# SEMIPARAMETRIC INFERENCE ON SOCIAL INTERACTIONS WITH HOMOPHILY\*

PAUL NIANQING LIU<sup>†</sup> AND HAIQING XU<sup>‡</sup>

ABSTRACT. This paper studies strategic social interactions among agents with correlated types. The type correlation, which represents the homophily principal in sociology, is not directly observed from the data. Such a correlation is our main object of interest, as well as the strategic components. By establishing the existence of a monotone pure strategy equilibrium, we represent players' equilibrium strategies as a single–index binary response model. Next, we establish identification constructively for both strategic components and the type homophily separately. Furthermore, we propose an estimation procedure that is computationally simple. Under regularity conditions, the strategic component estimator is shown to be  $\sqrt{n}$ -consistent and the kernel estimator of the type homophily converges uniformly at a nonparametric rate. Monte Carlo experiments show that our inference procedure works well in finite samples.

Keywords: Social interactions, homophily, copula, single-index, kernel estimation

Date: Thursday 28th January, 2016.

<sup>\*</sup>We gratefully acknowledge Quang Vuong for the guidance and advice. We also thank Jason Abrevaya, Chunrong Ai, Stefan Hoderlein, Shakeeb Khan, Brendan Kline, Qi Li, Isabelle Perrigne, Robin Sickles, Xun Tang, Yahong Zhou, and seminar participants at UT Austin, the Rice University and Shanghai University of Finance and Economics for providing helpful comments.

<sup>&</sup>lt;sup>+</sup>(corresponding author) Department of Economics, Shanghai University of Finance and Economics, P.R. China, nliu@shufe.edu.cn.

<sup>&</sup>lt;sup>‡</sup>Department of Economics, The University of Texas at Austin, h.xu@austin.utexas.edu.

# 1. INTRODUCTION

This paper studies semiparametric estimation of strategic social interactions among agents with correlated types. Our model can be viewed as a natural extension of Manski (1975)'s binary threshold crossing model to the game theoretic setting with asymmetric information. In particular, each player's decision depends on her expectations on other players' choices, and vice versa. In the presence of such a mutual dependence, inference could be difficult to infeasible (see Manski, 1993). We focus on a situation in which agents have complementary payoffs (i.e., agents benefit from choosing the same behavior) and their types are positively regression dependent.<sup>1</sup> Crucially, we show that the equilibrium strategy can be represented as a single–index binary response model. Therefore, we can apply the results in the single–index binary response model literature to our social interaction model. In particular, our identification and estimation strategy follows Klein and Spady (1993).

This paper contributes to the existing empirical game literature in two respects. First, our model does not require private information (i.e. type) to be independent (or conditionally independent) across players, which serves as the key assumption for identification strategies in the current literature, e.g., Aguirregabiria and Mira (2007) and Bajari, Hong, Krainer, and Nekipelov (2010).<sup>2</sup> Allowing type dependence is crucially important for empirical concerns, particularly in the social interaction context. The type independence assumption is convenient for inference but meanwhile it imposes strong model restrictions as well — players' choices must be conditionally independent, which conflicts with the herding observations in sociology. In sociology, the dependence structure of types is of interest by itself. As is well known, similarity breeds connection (see McPherson, Smith-Lovin, and Cook, 2001), which is formulated as the homophily principle. By such a notion, friendship-based interactions should occur among people who have similar/common characteristics and positively correlated private tastes. Because the correlation of private types is not directly

<sup>&</sup>lt;sup>1</sup>Following the convention, a player's private information is denoted as her type, see, e.g., Fudenberg and Tirole (1991).

<sup>&</sup>lt;sup>2</sup>Exceptions include Aradillas-Lopez (2010); Liu, Vuong, and Xu (2013); Wan and Xu (2014) and Xu (2014).

observed from the data, it becomes a challenge that whether we can identify it from those observed behaviors among friends. Note that both homophily and strategic effects can cause the herding behavior in a society group. In this paper, we develop a method that can identify both of them separately.

In this paper, we characterize the dependence of players' types by the copula function of the joint distribution. Under weak conditions, we establish the nonparametric identification of the copula function, which is the conditional probabilities of players' joint choices given their marginal choice probabilities. Moreover, suggested by Guerre, Perrigne, and Vuong (2000), we propose a nonparametric kernel estimator of the copula function, which is shown to be uniformly consistent at a suboptimal nonparametric rate. Investigating how the conditional probability of players' joint choices varies with their marginal choice probabilities is novel in the discrete game literature.

Second, we make no parametric assumptions on the joint distribution of types, which distinguish our paper from most of the current empirical discrete game literature, e.g., Xu (2014). In a semiparametric setup, Wan and Xu (2014) establish partial identification of payoff coefficients when types are positively regression dependent, and further achieve point identification under an additional full support condition on regressors. They suggest a maximum–score–type estimator that converges at  $\sqrt[3]{n}$ –rate. In contrast, we achieve point identification of structural parameters under much weaker support conditions. Furthermore, we develop a Klein–Spady type estimator for the strategic component that converges at the regular  $\sqrt{n}$ –rate. The key intuition for our point identification and faster convergence rate is due to the observation that the equilibrium beliefs can be nonparametrically identified as derivatives of conditional moments.

Similar to Liu, Vuong, and Xu (2013), a key to our method is to focus on the class of monotone pure strategy BNEs, which is a desirable solution concept for empirical applications. Monotonicity has been explicitly or implicitly imposed by the empirical game literature, see e.g. Guerre, Perrigne, and Vuong (2000), Brock and Durlauf (2001) and Bajari, Hong, Krainer, and Nekipelov (2010). In this paper, we first establish the existence

of such kind of equilibria involving conditions that naturally hold in the context of social interactions. Specifically, we require the payoff be complementary and players' types are positively (regression) dependent,<sup>3</sup> which have been endorsed by experimental observations in sociologies; see e.g. Easley and Kleinberg (2010). With monotonicity, then we show that the equilibrium strategies can be represented as a single–index binary response model.

The rest of the paper is organized as follows. In Section 2, we setup our game model and establish the existence of monotone pure strategy BNEs. In Section 3 and 4, we discuss the semiparametric identification and estimation of the structural model, respectively. Section 5 studies the finite sample performance of our estimator using Monte–Carlo experiments.

# 2. Model

Following Brock and Durlauf (2001), we consider an *I*-player binary game to model social interactions. Formally, each player,  $i \in \{1, \dots, I\}$ , simultaneously chooses a binary decision  $Y_i \in \{0, 1\}$ . Binary action space naturally fits a wide range of social phenomena, such as adolescent risky behaviors (e.g. substance use), staying in or dropping out of school, college attendance, entry or withdrawal from the labour force, *etc.* The payoff function is specified as follows:

$$\pi_i(Y, X_i, U_i) = \begin{cases} X'_i \beta_i + \alpha_i \sum_{j \neq i} Y_j - U_i, & \text{if } Y_i = 1; \\ 0, & \text{if } Y_i = 0. \end{cases}$$

In above payoff,  $X_i \in \mathbb{R}^d$  is a vector of individual characteristics that are publicly observed by all players; The error term  $U_i \in \mathbb{R}$  is player *i*'s private payoff shock, i.e., it is observed only by *i* but not by other players. Note that the zero payoff for action 0 is a normalization. For expositional simplicity, let  $X \equiv (X'_1, \dots, X'_I)'$  and  $U \equiv (U_1, \dots, U_I)'$ . Moreover, following the convention, we assume the (conditional) distribution  $F_{U|X}$  is assumed to be common knowledge of the game. Therefore, player *j* does not directly observe *i*'s private payoff shock  $U_i$ , but knows how  $U_i$  is distributed given *j*'s information  $U_i$  and *X*.

<sup>&</sup>lt;sup>3</sup>See Bulow, Geanakoplos, and Klemperer (1985) and Castro (2007) for the notion of strategic complements and positively regression dependent, respectively.

In the payoff function,  $\alpha_i \in \mathbb{R}^+$  and  $\beta_i \in \mathbb{R}^d$  are player–specific parameters of our interests. In particular, the coefficient  $\alpha_i$  measures the strength of the strategic interactions. We require  $\alpha_i$  to be non-negative, which implies *strategic complementarity* or the *super-modularity* (see e.g. Athey, 2001) of the game. Such a restriction naturally follows the observations in sociology on social interactions: friends benefit from choosing the same decision (see e.g. McPherson, Smith-Lovin, and Cook, 2001).

In the discrete game literature, it is commonly assumed that players' private information  $(U_1, \dots, U_I)$  are (conditionally) independent of each other. Such an assumption is convenient and effectively simplifies the identification and estimation of the structural model. Social interaction models, however, focus on the distinction between peer effects and homophily. The former explains the similarity of friends' decisions as a result of social interactions, while the latter justifies the similarity as the outcome of friendship selection — "Similarity breeds connection" (McPherson, Smith-Lovin, and Cook, 2001). Therefore, we make the following assumption.

**Assumption A** (Homophily). Conditional on X,  $(U_1, \dots, U_I)$  are positively regression dependent:  $\mathbb{P}(U_j \le u_j | X = x, U_i = u_i)$  is decreasing in  $u_i$  for  $i \ne j$  and  $(x, u_j) \in \mathscr{S}_{XU_j}$ .

Assumption A requires positive dependence among players' payoff shocks  $(U_1, \dots, U_I)$ . In particular, this condition holds if  $(U_1, \dots, U_I)$  are positive affiliated (see e.g. Castro, 2007). To characterize the statistical dependence, alternatively one could also use the correlation coefficient, which however does not fully disclose the dependence structure unless we focus on some particular parametric families of the joint distribution.

Combined with the non–negative peer effects, Assumption A implies the equilibrium strategies are monotone functions of errors, which simplifies the characterization of the equilibrium. In particular, the Bayesian Nash Equilibrium (BNE) solution concept requires each player i maximizes her expected payoff as follows:

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

where  $\mathbb{1}[\cdot]$  is the indicator function, and the conditional probability  $\mathbb{P}(Y_j = 1 | X, U_i)$  is called as player *i*'s equilibrium "*beliefs*" on *j*'s decision (conditional on *i*'s information). In an equilibrium solution, eq. (1) holds for  $i = 1, \dots, I$ , simultaneously.

Note that Athey (2001)'s *single crossing conditions (SCC)* are satisfied in our setting: For each player  $i = 1, \dots, I$ , whenever all other players use monotone strategies, player *i*'s expected payoff function satisfies Milgrom and Shannon (1994)'s single crossing property of incremental returns. To see this, let  $\{u_j^*(X) \in \mathbb{R} : j \neq i\}$  be an arbitrary vector in  $\mathbb{R}^{I-1}$ . Then, we can show that player *i*'s payoff difference under decision 1 and 0,  $X_i'\beta_i + \alpha_i \sum_{j\neq i} \mathbb{P}[U_j \leq u_j^*(X)|X, U_i] - U_i$ , is decreasing in  $U_i$ . Under additional weak conditions, Athey (2001) shows that the class of games with the SCC conditions hold admit monotone pure strategy BNEs. In our model, in particular, player *i*'s equilibrium strategy  $s_i^*$ is a threshold function:

$$s_i^*(X, U_i) = \mathbf{1} [U_i \le u_i^*(X)], \qquad (2)$$

for some function  $u_i^* : \mathbb{R}^{d \times I} \to \mathbb{R}$ . In the empirical game literature, threshold–type strategies have also been used in the seminal paper by Aradillas-Lopez (2010).

**Assumption B.** The conditional distribution of U given X is absolutely continuous w.r.t. the Lebesgue measure and has positive and continuous density function  $f_{U|X}$ .

Assumption B is weak and standard in the literature.

**Lemma 1.** Suppose assumptions A and B hold. The game admits a monotone pure strategy BNE, where player i's equilibrium strategy is characterized by eq. (2).

An important benchmark for the monotonicity of the equilibrium has been established under the assumption that  $(U_1, \dots, U_I)$  are mutually independent, which however rules out homophily effects in social actions.<sup>4</sup>

<sup>&</sup>lt;sup>4</sup>It is worthpointing out that Lemma 1 is silent about the existence of non-monotone strategy BNEs. The monotone pure strategy BNE, if exists, is a natural solution concept in many economics contexts, e.g., auction and nonlinear pricing, due to its great tractability. Given the existence of a monotone pure strategy equilibrium, it is less interesting to look at any non-monotone strategy equilibrium which is far more complicated, even if it exists. Throughout, we assume that a monotone pure strategy BNE is played under conditions in Lemma 1.

Under assumptions A and B, the threshold  $u_i^*(X)$  that defines the monotone pure strategy in the equilibrium should satisfy the following condition:

$$X'_{i}\beta_{i} + \alpha_{i}\sum_{j\neq i} \mathbb{P}[U_{j} \le u_{j}^{*}(X)|X, U_{i} = u_{i}^{*}(X)] - u_{i}^{*}(X) = 0.$$
(3)

Equation (3) is intuitive: player *i* with the threshold type  $u_i^*(X)$  should be indifferent between action 1 and 0. Let  $\phi_{ij}^*(X) = \mathbb{P}[U_j \leq u_j^*(X)|X, U_i = u_i^*(X)]$  and  $\phi_i^*(X) = \sum_{j \neq i} \phi_{ij}^*(X)$ . Then, the BNE solution can be rewritten as

$$Y_i = \mathbb{1} \big[ U_i \le X'_i \beta_i + \alpha_i \phi_i^*(X) \big], \quad \forall i = 1, \cdots, I.$$
(4)

In the next section, we establish the nonparametric identification of  $\phi_i^*(\cdot)$  under weak conditions. Hence, eq. (4) is essentially a single–index model, which serves as the basis for our identification and estimation analysis.

It is worthpointing out that multiple equilibria could exist (see e.g. Brock and Durlauf, 2001).<sup>5</sup> Following the convention, we assume that the same equilibrium is getting played in the data generating process. In other words, the equilibrium selection mechanism depends on the public state variables X in a deterministic manner.

#### 3. IDENTIFICATION

Recently, Liu, Vuong, and Xu (2013) establish nonparametric identification of discrete Bayesian games. With the linear–index payoff structure in our setting, we extend their results and develop a constructive identification strategy that leads to a simple estimation procedure. In our model, the structural parameters of interests include the coefficient  $\alpha_i$  in the strategic component and the copula function  $C_U$  of U. The recent empirical game literature has focused on the parametric or semiparametric identification and estimation of the former; see, e.g., Brock and Durlauf (2001), Aradillas-Lopez (2010), Bajari, Hong, Krainer, and Nekipelov (2010) and Wan and Xu (2014). In applications of sociology, the latter deserves more attention since it measures the homophily of players' unobserved preference shocks.

 $<sup>\</sup>overline{{}^{5}\text{Kasy}(2012)}$  studies identification and estimation of the number of equilibria in a general context.

We now propose a three–step identification strategy: First, we identify  $\beta_i$  by following the single–index model literature; see e.g. Powell, Stock, and Stoker (1989), Klein and Spady (1993) and Ichimura (1993). Next, we identify the equilibrium beliefs  $\phi_i^*(\cdot)$  nonparametrically as derivatives of conditional moments. Up to our knowledge, such an identification result is new in the literature. In the last step, we establish the identification of strategic component coefficient  $\alpha_i$  and the copula function  $C_{II}$ .

To begin with, we make the following assumptions.

**Assumption C.** The public state variable X is independent of payoff shocks, i.e.,  $X \perp U$ .

Assumption D.  $\|\beta_i\| = 1$  for  $i = 1, \dots, I$ .

**Assumption E.**  $\beta_{i1} \neq 0$  for  $i = 1, \dots, I$ . Moreover,  $X_{i1}$  is continuously distributed on an interval given  $X_{-i1} \equiv \{X_{i2} \cdots X_{id}; X_{-i}\}$ , which is a vector of either discrete and/or continuous random variables. Let  $f_{X_{i1}|X_{-i1}}$  be the density for the continuous variable  $X_{i1}$ conditional on  $X_{-i1}$ .

**Assumption F.** For  $i = 1, \dots, I$ , the matrix  $\mathbb{E}(X_i X'_i)$  has the full rank.

Assumption C is strong but indispensable. See e.g. Aguirregabiria and Mira (2007). Under assumption C, we rewrite the equilibrium condition (3) by

$$X'_{i}\beta_{i} + \alpha_{i}\sum_{j\neq i}F_{U_{j}|U_{i}}(u_{j}^{*}(X)|u_{i}^{*}(X)) - u_{i}^{*}(X) = 0, \text{ for } i = 1, \cdots, I,$$
(5)

where  $F_{U_j|U_i}$  denotes the conditional CDF of  $U_j$  given  $U_i$ . Therefore, we have  $u_i^*(X) = u_i^*(X'_1\beta_1, \cdots, X'_I\beta_I)$ . Here we abuse our notation by letting  $u_i^*(\cdot) : \mathbb{R}^I \to \mathbb{R}^{.6}$  Moreover, by assumption B and the implicit function theorem,  $u_i^*(\cdot)$  is continuously differentiable in the indices  $(X'_1\beta_1, \cdots, X'_I\beta_I)$ . Therefore, we have

$$\mathbb{E}(Y_i|X) = \int \mathbb{1}\{U_i \le u_i^*(X_1'\beta_1, \cdots, X_I'\beta_I)\}d\mathbb{P}_{U_i} = \mathbb{E}(Y_i|X_1'\beta_1, \cdots, X_I'\beta_I), \quad (6)$$

<sup>6</sup>Similarly, we also let  $\phi_{ii}^*(\cdot) : \mathbb{R}^I \to \mathbb{R}$  wherever it applies.

which is also continuously differentiable in all the indices.

Assumptions D to F are fairly standard in the binary response model literature; see e.g. Manski (1975); Klein and Spady (1993); Lewbel (1998). In particular, assumption D is a normalization. Assumption E requires one argument of  $X_i$  is continuously distributed given all the other state variables. This assumption can be relaxed at the expense of longer proofs, as discussed in Horowitz (1998). Assumption F is a standard rank condition.

**Lemma 2.** Suppose assumptions A to F hold. Then,  $\beta_i$  is identified for  $i = 1, \dots, I$ .

As a matter of fact, our proof follows the identification argument in Klein and Spady (1993). Next, we discuss identification of equilibrium beliefs  $\phi_i^*$ . For each  $s \in \mathbb{R}^I$  and  $j \neq i$ , let

$$m_i(s) = \mathbb{E}\left[Y_i|(X_1'\beta_1,\cdots,X_I'\beta_I)=s\right],$$
  
$$g_{ij}(s) = \mathbb{E}\left[Y_iY_j|(X_1'\beta_1,\cdots,X_I'\beta_I)=s\right].$$

By assumptions B and C, both  $m_i$  and  $g_{ij}$  are continuously differentiable functions.

**Assumption G.** For any  $j \neq i$ , we have

$$\frac{\partial}{\partial s_j} m_j(X_1'\beta_1, \cdots, X_I'\beta_I) \cdot \frac{\partial}{\partial s_i} m_i(X_1'\beta_1, \cdots, X_I'\beta_I)$$
  
$$\neq \frac{\partial}{\partial s_i} m_j(X_1'\beta_1, \cdots, X_I'\beta_I) \cdot \frac{\partial}{\partial s_j} m_i(X_1'\beta_1, \cdots, X_I'\beta_I), \ a.s..$$

Assumption G is a testable rank condition given that  $m_i$  and  $g_{ij}$  can be estimated from the data.

To motivate our identification strategy, note that for any  $i \neq j$ ,

$$\begin{aligned} \frac{\partial}{\partial s_i} g_{ij}(s) &= \mathbb{P} \left[ U_j \le u_j^*(s) | U_i = u_i^*(s) \right] \cdot f_{U_i}(u_i^*(s)) \cdot \frac{\partial}{\partial s_i} u_i^*(s) \\ &+ \mathbb{P} \left[ U_i \le u_i^*(s) | U_j = u_j^*(s) \right] \cdot f_{U_j}(u_j^*(s)) \cdot \frac{\partial}{\partial s_i} u_j^*(s) \\ &= \mathbb{P} \left[ U_j \le u_j^*(s) | U_i = u_i^*(s) \right] \cdot \frac{\partial}{\partial s_i} m_i(s) + \mathbb{P} \left[ U_i \le u_i^*(s) | U_j = u_j^*(s) \right] \cdot \frac{\partial}{\partial s_i} m_j(s) \end{aligned}$$

Therefore, we have

$$\frac{\partial}{\partial s_i}g_{ij}(s) = \phi_{ij}^*(s) \cdot \frac{\partial}{\partial s_i}m_i(s) + \phi_{ji}^*(s) \cdot \frac{\partial}{\partial s_i}m_j(s), \tag{7}$$

$$\frac{\partial}{\partial s_j} g_{ij}(s) = \phi_{ij}^*(s) \cdot \frac{\partial}{\partial s_j} m_i(s) + \phi_{ji}^*(s) \cdot \frac{\partial}{\partial s_j} m_j(s).$$
(8)

By assumption G, we can solve  $\phi_{ij}^*$  and  $\phi_{ji}^*$  as two unknowns from eqs. (7) and (8). Specifically,

$$\phi_i^*(X) = \sum_{j \neq i} \frac{\frac{\partial}{\partial s_i} g_{ij}(s) \cdot \frac{\partial}{\partial s_j} m_j(s) - \frac{\partial}{\partial s_j} g_{ij}(s) \cdot \frac{\partial}{\partial s_i} m_j(s)}{\frac{\partial}{\partial s_i} m_i(s) \cdot \frac{\partial}{\partial s_j} m_j(s) - \frac{\partial}{\partial s_j} m_i(s) \cdot \frac{\partial}{\partial s_i} m_j(s)} \bigg|_{s = (X_1' \beta_1, \cdots, X_l' \beta_l)}.$$
(9)

Note that above identification strategy is related to the copula approach developed in Liu, Vuong, and Xu (2013).

Moreover, the identification of  $\alpha_i$  directly follows eq. (3). Because  $\mathbb{E}(Y_i|X) = F_{U_i}(u_i^*(X))$  is a monotone function of  $u_i^*(X)$ , it follow that

$$\left\{X_i - \mathbb{E}\left[X_i | \mathbb{E}(Y_i | X)\right]\right\}' \beta_i + \alpha_i \left\{\phi_i^*(X) - \mathbb{E}\left[\phi_i^*(X) | \mathbb{E}(Y_i | X)\right]\right\} = 0,$$

Therefore, we have

$$\alpha_i = -\mathbb{E}\left\{\frac{X_i - \mathbb{E}\left[X_i \big| \mathbb{E}(Y_i | X)\right]}{\phi_i^*(X) - \mathbb{E}\left[\phi_i^*(X) \big| \mathbb{E}(Y_i | X)\right]}\right\}' \beta_i.$$

In the above argument, conditional on  $\mathbb{E}(Y_i|X)$ ,  $\phi_i^*(X)$  needs to have variations, which is true due to assumptions A and E. This rank condition can also be verified by the data.

Finally, by the monotonicity of the equilibrium strategies and assumption C, we have

$$C_U(p) \equiv \mathbb{P}\big[U_1 \leq F_{U_1}^{-1}(p_1), \cdots, U_I \leq F_{U_I}^{-1}(p_I)\big] = \mathbb{E}\big[\prod_{i=1}^I Y_i \big| \mathbb{E}(Y|X) = p\big],$$

for all  $p \in \mathscr{S}_{\mathbb{E}(Y|X)}$ . Note that  $\mathbb{E}(Y|X) = \mathbb{E}(Y|X'_1\beta_1, \cdots, X'_I\beta_I)$  in above expression.

# 4. ESTIMATION

In this section, we discuss the estimation of the game theoretic model with homophily. For expositional simplicity, we focus on a two-player game, i.e., I = 2, which has been focused in the empirical game literature (see e.g. Tamer, 2003; Aradillas-Lopez, 2010). It is straightforward to extend the proposed estimation procedure to the case of I > 2. Throughout, we will use subscript t to denote the t-th observation. Let  $X_t = (X'_{t1}, X'_{t2})'$ and  $Y_t = (Y_{t1}, Y_{t2})'$ . Moreover, let  $\{(X'_t, Y'_t)' : t = 1, \dots, n\}$  be an i.i.d. random sample.

Given the identification of equilibrium beliefs  $\phi_i^*$  by (9), our estimation takes a two-step procedure. In the first step we estimate  $\beta_i$  at the parametric  $\sqrt{n}$ -rate and then the belief function  $\phi_i^*(\cdot)$  at a uniform rate no slower than  $n^{1/4}$ . The  $\sqrt{n}$ -consistent estimator of  $\beta_i$ helps mitigate the curse of dimensionality that arises from X, while the uniform convergence rate,  $n^{1/4}$  or faster, of the nonparametric estimator is needed for obtaining the  $\sqrt{n}$ -consistent estimator of the strategic component  $\alpha_i$ . In the next step, we first use Klein and Spady (1993)'s pseudo MLE method to estimate  $\alpha_i$ , and then follow Guerre, Perrigne, and Vuong (2000) to estimate the copula function  $C_U$  nonparametrically.

4.1. Estimation of  $\beta_i$  and  $\phi_i^*$ . Let  $B_i \subseteq \mathbb{R}^d$  be the parameter space for  $\beta_i$ . For notional simplicity, we first make the following assumption on the distribution of *X*.

Assumption H.  $F_{X_{-i}|X_i} = F_{X_{-i}|X_i'\beta_i}$ .

Assumption H requires the dependence of  $X_{-i}$  on  $X_i$  through the linear index  $X'_i\beta_i$ . Under this condition, we can rewrite the multiple–index equation (6) to a single–index model, i.e.,

$$\mathbb{E}(Y_i|X_i) = \int \mathbb{E}(Y_i|X) dP_{X_{-i}|X_i} = \int \mathbb{E}(Y_i|X_1'\beta_1, X_2'\beta_2) dP_{X_{-i}|X_i'\beta_i} = \mathbb{E}(Y_i|X_i'\beta_i),$$

where the first and the last steps come form the law of iterated expectation. It should be noted that assumption H can be relaxed at the expositional expense of estimating the double index model (see e.g. Ichimura and Lee, 1991).

Following the single–index model literature, we can estimate  $\beta_i$  at the  $\sqrt{n}$ -rate. In particular, we use Klein and Spady (1993)'s pseudo MLE approach: for i = 1, 2, let

$$\tilde{\beta}_{i} = \operatorname{argmax}_{b_{i} \in B_{i}} \sum_{t=1}^{n} (\tilde{\tau}_{t}/2) \left[ Y_{ti} \ln \tilde{p}_{ti}^{2}(b_{i}) + (1 - Y_{ti}) \ln(1 - \tilde{p}_{ti}(b_{i}))^{2} \right],$$

where  $\tilde{\tau}_t$  is a trimming sequence introduced for technical reasons and

$$\tilde{p}_{ti}(b_i) = \frac{\sum_{s \neq t} Y_{si} \mathcal{K}_{\beta} (X'_{si} b_i - X'_{ti} b_i) + \tilde{\delta}_{1n}(b_i)}{\sum_{s \neq t} \mathcal{K}_{\beta} (X'_{si} b_i - X'_{ti} b_i) + \tilde{\delta}_n(b_i)},$$

in which  $\mathcal{K}_{\beta}(u) = K_{\beta}(u/h_{\beta})/h_{\beta}$  with  $K_{\beta} : \mathbb{R} \to \mathbb{R}$  and  $h_{\beta}$  as Parzen–Rosenblatt kernel and bandwidth, respectively, and  $\tilde{\delta}_{1n}$  and  $\tilde{\delta}_n$  are trimming sequences. Under the conditions in Klein and Spady (1993), we have

$$\tilde{\beta}_i = \beta_i + O_p(n^{-1/2}). \tag{10}$$

Note that one could also use alternative methods developed by e.g. Powell, Stock, and Stoker (1989) and Ichimura (1993) to estimate  $\beta_i$  at the same convergence rate.

It should also be noted that Klein and Spady (1993) requires a pilot estimator  $\hat{\beta}_i^p$  that converges to  $\beta_i$  no slower than  $n^{1/3}$ . One could apply Wan and Xu (2014)'s modified maximum score type estimator, which requires a strong condition that the first argument in the index have an unbounded support. A second approach is to replace the "likelihood trimming" in Klein and Spady (1993) with the high order moments restrictions suggested in e.g. Van de Geer (1990). The same asymptotic properties can be established for the untrimmed estimator under these additional moment conditions. This is also supported by the Monte Carlo evidence in Klein and Spady (1993) which is indeed obtained without any trimming.

Now we are ready to use (9) for estimating the equilibrium beliefs  $\phi_i^*$ . The estimation of derivatives of  $m_i$  and  $g_{ij}$  simply follows the kernel derivative estimation literature, albeit that their arguments contain the unknown parameter  $\beta_i$  that has been estimated at the  $\sqrt{n}$ -rate

in the first stage. Under conditions introduced later, we will show that the first stage bias does not matter for the asymptotics of the  $\phi_i^*$ 's estimator defined below.

Under assumption E,  $(X'_1\beta_1, X'_2\beta_2)$  is continuously distributed. Let  $f_{X'_1\beta_1, X'_2\beta_2} : \mathbb{R}^2 \to \mathbb{R}^+$  be the density function of  $(X'_1\beta_1, X'_2\beta_2)$ . We then estimate  $f_{X'_1\beta_1, X'_2\beta_2}(X'_{t1}\beta_1, X'_{t2}\beta_2)$  and its partial derivatives with respect to the *i*-th argument, respectively, by

$$\hat{f}_{t} = \frac{1}{n-1} \sum_{s \neq t} \mathcal{K}_{\phi} \big( (X_{s1} - X_{t1})' \tilde{\beta}_{1}, (X_{s2} - X_{t2})' \tilde{\beta}_{2} \big),$$
$$\hat{f}_{ti} = \frac{1}{n-1} \sum_{s \neq t} \frac{\partial}{\partial u_{i}} \mathcal{K}_{\phi} \big( (X_{s1} - X_{t1})' \tilde{\beta}_{1}, (X_{s2} - X_{t2})' \tilde{\beta}_{2} \big),$$

where  $\mathcal{K}_{\phi}(u) = K_{\phi}(u/h_{\phi})/h_{\phi}$  with  $K_{\phi} : \mathbb{R}^2 \to \mathbb{R}$  as a Parzen–Rosenblatt kernel and  $h_{\phi}$  as a bandwidth. Under standard kernel estimation conditions, we can establish the consistency of the estimators, i.e.,

$$\hat{f}_t \xrightarrow{p} f_{X_1'\beta_1, X_2'\beta_2}(X_{t1}'\beta_1, X_{t2}'\beta_2), \quad \hat{f}_{ti} \xrightarrow{p} \frac{\partial}{\partial s_i} f_{X_1'\beta_1, X_2'\beta_2}(X_{t1}'\beta_1, X_{t2}'\beta_2).$$

Next, we rewrite  $\phi_i^*(X_t)$  as follows:

$$\phi_i^*(X_t) = \frac{\frac{\partial}{\partial s_i} g_{12}(s) \cdot \frac{\partial}{\partial s_j} m_j(s) - \frac{\partial}{\partial s_j} g_{12}(s) \cdot \frac{\partial}{\partial s_i} m_j(s)}{\frac{\partial}{\partial s_i} m_i(s) \cdot \frac{\partial}{\partial s_j} m_j(s) - \frac{\partial}{\partial s_j} m_i(s) \cdot \frac{\partial}{\partial s_i} m_j(s)} \bigg|_{s = (X_{t1}' \beta_1, X_{t2}' \beta_2)} \equiv \frac{\xi_{ti}}{\lambda_t}, \quad (11)$$

where the sub-index j is denoted as the other player, i.e., j = -i, and  $\xi_{ti}$  and  $\lambda_t$  are defined as the numerator and denominator, respectively. Let  $\varphi_i(t) = \mathbb{E}[Y_i|(X'_1\beta_1, X'_2\beta_2) = t] \cdot f_{X'_1\beta_1, X'_2\beta_2}(t)$  and  $\psi(t) = \mathbb{E}[Y_1Y_2|(X'_1\beta_1, X'_2\beta_2) = t] \cdot f_{X'_1\beta_1, X'_2\beta_2}(t)$ . Then the partial derivatives in (11) can be rewritten as: for  $i, j \in \{1, 2\}$ ,

$$\begin{split} \frac{\partial m_i(s)}{\partial s_j} &= \frac{1}{f_{X_1'\beta_1, X_2'\beta_2}^2(s)} \cdot \Big[ \frac{\partial \varphi_i(s)}{\partial s_j} \cdot f_{X_1'\beta_1, X_2'\beta_2}(s) - \varphi_i(s) \cdot \frac{\partial f_{X_1'\beta_1, X_2'\beta_2}(s)}{\partial s_j} \Big], \\ \frac{\partial g_{12}(s)}{\partial s_j} &= \frac{1}{f_{X_1'\beta_1, X_2'\beta_2}^2(s)} \cdot \Big[ \frac{\partial \psi(s)}{\partial s_j} \cdot f_{X_1'\beta_1, X_2'\beta_2}(s) - \psi(s) \cdot \frac{\partial f_{X_1'\beta_1, X_2'\beta_2}(s)}{\partial s_j} \Big]. \end{split}$$

Moreover, we can estimate the partial derivatives  $\frac{\partial}{\partial s_j}\varphi_i(X'_{1t}\beta_1, X'_{2t}\beta_2)$  and  $\frac{\partial}{\partial s_i}\psi(X'_{1t}\beta_1, X'_{2t}\beta_2)$  respectively by

$$\hat{\varphi}_{tij} = \frac{1}{n-1} \sum_{s \neq t} Y_{si} \cdot \frac{\partial}{\partial u_j} \mathcal{K}_{\phi} \big( (X_{s1} - X_{t1})' \tilde{\beta}_1, (X_{s2} - X_{t2})' \tilde{\beta}_2 \big), 
\hat{\psi}_{ti} = \frac{1}{n-1} \sum_{s \neq t} Y_{s1} Y_{s2} \cdot \frac{\partial}{\partial u_i} \mathcal{K}_{\phi} \big( (X_{s1} - X_{t1})' \tilde{\beta}_1, (X_{s2} - X_{t2})' \tilde{\beta}_2 \big).$$

Thus, we define our estimator of  $\phi_i^*(X_t)$  by

$$\hat{\phi}_{ti}^* = rac{\hat{\xi}_{ti} \cdot \hat{f}_t^4 + \check{\delta}_{ti1}}{\hat{\lambda}_t \cdot \hat{f}_t^4 + \check{\delta}_t},$$

where  $\hat{\xi}_{ti}$  and  $\hat{\lambda}_t$  are nonparametric estimators of  $\xi_{ti}$  and  $\lambda_t$ , respectively, defined as: for j = -i,

$$\hat{\xi}_{ti} = \frac{1}{\hat{f}_t^4} \cdot \begin{vmatrix} \hat{\psi}_{ti}\hat{f}_t - \hat{\psi}_t\hat{f}_{ti} & \hat{\varphi}_{tji}\hat{f}_t - \hat{\varphi}_{tj}\hat{f}_{ti} \\ \hat{\psi}_{tj}\hat{f}_t - \hat{\psi}_t\hat{f}_{tj} & \hat{\varphi}_{tjj}\hat{f}_t - \hat{\varphi}_{tj}\hat{f}_{tj} \end{vmatrix}, \quad \hat{\lambda}_t = \frac{1}{\hat{f}_t^4} \cdot \begin{vmatrix} \hat{\varphi}_{t11}\hat{f}_t - \hat{\varphi}_{t1}\hat{f}_{t1} & \hat{\varphi}_{t21}\hat{f}_t - \hat{\varphi}_{t2}\hat{f}_{t1} \\ \hat{\varphi}_{t12}\hat{f}_t - \hat{\varphi}_{t1}\hat{f}_{t2} & \hat{\varphi}_{t22}\hat{f}_t - \hat{\varphi}_{t2}\hat{f}_{t2} \end{vmatrix},$$

and  $\check{\delta}_{ti1}$  and  $\check{\delta}_t$  are trimming sequences to be formally defined later. Note that we can drop the trimming sequences in the  $\hat{\phi}_{ti}^*$  if  $\lambda_t \cdot f_t^4$  is assumed to be uniformly bounded below from zero.

To establish the asymptotics for  $\hat{\phi}_{ij}^*$ , we make the following assumptions.

**Assumption I.**  $\tilde{\beta}_i = \beta_i + O_p(n^{-1/2})$  for i = 1, 2.

**Assumption J.** Functions  $m_i(s)$  and  $g_{ij}(s)$  are continuously differentiable for all  $i, j \in \{1, 2\}$ . Moreover, all partial derivatives of  $m_i(\cdot)$  and  $g_{ij}(\cdot)$  are uniformly bounded, i.e.

$$\{\|D_s m_i(\cdot)\|, \|D_s g_{ij}(\cdot)\|\} < c, \text{ for some } c > 0.$$

**Assumption K.** The bandwidth  $h_{\phi}$  satisfies  $h_{\phi} = (\ln n/n)^{1/12}$ .

Assumption L. The kernel function  $K_{\phi} : \mathbb{R}^2 \to \mathbb{R}$  is a symmetric Parzen–Rosenblatt kernel that satisfies:  $\max \left\{ \|D_u^r K(u)\|, \int \|D_u^r K(u)\| \, du \right\} < c \text{ for } r = 0, 1, 2, 3, 4 \text{ and}$  $\int u_1^{r_1} u_2^{r_2} K_{\phi}(u) \, dx = 0 \quad \text{if } 1 \leq r_1 + r_2 \leq 3,$  $< \infty \quad \text{if } r_1 + r_2 = 4.$ 

The support of  $K_{\phi}$  is a convex subset of  $\mathbb{R}$  with nonempty interior, with the origin as an interior point.

**Assumption M.** For some  $\epsilon > 0$ , let  $\check{\delta}_{ti1} \equiv h_{\phi}^{\epsilon} \cdot \frac{e^{z_{ti}}}{1+e^{z_{ti}}}$  and  $\check{\delta}_t \equiv h_{\phi}^{\epsilon} \cdot \frac{e^{z'_t}}{1+e^{z'_t}}$ , where  $z_{ti} \equiv h_{\phi}^{-\epsilon/4} \cdot (h_{\phi}^{\epsilon/5} - \hat{\xi}_{ti})$  and  $z'_t \equiv h_{\phi}^{-\epsilon/4} \cdot (h_{\phi}^{\epsilon/5} - \hat{\lambda}_t)$ .

Assumptions J to L are standard in the kernel derivative estimation literature (see e.g. Pagan and Ullah, 1999) and assumption M simply follows Klein and Spady (1993), which adjusts estimates for those observations at which numerator and denominator of estimated probabilities tend to zero. Note that assumption L requires a high order kernel since we demand the second stage estimator to uniformly converge at a rate faster than  $n^{1/4}$ , which is crucial for the final estimator of  $\alpha_i$  to have the regular  $\sqrt{n}$ -convergence rate.

**Lemma 3.** Suppose eq. (11), assumptions C, E, G and I to M hold. Let  $\mathscr{V}_n \equiv \{x : \lambda(x, \beta) > h_{\phi}^{\epsilon/5}\}$ . Then, for i = 1, 2,

$$\sup_{t} \left| \hat{\phi}_{ti}^* - \phi_i^*(X_t) \right| \cdot \mathbb{1}(X_t \in \mathscr{V}_n) = O_p \left( h_{\phi}^{-\epsilon} (\ln n/n)^{-1/3} \right).$$

Proof. See Appendix A.3

Note that the uniform convergence results only hold for  $X_t \in \mathcal{V}_n$ , which converges to the support of X as *n* goes large. Therefore, in the next stage estimation, we will downweight those observations outside of  $\mathcal{V}_n$ . Unlike the likelihood trimming in Klein and Spady (1993), this additional trimming does not depend on the unknown parameter to be estimated in the next stage. Hence, it does not cause essential difficulties to the next stage estimation. In

principal, we can choose  $\epsilon$  arbitrary small such that the uniform convergence of the beliefs estimator is close to the optimal rate.

4.2. Estimation of  $\alpha_i$  and  $C_{U}$ . We first discuss the estimation of the strategic component  $\alpha_i$ . Let  $A_i \subseteq \mathbb{R}$  be the parameter space for  $\alpha_i$ . Moreover, we denote  $\theta_i = (\alpha_i, \beta'_i)'$ . Let further  $\Theta_i = A_i \times B_i$  and  $\Theta = \Theta_1 \times \Theta_2$ .

Following Klein and Spady (1993), we define our objective function by

$$\hat{Q}_{ni}(c_i;\hat{\tau}) = \frac{1}{n} \sum_{t=1}^n (\hat{\tau}_t/2) \cdot \left\{ \left[ Y_{ti} \ln \hat{p}_{ti}^2(c_i) + (1 - Y_{ti}) \ln(1 - \hat{p}_{ti}(c_i))^2 \right] \right\}$$

where

$$\hat{p}_{ti}(c_i) = \frac{\sum_{s \neq t} \left[ Y_{si} \cdot \mathcal{K}_{\alpha} \left( (X_{si} - X_{ti})' b_i + a_i (\hat{\phi}_{si}^* - \hat{\phi}_{ti}^*) \right) \right] + \hat{\delta}_{t1}(c_i)}{\sum_{s \neq t} \mathcal{K}_{\alpha} \left( (X_{si} - X_{ti})' b_i + a_i (\hat{\phi}_{si}^* - \hat{\phi}_{ti}^*) \right) + \hat{\delta}_t(c_i)}$$

and  $\hat{\tau}_t$ ,  $\hat{\delta}_{t1}$ ,  $\hat{\delta}_t$  are trimming sequences defined similarly to Klein and Spady (1993), and  $\mathcal{K}_{\alpha} = K_{\alpha}(u/h_{\alpha})/h_{\alpha}$  with  $K_{\alpha} : \mathbb{R} \to \mathbb{R}$  and  $h_{\alpha}$  as Parzen–Rosenblatt kernel function and bandwidth, respectively. The proposed estimator is then given by

$$\hat{\theta}_i = \operatorname{argmax}_{c_i \in \Theta_i} \hat{Q}_{ni}(c_i; \hat{\tau}).$$
(12)

Note that our Klein–Spady type estimator uses the generated regressor  $\hat{\phi}_i^*$  that uniformly converges to equilibrium beliefs  $\phi_i^*$ ; see Lemma 3. Then, we can deal with the approximation errors in a similar way to e.g. Guerre, Perrigne, and Vuong (2000).

Here we introduce some notation and briefly motivate our estimator. For i = 1, 2, let  $v_i(c_i) = X'_i b_i + a_i \phi_i^*(X)$  and  $v_{ti}(c_i) = X'_{ti} b_i + a_i \phi_i^*(X_t)$ . For y = 0, 1, let  $g_i(y, v_i(c_i)) \equiv \mathbb{P}(Y_i = y | X'_i b_i + a_i \phi_i^*(X)) \cdot f_{v_i(c_i)}(v_i(c_i))$  and  $g_i(v_i(c_i)) \equiv \sum_{y \in \{0,1\}} g_i(y, v_i(c_i)) = f_{v_i(c_i)}(v_i(c_i))$ . Then we have

$$p_i(c_i) \equiv \mathbb{E}(Y_i|\nu_i(c_i)) = \frac{g_i(1,\nu_i(c_i))}{g_i(\nu_i(c_i))}.$$

Let further

$$\hat{g}_i(y, \nu_{ti}(c_i)) = \frac{1}{n-1} \sum_{s \neq t} \left[ \mathbb{1}(Y_{si} = y) \cdot \mathcal{K}_{\alpha} \left( (X_{si} - X_{ti})' b_i + a_i (\hat{\phi}_{si}^* - \hat{\phi}_{ti}^*) \right) \right],$$

and  $\hat{g}_i(\nu_{ti}(c_i)) = \sum_{y \in \{0,1\}} \hat{g}_i(y, \nu_{ti}(c_i))$ . By definition,  $\hat{g}_i(y, \nu_{ti}(c_i))$  and  $\hat{g}_i(\nu_{ti}(c_i))$  nonparametrically estimate  $g_i(y, \nu_{ti}(c_i))$  and  $g_i(\nu_{ti}(c_i))$ , respectively. Therefore, we estimate  $\mathbb{E}(Y_{ti}|\nu_{ti}(c_i))$  by

$$\hat{p}_{ti}(c_i) = \frac{\hat{g}_i(1, \nu_{ti}(c_i)) + \hat{\delta}_{t1}(c_i)}{\hat{g}_i(\nu_{ti}(c_i)) + \hat{\delta}_t(c_i)}.$$

The trimming sequences  $\hat{\delta}_t(c_i)$  and  $\hat{\delta}_{t1}(c_i)$  will be defined similarly as those in Klein and Spady (1993) to ensure the uniform convergence of  $\hat{p}_i(c_i)$  to  $p_i(c_i)$ .

**Assumption N.** The parameter space  $\Theta$  is compact and the support  $\mathscr{S}_X$  is bounded. Moreover, the true parameter  $\theta$  belongs to the interior of  $\Theta$ .

**Assumption O.** For i = 1, 2, there exists no proper linear subspace of  $\mathbb{R}^{d+1}$  have probability 1 under the probability distribution  $F_{X_i,\phi_i^*(X)}$ .

The first half of assumption N ensures that choice probabilities are bounded away from zero so that the likelihood function is bounded. The second half is standard in the literature. Assumption O strengths conditions in assumption E. In particular, if the support of  $F_{X_{i1}|X_{-i1}}$  is an interval with the length larger than 1, then assumption E implies assumption O. Since the variations of  $\phi_i^*(X)$  are less than 1.

**Assumption P.** Let  $h_{\phi}^{-\epsilon'} = o_p(n^{-1/6})$  with  $h_{\phi}$  and  $\epsilon'$  defined in assumption K and Lemma 3 respectively.

Assumption P implies that the beliefs estimator  $\hat{\phi}_i^*$  should uniformly (over  $\mathscr{V}_n$ ) converge to  $\phi_i^*$  at a rate no slower than  $n^{1/4}$ , which is required for the  $\sqrt{n}$ -consistency of  $\hat{\alpha}_i$ .

Assumption Q. The bandwidth  $h_{\alpha}$  satisfies  $n^{-1/(6+2\iota)} < h_{\alpha} < n^{-1/8}$  for some small  $\iota > 0$ .

**Assumption R.** The kernel  $K_{\alpha}(\cdot) : \mathbb{R} \to \mathbb{R}$  is a symmetric function that integrates to one, satisfies  $\int u^2 K_{\alpha}(u) du = 0$  and  $\int u^4 K_{\alpha}(u) du < \infty$ , and for some c > 0,

$$\max\left\{ \left| D_{u}^{r}K_{\alpha}(u) \right|, \int \left| D_{u}^{r}K_{\alpha}(u) \right| du \right\} < c, \quad (r = 0, 1, 2, 3, 4).$$

The support of  $K_{\alpha}$  is a convex subset of  $\mathbb{R}$  with nonempty interior, with the origin as an interior point.

**Assumption S.** Let  $\hat{\delta}_{ty}(c_i) \equiv h_{\phi}^{\iota} \cdot \frac{e^{\omega(c_i)}}{1+e^{\omega(c_i)}}$  with y = 0, 1, and  $\omega(c_i) = h_{\phi}^{-\iota/4} \cdot (h_{\phi}^{\iota/5} - \hat{g}_{yt}(c_i))$ . Let further  $\hat{\delta}_t = \hat{\delta}_{t0} + \hat{\delta}_{t1}$ .

For 
$$\iota > 0$$
 and  $z \in \mathbb{R}^+$ , let  $\tau(z, \iota) \equiv \left\{1 + \exp\left[(h_{\alpha}^{\iota/5} - z)/h_{\alpha}^{\iota/4}\right]\right\}^{-1}$ .

**Assumption T.** The trimming sequence employed to down weight observations has the form  $\hat{\tau}_t \equiv \hat{\tau}_{t0} \hat{\tau}_{t1} \hat{\tau}_t^{\phi}$ , where

$$\hat{\tau}_{ty} = \tau(\hat{g}_i(y, \hat{\nu}_{ti}(\theta_i^P)), \iota'), \quad \hat{\tau}_t^{\phi} = \tau(\hat{\lambda}(X_t; \tilde{\beta}_i), \iota'),$$

in which  $\theta_i^P$  is a preliminary consistent estimator satisfying  $\|\theta_i^P - \theta_i\| = O_p(n^{-1/3})$  and  $\iota' \in (0, \iota)$  with  $\iota$  defined in assumption S.

Assumptions Q to S simply follow Klein and Spady (1993). Assumption T is modified from Klein and Spady (1993): we introduce the trimming sequence  $\hat{\tau}_t^{\phi}$  for those observations with a small denominator (i.e. less than  $h_{\alpha}^{t'/5}$ ) in the estimation of  $\phi_i^*$ . It should be noted that this trimming does not depend on the current stage estimator  $\hat{\theta}_i$ .

Theorem 1. Suppose the conditions in Lemma 3 and assumptions N to T hold. Then

$$\sqrt{n}(\hat{\theta}_i - \theta_i) \stackrel{d}{\longrightarrow} N(0, \Sigma),$$

where

$$\Sigma \equiv \mathbb{E} \left\{ \frac{f_{U_i}^2(u_i^*(X)) \cdot (\phi_i^*(X), X_i')'(\phi_i^*(X), X_i')}{F_{U_i}(u_i^*(X)) \left[1 - F_{U_i}(u_i^*(X))\right]} \right\}^{-1}$$

Proof. See Appendix A.4

Lastly, we turn to the estimation of the copula function  $C_U$  by using ??. Our estimation strategy follows Guerre, Perrigne, and Vuong (2000)'s two–step approach: First, we estimate  $(\mathbb{E}(Y_{t1}|X_t), \mathbb{E}(Y_{t2}|X_t))$  by  $\hat{m}_t \equiv (\hat{m}_{t1}, \hat{m}_{t2})'$ :

$$\hat{m}_{ti} = rac{\hat{\hat{\phi}}_{ti} + \hat{\hat{\delta}}_{ti1}}{\hat{\hat{f}}_t + \hat{\hat{\delta}}_t}, \hspace{0.2cm} i = 1,2$$

where  $\hat{\delta}_{ti1}$  and  $\hat{\delta}_t$  as trimming sequences and

$$\hat{\phi}_{ti} = \frac{1}{n-1} \sum_{s \neq t} Y_{si} \cdot \mathcal{K}_m \big( (X_{s1} - X_{t1})' \hat{\beta}_1, (X_{s2} - X_{t2})' \hat{\beta}_2 \big),$$
$$\hat{f}_t = \frac{1}{n-1} \sum_{s \neq t} \mathcal{K}_m \big( (X_{s1} - X_{t1})' \hat{\beta}_1, (X_{s2} - X_{t2})' \hat{\beta}_2 \big).$$

in which  $\mathcal{K}_m = K_m(u/h_m)/h_m$  with a bandwidth  $h_m$  and a kernel function  $K_m : \mathbb{R}^2 \to \mathbb{R}$ . Note that we could use  $\tilde{\beta}_i$  instead of  $\hat{\beta}_i$  in the estimates  $\hat{\phi}_{ti}$  and  $\hat{f}_t$ , which does not change the asymptotic properties of  $\hat{C}_U$  established below as long as  $\tilde{\beta}_i$  is  $\sqrt{n}$ -consistent.

Next, we define our estimator of  $C_U$  as follows: for each  $p \in \mathscr{S}^{\circ}_{\mathbb{E}(Y|X)}$  such that  $f_{\mathbb{E}(Y|X)}(p) > 0$ , let

$$\hat{C}_{U}(p) = \frac{\sum_{t=1}^{n} Y_{t1} Y_{t2} \mathcal{K}_{c}(\hat{m}_{t} - p)}{\sum_{t=1}^{n} \mathcal{K}_{c}(\hat{m}_{t} - p)}$$

where  $\mathcal{K}_c = K_c(u/h_c)/h_c$  with a bandwidth  $h_c$  and a kernel function  $K_c : \mathbb{R}^2 \to \mathbb{R}$ . Clearly, the nonparametric estimator  $\hat{C}_U$  is an asymptotically biased estimator of  $C_U$  at the boundaries of the support  $\mathscr{S}_{\mathbb{E}(Y|X)}$ . Because this boundary effect, we need to exclude the estimates near the boundaries.

We follow Guerre, Perrigne, and Vuong (2000) to establish the asymptotics of the nonparametric estimator  $\hat{C}_U$ .

**Assumption U.** For some small positive number  $\epsilon''$  with  $0 < \epsilon'' < R - 2$ , let  $\hat{\delta}_{ti1} \equiv h_m^{\epsilon''} \cdot \frac{e^{\zeta_{ti1}}}{1 + e^{\zeta_{ti1}}}$  and  $\hat{\delta}_t \equiv h_m^{\epsilon''} \cdot \frac{e^{\zeta_t}}{1 + e^{\zeta_t}}$ , where  $\zeta_{ti1} = h_m^{-\epsilon''/4} \cdot (h_m^{\epsilon''/5} - \hat{\phi}_{ti})$  and  $\zeta_t = h_m^{-\epsilon''/4} \cdot (h_m^{\epsilon''/5} - \hat{f}_t)$ .

**Assumption V.** The copula function  $C_U$  and marginal quantile function  $F_{U_i}^{-1}$  is R + 1 times continuously differentiable with bounded derivatives on the support  $\mathscr{S}_{\mathbb{E}}(Y|X)$  and  $\mathscr{S}_{Y_i|X}$ , respectively, and with  $R \geq 2$ .

**Assumption W.** The kernels  $K_m(\cdot)$  and  $K_c(\cdot)$  are symmetric functions that integrate to one with bounded support and twice continuously bounded derivatives. Moreover,

$$\int u_1^{r_1} u_2^{r_2} K(u) dx = 0 \quad if \ 1 \le r_1 + r_2 \le R,$$
  
<\iims if \texts if \texts r\_1 + r\_2 = \texts R + 1.

for  $K = K_m$  and  $K = K_c$ .

**Assumption X.** The bandwidths  $h_m$  and  $h_c$  are of the form:  $h_m, h_c \propto (\ln n/n)^{1/(2R+4)}$ .

Assumption U defines the trimming sequence, which is needed when the denominator is close to zero. Alternatively, we can drop assumption U by focusing on a compact subset of the support  $\mathscr{S}_{\mathbb{E}(Y|X)}$  such that the density  $f_{X'_1\beta_1,X'_2\beta_2}$  is bounded away from zero. Assumptions V to X are standard in the kernel estimation literature; see e.g. Pagan and Ullah (1999). Note that applying the implicit function theorem to eq. (5), assumption V implies that  $m_i(\cdot)$  is also R + 1 times continuously differentiable with bounded derivatives on the support  $\mathscr{S}_{\mathbb{E}(Y|X)}$ .

**Theorem 2.** Suppose conditions in Theorem 1 and assumptions U to X hold. Then, for any closed inner subset P of  $\mathscr{S}_{\mathbb{E}(Y|X)}$  such that  $\inf_{p \in P} f_{\mathbb{E}(Y|X)}(p) > 0$ , we have

$$\sup_{p\in P} |\hat{C}_u(p) - C_u(p)| = O((\ln n/n)^{(R-\epsilon'')/(2R+4)}).$$

Proof. See Appendix A.5.

When the marginal choice probabilities  $m_i(X)$  are observed, the optimal uniform convergence rate for estimating  $C_U$  is  $(n/\ln n)^{R/(2R+I)}$  with I = 2; (see e.g. Stone, 1982). In our case, because the conditioning variable  $m_i(X_t)$  is nonparametrically estimated, then

we don't obtain the optimal uniform convergence rate for our estimator  $\hat{C}_U$ . Instead,  $\hat{C}_U$  converges at a rate (i.e.  $(n/\ln n)^{(R-\epsilon'')/(2R+I+2)}$ ) that is slightly slower than the optimal one. This result is also consistent with Guerre, Perrigne, and Vuong (2000).

### 5. MONTE CARLO SIMULATIONS

In this section, we use Monte Carlo experiments to illustrate the performance of our estimator in finite samples. Let I = 2,  $d_1 = d_2 = 2$  and  $X_i \in \mathbb{R}^2$  for i = 1, 2, where  $X_i$  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  $\rho = 0.5$ . All results are based on 1000 replications.

Moreover, let  $\beta_i = (1,1)'$  and  $\alpha_i = 1$  for i = 1,2. 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 i chooses 1 whenever her private type  $u_i \le u_i^*(x)$ . We compute  $u_i^*(\cdot)$  by solving the following equations for each  $X_t$  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,  $\beta_i$  is obtained by the recipe of Klein and Spady (1993). Specifically, we use second order biweight kernel and choose bandwidth according to Silverman's rule-of-thumb. Table 2 reports the summary statistics for  $\tilde{\beta}_1$ , including the sample mean (MEAN), median (MEDIAN), the standard deviation (SD), and root mean squared error (RMSE). It shows that the first–step estimator of  $\beta_1$  performs relatively well. As sample size increases, both the bias and standard deviation decreases.

п	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

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

For the estimation of  $\phi_i^*$ , we employ the fourth order biweight product kernel, i.e.,  $K_{\phi}(u_1, u_2) = k_{\phi}(u_1) \cdot k_{\phi}(u_2)$  where  $k_{\phi}(u_i) = \frac{7}{4}(1 - 3u_i^2) \cdot \frac{15}{16}(1 - u_i^2)^2 \cdot \mathbf{1}(|u_i| \le 1)$ and choose  $h_{\phi} = 4.40 \cdot \hat{\sigma} \cdot (n/\log(n))^{-1/10}$  where  $\hat{\sigma}$  is the estimated standard error of the regressor.

Figure 1 plots  $\phi_1^*$ ,  $\phi_2^*$  and their kernel estimates under sample size of n = 1000. For presentation purpose, we fix  $x_1 = (0,0)$ , but a similar pattern holds for other values of  $x_1$ . Figures 1a and 1b show functions  $\phi_1^*$  and  $\phi_2^*$ , and their estimates. Figure 1c provides the infeasible estimate of  $\phi_1^*$ , which uses the true value of  $(\beta_1, \beta_2)$  in estimation, and our estimate of  $\phi_1^*$ . It shows that the mean of these two estimates of belief  $\phi_1^*$  coincides.

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

Tables 3 and 4 report the finite sample performance of  $\hat{\alpha}_1$  and  $\hat{\beta}_1$ , respectively. The finite sample performance of  $\hat{\alpha}_2$  and  $\hat{\beta}_2$  are similar and hence omitted here. In Tables 3 and 4, the infeasible estimator of  $\alpha_1$  or  $\beta_1$  is obtained by plugging in the true values of equilibrium beliefs  $\phi_1^*$  in the second step estimation. From these two tables, our estimator behaves well under finite sample sizes. As sample size increases, both the bias and standard deviation decrease as expected, and the standard deviation decreases roughly at the  $\sqrt{n}$  rate. In addition, the performance of feasible estimator approaches the infeasible one when the sample size increases. Finally, the second step estimator of  $\beta_1$  has some improvement from the first step one when the sample size is small.



(A) True and estimated belief  $\phi_1^*(\cdot)$  (B) True and estimated belief  $\phi_2^*(\cdot)$  (C) Feasible v.s. infeasible  $\hat{\phi}_1^*(\cdot)$ 

FIGURE 1. Kernel estimates of  $\phi_1^*(\cdot)$  and  $\phi_2^*(\cdot)$  with  $x_1 = (0, 0)$ .

	Our Estimator				Infeasible Estimator					
п	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
250	1.00	0.9555	0.9409	0.4317	0.4337	1.00	0.9917	0.9778	0.3440	0.3440
500	1.00	0.9938	0.9927	0.3037	0.3037	1.00	1.0128	1.0164	0.2436	0.2438
1000	1.00	0.9857	0.9801	0.2075	0.2079	1.00	1.0054	1.0028	0.1678	0.1678
2000	1.00	0.9937	0.9953	0.1422	0.1422	1.00	1.0005	0.9992	0.1101	0.1101

TABLE 3. Mean, median, SD and RMSE for estimating  $\alpha_1$ 

TABLE 4. Mean, median, SD and RMSE for estimating  $\beta_1$  in last step

	Our Estimator				Infeasible Estimator					
п	TRUE	MEAN	MEDIAN	SD	RMSE	TRUE	MEAN	MEDIAN	SD	RMSE
250	1.00	1.0098	0.9995	0.1732	0.1734	1.00	1.0078	1.0013	0.1579	0.1580
500	1.00	1.0056	0.9957	0.1147	0.1148	1.00	1.0046	1.0007	0.1079	0.1080
1000	1.00	1.0037	0.9981	0.0832	0.0832	1.00	1.0042	0.9994	0.0777	0.0778
2000	1.00	1.0034	1.0010	0.0547	0.0548	1.00	1.0035	1.0013	0.0509	0.0510

We finally provide the estimate of copula  $C_U$  in Figure 2 with  $p_2 = 0.5$ . It is obtained under the sample size of n = 1000. It shows that our copula estimate behaves well. In particular, the point–wise average is close to the true copula, and the 90% confidence band roughly shows the shape of the true copula function. Clearly, our estimator is less precise when it comes close to the boundaries.



FIGURE 2. True and estimated copula  $C_U$  with  $p_2 = 0.5$ .

# REFERENCES

- AGUIRREGABIRIA, V., AND P. MIRA (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.
- 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.
- 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.
- BROCK, W., AND S. DURLAUF (2001): "Discrete choice with social interactions," *Review* of Economic Studies, 68(2), 235–260.
- BULOW, J. I., J. D. GEANAKOPLOS, AND P. D. KLEMPERER (1985): "Multimarket oligopoly: Strategic substitutes and complements," *The Journal of Political Economy*, pp. 488–511.
- CASTRO, L. I. D. (2007): "Affiliation, equilibrium existence and the revenue ranking of auctions," Discussion paper.
- EASLEY, D., AND J. KLEINBERG (2010): *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press.
- FUDENBERG, D., AND J. TIROLE (1991): Game Theory, 1991. MIT Press.
- GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.
- HOROWITZ, J. L. (1998): Semiparametric methods in econometrics. Springer.
- ICHIMURA, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58(1), 71–120.
- ICHIMURA, H., AND L.-F. LEE (1991): "Semiparametric least squares estimation of multiple index models: single equation estimation," in *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge*, pp. 3–49.

- KASY, M. (2012): "Nonparametric inference on the number of equilibria," Work. Pap., Harvard Univ., Cambridge, MA.
- KLEIN, R., AND R. SPADY (1993): "An efficient semiparametric estimator for binary response models," *Econometrica*, 61(2), 387–421.
- LEWBEL, A. (1998): "Semiparametric latent variable model estimation with endogenous or mismeasured regressors," *Econometrica*, pp. 105–121.
- LIU, N., Q. VUONG, AND H. XU (2013): "Rationalization and Identification of Discrete Games with Correlated Types," Discussion paper.
- MANSKI, C. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3(3), 205–228.
- MANSKI, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60(3), 531–542.
- MCPHERSON, M., L. SMITH-LOVIN, AND J. M. COOK (2001): "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444.
- MILGROM, P., AND C. SHANNON (1994): "Monotone comparative statics," *Econometrica*, pp. 157–180.
- NEWEY, W. K., AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press.
- POWELL, J., J. STOCK, AND T. STOKER (1989): "Semiparametric estimation of index coefficients," *Econometrica: Journal of the Econometric Society*, pp. 1403–1430.
- STONE, C. J. (1982): "Optimal global rates of convergence for nonparametric regression," *The Annals of Statistics*, pp. 1040–1053.
- TAMER, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, 70(1), 147–165.
- VAN DE GEER, S. (1990): "Estimating a regression function," *The Annals of Statistics*, pp. 907–924.

- WAN, Y., AND H. XU (2014): "Semiparametric identification of binary decision games of incomplete information with correlated private signals," *Journal of Econometrics*, 182(2), 235–246.
- XU, H. (2014): "Estimation of discrete games with correlated types," *The Econometrics Journal*.

### APPENDIX A. PROOFS OF IDENTIFICATION RESULTS

A.1. **Proof of Lemma 1.** The proof simply follows (Wan and Xu, 2014, Lemma 1).

#### A.2. Proof of Lemma 2.

*Proof.* Our proof follows the identification argument in Klein and Spady (1993). First, we show that  $\mathbb{P}(Y_i = 1 | X'_i \beta_i = s_i, X_{-i} = x_{-i})$  is strictly increasing in  $s_i$  for any  $x_{-i} \in \mathscr{S}_{X_{-i}}$ . Note that

$$\mathbb{P}(Y_i = 1 | X'_i \beta_i = s_i, X_{-i} = x_{-i}) = \mathbb{P}(Y_i = 1 | X'_i \beta_i = s_i, X'_{-i} \beta_{-i} = x'_{-i} \beta_{-i})$$
$$= F_{U_i}(u_i^*(s_i, x'_{-i} \beta_{-i})).$$

Moreover, by eq. (3) we have

$$s_i + \alpha_i \sum_{j \neq i} \mathbb{P}[U_j \le u_j^*(s) | X = x, U_i = u_i^*(s)] - u_i^*(s) = 0$$

which has an invertible Jacobian matrix under assumption A. To see this, note that the Jacobian matrix is strictly positive (resp. negative) when *I* is even (resp. odd). Therefore,  $\partial u_j^*(s)/\partial s_j > 0$ . It follows that  $\mathbb{P}(Y_i = 1 | X'_i \beta_i = s_i, X_{-i} = x_{-i})$  is strictly increasing in  $s_i$ .

Next, suppose  $\tilde{\beta}_i$  and  $\beta_i$  are observationally equivalent. Because

$$\mathbb{P}(Y_i = 1 | X = x) = \mathbb{P}(Y_i = 1 | X'_i \beta_i = x'_i \beta_i, X_{-i} = x_{-i})$$

where the LHS is identified. Then for any  $(x_i, x_{-i}), (\bar{x}_i, x_{-i}) \in \mathscr{S}_X$ , we have

$$x'_i \tilde{\beta}_i \geq \bar{x}'_i \tilde{\beta}_i$$
 if and only if  $x'_i \beta_i \geq \bar{x}'_i \beta_i$ .

In other words, there exists a strict increasing function  $T : \mathbb{R} \to \mathbb{R}$ , such that

$$x_i'\tilde{\beta}_i = T(x_i'\beta_i), \quad \forall x_i \in \mathscr{S}_{X_i}.$$

By assumption E, T has to be the identity function. Therefore, we have  $X'_i\beta_i = X_i\tilde{\beta}_i$ . By assumptions D and F,  $\tilde{\beta}_i = \beta_i$ .

#### A.3. Proof of Lemma 3.

*Proof.* Note that the kernel estimators are defined by leaving the *t*-th observation out, which is common in the literature. Such a modification does not make a difference for the asymptotic behavior of the estimator. For expositional brevity, in what follows we will denote our estimator of  $\phi_i^*(x)$  by the standard kernel estimator (i.e. without leaving one observation out).

Moreover, here we abuse our notation a little bit without causing any confusion. Let  $\phi_i^*(x; b)$  be the equilibrium beliefs under the generic parametric value  $b \in B$ , i.e.,

$$\phi_i^*(X;b) = \frac{\frac{\partial}{\partial s_i}g_{12}(s;b) \cdot \frac{\partial}{\partial s_j}m_j(s;b) - \frac{\partial}{\partial s_j}g_{12}(s;b) \cdot \frac{\partial}{\partial s_i}m_j(s;b)}{\frac{\partial}{\partial s_i}m_j(s;b) \cdot \frac{\partial}{\partial s_j}m_j(s;b) - \frac{\partial}{\partial s_j}m_i(s;b) \cdot \frac{\partial}{\partial s_i}m_j(s;b)}\Big|_{s=(X_1'b_1, X_2'b_2)}$$

where  $m_i(s;b) = \mathbb{E}(Y_i|X'_1b_1 = s_1;X'_2b_2 = s_2)$  and  $g_{12}(s;b) = \mathbb{E}(Y_1Y_2|X'_1b_1 = s_1;X'_2b_2 = s_2)$ . Let  $\hat{\phi}^*_i(x;b)$  be the nonparametric estimator  $\hat{\phi}_i(x)$  by replacing  $\tilde{\beta}$  with *b* (including the trimming part).

Next step is standard. Note that conditional on  $X_i \in \mathcal{V}_n$ , the denominator of  $\hat{\phi}_i^*(x; b)$  is large enough and the adjustment trimming sequence tends exponentially to zero, and hence negligible. By e.g. Klein and Spady (1993, Lemmas 2–4), we have uniformly on  $\mathcal{V}_n$  and B:

$$h_{\phi}^{\epsilon} |\hat{\phi}_{i}^{*}(x;b) - \phi_{i}^{*}(x;b)| = O_{p} \left( \left( \frac{\ln n}{nh_{\phi}^{4}} \right)^{1/2} + h_{\phi}^{4} \right) = O_{p} \left( (\ln n/n)^{1/3} \right).$$

Because  $\tilde{\beta}$  is a consistent estimator of  $\beta$  and the numerator of  $\psi_i^*$  is a smooth function with uniformly bounded derivatives (by assumption J), we have the following stochastic equicontinuity:

$$\hat{\phi}_i^*(x;\tilde{\beta}) = \hat{\phi}_i^*(x;\beta) + o_p(n^{-1/2})$$

uniformly on x. Thus, uniformly on  $\mathscr{V}_n$ ,

$$\left|\hat{\phi}_{i}^{*}(x;\tilde{\beta})-\phi_{i}^{*}(x;\beta)\right|=O_{p}\left(h_{\phi}^{-\epsilon}(\ln n/n)^{-1/3}\right).$$

It concludes the proof by noting that  $\hat{\phi}_{ti}(X_t) = \hat{\phi}_i^*(X_t; \tilde{\beta})$  and  $\phi_i^*(x; \beta) = \phi_i^*(x)$ .

A.4. **Proof of Theorem 1.** The consistency simply follows the uniform convergence of  $\hat{\phi}_i^*$  to  $\phi_i^*$  (noting that the trimming is negligible asymptotically) and the proof for Theorem 3 in Klein and Spady (1993), which is omitted here. We now focus on the asymptotic normality of its limiting distribution. By a similar argument, the Hessian metric converges in probability to its limit. Hence,

it remains to show the convergence of the gradient. For expositional simplicity, we suppress the subindex i in the following discussion.

Our proof is a modification of Klein and Spady (1993, Lemma 6). Let  $\bar{p}_t(\theta)$ ,  $\bar{\tau}_t$ ,  $\bar{g}(1, \nu_t(\theta))$ and  $\bar{g}(\nu_t(\theta))$  be infeasible estimators, which are defined as  $\hat{p}_t(\theta)$ ,  $\hat{\tau}_t$ ,  $\hat{g}(1, \nu_t(\theta))$  and  $\hat{g}(\nu_t(\theta))$ , respectively, with  $\hat{\phi}_t^*$  replaced by  $\phi_t^*$ . The gradient can be written as a weighted sum of residuals:

$$\hat{G}(\theta) = n^{-1} \sum_{t=1}^{n} \hat{\tau}_t \hat{r}_t \hat{w}_t$$

where  $\hat{r}_t \equiv [Y_t - \hat{p}_t(\theta)]/\hat{e}_t$ ,  $\hat{e}_t \equiv \hat{g}(\nu_t(\theta))[\hat{p}_t(\theta)(1 - \hat{p}_t(\theta))]$  and  $\hat{w}_t \equiv \hat{g}(\nu_t(\theta))[\partial \hat{p}_t(\theta)/\partial c]_{c=\theta}$ . Let further  $e_t \equiv [g(\nu_t(\theta)) + \delta_n(\nu_t(\theta))] \cdot [p_t(\theta)(1 - p_t(\theta))]$ ,  $r_t \equiv [Y_t - p_t(\theta)]/e_t$  and  $w_t \equiv g(\nu_t(\theta))[\partial p_t(\theta)/\partial c]_{c=\theta}$ . By definition,

$$\mathbb{E}(r_t|X_t) = \mathbb{E}(r_t|\nu_t(\theta)) = \mathbb{E}(w_t|\nu_t(\theta)) = 0.$$

Moreover, we define  $\bar{r}_t$ ,  $\bar{e}_t$  and  $\bar{w}_t$  as  $\hat{r}_t$ ,  $\hat{e}_t$  and  $\hat{w}_t$ , respectively, with  $\hat{p}_t(\theta)$ ,  $\hat{\tau}_t$ ,  $\hat{g}(1, \nu_t(\theta))$  and  $\hat{g}(\nu_t(\theta))$  replaced by  $\bar{p}_t(\theta)$ ,  $\bar{\tau}_t$ ,  $\bar{g}(1, \nu_t(\theta))$  and  $\bar{g}(\nu_t(\theta))$ , respectively.

Following Klein and Spady (1993), it suffices to show

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{\tau}_{t}\hat{\mu}_{t} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tau_{t}\mu_{t} = o_{p}(1).$$

Because

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{\tau}_{t} \hat{\mu}_{t} &- \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} \mu_{t} \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\hat{\mu}_{t} - \mu_{t}) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\hat{\tau}_{t} - \tau_{t}) \mu_{t} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\hat{\tau}_{t} - \tau_{t}) (\hat{\mu}_{t} - \mu_{t}) \equiv \mathbb{A} + \mathbb{B} + \mathbb{C}. \end{aligned}$$

Hence, it suffices to show  $\mathbb{A}$ ,  $\mathbb{B}$  and  $\mathbb{C}$  are all  $o_p(1)$ . Our proof is similar to Klein and Spady (1993) with small modification due to the generated regressor  $\hat{\phi}_t^*$ . Due to the similarity, here we only illustrate by showing  $\mathbb{A} = o_p(1)$ .

Note that  $\tau_i$  essentially restricts all densities to be no smaller than  $O(h_{\alpha}^{\iota''/5})$ ,  $o < \iota'' < \iota'$ . By definition,

$$\begin{split} \mathbb{A} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_t (\hat{\mu}_t - \mu_t) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_t (\hat{r}_t - r_t) w_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_t (\hat{r}_t - r_t) (\hat{w}_t - w_t) \\ &+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_t r_t (\hat{w}_t - w_t) \equiv \mathbb{A}_1 + \mathbb{A}_2 + \mathbb{A}_3. \end{split}$$

It follows that

$$\begin{split} \mathbb{A}_{1} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\hat{r}_{t} - \bar{r}_{t}) w_{t} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\bar{r}_{t} - r_{t}) w_{t} \equiv \mathbb{A}_{11} + \mathbb{A}_{12}, \\ \mathbb{A}_{3} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} r_{t} (\hat{w}_{t} - \bar{w}_{t}) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} r_{t} (\bar{w}_{t} - w_{t}) \equiv \mathbb{A}_{31} + \mathbb{A}_{32}, \end{split}$$

and

$$\begin{split} \mathbb{A}_{2} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\bar{r}_{t} - r_{t}) (\bar{w}_{t} - w_{t}) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\hat{r}_{t} - \bar{r}_{t}) (\bar{w}_{t} - w_{t}) \\ &+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\bar{r}_{t} - r_{t}) (\hat{w}_{t} - \bar{w}_{t}) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tau_{t} (\hat{r}_{t} - \bar{r}_{t}) (\hat{w}_{t} - \bar{w}_{t}) \equiv \mathbb{A}_{21} + \mathbb{A}_{22} + \mathbb{A}_{23} + \mathbb{A}_{24}. \end{split}$$

Klein and Spady (1993) show that  $\mathbb{A}_{11}$ ,  $\mathbb{A}_{21}$ ,  $\mathbb{A}_{31}$  are all  $o_p(1)$ . Hence, it suffices to show  $\mathbb{A}_{12}$ ,  $\mathbb{A}_{22}$ ,  $\mathbb{A}_{23}$ ,  $\mathbb{A}_{24}$ , and  $\mathbb{A}_{32}$  are all  $o_p(1)$ .

The proof for  $\mathbb{A}_{12}$  and  $\mathbb{A}_{32}$  simply follows Klein and Spady (1993), the proof for  $\mathbb{A}_1$  and  $\mathbb{A}_3$ , respectively. Moreover, note that

$$|\mathbb{A}_{22}| \leq n^{1/2} \cdot \sup_{t \leq n} \tau_t |\bar{w}_t - w_t| \cdot \sup_{t \leq n} \tau_t |\hat{r}_t - \bar{r}_t|.$$

By Klein and Spady (1993, Lemma 4), Lemma 4 and assumption Q, we have  $\mathbb{A}_{22} = o_p(1)$ . Similarly, we can show  $\mathbb{A}_{23} = o_p(1)$  and  $\mathbb{A}_{24} = o_p(1)$ .

All the rest simply follows Klein and Spady (1993, Theorem 3-4).

Lemma 4. Under the conditions stated in Theorem 1, we have

$$\sup_{t \le n} \tau_t |\hat{r}_t - \bar{r}_t| = o_p(n^{-1/2}h_{\alpha}^{-\iota-1}).$$

*Proof.* Without loss of generality, let t = 1. For expositional simplicity, in what follows we implicitly treat  $v_1$  as a constant and all expectations are conditioning on  $v_1$ .

Note that  $\tau_1 \in [0, 1]$ . By the definition of  $\hat{g}(y, \nu), \hat{g}(\nu)$  and assumption S, we have

$$\tau_1|\hat{r}_1-\bar{r}_1| \leq \frac{Ch_{\alpha}^{-\iota}}{(n-1)h_{\alpha}}\Big|\sum_{\ell=2}^n \Big[K_{\alpha}\Big(\frac{\hat{\nu}_{\ell}-\hat{\nu}_1}{h_{\alpha}}\Big)-K_{\alpha}\Big(\frac{\nu_{\ell}-\nu_1}{h_{\alpha}}\Big)\Big]\Big|,$$

for some constant C > 0.

Let  $\hat{\Delta}_{\ell} = \hat{\nu}_{\ell} - \nu_{\ell}$ . We now have

$$\begin{aligned} \tau_t |\hat{r}_t - \bar{r}_t| &\leq \frac{Ch_{\alpha}^{-\iota}}{(n-1)h_{\alpha}^2} \Big| \sum_{\ell=2}^n K_{\alpha}' \Big( \frac{\nu_{\ell} - \nu_1}{h_{\alpha}} \Big) \Big[ \hat{\Delta}_{\ell} - \hat{\Delta}_1 - \mathbb{E} (\hat{\Delta}_{\ell} - \hat{\Delta}_1 | \nu_1, \nu_{\ell}) \Big] \Big| \\ &+ \frac{Ch_{\alpha}^{-\iota}}{(n-1)h_{\alpha}^2} \Big| \sum_{\ell=2}^n K_{\alpha}' \Big( \frac{\nu_{\ell} - \nu_1}{h_{\alpha}} \Big) \mathbb{E} (\hat{\Delta}_{\ell} - \hat{\Delta}_1 | \nu_1, \nu_{\ell}) \Big| \\ &+ \frac{Ch_{\alpha}^{-\iota}}{2(n-1)h_{\alpha}^3} \sum_{\ell=2}^n \left\{ \Big| K_{\alpha}'' \Big( \frac{\nu_{\ell} - \nu_1 + \kappa(\hat{\Delta}_{\ell} - \hat{\Delta}_1)}{h_{\alpha}} \Big) \Big| \cdot (\hat{\Delta}_{\ell} - \hat{\Delta}_1)^2 \right\} \equiv \mathbb{C}_1 + \mathbb{C}_2 + \mathbb{C}_3. \end{aligned}$$

By Newey and McFadden (1994), let  $\mu_0(\nu_1) \equiv \mathbb{E}\left\{\frac{1}{h_{\alpha}^2}K'_{\alpha}\left(\frac{\nu_{\ell}-\nu_1}{h_{\alpha}}\right)\mathbb{E}(\hat{\Delta}_{\ell}-\hat{\Delta}_1|\nu_1,\nu_{\ell})|\nu_1\right\}.$ 

$$\begin{split} \frac{1}{(n-1)h_{\alpha}^{2}} \sum_{\ell=2}^{n} K_{\alpha}^{\prime} \Big( \frac{\nu_{\ell} - \nu_{1}}{h_{\alpha}} \Big) \Big[ \hat{\Delta}_{\ell} - \hat{\Delta}_{1} - \mathbb{E}(\hat{\Delta}_{\ell} - \hat{\Delta}_{1} | \nu_{1}, \nu_{\ell}) \Big] \\ &= \frac{1}{(n-1)} \sum_{\ell=2}^{n} \left[ \frac{1}{h_{\alpha}^{2}} K_{\alpha}^{\prime} \Big( \frac{\nu_{\ell} - \nu_{1}}{h_{\alpha}} \Big) \mathbb{E}(\hat{\Delta}_{\ell} - \hat{\Delta}_{1} | \nu_{1}, \nu_{\ell}) - \mu_{0}(\nu_{1}) \right] + o_{p}(n^{-1/2}h_{\alpha}^{-1}). \end{split}$$

Because

$$\mathbb{E}\left\{\frac{1}{(n-1)}\sum_{\ell=2}^{n}\left[\frac{1}{h_{\alpha}^{2}}K_{\alpha}'\left(\frac{\nu_{\ell}-\nu_{1}}{h_{\alpha}}\right)\mathbb{E}(\hat{\Delta}_{\ell}-\hat{\Delta}_{1}|\nu_{1},\nu_{\ell})-\mu_{0}(\nu_{1})\right]\right\}^{2}$$

$$=\frac{1}{n-1}\mathbb{E}\left\{\frac{1}{h_{\alpha}^{2}}K_{\alpha}'\left(\frac{\nu_{2}-\nu_{1}}{h_{\alpha}}\right)\mathbb{E}(\hat{\Delta}_{2}-\hat{\Delta}_{1}|\nu_{1},\nu_{2})\right\}^{2}$$

$$\leq\frac{1}{(n-1)}\int\frac{1}{h_{\alpha}^{3}}\left[K_{\alpha}'(u)\right]^{2}f_{\nu}(\nu_{1}+h_{\alpha}u)du\cdot O_{p}(\|\hat{\Delta}_{\ell}\|^{2})=o_{p}(n^{-1}h_{\alpha}^{-2}).$$

Thus,  $\mathbb{C}_1 = o_p(n^{-\frac{1}{2}}h_{\alpha}^{-\iota-1})$ . Moreover,

$$\begin{split} \frac{1}{(n-1)h_{\alpha}^{2}} \sum_{\ell=2}^{n} K_{\alpha}^{\prime} \Big( \frac{\nu_{\ell} - \nu_{1}}{h_{\alpha}} \Big) \mathbb{E}(\hat{\Delta}_{\ell} - \hat{\Delta}_{1} | \nu_{1}, \nu_{\ell}) \\ &= \frac{1}{(n-1)} \sum_{\ell=2}^{n} \left[ \frac{1}{h_{\alpha}^{2}} K_{\alpha}^{\prime} \Big( \frac{\nu_{\ell} - \nu_{1}}{h_{\alpha}} \Big) \mathbb{E}(\hat{\Delta}_{\ell} - \hat{\Delta}_{1} | \nu_{1}, \nu_{\ell}) - \mu_{0}(\nu_{1}) \right] + \mu_{0}(\nu_{1}) \\ &= \int \frac{1}{h_{\alpha}} K_{\alpha}^{\prime}(u) q(\nu_{1} + h_{\alpha}u) du + o_{p}(n^{-1/2}h_{\alpha}^{-1}) \\ &= \int K_{\alpha}(u) q^{\prime}(\nu + h_{\alpha}u) du + o_{p}(n^{-1/2}h_{\alpha}^{-1}). \end{split}$$

where  $q(\nu_2) = \mathbb{E}(\hat{\Delta}_{\ell} - \hat{\Delta}_1 | \nu_1, \nu_2) \cdot f_{\nu}(\nu_2)$ . Clearly,  $q(\nu_1 + h_{\alpha}u) = O_p(\|\hat{\Delta}_{\ell}\| \cdot h_{\alpha}) = o_p(n^{-1/2}h_{\alpha}^{-1})$ and we have  $\int |uK_{\alpha}(u)| du < +\infty$ . Hence,  $\mathbb{C}_2 = o_p(n^{-\frac{1}{2}}h_{\alpha}^{-\iota-1})$ .

Next, note that

$$\begin{split} \mathbb{E} \left| \frac{1}{(n-1)h_{\alpha}^{3}} \sum_{\ell=2}^{n} \left\{ \left| K_{\alpha}^{\prime\prime} \left( \frac{\nu_{\ell} - \nu_{1} + \kappa(\hat{\Delta}_{\ell} - \hat{\Delta}_{1})}{h_{\alpha}} \right) \right| (\hat{\Delta}_{\ell} - \hat{\Delta}_{1})^{2} \right\} \right| \\ = \frac{1}{h_{\alpha}^{3}} \mathbb{E} \left| K_{\alpha}^{\prime\prime} \left( \frac{\nu_{\ell} - \nu_{1} + \kappa(\hat{\Delta}_{\ell} - \hat{\Delta}_{1})}{h_{\alpha}} \right) (\hat{\Delta}_{\ell} - \hat{\Delta}_{1})^{2} \right| = \frac{1}{h_{\alpha}^{3}} \mathbb{E} \sup_{|\epsilon| \le n^{-\frac{1}{4}}} \left| K_{\alpha}^{\prime\prime} \left( \frac{\nu_{\ell} - \nu_{1} + \epsilon}{h_{\alpha}} \right) (\hat{\Delta}_{\ell} - \hat{\Delta}_{1})^{2} \right| \\ & \le \frac{1}{h_{\alpha}^{2}} \int \sup_{|\epsilon| \le n^{-1/4}} \left| K_{\alpha}^{\prime\prime} \left( u + \frac{\epsilon}{h_{\alpha}} \right) \right| \cdot T(\nu_{1} + h_{\alpha}u) du \end{split}$$

where  $\kappa \in [0,1]$  and  $T(\nu_2) = \mathbb{E}[(\hat{\Delta}_{\ell} - \hat{\Delta}_1)^2 | \nu_1, \nu_2] \cdot f_{\nu}(\nu_2)$ . Note that  $T(\nu_1 + h_{\alpha}u) = T(\nu_1) + h_{\alpha}u \cdot T'(\nu_1^{\dagger})$  and  $T(\nu_1) = 0$ , where  $\nu_1^{\dagger}$  is between  $\nu_1$  and  $\nu_1 + h_{\alpha}u$ . Because  $\sup_{\nu} ||T'(\nu)|| = o_p(n^{-1/2})$  by Lemma 3. Therefore,

$$\mathbb{E}|\mathbb{C}_3| \le Ch_{\alpha}^{-\iota} \cdot \left[ \int \left| K_{\alpha}''(u)u \right| du + o(1) \right] \cdot o(n^{-1/2}h_{\alpha}^{-1}) = o(n^{-1/2}h_{\alpha}^{-\iota-1}). \quad \Box$$

A.5. **Proof of Theorem 2.** Let  $\mathcal{V}'_n = \{x : f_{X'_1\beta_1, X'_2\beta_2}(x'_1\beta_1, x'_2\beta_2) > h_m^{\epsilon/5}\}$ . Moreover, we denote  $\hat{m}_i(x; b)$  be the nonparametric estimator  $\hat{m}_i(x)$  by replacing  $\tilde{\beta}$  with *b* (including the trimming part). Let further  $\tilde{C}_U$  be the infeasible estimator defined by

$$\widetilde{C}_{U}(p) = \frac{\sum_{t=1}^{n} Y_{t1} Y_{t2} \mathcal{K}_{c} (m(X_{t}) - p)}{\sum_{t=1}^{n} \mathcal{K}_{c} (m(X_{t}) - p)}.$$

Similarly to the proof of Lemma 3, conditionally on  $X_i \in \mathscr{V}'_n$ , the denominator of  $\hat{m}_{ti}$  is large enough and the adjustment trimming sequence tends exponentially to zero, and hence negligible. By a similar argument to Klein and Spady (1993, Lemmas 2–4), we have uniformly on  $\mathscr{V}'_n$  and B,

$$h_m^{\epsilon} |\hat{m}_i^*(x;b) - m_i(x;b)| = O_p \Big( \Big( \frac{\ln n}{nh_m^2} \Big)^{1/2} + h_m^{2R} \Big).$$

By assumption X, we have

$$\sup_{x\in\mathscr{V}'_n,b\in B}\left|\hat{m}^*_i(x;b)-m_i(x;b)\right|=O_p\left(\left(\frac{\ln n}{n}\right)^{\frac{R+1-\epsilon}{2R+4}}\right).$$

The rest of proof follows Guerre, Perrigne, and Vuong (2000): First, by the standard kernel estimation literature, we have the uniform convergence of  $\tilde{C}_U$  to  $C_U$  on any subset P at the rate  $(n/\ln n)^{R/(R+4)}$ , which is slower than the optimal uniform convergence rate due to the suboptimal bandwidth  $h_m$ . Therefore, it suffices to show

$$\sup_{p\in P} |\hat{C}_u(p) - \widetilde{C}_U(p)| = O_p((\ln n/n)^{(R-\epsilon'')/(2R+4)}).$$

Because for each  $p \in P$ , the denominator of the estimator converges to  $f_{\mathbb{E}}(Y|X)(p)$ , which is larger than c > 0, then it suffices to show

$$\sup_{p\in P} \left| n^{-1} \sum_{t=1}^{n} Y_{t1} Y_{t2} \mathcal{K}_c(\hat{m}_t - p) - n^{-1} \sum_{t=1}^{n} Y_{t1} Y_{t2} \mathcal{K}_c(m(X_t) - p)) \right| = O_p((\ln n/n)^{(R-\epsilon'')/(2R+4)}),$$

and

$$\sup_{p\in P} \left| n^{-1} \sum_{t=1}^{n} \mathcal{K}_{c}(\hat{m}_{t}-p) - n^{-1} \sum_{t=1}^{n} \mathcal{K}_{c}(m(X_{t})-p)) \right| = O_{p}((\ln n/n)^{(R-\epsilon'')/(2R+4)}).$$

For expositional simplicity, we only show the latter.

Following the proof in Guerre, Perrigne, and Vuong (2000, Theorem 3), we have

$$\begin{split} \left| n^{-1} \sum_{t=1}^{n} \mathcal{K}_{c}(\hat{m}_{t} - p) - n^{-1} \sum_{t=1}^{n} \mathcal{K}_{c}(m(X_{t}) - p)) \right| \\ &\leq O_{p} \left( (\ln n/n)^{(R - \epsilon'')/(2R + 4)} \right) \times \frac{1}{nh_{c}^{2}} \sum_{t=1}^{n} \sum_{i=1}^{2} \left| \frac{\partial \mathcal{K}_{c}}{\partial p_{i}} \left( \frac{m(X_{t}) - p}{h_{c}} \right) \right| \\ &+ O_{p} \left( (\ln n/n)^{(2R - 2\epsilon'' - 2)/(2R + 4)} \right) \times \sup_{p} \left| \frac{\partial^{2} \mathcal{K}_{c}}{\partial p \partial p'}(p) \right|. \end{split}$$

Because  $R - 2 - \epsilon'' > 0$  and  $\frac{1}{nh_c^2} \sum_{t=1}^n \sum_{i=1}^2 \left| \frac{\partial K_c}{\partial p_i} \left( \frac{m(X_t) - p}{h_c} \right) \right|$  uniformly converges to  $f_{\mathbb{E}(Y|X)}(p) \sum_{i=1}^2 \int \left| \frac{\partial K_c}{\partial p_i}(u) \right| du$ , then  $\sup_{p \in P} \left| n^{-1} \sum_{t=1}^n \mathcal{K}_c(\hat{m}_t - p) - n^{-1} \sum_{t=1}^n \mathcal{K}_c(m(X_t) - p)) \right| = O_p((\ln n/n)^{(R - \epsilon'')/(2R + 4)})$ .