X}$. The definition of identification of parameters in a structural model follows Hurwicz (1950) and Koopmans and Reiersol (1950), i.e. given the conditional distribution $\mathbb{P}_{Y|X}$ that is generated from a structure with parameter θ_0, the structural parameter θ_0 is identified if there exists a function $\mathscr{G}$ such that $\theta_0 = \mathscr{G}(\mathbb{P}_{Y|X})$.

Our identification strategy takes two steps: first, we establish nonparametric identification of the function φ_{ij} and the (conditional) copula function of the distribution of U; second, we identify (α_i, β_i) and F_{U_i} under an additional location–scale normalization of the payoff function. To proceed, we first make the following assumptions.

Assumption C. *Let $X_i = (W_i, Z_i) \in \mathbb{R}^{d_{W_i}} \times \mathbb{R}^{d_{Z_i}}$ where $d_{W_i} + d_{Z_i} = d_i$. Conditional on $W = (W_1, \cdots, W_I)$, U and $Z = (Z_1, \cdots, Z_I)$ are independent of each other.*

Assumption C assumes the conditional independence between U and Z given W, which-has been frequently made in the empirical discrete game literature. See, e.g. Aradillas-Lopez (2010), Bajari, Hong, Krainer, and Nekipelov (2010), and Lewbel and Tang (2011).

Fix $W = w$. For any $i \neq j$ and $(v_i, v_j) \in [0,1]^2$, define a copula function $C_{ij}(\cdot|w) : [0,1]^2 \to [0,1]$ as follows:

$$C_{ij}(v_i, v_j; w) = \mathbb{P}\left(U_i \leq F_{U_i}^{-1}(v_i), U_j \leq F_{U_j}^{-1}(v_j) \big| W = w\right).$$

By definition, $C_{ij}(v; w) = C_{ji}(v'; w)$, where v' is the transpose of the vector $v \in [0,1]^2$. Let further $V_i = \mathbb{E}(Y_i|X)$. Note that $C_{ij}(\cdot; w)$ can be identified on the support for all $(v_i, v_j) \in \mathscr{S}_{V_i V_j | W = w}$, by

$$C_{ij}(v_i, v_j; w) = \mathbb{E}(Y_i Y_j | V_i = v_i, V_j = v_j, W = w).$$

Assumption D. *For some $w \in \mathscr{S}_W$, the support $\mathscr{S}_{V_i V_j | W = w}$ is convex and compact subset of $[0,1]^2$, and has full rank, i.e., dim $\left(\mathscr{S}_{V_i V_j | W = w}\right) = 2$.*

The second half of Assumption D is a restriction similar to the exclusion restriction which requires a rich support for Z conditional on X (see, e.g. Bajari, Hong, Krainer, and Nekipelov, 2010). The first part is restrictive, but can relaxed significantly. For the brevity of notation, we will not pursue this direction. Please note, however, that the support of (V_i, V_j) given W needs not to be $[0,1]^2$ and, as a consequence, the conditional distribution of $F_{U_i|W}(\cdot|w)$ is only disclosed on a subset of $[0,1]$. It should also be noted that the support restriction on (V_i, V_j) given W is only required for some w in the support, instead of the whole support of W.

Assumptions C and D allow us to identify φ_{ij} on the support $\mathscr{S}_{X|W=w}$.

Lemma 3. *Suppose that Assumptions A and B hold and that monotone pure strategy BNEs, $s^* = (s_1^*, \cdots, s_I^*)$, are played. In addition, suppose that Assumptions C and D hold. Then for any $i \neq j$, $\varphi_{ij}(\cdot)$ is identified on the support $\mathscr{S}_{X|W=w}$.*

Proof. See Appendix A.1 □

The identification of (α_i, β_i) is similar to the single agent binary response model. By Lemma 2,

$$F^{-1}_{U_i|W}(V_i|W) = X_i'\beta_i + P_i'\alpha_i \tag{4}$$

Let $T_i = [X_i' - \mathbb{E}(X_i'|V_i, W), P_i' - \mathbb{E}(P_i'|V_i, W)]'$. Thus we can define a hyperplane in terms of T_i and payoff coefficients (α_i, β_i):

$$T_i' \times \begin{pmatrix} \beta_i \\ \alpha_i \end{pmatrix} = 0,$$

from which we identify (α_i, β_i) under a scale normalization and a rank condition. Moreover, given the identification of (α_i, β_i) and φ_i $\mathscr{S}_{X|W=w}$, we can identify $F_{U_i|W}(\cdot|w)$ using the fact that $F_{U_i|W}(X_i'\beta_i + P_i'\alpha_i|W) = \mathbb{E}(Y_i|X)$.

Assumption E. $\|\beta_i\| = 1$.

Assumption E normalizes the scale of β_i only, instead of (α_i, β_i), because in Section 5 we will estimate β_i up to scale in the first stage, therefore this normalization will simplify our estimation analysis.

Assumption F. *For some $w \in \mathscr{S}_W$ satisfying Assumption D, the matrix $\mathbb{E}(T_i T_i' | W = w)$ has full rank which equals to $d_i + I - 1$.*

In addition to Assumption D, Assumption F is another rank condition, which implicitly excludes the constant term in X_i and serves as a location normalization. Assumption F is not a primitive restriction because P_i obtains from the equilibrium. Please note, however, it's not difficult to view that a full rank condition on $X_i' - \mathbb{E}(X_i' | V_i, W)$ and a rich support of $X_{-i}' \beta_{-i}$ given X_i will imply Assumption F.

Theorem 1. *Suppose that Assumptions A and B hold and that monotone pure strategy BNEs, $s^* = (s_1^*, \cdots, s_I^*)$, are played. In addition, suppose that Assumptions C to F hold. Then (α_i, β_i) is identified. Moreover, $F_{U_i|W}(\cdot|w)$ is also identified on $\mathscr{S}_{X_i'\beta_i + Z_i'\alpha_i | W = w}$.*

The proof of Theorem 1 is straightforward under above discussion and, therefore, omitted.

5. Semiparametric Estimation of Index Payoffs

In this section, we discuss the estimation of (α_i, β_i) coefficients in the payoff function and leave $F_{U|X}$ as a nuisance parameter. For the brevity of notation, we illustrate our method in a two–player setup, i.e. $I = 2$. Our estimation procedure takes three steps: First, we estimate β_i up to scale at a $\sqrt{N}$ rate. Next, we estimate the belief function φ_i at a uniform non–parametric rate using kernel method. Finally, we propose a simple estimator for α_i and show that $\widehat{\alpha}_i$ converges at a $\sqrt{N}$ rate. We also establish asymptotic distributions for $\widehat{\beta}_i$ and $\widehat{\alpha}_i$.

Without causing any confusion, we denote by subscript n (or ℓ, alternatively) the index of observation in a sample and by N the sample size. In contrast, we use subscript i (or j, k, alternatively) to denote the index of player. Let $X_n = (X_{1n}, X_{2n})$ and $Y_n = (Y_{1n}, Y_{2n})$.

Assumption G. *Let* $\{(X_n, Y_n) : n = 1, \cdots, N\}$ *be an i.i.d. random sample.*

5.1. **Estimation of β_i.** In a two–player game, the payoff function for player i becomes

$$\pi_i(y, x_i, u_i) = \begin{cases} x_i'\beta_i + \alpha_i y_{-i} - u_i, & \text{if } y_i = 1, \\ 0, & \text{if } y_i = 0, \end{cases}$$

where the strategic effects coefficient is a scale. Suppose that the conditions in Lemmas 1 hold and that the equilibrium adopted is a monotone pure strategy BNE, (s_1^*, s_2^*), where $s_i^*(x, u_i) = \mathbf{1}\{u_i \leq u_i^*(x)\}$. Then the mutual consistency restriction requires that

$$x_1'\beta_1 + \alpha_1 \mathbb{P}(U_2 \leq u_2^* | X = x, U_1 = u_1^*) - u_1^* = 0, \tag{5}$$

$$x_2'\beta_2 + \alpha_2 \mathbb{P}(U_1 \leq u_1^* | X = x, U_2 = u_2^*) - u_2^* = 0. \tag{6}$$

Note that there could be multiple solution to eqs. (5) and (6) and we assume that only one solution contributes the equilibrium played. We also maintain the following assumption throughout this section, which strengthens Assumption C.

Assumption H. *Let X and U be independent of each other.*

Under Assumption H, $F_{U|X} = F_U$ and $u_i^*(x) = u_i^*(x_1'\beta_1, x_2'\beta_2)$. Therefore, $\mathbb{E}(Y_i|X) = G_i(X_1'\beta_1, X_2'\beta_2)$, where $G_i(t_1, t_2) = F_{U_i}(u_i^*(t_1, t_2))$. Following the literature on the index models, β_i can be estimated up to scale at a $\sqrt{N}$ rate, which is well discussed (see, e.g. Bierens, 2011; Ichimura, 1993; Klein and Spady, 1993; Powell, Stock, and Stoker, 1989). For example, here we simply describe a procedure to estimate β by following Klein and Spady (1993).

Let $\beta = (\beta_1', \beta_2')'$ and B be the parameter space for β such that Assumption E is satisfied for all its elements. For $y \in \mathcal{A}$, $x \in \mathscr{S}_X$ and $b \in B$, let . Let further $P(y|x; b) = \mathbb{E}(Y = y | X_1'b_1 = x_1'b_1, X_2'b_2 = x_2'b_2)$ and $\tilde{P}(y|x_n; b)$ be a Kernel estimator for the conditional probabilities $P(y|x_n;b)$ given the n–th observation $X_n = x_n$, i.e.

$$\tilde{P}(y|x_n;b) = \frac{\sum_{\ell \neq n} \mathbf{1}(Y_\ell = y) K_p\left(\frac{X'_{1\ell}b_1 - x'_{1n}b_1}{h_p}, \frac{X'_{2\ell}b_2 - x'_{2n}b_2}{h_p}\right) + \tilde{\delta}_{1n}(b)}{\sum_{\ell \neq n} K_p\left(\frac{X'_{1\ell}b_1 - x'_{1n}b_1}{h_p}, \frac{X'_{2\ell}b_2 - x'_{2n}b_2}{h_p}\right) + \tilde{\delta}_n(b)},$$

where $K_p(\cdot) : \mathbb{R}^2 \to R$ denotes a Parzen–Rosenblatt kernel function and h_p is a bandwidth, and $\tilde{\delta}_{1n}$ and $\tilde{\delta}_n$ are trimming sequences introduced for technical reasons, see Klein and Spady (1993) for more detail.

Therefore, we define a Klein–Spady type estimator as follows:

$$\tilde{\beta} = \text{argmax}_{b \in B} \sum_{n=1}^{N} (\tilde{\tau}_n/2) \left\{ \sum_{y \in \mathcal{A}} \left[\mathbf{1}\{Y_n = y\} \ln \tilde{P}^2(y|X_n;b) \right] \right\},$$

in which $\tilde{\tau}_n$ is a trimming sequence. Given the rich literature on the asymptotic properties of such kind of index estimators, in the following analysis, we simply assume a pilot $\sqrt{N}$–consistent estimator $\tilde{\beta} = \beta + O_p(N^{-1/2})$.

5.2. **Estimation of Belief Function φ_i.** Now we establish a nonparametric estimator for the equilibrium belief function $\varphi_i(\cdot)$. Rather than following the identification strategy in Section 4, here we derive a similar expression for (φ_1, φ_2). For $t \in \mathbb{R}^2$ and $i = 1, 2$, let $m_i(t) = \mathbb{E}(Y_i | X'_1\beta_1 = t_1, X'_2\beta_2 = t_2)$. Let further $M(t) = \mathbb{E}(Y_1 Y_2 | X'_1\beta_1 = t_1, X'_2\beta_2 = t_2)$. Then

$$\varphi_1(x) = \frac{\frac{\partial M(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial m_2(x'_1\beta_1, x'_2\beta_2)}{\partial t_2} - \frac{\partial m_2(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial M(x'_1\beta_1, x'_2\beta_2)}{\partial t_2}}{\frac{\partial m_1(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial m_2(x'_1\beta_1, x'_2\beta_2)}{\partial t_2} - \frac{\partial m_2(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial m_1(x'_1\beta_1, x'_2\beta_2)}{\partial t_2}}, \tag{7}$$

$$\varphi_2(X) = \frac{\frac{\partial m_1(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial M(x'_1\beta_1, x'_2\beta_2)}{\partial t_2} - \frac{\partial M(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial m_1(x'_1\beta_1, x'_2\beta_2)}{\partial t_2}}{\frac{\partial m_1(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial m_2(x'_1\beta_1, x'_2\beta_2)}{\partial t_2} - \frac{\partial m_2(x'_1\beta_1, x'_2\beta_2)}{\partial t_1} \times \frac{\partial m_1(x'_1\beta_1, x'_2\beta_2)}{\partial t_2}}, \tag{8}$$

which comes from the fact that

$$\frac{\partial M(x_1'\beta_1, x_2'\beta_2)}{\partial t_1} = \varphi_1(X) \times \frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_1} + \varphi_2(X) \times \frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_1},$$
$$\frac{\partial M(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} = \varphi_1(X) \times \frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} + \varphi_2(X) \times \frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_2}.$$

Therefore, we estimate $\varphi_i(X_n)$ for each observation X_n by plugging into the leave–one–out Nadaraya–Watson estimator for each term in equations (7) and (8).

Let

$$\hat{f}_X(x_n) = \sum_{\ell \neq n} K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\tilde{\beta}_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\tilde{\beta}_2}{h_\varphi} \right) \Big/ N h_\varphi^2,$$
$$\hat{q}_i(x_n) = \sum_{\ell \neq n} Y_{i\ell} K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\tilde{\beta}_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\tilde{\beta}_2}{h_\varphi} \right) \Big/ N h_\varphi^2,$$
$$\hat{Q}(x_n) = \sum_{\ell \neq n} Y_{1\ell} Y_{2\ell} K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\tilde{\beta}_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\tilde{\beta}_2}{h_\varphi} \right) \Big/ N h_\varphi^2,$$

where $K_\varphi(\cdot) : \mathbb{R}^2 \to R$ denotes a Parzen–Rosenblatt kernel function and h_φ is a bandwidth. Thus, $M(X_{1n}'\beta_1, X_{2n}'\beta_2)$ and $m_i(X_{1n}'\beta_1, X_{2n}'\beta_2)$ can be estimated by $\hat{Q}(X_n)/\hat{f}_X(X_n)$ and $\hat{q}_i(X_n)/\hat{f}_X(X_n)$, respectively. For notational brevity, we denote $\hat{M}(X_n) = \hat{Q}(X_n)/\hat{f}_X(X_n)$ and $\hat{m}_i(X_n) = \hat{q}_i(X_n)/\hat{f}_X(X_n)$.

Moreover, let

$$\hat{a}_i(x_n) = \frac{1}{N h_P^3} \sum_{\ell \neq n}^{N} \partial K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\tilde{\beta}_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\tilde{\beta}_2}{h_\varphi} \right) \Big/ \partial t_i,$$
$$\hat{b}_{ji}(x_n) = \frac{1}{N h_P^3} \sum_{\ell \neq n}^{N} Y_{j\ell} \times \partial K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\tilde{\beta}_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\tilde{\beta}_2}{h_\varphi} \right) \Big/ \partial t_i,$$
$$\hat{c}_i(x_n) = \frac{1}{N h_P^3} \sum_{\ell \neq n}^{N} Y_{1\ell} Y_{2\ell} \times \partial K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\tilde{\beta}_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\tilde{\beta}_2}{h_\varphi} \right) \Big/ \partial t_i.$$

Thus we estimate $\partial M(X_{1n}'\beta_1, X_{2n}'\beta_2)/\partial t_1$ by $\hat{f}_X^{-2}(X_n) \left[\hat{c}_i(X_n) \hat{f}_X(X_n) - \hat{a}_i(X_n) \times \hat{M}(X_n) \right]$, and $\partial m_j(X_{1n}'\beta_1, X_{2n}'\beta_2)/\partial t_i$ by $\hat{f}_X^{-2}(X_n) \left[\hat{b}_{ji}(X_n) \hat{f}_X(X_n) - \hat{a}_i(X_n) \times \hat{m}_j(X_n) \right]$. Hence, we obtain an estimator for $\varphi_i(x_n)$,

$$\widehat{\varphi}_i(x_n) = \frac{\hat{A}_i(x_n)}{\hat{A}(x_n)}, \tag{9}$$

in which for $j \neq i$

$$\begin{aligned}
\hat{A}_i(x_n) \equiv & \left[\hat{c}_i(x_n)\hat{f}_X(x_n) - \hat{a}_i(x_n)\hat{Q}(x_n)\right] \times \left[\hat{b}_{jj}(x_n)\hat{f}_X(x_n) - \hat{a}_j(x_n)\hat{q}_j(x_n)\right] \\
& - \left[\hat{c}_j(x_n)\hat{f}_X(x_n) - \hat{a}_j(x_n)\hat{Q}(x_n)\right] \times \left[\hat{b}_{ji}(x_n)\hat{f}_X(x_n) - \hat{a}_i(x_n)\hat{q}_j(x_n)\right], \\
\hat{A}(x_n) \equiv & \left[\hat{b}_{11}(x_n)\hat{f}_X(x_n) - \hat{a}_1(x_n)\hat{q}_1(x_n)\right] \times \left[\hat{b}_{22}(x_n)\hat{f}_X(x_n) - \hat{a}_2(x_n)\hat{q}_2(x_n)\right] \\
& - \left[\hat{b}_{12}(x_n)\hat{f}_X(x_n) - \hat{a}_2(x_n)\hat{q}_1(x_n)\right] \times \left[\hat{b}_{21}(x_n)\hat{f}_X(x_n) - \hat{a}_1(x_n)\hat{q}_2(x_n)\right].
\end{aligned}$$

To guarantee a uniform convergence, we further impose a convenient assumption that restricts the denominator in eq. (9) to be bounded away from zero almost surely.

Assumption I. *There exists a constant $c_0 > 0$ such that*

$$\inf_{x \in \mathscr{S}_X} \left| \frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_1} \times \frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} - \frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} \times \frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_1} \right| > c_0,$$

almost surely.

Assumption I could be replaced by introducing trimming adjustments to the denominator of the estimator (see, e.g., Klein and Spady (1993)).

We make further assumptions, which are standard for the uniform convergence of kernel estimator.

Assumption J. *Let $R \geq 1$. For some $\delta > 0$ and all $\beta^\delta \in \{b \in \mathbb{B} : \|b - \beta\| \leq \delta\}$, $f_{X_1'\beta_1^\delta, X_2'\beta_2^\delta}(\cdot)$ is (R+1)–times continuously differentiable on $\mathbb{R}^2$ with bounded (R+1)th–partial derivatives on $\mathbb{R}^2$. Further, $\mathbb{E}(Y_i|(X_1'\beta_1^\delta, X_2'\beta_2^\delta) = \cdot)$ and $\mathbb{E}\left(Y_1 Y_2|(X_1'\beta_1^\delta, X_2'\beta_2^\delta) = \cdot\right)$ are (R+1)–times continuously differentiable on $\mathbb{R}^2$ with bounded (R+1)th–partial derivatives on $\mathbb{R}^2$.*

In particular, $f_{X_1'\beta_1^\delta, X_2'\beta_2^\delta}(\cdot)$ is uniformly continuous on $\mathbb{R}^2$ and integrable. Thus $f_{X_1'\beta_1^\delta, X_2'\beta_2^\delta}(\cdot)$ is bounded, i.e., $\sup_{t\in\mathbb{R}^2} f_{X_1'\beta_1^\delta, X_2'\beta_2^\delta}(t) < \infty$. Moreover, a similar argument also applies to functions $\mathbb{E}(Y_i|(X_1'\beta_1^\delta, X_2'\beta_2^\delta) = \cdot)$ and $\mathbb{E}(Y_1Y_2|(X_1'\beta_1^\delta, X_2'\beta_2^\delta) = \cdot)$.

Assumption K. *Let* $\kappa_N \propto N^\iota$ *for some* $\iota > 0$ *and* $\inf_{x\in\mathscr{S}_X} f_X(x) > 0$.

Note that we can let η_N go to zero at an arbitrary slow rate by choosing small ι. We will derive the uniform convergence of $\widehat{\varphi}_i(x)$ with respect to the compact sub–support $\{x : \|x\| \le \kappa_N\}$. Let $\eta_N \equiv \inf_{\|x\|\le\kappa_N} f_X^4(x)$. If the second half condition in Assumption K does not hold, then the observations in the compact sub–support with $f_X^4(x) \le \eta_N$ need to be trimmed.

Assumption L. *Let* $\mathbb{E}|X| < \infty$.

Assumptions K and L could be replaced by the simpler conditions that the support of X is compact and f_X is bounded away from zero.

Assumption M. $K_\varphi(u) : \mathbb{R}^2 \to \mathbb{R}$ *is (R+1)–continuously differentiable on* $\mathbb{R}^2$ *with bounded (R+1)th–partial derivatives on* $\mathbb{R}^2$*. The support of* $K_P(\cdot)$ *is a convex subset of* $\mathbb{R}^2$ *with nonempty interior, with the origin as an interior point.* $K_\varphi(u)$ *satisfies*

$$\int u_1^{r_1} u_2^{r_2} K_\varphi(u)dx = 0 \quad \text{if } r_1 + r_2 = R,$$

$$< \infty \quad \text{if } r_1 + r_2 = R + 1.$$

Assumption N. *Setting* $h_\varphi = (\ln N/N)^{1/(2R+4)}$.

Proposition 1. *Suppose that* $\tilde{\beta} = \beta + O_p(N^{-1/2})$*. If Assumption G through N hold, then*

$$\sup_{\|x_n\|\le\kappa_N} \|\widehat{\varphi}_i(x_n) - \varphi_i(x_n)\| = O_p\left(\eta_N^{-1}\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right).$$

Proof. See Appendix B.1 □

Note that our choice of h_φ implies over smoothing for the nonparametric estimation of functions m_i and M and would be sub–optimal in this sense. However, this sub–optimality will not affect the fact that $\widehat{\varphi}_i$ converges uniformly at the best possible rate, which mainly relies on the optimal convergence rate for the derivate estimator of functions m_i and M.

5.3. **Estimation of Strategic Component α_i.** Our final step is to estimate α_i (together with β_i) at a $\sqrt{N}$–convergence rate. Since in equilibrium, $Y_i = \mathbf{1}\left\{U_i \leq X_i'\beta_i + \alpha_i\varphi_i(X)\right\}$ is a single index model on $(X_i, \varphi_i(X))$, here we simply follow the approach proposed by Klein and Spady (1993),[5] which achieves the semi–parametric efficiency bound. To simplify our discussion and the notation, we use the marginal distribution of Y_i to derive the quasi–likelihood function indexed by $(a_i, b_i) \in A_i \times B_i$, instead of employing the joint distribution of (Y_1, Y_2). Thus, our estimator is defined by

$$(\widehat{\alpha}_i, \widehat{\beta}_i) = \underset{(a_i,b_i)\in A_i\times B_i}{\text{argsup}} \quad \hat{L}_i(a_i, b_i; \hat{\tau}), \tag{10}$$

where

$$\hat{L}_i(a_i, b_i; \hat{\tau}) \equiv \sum_{n=1}^{N} (\hat{\tau}_n/2)\left\{Y_{in}\ln\left[\hat{P}_i(X_n; a_i, b_i)\right]^2 + (1-Y_{in})\ln\left[1-\hat{P}_i(X_n; a_i, b_i)\right]^2\right\},$$

and

$$\hat{P}_i(X_n; a_i, b_i) = \frac{\sum_{\ell\neq n}\left[Y_{i\ell}\times K_P\left(\frac{(X_{i\ell}-X_{in})'b_i + a_i[\widehat{\varphi}_i(X_\ell)-\widehat{\varphi}_i(X_n)]}{h_P}\right)\right] + \hat{\delta}_{1n}(a_i, b_i)}{\sum_{\ell\neq n} K_P\left(\frac{(X_{i\ell}-X_{in})'b_i + a_i[\widehat{\varphi}_i(X_\ell)-\widehat{\varphi}_i(X_n)]}{h_P}\right) + \hat{\delta}_n(a_i, b_i)},$$

and $\hat{\tau}_n$, $\hat{\delta}_{1n}$ and $\hat{\delta}_n$ are trimming sequences (see Klein and Spady (1993)).

Note that the only difference with the estimator defined Klein and Spady (1993) is the fact that we replace the unobserved belief $\varphi_i(X)$ with the belief estimator $\hat{\varphi}_i(X)$. By proposition 1 and under a similar argument as in Klein and Spady (1993), we also show that $(\hat{\alpha}_i', \hat{\beta}_i')$ is a $\sqrt{N}$–consistent estimator of (α_i', β_i').

Assumption O. *The parameter vector (α_i', β_i') lies in the interior of a compact space $A_i \times B_i \subseteq \mathbb{R} \times \mathbb{R}^{d_i}$.*

[5] Because P_i is bounded between $[0, 1]$ and has positive density close to the boundary, which violate the conditions in Powell, Stock, and Stoker (1989).

Assumption P. *Let X be distributed in a compact support, and* η_N *be a strict positive constant by setting* $\kappa_N \equiv \sup_{x \in \mathscr{S}_X} \|x\|$*. Let* $f_{X_{i1|o}}(x_{i1})$ *be the density for some continuous variable, denoted as* X_{i1}*, conditioned on the remaining exogenous variables (including* X_{-i}*), and U. This conditional density is smooth in that for all* $x \in \mathscr{S}_X$*, there exists a constant* $c_1 \in \mathbb{R}_+$ *such that*

$$\left| D^r_{x_{i1}} f_{X_{i1}|o}(x_{i1}) \right| < c_1, \quad (r = 1, 2, 3, 4).$$

Assumption Q. *With* $h_P \to 0$*, the trimming function employed to down weight observations has the form*

$$\tau(t, \varepsilon) \equiv \left\{ 1 + \exp\left[(h_P^{\varepsilon/5} - t)/h_P^{\varepsilon/4} \right] \right\}^{-1},$$

where $\varepsilon > 0$ *and t is to be interpreted as a density estimator (e.g.* $\hat{f}_{X_i' b_i + a_i \varphi_i(X)}$*.) Let*

$$\hat{\delta}_{dn} \equiv \tau(\hat{g}_{idn}(\widetilde{a}_{iP}, \widetilde{b}_{iP}), \varepsilon), \quad \textit{for } d = 0, 1,$$

$$\textit{and} \qquad \hat{\delta}_n \equiv \hat{\delta}_{0n} + \hat{\delta}_{1n},$$

where for $d = 0, 1$*,*

$$\hat{g}_{idn}(\widetilde{a}_{iP}, \widetilde{b}_{iP}) \equiv \sum_{\ell \neq n}^{N} \frac{\boldsymbol{1}(Y_{i\ell} = d)}{h_P} K_P \left(\frac{(X_{in} - X_{i\ell})' \widetilde{b}_{iP} + \widetilde{a}_{iP} \left[\widehat{\varphi}_i(X_n) - \widehat{\varphi}_i(X_\ell) \right]}{h_P} \right) \Big/ (N - 1),$$

and $(\widetilde{a}_{iP}, \widetilde{b}_{iP})$ *is a preliminary consistent estimator for which* $\|(\widetilde{a}_{iP}, \widetilde{b}_{iP}) - (a_i, b_i)\|$ *is* $O_p(N^{-1/3})$*.*

Assumption R. *The kernel function,* $K_P(\cdot) : \mathbb{R} \to \mathbb{R}$*, is a symmetric function that integrates to one, has bounded third moment, and for some* $c_2 > 0$*,*

$$\max\left\{ |D^r_u K_P(u)|, \int |D^r_u K_P(u)|\, du \right\} < c_2, \quad (r = 0, 1, 2, 3, 4),$$

$$\int u^2 K_P(u) du = 0.$$

Moreover, let h_P *be a bandwidth sequence satisfying (i)* $N^{-R/(2R+4)} \times h_P^{-2} \to 0$*; (ii)* $N^{-1/4} < h_P < N^{-1/8}$*.*

Note that we apply a stronger result of uniform convergence in Hansen (2008), which modifies the lower bound of h_P from $N^{-1/6}$ in Klein and Spady (1993) to $N^{-1/4}$ in our Assumption R, (ii). Assumption R implies that $R > 2$, a restriction to the order of kernel in our first–step estimation.

Assumption S. *For $i = 1, 2$, there exists no proper linear subspace of $\mathbb{R}^d$ having probability 1 under $\mathbb{P}_X$.*

Theorem 2. *Suppose that $\sup_x \|\hat{\varphi}_i(x) - \varphi_i(x)\| = O_p\left((\ln N/N)^{-R/(2R+4)}\right)$ for some $R \geq 1$. If Assumption G through S hold. Then*

$$\sqrt{N}\begin{pmatrix} \hat{\alpha}_i - \alpha_i \\ \hat{\beta}_i - \beta_i \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma \equiv \mathbb{E}\left\{\frac{f^2_{U_i}\left(u_i^*(X)\right) \times (\varphi_i(X), X_i')'(\varphi_i(X), X_i')}{F_{U_i}\left(u_i^*(X)\right)\left[1 - F_{U_i}\left(u_i^*(X)\right)\right]}\right\}^{-1}.$$

Proof. See Appendix C □

5.4. **A sketch of semiparametric estimation in I–player games.** Now we consider a discrete game with I players. In the setup specified in section 1, we make the following parametric assumption on the payoff functions:

$$\pi_i = X_i'\beta_i + \sum_{j \neq i} \alpha_{ij} Y_j.$$

Now the structural parameters of interest are $(\alpha_i', \beta_i')'$. In this setup, the equilibrium strategy can be written as

$$Y_i = \mathbf{1}\left\{U_i \leq X_i'\beta_i + \sum_{j \neq i} \alpha_{ij} \mathbb{P}\left(U_j \leq u_j^*(X) | U_i = u_i^*(X)\right)\right\}.$$

In our first–step estimation, similarly, we estimate β by $\tilde{\beta}$ in an I–index model. Second, let $\varphi_{i,j}(x) = \mathbb{P}\big(U_j \leq u_j^*(x) | U_i = u_i^*(x)\big)$, and similar to equation (7) and (8), we derive an expression for φ_{ij},

$$\varphi_{ij}(X) = \frac{\frac{\partial \mathbb{E}(Y_iY_j|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_i} \times \frac{\partial \mathbb{E}(Y_j|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_j} - \frac{\partial \mathbb{E}(Y_j|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_i} \times \frac{\partial \mathbb{E}(Y_iY_j|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_j}}{\frac{\partial \mathbb{E}(Y_i|X_1'\beta_1\cdots,X_I'\beta_I)}{\partial t_i} \times \frac{\partial \mathbb{E}(Y_j|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_j} - \frac{\partial \mathbb{E}(Y_j|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_i} \times \frac{\partial \mathbb{E}(Y_i|X_1'\beta_1,\cdots,X_I'\beta_I)}{\partial t_j}}. \tag{11}$$

Hence, we obtain a nonparametric estimator $\widehat{\varphi}_{ij}$ by plugging into the leave–one–out Nadaraya–Watson estimator for each term on the RHS of equation (11). By a similar argument as that for Proposition 1, it can be shown that under similar set of conditions, there is

$$\sup_x \left\| \widehat{\varphi}_{ij}(x) - \varphi_{ij}(x) \right\| = O_p\left(\eta_N^{-1} \left(\frac{\ln N}{N} \right)^{R/(2R+I+2)} \right).$$

Finally, by an analogous analysis, we follow Klein and Spady (1993) to obtain a $\sqrt{N}$–consistent estimator for (α_i', β_i') under a similar set of conditions, for which we require $R > 1 + I/2$.

6. Monte Carlo Simulations

In this section, we use a numerical experiment to illustrate the performance of our estimator in a finite–size sample. Let $I = 2$, $d_1 = d_2 = 2$ and $X_1 = (X_{11}, X_{12})$ and $X_2 = (X_{21}, X_{22})$, where $X \equiv (X_1, X_2)$ has a mean zero normal distribution with identity covariance matrix. Let U_1 and U_2 be independent of X and conform to a joint mean zero normal distribution with unit variances and correlation parameter $\rho = 0.5$.

Moreover, let $\beta_1 = \beta_2 = (1,1)'$, $\alpha_1 = \alpha_2 = 1$. It can be shown that a (unique) monotone strategy BNE exists under this design, i.e., for each x, there exist cutoff values $u_1^*(x)$ and $u_2^*(x)$, such that player j chooses 1 whenever her private signal $u_j \leq u_j^*(x)$. We compute $u_j^*(x)$ by solving the following equations for each X_n in the sample:

$$u_1^* = \beta_{11}x_{11} + \beta_{12}x_{12} + \alpha_1\Phi\left(\frac{u_2^* - \rho u_1^*}{\sqrt{1-\rho^2}}\right), u_2^* = \beta_{21}x_{21} + \beta_{22}x_{22} + \alpha_2\Phi\left(\frac{u_1^* - \rho u_2^*}{\sqrt{1-\rho^2}}\right).$$

where $\Phi(\cdot)$ is the c.d.f of standard normal distribution.

Table 1 shows the composition of one random sample with $N = 500$. In our first–step

TABLE 1. Sample composition

Choice profile	Percentage
$Y = (1,1)$	46.0%
$Y = (1,0)$	15.8%
$Y = (0,1)$	17.8%
$Y = (0,0)$	20.4%

estimation, β_i obtains by the recipe of Klein and Spady (1993). Specifically, we use second order biweight kernel and choose bandwidth according to rule of thumb. Table 2 reports summary statistics for $\tilde{\beta}_1$, including the sample mean(MEAN) and median (MEDIAN), as well as the standard deviation (SD), and root–mean–squared–error (RMSE).

TABLE 2. Finite–Sample Behavior of $\tilde{\beta}_1$

N	TRUE	MEAN	MEDIAN	SD	RMSE
250	1.00	1.0109	0.9969	0.1739	0.1742
500	1.00	1.0063	0.9984	0.1160	0.1161
1000	1.00	1.0038	0.9987	0.0829	0.0830
2000	1.00	1.0037	1.0018	0.0547	0.0548

For the estimation of φ_i, we employ the fourth order biweight product kernel, i.e., $K(u_1, u_2) = k(u_1) \cdot k(u_2)$ where $k(u_i) = \frac{7}{4}(1 - 3u_i^2) \cdot \frac{15}{16}(1 - u_i^2)^2 \cdot \mathbf{1}(|u_i| \leq 1)$ and choose $h_P = 4.40 \cdot \widehat{\sigma} \cdot \big(N/log(N)\big)^{-1/10}$ where $\widehat{\sigma}$ is the estimated standard error of the regressor.

Figure 1 plots φ_1, φ_2 and their kernel estimates. For presentation purpose, we fix $x_1 = (0,0)$, but a similar pattern holds for other values of x_1. The upper panel shows functions φ_1 and φ_2 and their estimates. The lower–left panel shows the estimate of φ_1 and the infeasible estimate of φ_1 when (β_1, β_2) are known. Further, the lower–right panel shows the the marginal distribution of $\varphi_1(X)$, $f_{\varphi_1(X)}$, and its estimate.

In last step, we use second order biweight kernel and rule of thumb bandwidth again to implement the Klein and Spady (1993) estimation procedure.

Table 3 reports the finite sample performance for estimating α_1 by our three-step estimation procedure. The case of estimating α_2 has similar result. There are five numbers reported

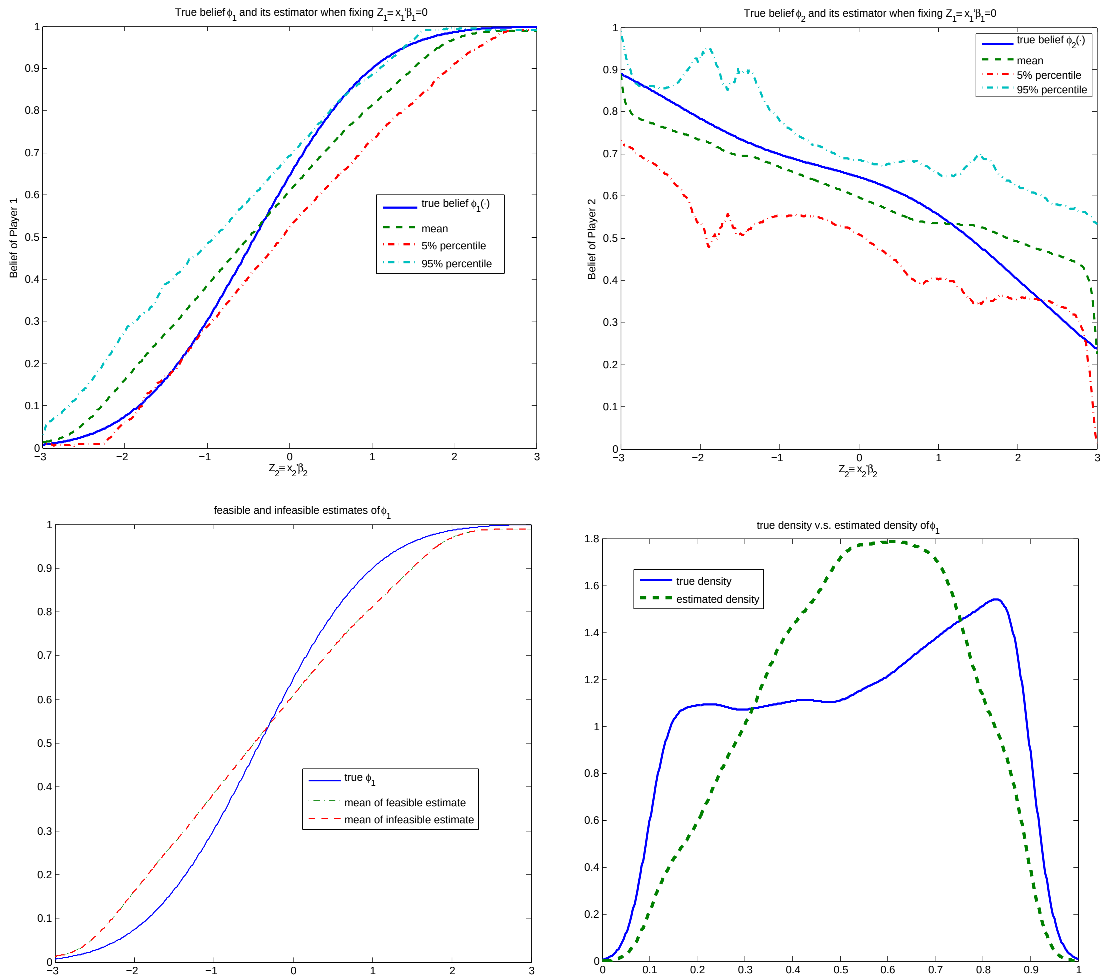

FIGURE 1. Kernel estimates of φ_1, φ_2 and $f_{\varphi_1(X)}$

for each type of estimator with a certain sample size. The first number refers to the true value of the parameter, the second number refers to the mean, the third one refers to the median, the fourth one refers to Standard Deviation (SD) and the last one refers to the Root Mean Square Error (RMSE).

Table 4 reports the finite sample performance for estimating β_1 in the last step of our estimation procedure. The case of estimating β_2 yields similar result. Similar to table 3, there are five numbers reported in the table.

TABLE 3. Mean, median, SD and RMSE for estimating α_1

Sample size	Our Estimator					Infeasible Estimator				
	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
250	1.00	0.946	0.926	0.4314	0.4346	1.00	0.988	0.984	0.3347	0.3348
500	1.00	0.988	0.988	0.3022	0.3022	1.00	1.0103	1.0168	0.2366	0.2367
1000	1.00	0.984	0.979	0.2072	0.2078	1.00	1.0032	1.0050	0.1628	0.1628
2000	1.00	0.993	0.994	0.1425	0.1426	1.00	0.999	0.995	0.1067	0.1067

TABLE 4. Mean, median, SD and RMSE for estimating β_1 in last step

Sample size	Our Estimator					Infeasible Estimator				
	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
250	1.00	1.0197	0.9996	0.1853	0.1861	1.00	1.0163	0.9963	0.1646	0.1652
500	1.00	1.0045	1.0048	0.1161	0.1161	1.00	1.0049	0.9968	0.1114	0.1114
1000	1.00	0.9970	0.9942	0.0826	0.0826	1.00	0.9953	0.9902	0.0774	0.0775
2000	1.00	1.0008	1.0017	0.0557	0.0557	1.00	1.0003	1.0007	0.0518	0.0518

REFERENCES

AGUIRREGABIRIA, V., AND P. MIRA (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.

ANDREWS, D. (1992): "Generic uniform convergence," *Econometric Theory*, 8(02), 241–257.

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

ATHEY, S. (2001): "Single crossing properties and the existence of pure strategy equilibria in games of incomplete information," *Econometrica*, 69(4), 861–889.

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

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.

BANKS, J., R. BLUNDELL, AND M. CASANOVA RIVAS (2010): "The dynamics of retirement behavior in couples: reduced-form evidence from England and the US," *Working Paper*.

BERRY, S. (1992): "Estimation of a model of entry in the airline industry," *Econometrica: Journal of the Econometric Society*, pp. 889–917.

BIERENS, H. (2011): "The hilbert space theoretical foundation of semi–nonparametric modeling," *working paper*.

BJORN, P., AND Q. VUONG (1984): "Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation," *Working Papers*.

BRESNAHAN, T., AND P. REISS (1990): "Entry in monopoly market," *The Review of Economic Studies*, 57(4), 531.

BRESNAHAN, T. F., AND P. C. REISS (1991): "Empirical models of discrete games," *Journal of Econometrics*, 48(1-2), 57–81.

BROCK, W., AND S. DURLAUF (2001a): "Discrete choice with social interactions," *Review of Economic Studies*, 68(2), 235–260.

——— (2001b): "Interactions-based models," *Handbook of econometrics*, 5, 3297–3380.

CASANOVA, M. (2010): "Happy together: a structural model of couples' joint retirement choices," Discussion paper, Mimeo, UCLA.

DARSOW, W., B. NGUYEN, AND E. OLSEN (1992): "Copulas and Markov processes," *Illinois Journal of Mathematics*, 36(4), 600–642.

DE PAULA, A., AND X. TANG (2010): "Inference of signs of interaction effects in simultaneous games with incomplete information, second version," *forthcoming, Econometrica*.

GRIECO, P. (2011): "Discrete games with flexible information structures: an application to local grocery markets," *Working paper*.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal nonparametric estimation of first–price auctions," *Econometrica*, 68(3), 525–574.

HANSEN, B. (2008): "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24(3), 726.

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

ICHIMURA, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58(1-2), 71–120.

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

KLEIN, R., AND R. SPADY (1993): "An efficient semiparametric estimator for binary response models," *Econometrica*, 61(2), 387–421.

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.

LEWBEL, A., AND X. TANG (2011): "Identification and estimation of discrete Bayesian games with multiple equilibria using excluded regressors," Discussion paper, Working paper, Boston College.

LIU, N., Q. VUONG, AND H. XU (2012): "Nonparametric analysis of binary games of incomplete information," *Working paper*.

MANSKI, C. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3(3), 205–228.

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

MCADAMS, D. (2003): "Isotone equilibrium in games of incomplete information," *Econometrica*, 71(4), 1191–1214.

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2003): "Identification and estimation of dynamic games," Discussion paper, National Bureau of Economic Research.

POWELL, J., J. STOCK, AND T. STOKER (1989): "Semiparametric estimation of index coefficients," *Econometrica: Journal of the Econometric Society*, pp. 1403–1430.

RENY, P. (2011): "On the existence of monotone pure–strategy equilibria in Bayesian games," *Econometrica*, 79(2), 499–553.

SEIM, K. (2006): "An empirical model of firm entry with endogenous product–type choices," *The RAND Journal of Economics*, 37(3), 619–640.

SOETEVENT, A., AND P. KOOREMAN (2007): "A discrete-choice model with social interactions: with an application to high school teen behavior," *Journal of Applied Econometrics*, 22(3), 599–624.

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

VIVES, X. (1990): "Nash equilibrium with strategic complementarities," *Journal of Mathematical Economics*, 19(3), 305–321.

WAN, Y., AND H. XU (2010): "Semiparametric estimation of binary decision games of incomplete information with correlated private signals," *Working paper*.

XU, H. (2010): "Estimation of discrete games with correlated private signals," *Working paper*.

XU, H. (2011): "Social interactions: a game theoretical apporoach," .

APPENDIX A. PROOFS OF IDENTIFICATION RESULTS

A.1. **Proof of Lemma 3.**

Proof. Let $\nu_k(x) = \mathbb{E}(Y_k|X = x)$ for $k \in \mathcal{I}$. By definition and Assumption C,

$$\varphi_{ij}(x) = \mathbb{P}\left(U_j \leq u_j^*(x)|U_i = u_i^*(x), X = x\right) \\ = \mathbb{P}\left(U_j \leq u_j^*(x)|U_i = u_i^*(x), W = w\right).$$

Then, it follows from Darsow, Nguyen, and Olsen (1992) that

$$\mathbb{P}(U_j \leq u_j^*(x)|U_i = u_i^*(x), W = w) = \frac{\partial C_{ij}(v_i, v_j; w)}{\partial v_i}\Big|_{v_i = F_{U_i|W}(u_i^*(x)|w), v_j = F_{U_j|W}(u_j^*(x)|w)} \\ = \frac{\partial C_{ij}(v_i, v_j; w)}{\partial v_i}\Big|_{v_i = \mathbb{E}(Y_i|X=x), v_j = \mathbb{E}(Y_j|X=x)}.$$

which is identified by the fact that $C_{ij}(\cdot; w)$ can be identified on the support for all $(v_i, v_j) \in \mathscr{S}_{V_i V_j|W=w}$.

□

APPENDIX B. PROOFS OF STATISTICAL PROPERTIES

B.1. **Proof of Proposition 1.**

Proof. Our proofs follow Guerre, Perrigne, and Vuong (2000). For the notation brevity, here we ignore the difference cased by leaving–one–observation–out in the estimator $\widehat{\varphi}_i$. Moreover, let

$$a_{iN}(x_n) = \frac{1}{Nh_P^3} \sum_{\ell \neq n}^{N} \partial K_\varphi \left(\frac{(X_{1\ell} - x_{1n})'\beta_1}{h_\varphi}, \frac{(X_{2\ell} - x_{2n})'\beta_2}{h_\varphi} \right) \Big/ \partial t_i$$

be the (infeasible) nonparametric estimator of the derivative

$$\frac{\partial}{\partial t_1} \left\{ \mathbb{E}(Y_1 Y_2 | (X_1'\beta_1, X_2'\beta_2) = t) \times f_{X_1'\beta_1, X_2'\beta_2}(t) \right\} \Big|_{t = x_n}$$

using the true parameters β. Similarly, we define b_{iN}, c_{jiN}, q_{iN}, Q_N and f_{XN}. By plugging these infeasible estimators, we define our infeasible estimator $A_{iN}(x_n)$ and $A_N(x_n)$. Further, let

$\varphi_{iN}(X_n) = A_{iN}(X_n)/\ A_N(X_n)$ and

$$A_1(x) = f^4_{X_1'\beta_1, X_2'\beta_2}\left[\frac{\partial M(x_1'\beta_1, x_2'\beta_2)}{\partial t_1}\frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} - \frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_1}\frac{\partial M(x_1'\beta_1, x_2'\beta_2)}{\partial t_2}\right],$$

$$A_2(x) = f^4_{X_1'\beta_1, X_2'\beta_2}\left[\frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_1}\frac{\partial M(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} - \frac{\partial M(x_1'\beta_1, x_2'\beta_2)}{\partial t_1}\frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_2}\right],$$

$$A(x) = f^4_{X_1'\beta_1, X_2'\beta_2}\left[\frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_1}\frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_2} - \frac{\partial m_2(x_1'\beta_1, x_2'\beta_2)}{\partial t_1}\frac{\partial m_1(x_1'\beta_1, x_2'\beta_2)}{\partial t_2}\right].$$

Note that

$$\hat{\varphi}_i(x) = \frac{\hat{A}_i(x)}{\hat{A}(x)} = \frac{\hat{A}_i(x)/A(x)}{\hat{A}(x)/A(x)}$$

$$= \frac{A_{iN}(x)/A(x) + \left[\hat{A}_i(x) - A_{iN}(x)\right]/A(x)}{1 + \left[A_N(x)/A(x) - 1\right] + \left[\hat{A}(x) - A_N(x)\right]/A(x)}.$$

Hence it suffices to show

$$\sup_{\|x\|\le\kappa_N} \|A_{iN}(x)/A(x) - A_i(x)/A(x)\| = O_p\left(\eta_N^{-1}\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right), \tag{12}$$

$$\sup_{\|x\|\le\kappa_N} \|A_N(x)/A(x) - 1\| = O_p\left(\eta_N^{-1}\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right), \tag{13}$$

$$\sup_{\|x\|\le\kappa_N} \left\|\hat{A}_i(x)/A(x) - A_{iN}(x)/A(x)\right\| = o_p\left(\eta_N^{-1}\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right), \tag{14}$$

$$\sup_{\|x\|\le\kappa_N} \left\|\hat{A}(x)/A(x) - A_N(x)/A(x)\right\| = o_p\left(\eta_N^{-1}\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right). \tag{15}$$

Equations (12) and (13) are satisfied under Lemmas 5 and 6. We illustrate the argument for eq. (14), and then eq. (15) is proved analogously. By Lemma 5, it suffice to show

$$\sup_{\|x\|\le\kappa_N} \left\|\hat{A}_i(x) - A_{iN}(x)\right\| = o_p\left(\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right).$$

We will show that $\sup_{\|x\|\le\kappa_N} |a_{iN}(x) - \hat{a}_i(x)|$, $\sup_{\|x\|\le\kappa_N} \left|b_{iN}(x) - \hat{b}_{ij}(x)\right|$, $\sup_{\|x\|\le\kappa_N} |c_{iN}(x) - \hat{c}_i(x)|$, $\sup_{\|x\|\le\kappa_N} |q_{iN}(x) - \hat{q}_i(x)|$, $\sup_{\|x\|\le\kappa_N} \left|Q_N(x) - \hat{Q}(x)\right|$ and $\sup_{\|x\|\le\kappa_N} \left|f_{XN}(x) - \hat{f}_X(x)\right|$ all converge to zero at the $\sqrt{N}$ rate. Since the arguments for all other terms are quite similar to or simpler than those for $\sup_{\|x\|\le\kappa_N}|c_{iN}(x)-\hat{c}_i(x)|$, here we only provide a detailed proof for the latter.

Because $\tilde{\beta}_i=\beta_i+O_p(N^{-1/2})$, then for any fixed $\epsilon>0$, the following inequality holds with probability approaching to 1,

$$\begin{aligned}|c_{iN}(x_n)-\hat{c}_i(x_n)| &= \left|\frac{1}{Nh_\varphi^3}\sum_{\ell\neq n}^N Y_{1\ell}Y_{2\ell}\times\left\{\partial K_\varphi\left(\frac{(X_{1\ell}-x_{1n})'\tilde{\beta}_1}{h_\varphi},\frac{(X_{2\ell}-x_{2n})'\tilde{\beta}_2}{h_\varphi}\right)\Big/\partial t_i\right.\right.\\ &\qquad\left.\left.-\partial K_\varphi\left(\frac{(X_{1\ell}-x_{1n})'\beta_1}{h_\varphi},\frac{(X_{2\ell}-x_{2n})'\beta_2}{h_\varphi}\right)\Big/\partial t_2\right\}\right|\\ &\le \sup_{\|\beta^\dagger-\beta\|\le\epsilon}\left|\frac{1}{Nh_\varphi^4}\sum_{\ell\neq n}^N Y_{1\ell}Y_{2\ell}\times\partial^2K_\varphi\left(\frac{(X_{1\ell}-x_{1n})'\beta_1^\dagger}{h_\varphi},\frac{(X_{2\ell}-x_{2n})'\beta_2^\dagger}{h_\varphi}\right)\Big/\partial t_i\partial t_1\times(X_{1\ell}-x_{1n})'(\tilde{\beta}_1-\beta_1)\right|\\ &+\sup_{\|\beta^\dagger-\beta\|\le\epsilon}\left|\frac{1}{Nh_P^4}\sum_{\ell\neq n}^N Y_{1\ell}Y_{2\ell}\times\partial^2K_\varphi\left(\frac{(X_{1\ell}-x_{1n})'\beta_1^\dagger}{h_\varphi},\frac{(X_{2\ell}-x_{2n})'\beta_2^\dagger}{h_\varphi}\right)\Big/\partial t_i\partial t_2\times(X_{2\ell}-x_{2n})'(\tilde{\beta}_2-\beta_2)\right|.\end{aligned}$$

By lemma 4, we have

$$\sup_x|a_{iN}(x)-\hat{a}_i(x)|\le\left\|\tilde{\beta}_1-\beta_1\right\|\times O_p(1)+\left\|\tilde{\beta}_2-\beta_2\right\|\times O_p(1)=O_p(N^{-1/2}).\quad\square$$

Lemma 4. *Suppose that Assumptions G, H, J, L and M hold. Thus,*

$$\sup_{\|x\|\le\kappa_N}\sup_{\|b-\beta\|\le\delta}\left\|\frac{1}{Nh_P^4}\sum_{\ell=1}^N Y_{1\ell}Y_{2\ell}\times\partial^2K_\varphi\left(\frac{(X_{1\ell}-x_1)'b_1}{h_\varphi},\frac{(X_{2\ell}-x_2)'b_2}{h_\varphi}\right)\Big/\partial t_i\partial t_j\times(X_{j\ell}-x_j)'\right\|=O_p(1),$$

Proof. Fix i,j. Let $S_\ell(x,b)=\frac{1}{h_\varphi^4}Y_{1\ell}Y_{2\ell}\times\partial^2K_\varphi\left(\frac{(X_{1\ell}-x_1)'b_1}{h_\varphi},\frac{(X_{2\ell}-x_2)'b_2}{h_\varphi}\right)\Big/\partial t_i\partial t_j\times(X_{j\ell}-x_j)'$ as a random vector indexed by x and b. Let further $\psi_{x,b}(t)=\mathbb{E}\left[\mathbb{E}\left(Y_1Y_2|X\right)\times(X_j-x_j')\middle|(X_1'b_1,X_2'b_2)=t\right]$ and $\phi_{x,b}(t)=\psi_{x,b}(t)\times f_{X_1'b_1,X_2'b_2}(t)$.[6] Then we have

$$\begin{aligned}\sup_{\|x\|\le\kappa_N}\sup_{\|b-\beta\|\le\delta}\left\|\frac{1}{N}\sum_{\ell=1}^N S_\ell(x,b)\right\| &\le \sup_{\|x\|\le\kappa_N}\sup_{\|b-\beta\|\le\delta}\left\|\frac{1}{N}\sum_{\ell=1}^N S_\ell(x,b)-\mathbb{E}S_\ell(x,b)\right\|\\ &+\sup_{\|x\|\le\kappa_N}\sup_{\|b-\beta\|\le\delta}\left\|\frac{1}{N}\sum_{\ell=1}^N\mathbb{E}S_\ell(x,b)-\partial^2\phi_{x,b}(x_1'b_1,x_2'b_2)/\partial t_1\partial t_2\right\|\\ &+\sup_{\|x\|\le\kappa_N}\sup_{\|b-\beta\|\le\delta}\left\|\partial^2\phi_{x,b}(x_1'b_1,x_2'b_2)/\partial t_1\partial t_2\right\|.\end{aligned}$$

[6]Note that we suppress a subscript j in the notation for $\psi_{x,b}$ and $\phi_{x,b}$.

By Theorem 1 in Andrews (1992), the first term of the RHS is $o_p(1)$; and by Assumptions J and L, the last term is $O_p(1)$.

Moreover, for the second term, we have

$$\begin{aligned}\mathbb{E}S_\ell(x,b) &= \frac{1}{h_P^4}\mathbb{E}\left[\psi(X_{1\ell}'b_1, X_{2\ell}'b_2)\times\partial^2 K_\varphi\left(\frac{(X_{1\ell}-x_1)'\beta_1}{h_\varphi},\frac{(X_{2\ell}-x_2)'\beta_2}{h_\varphi}\right)\right]\\ &= \frac{1}{h_\varphi^2}\int_{\mathbb{R}^2}\phi_{x,b}(x_1'b_1-h_\varphi u_1, x_2'b_2-h_\varphi u_2)\times\partial^2 K_\varphi(u)/\partial t_1\partial t_2 du\\ &= \int_{\mathbb{R}^2}\partial^2\phi_{x,b}(x_1'b_1-h_\varphi u_1, x_2'b_2-h_\varphi u_2)/\partial t_1\partial t_2\times K_\varphi(u)\,du.\end{aligned}$$

By Taylor expansion of order 2 with integral remainder,

$$\begin{aligned}&\partial^2\phi_{x,b}(x_1'b_1-h_\varphi u_1, x_2'b_2-h_\varphi u_2)/\partial t_1\partial t_2 = \partial^2\phi_{x,b}(x_1'b_1, x_2'b_2)/\partial t_1\partial t_2\\ &-h_\varphi\sum_{k=1}^{2}\frac{\partial^3\phi_{x,b}(x_1'b_1,x_2'b_2)}{\partial t_1\partial t_2\partial t_k}u_k+\frac{h_\varphi^2}{2}\int_0^1(1-s)\sum_{k_1=1}^{2}\sum_{k_2=1}^{2}\frac{\partial^4\phi_{x,b}(x_1'b_1-th_\varphi u_1, x_2'b_2-th_\varphi u_2)}{\partial t_1\partial t_2\partial t_{k_1}\partial t_{k_2}}u_{k_1}u_{k_2}ds.\end{aligned}$$

By Assumption J, $\phi_{x,b}$ has bounded fourth order derivative. Thus, uniformly over x and b

$$\begin{aligned}&\left\|\mathbb{E}S_\ell(x,b)-\partial^2\phi_{x,b}(x_1'b_1,x_2'b_2)/\partial t_1\partial t_2\right\|\\ &=\frac{h_\varphi^2}{2}\left\|\int_{\mathbb{R}^2}\int_0^1(1-t)\sum_{k_1=1}^{2}\sum_{k_2=1}^{2}\frac{\partial^4\phi_{x,b}(x_1'b_1-th_\varphi u_1, x_2'b_2-th_\varphi u_2)}{\partial t_1\partial t_2\partial t_{k_1}\partial t_{k_2}}u_{k_1}u_{k_2}K_\varphi(u)\,dtdu\right\|\le C\times\frac{h_\varphi^2}{2},\end{aligned}$$

for some $C\in\mathbb{R}_+$. Thus the second term is $o_p(1)$. □

Lemma 5. *Suppose that Assumptions I and K hold. Then*

$$\inf_{\|x\|\le\kappa_N}\|A(x)\| = O(\eta_N).$$

Proof. By Assumptions I and K

$$\begin{aligned}&|A(x)| = f^4_{X_1'\beta_1,X_2'\beta_2}(x_1'\beta_1,x_2'\beta_2)\\ &\times\left|\frac{\partial m_1(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\times\frac{\partial m_2(x_1'\beta_1,x_2'\beta_2)}{\partial t_2}-\frac{\partial m_1(x_1'\beta_1,x_2'\beta_2)}{\partial t_2}\times\frac{\partial m_2(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\right|\ge c_0\eta_N.\end{aligned}$$ □

Lemma 6. *Suppose that assumptions in Proposition 1 hold. Then*

$$\sup_{\|x\|\leq\kappa_N}\|A_{iN}(x)-A_i(x)\| = O_p\left(\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right),$$

$$\sup_{\|x\|\leq\kappa_N}\|A_N(x)-A(x)\| = O_p\left(\left(\frac{\ln N}{N}\right)^{R/(2R+4)}\right),$$

Proof. We only illustrate the argument for $\sup_{\|x\|\leq\kappa_N}\|A_{1N}(x)-A_1(x)\|$; other results can be established analogously. It suffices to show that

$$\sup_{\|x\|\leq\kappa_N}\left\|c_{1N}(x)f_{XN}(x)-a_{1N}(x)Q_N(x)-f^2_{X_1'\beta_1,X_2'\beta_2}\times\frac{\partial M(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\right\| = O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right),$$

$$\sup_{\|x\|\leq\kappa_N}\left\|b_{22N}(x)f_{XN}(x)-a_{2N}(x)q_{2N}(x)-f^2_{X_1'\beta_1,X_2'\beta_2}\times\frac{\partial m_2(x_1'\beta_1,x_2'\beta_2)}{\partial t_2}\right\| = O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right),$$

$$\sup_{\|x\|\leq\kappa_N}\left\|c_{2N}(x)f_{XN}(x)-a_{2N}(x)Q_N(x)-f^2_{X_1'\beta_1,X_2'\beta_2}\times\frac{\partial M(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\right\| = O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right),$$

$$\sup_{\|x\|\leq\kappa_N}\left\|b_{21N}(x)f_{XN}(x)-a_{1N}(x)q_{2N}(x)-f^2_{X_1'\beta_1,X_2'\beta_2}\times\frac{\partial m_2(x_1'\beta_1,x_2'\beta_2)}{\partial t_2}\right\| = O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right).$$

Again, we provide a detailed proof only for the first term due to the similarity. Let $Q(t) = M(t)\times f_{X_1'\beta_1,X_2'\beta_2}(t)$. Because

$$\frac{\partial Q(t)}{\partial t_1} = \frac{\partial M(t)}{\partial t_1}\times f_{X_1'\beta_1,X_2'\beta_2}(t)+\frac{\partial f_{X_1'\beta_1,X_2'\beta_2}(t)}{\partial t_1}\times M(t),$$

then

$$f^2_{X_1'\beta_1,X_2'\beta_2}\times\frac{\partial M(x_1'\beta_1,x_2'\beta_2)}{\partial t_1} = \frac{\partial Q(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\times f_{X_1'\beta_1,X_2'\beta_2}(x_1'\beta_1,x_2'\beta_2) - \frac{\partial f_{X_1'\beta_1,X_2'\beta_2}(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\times Q(x_1'\beta_1,x_2'\beta_2).$$

Thus, it suffices to show that

$$\sup_{\|x\|\leq\kappa_N}\left\|c_{1N}(x)-\frac{\partial Q(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\right\|=O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right), \tag{16}$$

$$\sup_{\|x\|\leq\kappa_N}\left\|f_{XN}(x)-f_{X_1'\beta_1,X_2'\beta_2}(x_1'\beta_1,x_2'\beta_2)\right\|=O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right), \tag{17}$$

$$\sup_{\|x\|\leq\kappa_N}\left\|a_{1N}(x)-\frac{\partial f_{X_1'\beta_1,X_2'\beta_2}(x_1'\beta_1,x_2'\beta_2)}{\partial t_1}\right\|=O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right), \tag{18}$$

$$\sup_{\|x\|\leq\kappa_N}\left\|Q_N(x_1'\beta_1,x_2'\beta_2)-Q(x_1'\beta_1,x_2'\beta_2)\right\|=O_p\left(\left(\frac{\ln N}{N}\right)^{\frac{R}{2R+4}}\right). \tag{19}$$

Equations (17) and (18) directly follows Hansen (2008), Theorem 6, and by following its proof, eqs. (16) and (19) also obtain, which is straightforward, and hence omitted here.

□

APPENDIX C. PROOF OF THEOREM 2

Our proof follows Klein and Spady (1993). Throughout appendix C, we introduce some notation, which is consistent with Klein and Spady (1993).

Let $v_i(X; a_i, b_i) \equiv X_i' b_i + a_i \varphi_i(X)$ and $\bar{v}_i(X; a_i, b_i) \equiv X_i' b_i + a_i \widehat{\varphi}_i(X)$. Through, we suppress the subscript for player i in v_i and $\hat{v}_i$, i.e., we use $v(x; a_i, b_i)$ and $\bar{v}(x; a_i, b_i)$ to denote $v_i(x; a_i, b_i)$ and $\bar{v}_i(x; a_i, b_i)$, respectively. Similarly, we will suppress subscript i in the following discussion. Let $v_n(a_i, b_i) \equiv v(X_n; a_i, b_i)$ and $\bar{v}_n(a_i, b_i) \equiv \bar{v}(X_n; a_i, b_i)$. Similarly, by replacing $\hat{\varphi}_i$ with the underlying belief φ_i, we can define $\bar{\tau}_n$, $\bar{\tau}_{0n}$, $\bar{\tau}_{1n}$, $\bar{\delta}_n$, $\bar{\delta}_{0n}$, $\bar{\delta}_{1n}$. Let $g_v(v_n; a_i, b_i)$ be the density of $v_n(a_i, b_i)$. Moreover, for $d = 0, 1$ let $g_{dv}(v_n; a_i, b_i) \equiv \mathbb{P}(Y_i = d | v(a_i, b_i) = v_n(a_i, b_i)) g_v(v_n; a_i, b_i)$ and for $d = 0, 1$

$$\bar{g}_{dv}(v_n; a_i, b_i) \equiv \sum_{\ell \neq n}^{N} \frac{\mathbf{1}(Y_{i\ell} = d)}{h_P} K\left(\frac{v_\ell - v_n}{h_P}\right) \Big/ (N-1),$$

$$\hat{g}_{dv}(v_n; a_i, b_i) \equiv \sum_{\ell \neq n}^{N} \frac{\mathbf{1}(Y_{i\ell} = d)}{h_P} K\left(\frac{\bar{v}_\ell - \bar{v}_n}{h_P}\right) \Big/ (N-1),$$

Let further

$$\bar{L}_N(a_i, b_i; \bar{\tau}) \equiv \sum_{n=1}^{N} (\bar{\tau}_n/2) \left\{ Y_{in} \ln\left[\bar{P}_i(v_n; a_i, b_i)^2\right] + (1 - Y_{in}) \ln\left[1 - \bar{P}_i(v_n; a_i, b_i)\right]^2 \right\} \Big/ N$$

and

$$\bar{P}_i(v_n; a_i, b_i) \equiv \left[\bar{g}_{i1v}(v_n; a_i, b_i) + \bar{\delta}_{1n}(v_n; a_i, b_i)\right] \Big/ \left[\bar{g}_{iv}(v_n; a_i, b_i) + \bar{\delta}_n(v_n; a_i, b_i)\right],$$

$$P(v_n; a_i, b_i) \equiv g_{1v}(v_n; a_i, b_i) / g_v(v_n; a_i, b_i).$$

Also, we define the r–th order derivative of any function g with respect to z by

$$D_z^r[g] = \begin{cases} g, & r = 0, \\ \partial^r g / (\partial z)^r, & r = 1, 2, \cdots. \end{cases}$$

Further, we use $\|\cdot\|$ to denote the Euclidean norm.

Let

$$\hat{G}(\alpha_i, \beta_i) \equiv \left[\partial \hat{L}_i / \partial(a_i, b_i)\right]\Big|_{(a_i, b_i) = (\alpha_i, \beta_i)} = \sum_{n=1}^{N} \hat{\tau}_n \hat{r}_n \hat{w}_n / N,$$

where

$$\hat{r}_n \equiv \left[Y_{in} - \hat{P}_i(X_n;\alpha_i,\beta_i)\right] \Big/ \hat{c}_n, \;\; \hat{c}_n \equiv \hat{g}_v(v_n;\alpha_i,\beta_i)\left[\hat{P}_i(X_n;\alpha_i,\beta_i)(1-\hat{P}_i(X_n;\alpha_i,\beta_i))\right],$$

$$\hat{w}_n \equiv \hat{g}_v(v_n;\alpha_i,\beta_i)\left[\partial\hat{P}_i(X_n;\alpha_i,\beta_i)/\partial(a_i,b_i)\right].$$

Let further

$$G_N(\alpha_i,\beta_i) \equiv [\partial L_N/\partial(a_i,b_i)]\Big|_{(a_i,b_i)=(\alpha_i,\beta_i)} = \sum_{n=1}^{N} \tau_n r_n w_n/N,$$

where

$$r_n \equiv [Y_{in} - P_i(v_n;\alpha_i,\beta_i)]/c_n, \;\; c_n \equiv [g_v(v_n;\alpha_i,\beta_i) + \delta_n(v_n;\alpha_i,\beta_i)]$$

$$\times[P_i(v_n;\alpha_i,\beta_i)(1-P_i(v_n;\alpha_i,\beta_i))], \quad w_n \equiv g_v(v_n;\alpha_i,\beta_i)\left[\partial P(v_n;\alpha_i,\beta_i)/\partial(a_i,b_i)\right]$$

C.1. **Proof for Theorem 2.**

Proof. The consistency simply follows the uniform convergence of $\widehat{\varphi}_i$ to φ_i and the proof for Theorem 3 in Klein and Spady (1993), which is omitted here.

For asymptotic normality, it suffices to show $N^{1/2}\widehat{G}(\alpha_i,\beta_i) - N^{1/2}G_N(\alpha_i,\beta_i) = o_p(1)$, and all the left simply follows Klein and Spady (1993), Theorem 4.

$$\begin{aligned} N^{1/2}\hat{G}(\alpha_i,\beta_i) - N^{1/2}G_N(\alpha_i,\beta_i) \\ = N^{-1/2}\sum_{n=1}^{N}\tau_n(\hat{r}_n\hat{w}_n - r_n w_n) + N^{-1/2}\sum_{n=1}^{N}(\hat{\tau}_n - \tau_n)r_n w_n \\ + N^{-1/2}\sum_{n=1}^{N}(\hat{\tau}_n - \tau_n)(\hat{r}_n\hat{w}_n - r_n w_n). \end{aligned} \tag{20}$$

For the first term in equation (20), denoted as $\mathbf{A}$,

$$\mathbf{A} = N^{-1/2}\sum_{n=1}^{N}\tau_n(\hat{r}_n - r_n)w_n + N^{-1/2}\sum_{n=1}^{N}\tau_n(\hat{r}_n - r_n)(\hat{w}_n - w_n) + N^{-1/2}\sum_{n=1}^{N}\tau_n r_n(\hat{w}_n - w_n). \tag{21}$$

For the first term, similar to the arguments for $\mathbf{A}_1$ in Lemma 6 of Klein and Spady (1993), it is $o_p(1)$. For the second term, because

$$\left|N^{-1/2}\sum_{n=1}^{N}\tau_n(\hat{r}_n - r_n)(\hat{w}_n - w_n)\right| \le N^{1/2}\sup|\tau_n(\hat{r}_n - r_n)|\sup|\tau_n(\hat{w}_n - w_n)|$$

By definition

$$\hat{r}_n - r_n = \frac{Y_{in}}{\hat{g}_{1vn}} - \frac{Y_{in}}{g_{1vn}} + \frac{1-Y_{in}}{\hat{g}_{0vn}} - \frac{1-Y_{in}}{g_{0vn}}.$$

By Lemma 7, we have

$$\tau_n \frac{Y_{in}}{\hat{g}_{1vn}} = \tau_n \frac{Y_{in}/g_{1vn}}{\hat{g}_{1vn}/g_{1vn}} = \tau_n \frac{Y_{in}/g_{1vn}}{1+(\hat{g}_{1vn}-g_{1vn})/g_{1vn}} = \tau_n \frac{Y_{in}}{g_{1vn}} + \tau_n \frac{O_p\left(\sqrt{\ln N/Nh_P} \vee h^2\right)}{g_{1vn}},$$

then

$$\sup \left| \tau_n \frac{Y_{in}}{\hat{g}_{1vn}} - \tau_n \frac{Y_{in}}{g_{1vn}} \right| = O_p\left(\sqrt{\ln N/Nh_P} \vee h^2\right).$$

Similarly,

$$\sup \left| \tau_n \frac{1-Y_{in}}{\hat{g}_{0vn}} - \tau_n \frac{1-Y_{in}}{g_{0vn}} \right| = O_p\left(\sqrt{\ln N/Nh_P} \vee h^2\right).$$

Then we have $\sup|\tau_n(\hat{r}_n - r_n)| = O_p\left((\ln N/N)^{2/(2p+3)}\right)$. By a similar argument, $\sup|\tau_n(\hat{w}_n - w_n)| = O_p\left(\sqrt{\ln N/Nh_N^3} \vee h^2\right)$. Further, by the condition (ii) in assumption R,

$$N^{1/2} \sup|\tau_n(\hat{r}_n - r_n)| \sup|\tau_n(\hat{w}_n - w_n)| = o_p(1).$$

For the last term in the RHS of equation (21), denoted by $\mathbf{A}_3$, we have

$$\mathbb{E}\left(\mathbf{A}_3^2\right) = \sum_{n=1}^{N} \mathbb{E}\left[\tau_n^2 r_n^2 (\hat{w}_n - w_n)^2\right] \Big/ N + \mathbb{E}\sum_{\ell \neq n} r_n r_\ell \tau_n \tau_\ell (\hat{w}_n - w_n)(\hat{w}_\ell - w_\ell) \Big/ N.$$

By lemma 7 and Chung(1974, Thm. 4.5.2), the first term is $o_p(1)$, Note that the second term is more complicated than the corresponding part in Klein and Spady (1993). Recall that, by definition, $\widehat{\varphi}_i(X_n)$ is estimated by leaving out one observation Y_n. Similarly, we define $\overline{\varphi}_i(X_n;\ell)$ by leaving out two observations Y_n and Y_ℓ. Thus we can define $\overline{w}_n$ by replacing $\widehat{\varphi}_i(X_k)$ with $\overline{\varphi}_i(X_k;n)$ for all $k \neq n$ and $\widehat{\varphi}_i(X_n)$ with $\overline{\varphi}_i(X_n;\ell)$ in $\widetilde{w}_n$. Note that $\overline{w}_n$ depends neither on Y_{in} and $Y_{i\ell}$, then by a similar argument as in Klein and Spady (1993), Lemma 6, we have $\mathbb{E}\sum_{\ell\neq n} r_n r_\ell \tau_n \tau_\ell (\overline{w}_n - w_n)(\overline{w}_\ell - w_\ell)\Big/N = o_p(1)$. It should also be noted that $\overline{w}_n - \widetilde{w}_n = O(N^{-1})$ uniformly over x, since $\overline{\varphi}_i(X_n;k) - \widehat{\varphi}_i(X_n) = O_p(N^{-1})$ uniformly. Therefore the second term in the RHS of above equation is also $o_p(1)$.

Turning to the second term in (20) above, under a similar argument used to analysis $\mathbf{A}_3$, it is $o_p(1)$. The proof for the last term in equation (20) being $o_p(1)$ simply follows the corresponding part of the arguments in Klein and Spady (1993).

□

Lemma 7. *Suppose that assumptions in Theorem 2 hold. Then for $y = 0, 1$,*

$$\sup \left|\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - g_{yv}(v_n; \alpha_i, \beta_i)\right| = O_p\left(\sqrt{\ln N / N h_P} \vee h^2\right)$$
$$\sup \left|D^1_{(a_i,b_i)}\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - D^1_{(a_i,b_i)} g_{yv}(v_n; \alpha_i, \beta_i)\right| = O_p\left(\sqrt{\ln N / N h_N^3} \vee h^2\right).$$

Proof. First, let $\mathbb{P}_N \widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) \equiv \int \frac{\mathbf{1}(Y_i=y)}{h_P} K\left(\frac{\widehat{v}(x;\alpha_i,\beta_i) - \widehat{v}(X;\alpha_i,\beta_i)}{h_P}\right) dF_{XY}$ and $\mathbb{P}_N g_{yv}(v_n; \alpha_i, \beta_i) \equiv \int \frac{\mathbf{1}(Y_i=y)}{h_P} K\left(\frac{v(x;\alpha_i,\beta_i) - v(X;\alpha_i,\beta_i)}{h_P}\right) dF_{XY}$. By triangular inequality,

$$\begin{aligned}
&\sup_x \left|\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - \mathbb{P}_N \widetilde{g}_{yv}(v_n; \alpha_i, \beta_i)\right| \\
&\leq \sup_x \sup_{\|\widehat{\varphi} - \varphi\| \downarrow 0} \left|\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - \mathbb{P}_N \widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - \left[\widehat{g}_{yv}(v_n; \alpha_i, \beta_i) - \mathbb{P}_N g_{yv}(v_n; \alpha_i, \beta_i)\right]\right| \\
&\qquad + \sup_x \left|\widehat{g}_{yv}(v_n; \alpha_i, \beta_i) - \mathbb{P}_N g_{yv}(v_n; \alpha_i, \beta_i)\right|,
\end{aligned}$$

where the first term is $o_p(N^{-1/2})$, referred as the stochastic equicontinuity condition, by Theorem 11.16 in Kosorok (2008).

Next, $\sup \left|D^r_{(a_i,b_i)}\widehat{g}_{yv}(v_n; \alpha_i, \beta_i) - \mathbb{P}_N g_{yv}(v_n; \alpha_i, \beta_i)\right| = O_p\left(\sqrt{\ln N / N h_P^{1+2r}}\right)$ by Hansen (2008), Theorem 8. Hence,

$$\sup \left|\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - g_{yv}(v_n; \alpha_i, \beta_i)\right| \leq \sup \left|\mathbb{P}_N \widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - g_{yv}(v_n; \alpha_i, \beta_i)\right| + O_p\left((\ln N / N h_P)^{1/2}\right)$$

Let $\Delta(x) = (\widehat{\varphi}_i(x) - \varphi_i(x))/h_P$. Note that, uniformly on x

$$\begin{aligned}
\mathbb{P}_N \widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) &= \int_{\mathbb{R}^{2d}} \frac{1}{h} K\left(\frac{v(x; \alpha_i, \beta_i) - v(t; \alpha_i, \beta_i)}{h_P} + \Delta(x) - \Delta(t)\right) g(v(t; \alpha_i, \beta_i)) dt \\
&= \int K(u) g\Big[(v(t; \alpha_i, \beta_i) - (u - \Delta(x) + \Delta(t)) h_P\Big] dt \\
&\qquad = g\left[v(t; \alpha_i, \beta_i)\right] + O_p\left((\ln N / N)^{R/(2R+4)}\right) + O(h_N^2).
\end{aligned}$$

By assumption R,

$$\sup \left|\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - g_{yv}(v_n; \alpha_i, \beta_i)\right| = O_p\left(\sqrt{\ln N / N h_P} \vee h^2\right).$$

Similarly,

$$\sup \left| D^1(a_i, b_i)\widetilde{g}_{yv}(v_n; \alpha_i, \beta_i) - D^1(a_i, b_i)g_{yv}(v_n; \alpha_i, \beta_i) \right| = O_p\left(\sqrt{\ln N / N h_N^3} \vee h^2 \right).$$

□