SEMIPARAMETRIC IDENTIFICATION OF BINARY DECISION GAMES OF INCOMPLETE INFORMATION WITH CORRELATED PRIVATE SIGNALS*

YUANYUAN WAN† AND HAIQING XU‡

ABSTRACT. This paper studies the identification and estimation of a static binary decision game of incomplete information. We make no parametric assumptions on the joint distribution of private signals and allow them to be correlated. We show that the parameters of interest can be point–identified subject to a scale normalization under mild support requirements for the regressors (publicly observed state variables) and errors (private signals). Following Manski and Tamer (2002), we propose a maximum score type estimator for the structural parameters and establish the asymptotic properties of the estimator.

Keywords: Semiparametric Identification and Estimation, Incomplete Information Games, Modified Maximum Score Estimator, $\mathcal{U}$ Process

JEL: C14, C35, C62 and C72

Date: Monday 25th June, 2012.

*Yuanyuan Wan and Haiqing Xu gratefully acknowledge Joris Pinkse and Sung Jae Jun for their advice and suggestions; we thank Jason Abrevaya, Victor Aguirregabiria, Gaurab Aryal, Herman Bierens, Frank Erickson, Paul Grieco, Hiroyuki Kasahara, Vajay Krishna, Aureo de Paula, Isabelle Perrigne, Youngki Shin, Xun Tang, James Tybout, Quang Vuong, Neil Wallace, Hal White, the participants of the Cornell–PSU Macro Workshop (Spring 2010) and the 27th CESG conference at UBC for their valuable comments. Yuanyuan Wan thanks the Human Capital Foundation for their support.

†Department of Economics, University of Toronto, yuanyuan.wan@utoronto.ca.

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

1. Introduction

This paper studies the identification and estimation of a static binary decision game of incomplete information. We make no parametric assumptions on the joint distribution of private signals and allow them to be correlated. We show that the parameters of interest can be point–identified subject to a scale normalization under mild support requirements for the regressors (publicly observed state variables) and errors (private signals). Following Manski and Tamer (2002), we propose a maximum score type estimator for the structural parameters and establish its asymptotic properties.

Static binary decision games have many applications. Bjorn and Vuong (1984), for example, studies labor force participation. Recently, this class of games are more widely adopted in the empirical industrial organization literature for studying firms' entry behavior (e.g. Berry, 1992; Bresnahan and Reiss, 1990, 1991a,b; Ciliberto and Tamer, 2009; Jia, 2008). In much of this literature, an agent's payoff often depends on not only his covariates, but also other agents' choices. Therefore, the strategic effects are embedded in the equilibrium of the game model, which usually is a solution to a set of simultaneous equations. Recent contributions in static game of incomplete information includes Aguirregabiria and Mira (2002), Aradillas-Lopez (2010), Bajari, Hong, Krainer, and Nekipelov (2010), de Paula and Tang (2010), Lewbel and Tang (2012), Pesendorfer and Schmidt-Dengler (2003) and Tang (2010).

In this paper, we propose a new methodology that contributes to the literature in two respects. First, we do not require the (conditional) independence of private signals across players, as e.g., Aguirregabiria and Mira (2002), Bajari, Hong, Krainer, and Nekipelov (2010), and de Paula and Tang (2010) do. Instead, we assume that the private signals are positively regression dependent, conditional on the publicly observed states, which includes (conditional) independence as a special case. Positive regression dependence is a weaker notion of positive dependence than alternatives such as positive affiliation and decreasing inverse hazard rates, but stronger than positive correlation. We provide a numerical example showing that ignoring the correlation between the private signals results in inconsistent estimates.

Allowing correlated private signals in discrete games is motivated primarily by empirical concerns. The (conditional) independence assumption, commonly assumed in the literature, is convenient but also implies that equilibrium choices must be conditionally independent, an implication which could be invalidated by data. In many empirical applications, e.g. in the study of oligopolies' strategic entry, there is indeed not much of theoretical justification that the firms' private profit shocks should be independent of each other. On the contrary, one would expect the shocks to be positively correlated as they may originate from demand and cost shifters of the same market. Bajari, Hong, Krainer, and Nekipelov (2010) studies stock recommendations, including "strong buy", "buy", "hold", and "sell", for high technology stocks from equity market analysts. The payoff relevant private signals received by analysts would be correlated with each other, if these signals reflect their customers preference shocks and their customer groups overlap or interact with each other.

Second, we make no parametric assumptions on the joint distribution of private signals, which distinguishes our paper from Xu (2011). We provide a semiparametric identification procedure without explicitly solving an equilibrium. Our model also accommodates heteroskedasticity of unknown form, as we assume the conditional median independence of the private signals given the publicly observed states. It turns out that the point identification of structural parameters only requires a similar set of weak conditions as in single agent binary decision models (Manski, 1975). The method developed in this paper hence sheds some lights on the comparison between the identification of a semiparametric binary response model with and without strategic interactions.

Aradillas-Lopez (2010) also proposes a semiparametric approach in a similar setup of a discrete choice game of incomplete information, in which private signals are allowed to be correlated. Aradillas-Lopez (2010) adopts an equilibrium notation that uses the beliefs concept suggested by Aumann (1987). Our paper, on the other hand, focuses on the standard Bayesian Nash Equilibrium (BNE) concept (see, e.g., Fudenberg and Tirole, 1991, Chapter

6), which delivers a different set of structural equations from Aradillas-Lopez (2010)'s. We provide a more detailed comparison in Section 7.

We achieve point identification at infinity (of regressor values). With a Monotone Strategy Bayesian Equilibrium (MSBE) being played,[1] we show that the equilibrium strategies can be represented by a semiparametric binary response model with an unobserved regressor. We derive (estimable) bounds for the unobserved regressor. Under further mild assumptions, mainly a support condition about the regressors (similar to that in Manski, 1975), the upper and lower bounds on the unobserved regressor can be shown to converge to each other at the infinity of regressor values. Point identification then follows in the same way as with the maximum score estimator. The strategy of identification at infinity is first adopted by Chamberlain (1986), Heckman (1990) and Manski (1975, 1985, 1988). In the empirical game literature, it can also be found in Bajari, Hong, and Ryan (2010) and Tamer (2003) in their studies of complete information games with multiple equilibria.

Because of the nature of our identification and weak model restrictions imposed, our estimator follows the principles of the maximum score estimator (see Manski, 1975, 1985, 1988; Manski and Tamer, 2002). In particular, we extend Manski and Tamer (2002) and propose a two–step modified maximum score estimator in which the bounds of unobserved beliefs are nonparametrically estimated in the first step. By extending Kim and Pollard (1990); Nolan and Pollard (1988), we also show our estimator is $\sqrt[3]{n}$ consistent and has a non–normal limiting distribution. Our estimation approach is thus different from most of the literature, e.g. Aguirregabiria and Mira (2002), Aradillas-Lopez (2010) and Bajari, Hong, Krainer, and Nekipelov (2010), where (pseudo) likelihood based approaches are used. Our estimation method is also different from Tang (2010) who takes a single index approach.

The rest of this paper is organized as follows. Section 2 introduces the game model, explains the set of structural equations delivered by BNE solution and characterizes sufficient conditions for the existence of MSBE. In Section 3 we derive the bounds for the equilibrium

[1] Athey (2001) provides the pioneering result that an MSBE generally exists in a large class of games, which are referred as *supermodular* or *log–supermodular* games.

strategy and show the point identification of the structural parameters. Further, in Section 4, we propose a two-step modified maximum score estimator and establish its $\sqrt[3]{n}$–consistency and the limiting distribution. Section 5 contains an extension to the partial identification of structural parameters under weaker conditions. Section 6 gives a simulation example to illustrate the performance of our estimator in finite samples. Section 7 provides discussions.

2. MODEL

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

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

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

In this structure, $\beta_j \in \mathbb{R}^{d_j}$ and $\alpha_j \in \mathbb{R}$ are the parameters of interest. α_j measures the strategic effect: how the action of the other player $(-j)$ affects the payoff of player j when choosing $Y_j = 1$. Let $\beta = (\beta_1', \beta_2')'$, $\alpha = (\alpha_1, \alpha_2)'$. We assume $\alpha_j \geq 0$ for the brevity of our notation.[2]

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

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

[2] See Section 7 for an extension to the case of $\alpha_j \leq 0$. In this paper, we consider the case where the strategic component is a constant and there are no exogenous regressors affecting α_j. Such a specification restriction simplifies our discussion and notation.

where $\mathbf{1}[\cdot]$ is the indicator function. The term $\mathbb{P}(Y_{-j} = 1|X, U_j)$ is player j's expectation about the other player's action (based on player j's information). Equation (1) defines a set of simultaneous equations which are our econometric model. Aradillas-Lopez (2010) adopts a different belief concept suggested by Aumann (1987), which delivers a different set of simultaneous equations. (See our Section 7 for a more detailed discussion.)

A special type of BNE that we are interested in is the MSBE. Monotonicity of equilibrium strategies is a desirable property in many other applications of incomplete information games, e.g. auctions, differentiated-product price-competition and global games. With an MSBE, the equilibrium strategies are weakly monotone functions, i.e., there exists a cutoff–value function profile $u^* = (u_1^*, u_2^*) : \mathscr{X} \to \mathbb{R}^2$ such that for each j,

$$Y_j = \mathbf{1}\left[U_j \le u_j^*(X)\right]. \tag{2}$$

The MSBE equilibrium concept used here is somewhat restrictive, but is widely employed in the empirical game literature, explicitly or implicitly. Note that the usual assumption of the independent private signals necessarily implies that any equilibrium in this binary decision game has to be a monotone strategy equilibrium.

Athey (2001) provides the pioneering result that a monotone pure-strategy equilibrium exists whenever a Bayesian game obeys a Spence-Mirlees single-crossing condition (SCC). In the model we are considering here, the SCC is satisfied if Assumption A holds.

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

Assumption A implies that when his opponent plays a monotone strategy, a player's best response is nondecreasing in his private signal. It imposes a restriction on the curvature of the conditional copula function $C_{U|X}$ of U_1 and U_2 given X.[3] In a special case where

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

$$u_1 + \alpha_1 \mathbb{P}(U_2 \le u_2 | X = x, U_1 = u_1) = u_1 + \frac{\alpha_1}{f_{U_1}(u_1)} \frac{\partial C\left(F_{U_1}(u_1), F_{U_2}(u_2)\right)}{\partial u_1}.$$

U is parametrized to be bivariately normal distributed with mean zero and unit variance, Assumption A implies that the correlation between U_1 and U_2 is bounded above (below one), vis–a–vis the magnitude of α. Assumption A is trivially satisfied when U_1 and U_2 are independent, conditional on X.

Lemma 1. *Suppose that Assumption A holds. Then for any $X = x \in \mathscr{X}$, there exists an MSBE with equilibrium strategies characterized by Equation* (2).

Proof. See Appendix A.1. □

Note that Assumption A does not rule out the possibility of the existence of non–monotone strategy Bayesian equilibria. Under Assumption A, however, we implicitly assume that only one MSBE is played in equilibrium.

Furthermore, Assumption A is a sufficient condition for the existence of an MSBE. Our analysis stays valid as long as an MSBE exists and is adopted. In addition, if there is an identifiable subset of $\mathscr{X}$ on which an MSBE is played, our analysis can still be applied by using data belonging to the subset.[4]

3. Identification

In this section we study the identification of (α, β). Researchers observe X and Y, but not the private signals U. Our identification strategy is as follows. We first derive (estimable) bounds for the equilibrium strategies of the players under the assumption that U_1 and U_2 are positively regression dependent conditional on X (Assumption B below). Then we show that the bounds can be arbitrarily close to each other when some argument of the regressors goes to infinity (Lemma 2). These results deliver the point–identification of (α, β).

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

Suppose that $\alpha_1 > 0$ and the copula function is twice differentiable, then the condition stated in the assumption says that $\frac{\partial^2 C(s_1,s_2)}{\partial s_1^2} \ge -\frac{1}{\alpha_1 f_{U_1}(F_{U_1}^{-1}(s_1))}$. Similarly, we have $\frac{\partial^2 C(s_1,s_2)}{\partial s_2^2} \ge -\frac{1}{\alpha_2 f_{U_2}(F_{U_2}^{-1}(s_2))}$.

[4] Xu (2011) proposes a method to identify such a subset in a parametric framework.

Positive regression dependence is a weaker notion of positive dependence than other alternatives such as positive affiliation and decreasing inverse hazard rate.[5] Weaker correlation between U_1 and U_2 implies less restrictions on α_j. Note that, together with Assumption A, Assumption B further imposes restrictions on the scale of the strategic components: $\alpha_j \leq -\frac{1}{\partial \mathbb{P}\left(U_{-j} \leq t | X=x, U_j=u_j\right)/\partial u_j}$ for all x, t and u_j.

Under Assumption B, we are able to represent the equilibrium strategies by a semiparametric binary regression model with interval–observed data (see Manski and Tamer, 2002). To motivate, suppose that $\mathbb{P}(U_{-j} \leq t | X = x, U_j = u_j)$ is continuous in u_j for all $t \in \mathbb{R}$ and $x \in \mathscr{X}$ (we actually do not assume this in Theorem 1). Given the equilibrium is monotone, and conditional on any $x \in \mathscr{X}$, player j receives zero expected payoff when the value of his private signal equals to $u_j^*(x)$,

$$x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right) - u_j^*(x) = 0. \tag{3}$$

By the definition of MSBE, it follows that

$$\begin{aligned} Y_j &= 1[U_j \leq u_j^*(x)] \\ &= 1\left[U_j \leq x_j'\beta_j - \alpha_j \mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right)\right]. \end{aligned} \tag{4}$$

Note that the term $\mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right)$ in Equation (4) is an unobservable regressor of the binary response model. Let $\nu_j^0(x) = \mathbb{P}\left(Y_{-j} = 1 | X = x, Y_j = 0\right)$ and $\nu_j^1(x) = \mathbb{P}\left(Y_{-j} = 1 | X = x, Y_j = 1\right)$. Assumption B implies that,

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

[5] de Castro (2007) provides an example. Let the density of $U \in [0,1]^2$ be $f(u_1, u_2) = k/[1 + (u_1 - u_2)^2]$ for some $k > 0$. Then U_1 and U_2 are positive regression dependent, but neither affiliated nor with the decreasing inverse hazard rate.

Hence the unobserved regressor, as a function of x, is bounded by a pair of estimable functions. The result is formally summarized in Theorem 1.

Theorem 1. *Suppose that Assumptions A and B are satisfied. Then for player $j = 1, 2$, the equilibrium strategy of player j is represented by the structure*

$$Y_j = \mathbf{1}\left[U_j \leq X_j'\beta_j - \alpha_j \nu_j(X)\right]; \quad \mathbb{P}\left(\nu_j^0(X) \leq \nu_j(X) \leq \nu_j^1(X)\right) = 1, \tag{5}$$

where

$$\nu_j(x) = \mathbb{P}\left(U_{-j} \leq u_{-j}^*(x) | X = x, U_j = u_j^*(x)\right). \tag{6}$$

Proof. See Appendix A.2. □

We have a few comments on Theorem 1. First, the bounds ν_j^0 and ν_j^1 are estimable. We study the identification taking ν_j^0 and ν_j^1 as known. Second, if U_1 and U_2 are independent conditional on X, then $\nu_j^1(x) = \nu_j^0(x)$ for all $x \in \mathscr{X}$. If the inequality in Assumption B holds strictly on a subset $\tilde{\mathscr{X}_j}$ of $\mathscr{X}_j$ (a violation to conditional independence), then $\nu_j^1(x) > \nu_j^0(x)$ for all $x \in \tilde{\mathscr{X}_j}$. Theorem 1 thus provides a testable implication for conditional independence provided players play one MSBE.[6] Further, we can relax Assumption B by requiring $\mathbb{P}(U_{-j} \leq t | X = x, U_j = u_j)$ be monotonic in u_j for all t and x, i.e., our model allows for both positive and negative regression dependence. Similar results as stated in Theorem 1 follow by redefining ν_j^1 and ν_j^0. We maintain Assumption B throughout this paper for the ease of notation.

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

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

Assumptions C and D are also made in Manski (1985) for a single agent binary response model. Assumption C imposes a conditional median independence restriction on the private

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

signals. Normalizing the median to be zero is innocuous as long as a constant term is included in X_j. This assumption allows for heteroskedasticity of unknown form. Assumption D excludes multicollinearity.

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

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

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

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

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

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

Lemma 2. *Suppose that Assumptions A to F hold. Then for all* $\epsilon > 0$ *and all* $x_j \in \mathscr{X}_j$,

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

Proof. See Appendix A.3. □

Note that since $\nu_j^0(x) \leq \nu_j^1(x) \leq 1$ for all x, it follows from Lemma 2 that both ν_j^0 and ν_j^1 converge to 1 in probability as $x_{-j}'\beta_{-j}$ goes to $+\infty$. Meanwhile, both bounds converge to 0 in probability as $x_{-j}'\beta_{-j}$ goes to $-\infty$.

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

Theorem 2. *Suppose that Assumptions A to F hold. Then (α, β) is point identified.*

Proof. See Appendix A.4. □

In our model, the support condition on regressor $X_{j,1}$ plays an extra role under Assumption F other than in Manski (1985): when $x_1'\beta_1$ goes to $+\infty$ $(-\infty)$, both ν_2^1 and ν_2^0 converge to 1 (0). Thus we take advantage of both infinities, at which the unobserved regressor behaves as an observed 0–1 variable. We can then achieve point identification of (α, β) in the same way as that in a single agent binary response model.

4. Estimation

We now propose an estimator which is motivated by the modified maximum score estimator (MMSE) of Manski and Tamer (2002). We modify the objective function in Manski and Tamer (2002) by using the density function f_X as weighting factors of scores. Our estimation consists of two steps. First, we non–parametrically estimate of our bounds $(\nu_j^0(\cdot), \nu_j^1(\cdot))$, f_X and a conditional choice probability function $p_j(x)$ that will be defined later; Second, by plugging the nonparametric estimates into the sample analog of the objective function, we rewrite it as a $\mathcal{U}$–process, the maximizer of which is shown to be $\sqrt[3]{n}$–consistent. We also establish its limiting distribution by extending Kim and Pollard (1990, Theorem 1.1) to our $\mathcal{U}$–process sample analog.

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

Let $p_j(x) = \mathbb{P}\left(Y_j = 1 | X = x\right)$ and $\delta_j(x) = 1\left[p_j(x) \geq 1/2\right]$. To simplify the notation, we denote $\mathbf{v}_j = (v_j^1, v_j^0)'$, $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$ and $\delta = (\delta_1, \delta_2)'$. We also denote $\theta_j = (a_j, b_j')'$, $\theta = (\theta_1', \theta_2')'$, as generic parameter values, $\theta_{0,j} = (\alpha_j, \beta_j')'$, and $\theta_0 = (\alpha', \beta')'$. Let further

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

and $\Theta = \Theta_1 \times \Theta_2$, where $\mathbb{M}$ is a large positive number to ensure Θ is compact and $(\theta_{0,j}) \in \Theta_j$. Moreover, let $\operatorname{sgn}(\cdot)$ be the sign function, i.e. $\operatorname{sgn}(x) = 1$, 0 and -1 respectively when $x > 0$, $x = 0$ and $x < 0$.

Lemma 3. *Suppose that assumptions A through F are satisfied, then*

$$\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta),$$

where $L(\theta) = \sum_{j=1}^{2} \mathbb{E}\left[g_j\left(Z; \theta_j, \boldsymbol{v}_j, \delta_j, f_X\right)\right]$ *and for any arbitrarily fixed bounded and strictly positive function* $w_j(\cdot) : \mathbb{R}^d \to \mathbb{R}$, *the function* g_j *is defined by*

$$\begin{aligned} g_j\left(Z; \theta_j, \boldsymbol{v}_j, \delta_j, f_X\right) &= \left(2Y_j - 1\right) \times f_X(X) \times w_j(X) \\ &\times \left\{\delta_j(X) \operatorname{sgn}\left[X_j' b_j - a_j v_j^0(X)\right] + \left(1 - \delta_j(X)\right) \operatorname{sgn}\left[X_j' b_j - a_j v_j^1(X)\right]\right\}. \end{aligned}$$

Proof. See Appendix B.1. □

The function $w_j : \mathscr{X} \to \mathbb{R}_{++}$ is known and serves as a weighting factor. For the brevity of notation, we set $w_j(x) = 1$ for all $x \in \mathscr{X}$. It should be noted that our analysis carries through as long as w_j is bounded from above.

The population objective function defined above is similar to Manski and Tamer (2002) except for the weighting factors f_X of scores. We could define an estimator for θ_0 as the maximizer of a sample analog L_n^* of L:

$$L_n^*(\theta;\hat{\mathbf{v}},\hat{\delta},\hat{f}_X) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} g_j(Z_i;\theta_j,\hat{\mathbf{v}}_j,\hat{\delta}_j,\hat{f}_X),$$

where $\hat{\mathbf{v}} = (\hat{\mathbf{v}}_1',\hat{\mathbf{v}}_2')'$, $\hat{\delta} = (\hat{\delta}_1,\hat{\delta}_2)'$ and $\hat{f}_X$ are nonparametric estimators of $\mathbf{v}$, δ and f_X, respectively. Unfortunately, it is difficult to establish the asymptotic properties for such an estimator, partly due to the first–stage nonparametric estimates in indicator functions. In contrast, we follow Wan and Xu (2012) and construct a $\mathcal{U}$–process sample analog to define our estimator.

We define some notation to begin with. Let $\lambda_j = (\delta_j,\mathbf{v}_j')'$ and $\lambda = (\lambda_1',\lambda_2')'$. Let further $\vartheta_j(x;\theta_j,\lambda_j) = \delta_j(x)\,\mathrm{sgn}[x_j'b_j - a_j\nu_j^0(x)] + (1-\delta_j(x))\,\mathrm{sgn}[x_j'b_j - a_j\nu_j^1(x)]$. Then we define

$$\mathcal{U}_n(\theta;\lambda) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\ell\neq i}^{n}\sum_{j=1}^{2}\left\{(2Y_{j\ell}-1)K_h(X_\ell - X_i)\vartheta_j(X_i;\theta_j,\lambda_j)\right\}, \qquad (7)$$

where $K_h(\cdot) = K(\cdot/h)/h^d$, K and h are the kernel function and the smoothing bandwidth, respectively. As mentioned above, we let $w_j(x) = 1$ for all x for the notational brevity.[7]

$\mathcal{U}_n(\theta;\lambda)$ is not readily usable because λ is unknown. Note $\delta_j(x) = \mathbf{1}[p_j(x) \geq 1/2] \overset{a.e.}{=} \mathbf{1}[\big(2p_j(x)-1\big)\times f_X(x) \geq 0]$, we estimate δ_j by

$$\hat{\delta}_j(x_i) = \mathbf{1}\left[\frac{1}{(n-1)}\sum_{\ell\neq i}^{n}\Big\{(2Y_{j\ell}-1)\times K_h(X_\ell - x_i)\Big\} \geq 0\right].$$

Further, we estimate our bounds ν_j^1 and ν_j^0 by

$$\hat{\nu}_j^1(x_i) = \frac{\sum_{\ell\neq i}^{n} Y_{-j\ell}Y_{j\ell}K_h(X_\ell - x_i)}{\sum_{\ell\neq i}^{n} Y_{j\ell}K_h(X_\ell - x_i)} + n^{-\gamma}, \quad \hat{\nu}_j^0(x_i) = \frac{\sum_{\ell\neq i}^{n} Y_{-j\ell}(1-Y_{j\ell})K_h(X_\ell - x_i)}{\sum_{\ell\neq i}^{n}(1-Y_{j\ell})K_h(X_\ell - x_i)} - n^{-\gamma},$$

[7] If we choose a weighting function $w_j(\cdot) \neq 1$, we could write our sample objective function as

$$\mathcal{U}_n(\theta;\lambda) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\ell\neq i}^{n}\sum_{j=1}^{2}\Big\{(2Y_{j\ell}-1)K_h(X_\ell - X_i)\vartheta_j(X_i;\theta_j,\lambda_j)w_j(X_i)\Big\}.$$

where $\gamma > 1/3$. Later we will make assumptions on K, h, $\mathbf{v}$ and the underlying distribution functions such that the first term in $\hat{v}_j^0$ (reps. $\hat{v}_j^1$) converges to v_j^0 (reps. $\hat{v}_j^1$) at a rate faster than $n^{-\gamma}$. Thus, the second term $n^{-\gamma}$ ensures that $\hat{v}_j^0(x) \le \hat{v}_j^1(x)$ with probability approaching one (faster than any polynomial rates).

We hence define our estimator of θ_0 as: $\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \mathcal{U}_n\left(\theta;\hat{\lambda}\right)$.

To compare with Manski and Tamer (2002)'s MMSE, note that

$$\mathcal{U}_n(\theta;\hat{\lambda}) = L_n(\theta;\hat{p},\hat{f}_X,\hat{\lambda}) \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\left[2\hat{p}_j(X_i)-1\right]\hat{f}_X(X_i)\vartheta_j(X_i;\theta_j,\hat{\lambda}_j), \tag{8}$$

where $\hat{f}_X(x_i) = \sum_{\ell\neq i}^{n} K_h(X_\ell - x_i)/(n-1)$ and $\hat{p}_j(x_i) = \sum_{\ell\neq i}^{n} Y_{j\ell}K_h(X_\ell - x_i)/\sum_{\ell\neq i}^{n} K_h(X_\ell - x_i)$. It is straightforward to note the difference is that we replace $Y_{j,i}$ with $\hat{p}_j(X_i)$ in our sample analog. The advantage of such modification is that, under additional weak conditions, it can be shown that the first–stage nonparametric estimates $\hat{\delta}$ in our sample analog only causes a small approximation error of order $o_p(n^{-2/3})$.

To establish the consistency of $\hat{\theta}$, we impose the following assumptions.

Assumption H. *θ_0 is in the interior of the compact parameter space Θ.*

Assumption I. $\mathbb{P}\left[p_j(X) = 1/2\right] = 0$. *Furthermore,* $\forall\theta\in\Theta$, $\mathbb{P}[X_j'b_j - a_j v_j^0(X) = 0] = \mathbb{P}[X_j'b_j - a_j v_j^1(X) = 0] = 0$.

The first equation of Assumption I is also assumed in Manski and Tamer (2002).[8] The second equation is a rank condition imposed on the augmented random vector $(X, v_j^0(X))$ and $(X, v_j^1(X))$ respectively.

Assumption J. *X is a continuous random vector and the density function $f_X(\cdot)$ are continuous in the support.*

[8]In Manski and Tamer (2002), it is assumed that $\mathbb{P}\left\{(x,v_1,v_0) : \mathbb{P}\left(y=1|x,v_1,v_0\right) = 1-\alpha\right\} = 0$, where v_0 and v_1 are bounds of the unobserved regressor.

Assumption K. *The kernel function K is a symmetric Parzen Rosenblatt kernel. i.e., (1) $\int_{\mathbb{R}^d} K(u)du = 1$; (2) $\sup_u |K(u)| = \overline{K} < \infty$; (3) $\int_{\mathbb{R}^d} |K(u)|du < \infty$; (4) $K(u) = K(-u)$ and (5) $\|u\|^d \times |K(u)| \to 0$ as $\|u\| \to \infty$. Moreover, $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$.*

Assumptions J and K are standard in nonparametric estimation literature (e.g. Pagan and Ullah, 1999).[9]

Theorem 3. *Suppose that assumptions A through K hold. Then $\hat{\theta} \xrightarrow{p} \theta_0$.*

Proof. See Appendix B.2. □

The sample objective function in Equation (8) is "irregular" in the sense that it belongs to the class of Kim and Pollard (1990)–type objective functions which do not allow for quadratic expansions. As a consequence, this class of estimators, including the one we propose here, have slower convergence rate ($\sqrt[3]{n}$) and non–normal limiting distributions. We proceed with smoothness conditions on the nonparametric functions.

Assumption L. *$p_j(\cdot)$, $f_X(\cdot)$, $\nu_j^0(\cdot)$ and $\nu_j^1(\cdot)$ are everywhere R times continuously differentiable with bounded R–th partial derivatives.*

Assumption M. *For $x \in \mathscr{X}$, let $\xi_j(x) = \left[2P_j(x) - 1\right] f_X(x)$. Assume that there exists an ϵ–neighborhood around zero, denoted as $\mathcal{N}_\epsilon$ and a constant $C_\xi > 0$, such that for any subset $S \subseteq \mathcal{N}_\epsilon$, there is*

$$\mathbb{P}\left(\xi_j(X) \in S\right) \leq C_\xi \times \mu(S),$$

where μ is the Lebesgue measure. Moreover, for $m = 0, 1$, there exists an ϵ_m–neighborhood around zero, denoted as $\mathcal{N}_{\epsilon_m}$, an η_m–neighborhood around θ_0, denoted as $\mathcal{N}_{\eta_m}$, and a constant $C_m > 0$, such that for any subset $S \subseteq \mathcal{N}_{\epsilon_m}$ and $\theta \in \mathcal{N}_{\eta_m}$, there is

$$\mathbb{P}\left(X_j' b_j - a_j \nu_j^m(X) \in S\right) \leq C_m \times \mu(S).$$

Note that Assumptions L and M imply Assumptions I and J, respectively.

[9] It is not necessary that all the regressors are continuous random variables. Similar arguments can be carried through with additional notation.

Assumption N. *The kernel function K is a symmetric Parzen Rosenblatt kernel of order R. i.e., (1) Assumption K holds; (2) $\int_{\mathbb{R}^d} u_1^{r_1},...,u_d^{r_d} K(u)du = 0$ if $1 \leq \sum_{k=1}^{d} r_k \leq R-1$; (3) $\int_{\mathbb{R}^d} u_1^{r_1},...,u_d^{r_d} K(u)du = 1$ if $\sum_{k=1}^{d} r_k = R$, where $r_k \in \mathbb{N}^+$ for $k = 1,...,d$.*

Assumption O. *Let $\gamma > 1/3$ and h satisfies $n^{\gamma}(nh^d)^{-1/2} = o(n^{-r_0})$ and $n^{\gamma}h^R = o(n^{-r_0})$ for some $r_0 > 0$.*

Assumption O implies that the high order kernel we use should satisfy $R > d$, a similar condition is also assumed in Powell, Stock, and Stoker (1989). To see this, let $h \propto n^{-1/(2R+d)}$ be the optimal choice of bandwidth. By Assumption O, we have $R/(2R+d) > 1/3$ and, therefore, $R > d$. It should also be noted that Assumptions G and L through O guarantee $\hat{\xi}_j(x)$ converges to $\xi_j(x)$ faster than $n^{-1/3}$ for all $x \in \mathscr{X}$.

Moreover, for $t \in \mathbb{R}$, let $\mathbf{v}_j(t) = (v_j^0(\cdot) - t, v_j^1(\cdot) + t)'$. Let further $\mathbf{v}(t) = (\mathbf{v}_1'(t), \mathbf{v}_2'(t))'$. By definition, $\mathbf{v} = \mathbf{v}(0)$.

Lemma 4. *Suppose that Assumptions A through G, I, and L through O hold. Then there exist random vectors $\{(\Gamma_n, T_n)\}$ of order $o_p(n^{-2/3})$, which are independent with θ, such that*

$$\mathcal{U}_n(\theta;\delta,\boldsymbol{v}(0)) - T_n \leq \mathcal{U}_n(\theta;\hat{\delta},\hat{\boldsymbol{v}}) \leq \mathcal{U}_n(\theta;\delta,\boldsymbol{v}(2n^{-\gamma})) + \Gamma_n.$$

Proof. See Appendix B.3 □

Lemma 4 is crucial because it shows that the first stage nonparametric estimates approximation error is negligible and allows us to focus on the infeasible sample analog $\mathcal{U}_n(\theta;\delta,\mathbf{v}(t))$, which is a $\mathcal{U}$–process. Kim and Pollard (1990) establish the cube–root convergence rate for the maximum score estimator (see, e.g., Manski, 1975). We extend their results to our $\mathcal{U}$–process sample analog.

Let $\bar{\vartheta}_j(X;\theta_j) = \vartheta_j(X;\theta_j,\lambda_j) - \vartheta_j(X;\theta_{0,j},\lambda_j)$. Let further $\kappa_{0,j}(x_j,v_j) = 2\mathbb{E}\{\xi_j^+(X)|X_j = x_j, v_j^0(X) = v_j\}$ and $\kappa_{1,j}(x_j,v_j) = 2\mathbb{E}\{\xi_j^-(X)|X_j = x_j, v_j^1(X) = v_j\}$ where $\xi_j^+ = \max\{\xi_j,0\}$ and $\xi_j^- = -\min\{\xi_j,0\}$.

Theorem 4. *Suppose that Assumptions A through O hold. Then*

$$n^{1/3}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \operatorname{argmax}_{t\in\mathbb{R}^d}\{W(t)-\frac{1}{2}t'Vt\},$$

where W is a mean zero Gaussian process with covariance kernel H:

$$H(s,t)=2\lim_{\eta\to\infty}\eta\sum_{j=1,2}\mathbb{E}\left\{\xi_j^2(X)\bar{\vartheta}_j(X;\theta_{0,j}+s/\eta)\bar{\vartheta}_j(X;\theta_{0,j}+t/\eta)\right\};\ \textit{and}$$

$$V=2\sum_{m=0,1}\sum_{j=1,2}\int \mathbf{1}\{x_j'\beta_j-\alpha_j v_j=0\}\left[\dot{\kappa}_{m,j}(x_j,v_j)'\theta_{0,j}\right]f_{X_j,\nu_j(X)}(x_j,v_j)(x',v_j)'(x',v_j)d\sigma_j^m,$$

where σ_j^m denotes surface measure of $(X_j',\nu_j^m(X))$ on $x_j'\beta_j+\alpha_j\nu_j^m=0$. Moreover, $W(t)-\frac{1}{2}t'Vt$ has a unique maximizer almost surely on all its sample path.

Proof. See Appendix B.4. □

5. Partial Identification under Weaker conditions

In this section we extend our approach under weaker conditions such that the structural parameters are partially identified. Without assuming the support conditions (Assumptions D to F), we show that parameters of interest are set–identified in general, and the identified set can be estimated by a level set of the modified maximum score objective function.

For $j=1,2$, let Θ_j^I be the collection of θ_j such that for all $x\in\mathscr{X}$,

$$\mathbf{1}[x_j'b_j-a_j\nu_j^1(x)\geq 0]\leq \mathrm{Med}(Y_j|X=x)\leq \mathbf{1}[x_j'b_j-a_j\nu_j^0(x)\geq 0].$$

Let further $\Theta^I=\Theta_1^I\times\Theta_2^I$. We call Θ^I the identified set. Theorem 1 ensures Θ^I be nonempty as $\theta_0\in\Theta^I$ under Assumptions A and B. Note that Θ^I would degenerate to a singleton if we impose Assumptions D to F.

Following the modified maximum score estimator proposed by Manski and Tamer (2002), we define the set estimator $\widehat{\Theta}^I$ for Θ^I as

$$\widehat{\Theta}^I=\left\{\theta^*\in\Theta:\mathcal{U}_n\left(\theta^*;\hat{\lambda}\right)\geq\sup_{\theta\in\Theta}\mathcal{U}_n\left(\theta;\hat{\lambda}\right)-\kappa_n\right\} \tag{9}$$

for some $\kappa_n \to 0$. The estimator is a level (of κ_n) set of the sample objective function.

To establish consistency, we adopt the directional Hausdorff distance measure:

$$\rho(A, B) = \sup_{a \in A} \inf_{b \in B} \|a - b\|,$$

where $\|\cdot\|$ is the Euclidean norm. When either A or B is empty set, the distance is $+\infty$.

Theorem 5. *Suppose that Assumptions A though C, and G through O are satisfied, then* $\rho(\widehat{\Theta}^I, \Theta^I) \xrightarrow{p} 0$. *Suppose in addition that* $\sup_{\theta \in \Theta} |\mathcal{U}_n(\theta; \hat{\lambda}) - \mathcal{U}(\theta; \lambda)| = o_p(\kappa_n)$, *then* $\rho(\Theta^I, \widehat{\Theta}^I) \xrightarrow{p} 0$.

Proof. See Appendix B.5. □

It is also possible to conduct inference on Θ^I under partial identification conditions, following the methods proposed in Chernozhukov, Hong, and Tamer (2007); Blevins (2012), among others. We do not further investigate it in the present paper.

6. Experiments

In this section, we provide a numerical example to illustrate that ignoring the correlation between the private signals results in inconsistent estimates and possibly misleading inference. In particular, we investigate the performance of a two–step Maximum Likelihood Estimator (MLE) when the belief is mis–specified (while the distribution of signals are corrected specified).

Recall that Theorem 1 implies that the structural model can be represented as a semiparametric binary regression model

$$Y_j = \mathbf{1}[\beta_{j,1} X_{j,1} + \beta_{j,2} - \alpha_j \nu_j(X) - U_j \geq 0],$$

where $\nu_j(x) = \mathbb{P}(U_{-j} \leq u^*_{-j}(x) | X = x, U_j = u^*_j(x))$ is the "interval–observed regressor".

If the correlation between U_1 and U_2 were ignored, a two–step MLE would be based on the following specification,

$$Y_j = \mathbf{1}[\beta_{j,1}X_{j,1} + \beta_{j,2} - \alpha_j \mathbb{P}(Y_{-j} = 1|X) - U_j \geq 0]. \tag{10}$$

Two–step MLE could be inconsistent because $\mathbb{P}(Y_{-j} = 1|X) \neq \mathbb{P}(Y_{-j} = 1|X, U_j)$ in general.

We evaluate the performance of both estimators in following examples. Let $d_1 = d_2 = 2$, $X_1 = (X_{1,1}, 1)'$ and $X_2 = (X_{2,1}, 1)'$. $(X_{1,1}, X_{2,1})$ is drawn from a mixture of two normal distributions[10]

$$\begin{cases} X \sim N(0, 0.16), & \text{with prob. } \frac{3}{4} \\ X \sim N(0, 100), & \text{with prob. } \frac{1}{4} \end{cases}$$

We let U be independent of X. U_1 and U_2 have a mean zero bivariate normal distribution with variance $\sigma_j^2 = 1$ and correlation coefficient $r \in \{0, 0.1, \cdots, 0.7\}$. We consider different values of correlation and illustrate how the "magnitude" of ignored correlation affects bias.

The parameters in the profit function are set as $\beta_{1,2} = \beta_{2,2} = 0$, $\beta_{1,1} = \beta_{2,1} = 1$, $\alpha_1 = \alpha_2 = 1$. It can be shown that an MSBE exists under these designs, *i.e.*, for each x, there exist cutoff values $u_1^*(x)$ and $u_2^*(x)$, such that player j chooses 1 whenever his private signal $U_j \leq u_j^*(x)$. We compute $u_j^*(x)$ by solving the following equations for each realization of X in the sample:

$$u_1^* = \beta_{1,1}x_{1,1} + \beta_{1,2} - \alpha_1 \Phi\Big(\frac{\sigma_2 u_2^* - \rho\sigma_1 u_1^*}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\Big),$$
$$u_2^* = \beta_{2,1}x_{2,1} + \beta_{2,2} - \alpha_2 \Phi\Big(\frac{\sigma_1 u_1^* - \rho\sigma_2 u_2^*}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\Big),$$

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

[10]Large variance generates X with big absolute values, which mimic extremely "large" or "small" markets.

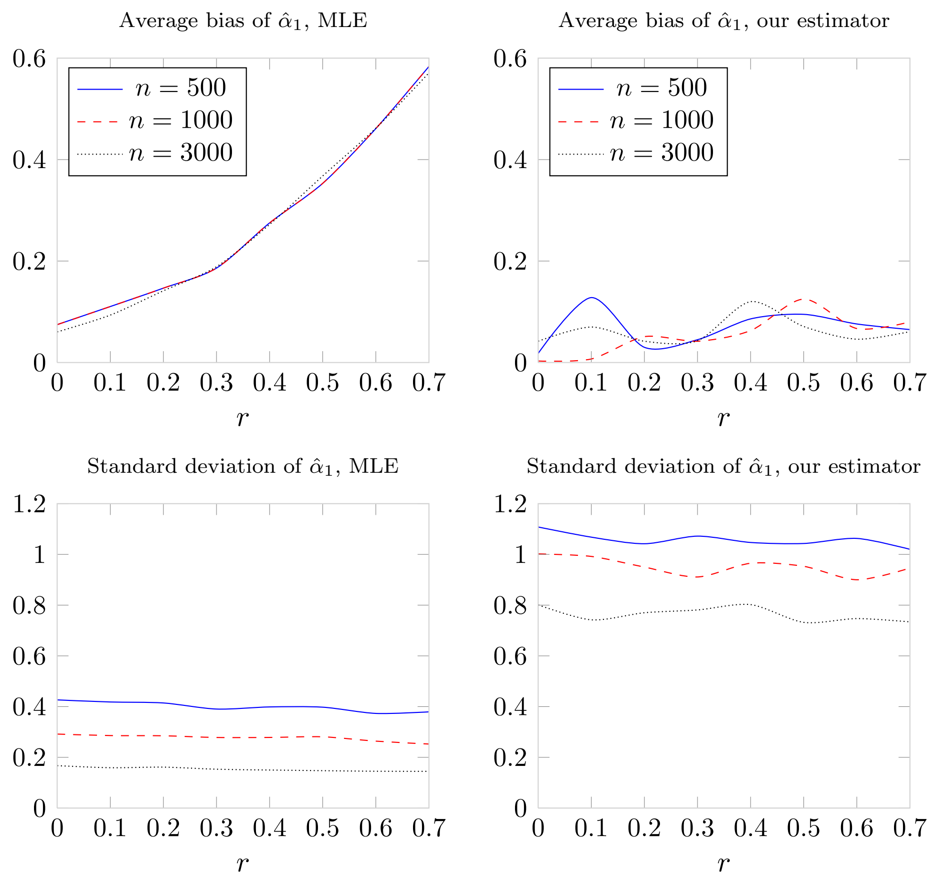

FIGURE 1. Average bias and standard deviation

Since U_1 and U_2 are jointly normally distributed with positive correlation, they are positively regression dependent and Assumptions D and E are satisfied as well. Moreover, Assumption F holds since U and X are independent.

Figure 1 shows the average bias (the two top panels) and standard deviation (the two bottom panels) of both estimators under different sample sizes. All results are based on 500 replications.[11] Detailed results are reported in Appendix D. We can see from the figure that the performance of our estimator is robust for different values of r. The finite sample bias decreases as the sample size increases. The standard deviation of our estimator decreases

[11]We estimate the unknown functions using a third order kernel proposed in Pagan and Ullah (1999, Section 2.7.2). Bandwidth is chosen to be propositional to $n^{-1/8}$. We also use a weighting function $w(x) = \min\{K, \|x\|^3\}$ for some large K.

roughly at rate $n^{-1/3}$. On the other hand, although the mis–specified MLE converges faster (at rate $n^{-1/2}$), it does not converge to the true value. The bias has significant size at all sample sizes. In particular, the magnitude of the bias is increasing in r. When r is large, hypothesis tests based on mis–specified MLE could be misleading, e.g when $r = 0.7$ and $n = 3,000$, the bias is almost five times of the standard deviation.

7. Dicussions

7.1. A Comparison with Aradillas-Lopez (2010). A similar two–by–two static game of incomplete information is also studied in Aradillas-Lopez (2010), which adopts a different set of assumptions and proposes a novel identification and estimation strategy.

Aradillas-Lopez (2010) assumes a similar linear payoff structure and allows private signals to be correlated, but differs from the present paper in several aspects. First, Aradillas-Lopez (2010) and the present paper focus on two different econometric models due to different equilibrium solution concepts adopted. Aradillas-Lopez (2010) applies a belief concept introduced by Aumann (1987), which gives the following simultaneous equations structure:[12]

$$\begin{aligned} Y_1 &= \mathbf{1}\left\{X_1'\beta_1 + \alpha_1 \mathbb{P}_1(Y_2 = 1|Y_1 = 1, X) - U_1 \geq 0\right\}, \\ Y_2 &= \mathbf{1}\left\{X_2'\beta_2 + \alpha_2 \mathbb{P}_2(Y_1 = 1|Y_2 = 1, X) - U_2 \geq 0\right\}, \end{aligned} \tag{11}$$

where $\mathbb{P}_j(Y_{-j} = 1|Y_j = 1, X)$ are equilibrium beliefs of player j about his rival's equilibrium action. In contrast, we follow BNE solution concept, which requires player j's beliefs to be expectations conditioning on his own private signal, U_j (a detailed definition of BNE for general Bayesian games can be found in, e.g., Harsanyi, 1967–68; Fudenberg and Tirole, 1991). In the similar game setup, therefore, the BNE solution concept delivers us a different set of simultaneous equations, which are our Equation (1).

The BNE solution concept has been widely used in many empirical applications of incomplete information games, e.g. auctions. Regarding to the literature of discrete games, for

[12] To simplify the comparison, we assume that all elements in X are publicly observed. Aradillas-Lopez (2010) allows the publicly observed vector Z to be different from X.

instance, Bajari, Hong, Krainer, and Nekipelov (2010) studies the strategic stock recommendation behaviors of equity market analysts. In their paper, the utility an analyst receives from issuing a recommendation is a function of both recommendations issued by other analysts and his own private payoff shock. If private shocks are correlated and the dependence pattern is known to all the analysts, then they will exploit their private information as well as the dependence pattern to form a rational expectation on rivals' choices.

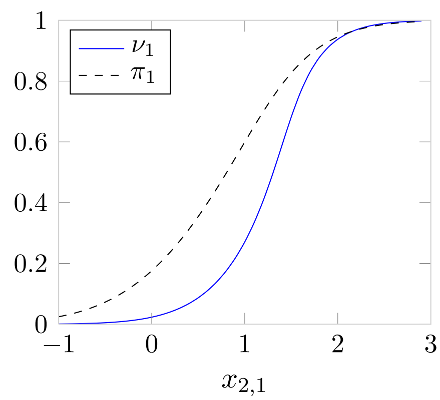

FIGURE 2. $\nu_1(1.5, x_{2,1}; \theta_0)$ and $\pi_1(1.5, x_{2,1}; \theta_0)$

Figure 2 plots the equilibrium beliefs considered by Aradillas-Lopez (2010) and the present paper in a game with payoff structure similar to the one specified in Section 6 ($r = 0.5$ case). Here π_1 denotes the equilibrium beliefs $\mathbb{P}_1(Y_2 = 1 | Y_1 = 1, X)$ in Aradillas-Lopez (2010). Because $\nu_1 \neq \pi_1$, Aradillas-Lopez (2010) and the present paper would generate different conditional probability distributions of players' choices $Y = (Y_1, Y_2)$ even in the games with the same setup. Not surprisingly, when we generate data based on Equation (1), the estimator suggested by Aradillas-Lopez (2010) could be inconsistent.[13] For illustration, see Figure 3.[14]

[13]Not vice versa, however, our estimator is still consistent when the data are generated from the econometric model in Aradillas-Lopez (2010), since our bounds are still valid. See Table 3 in Appendix D for more detailed results.

[14]We generate covariates from the following distribution (other configurations are the same as in Section 6)

$$\begin{cases} X_{1,1} \sim U[0,2], \quad X_{2,1} \sim U[-1,1], & \text{with p. 0.8,} \\ X_{1,1} \sim N(0,4), \quad X_{2,1} \sim N(0,4), & \text{with p. 0.2.} \end{cases}$$

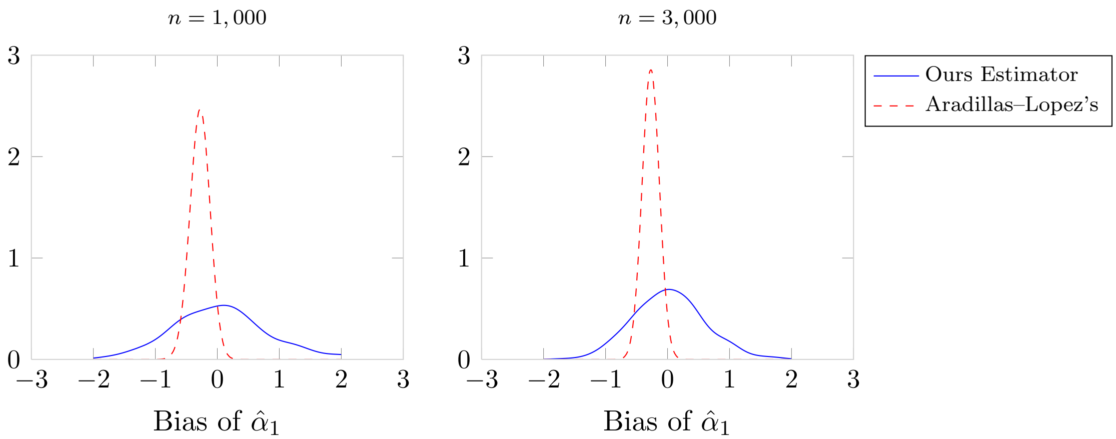

FIGURE 3. Data generated from a BNE

Second, Aradillas-Lopez (2010) assumes that the private signals are independent of publicly observed X variables, which is stronger than what is assumed in the present paper — the conditional median independence of private signals given X. As a tradeoff, our maximum score type estimator is endowed with a slower convergence rate than the one proposed by Aradillas-Lopez (2010). As pointed out by Liu, Vuong, and Xu (2012), the independence between X and private signals imposes testable restrictions on the observed choices, i.e.,

$$\mathbb{P}(Y_j|X = x) \geq \mathbb{P}(Y_j|X = x') \text{ for } j = 1, 2.$$
$$\Longrightarrow \mathbb{P}(Y_1 = 1, Y_2 = 1|X = x) \geq \mathbb{P}(Y_1 = 1, Y_2 = 1|X = x'). \quad (12)$$

In contrast, the conditional median independence assumption is somewhat the "minimum" requirement for the identification of the binary games without imposing any restriction on the data observed. In a real data application, if the restriction (12) is rejected by statistical tests, then our conditional median dependence assumption would be more applicable.

7.2. **Strategic Complementarity.** Throughout this paper, we focus on the negative strategic effect case ($\alpha_j \geq 0$). However, our method can easily extend to binary games of strategic complementarity ($\alpha_j \leq 0$).[15]

When $\alpha_j \leq 0$, the existence of MSBE is ensured by Assumption B only, since Assumption A is implied. Moreover, the MSBE can be written as

$$Y_j = \mathbf{1}\left[U_j \leq X_j'\beta_j - \alpha_j \nu_j\left(X\right)\right],$$

and ν_j satisfies $\mathbb{P}\left(\nu_j^1\left(X\right) \leq \nu_j\left(X\right) \leq \nu_j^0\left(X\right)\right) = 1$. Note that we switch the positions of upper and lower bounds in Equation (5) due to the fact $\alpha_j \leq 0$. It means that the sign of α_j can be revealed by the relative order of the two bounds. Hereafter, the analysis of identification and estimation of the structural parameters similarly follows Sections 3 and 4.

REFERENCES

AGUIRREGABIRIA, V., AND P. MIRA (2002): "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 70(4), 1519–1543.

ARADILLAS-LOPEZ, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," *Journal of Econometrics*, 157(2), 409–431.

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

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

BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2009): "Semiparametric Estimation of Dynamic Games of Incomplete Information," *Working Paper*.

[15] As an anonymous referee correctly pointed out, in this case the positive strategic effect may be confounded with positive signal correlation.

BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics*, 28(4), 469–482.

BAJARI, P., H. HONG, AND S. RYAN (2010): "Identification and estimation of discrete games of complete information," *Econometrica*, 78(5), 1529–1568.

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

BJORN, P. A., AND Q. H. VUONG (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," Working Paper 537, California Institute of Technology, Division of the Humanities and Social Sciences.

BLEVINS, J. R. (2012): "Non-Standard Rates of Convergence of Criterion-Function-Based Set Estimators," Unpublished manuscript, Ohio State University.

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 (1991a): "Empirical models of discrete games," *Journal of Econometrics*, 48(1-2), 57–81.

——— (1991b): "Entry and Competition in Concentrated Markets," *The Journal of Political Economy*, 99(5), 977–1009.

CHAMBERLAIN, G. (1986): "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32(2), 189–218.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

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

DE CASTRO, L. I. (2007): "Affiliation, equilibrium existence and the revenue ranking of auctions," Economics working papers, Universidad Carlos III, Departamento de Economï£¡a.

DE PAULA, A., AND X. TANG (2010): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," Working papers.

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. MIT Press.

HARSANYI, J. (1967–68): "Games with incomplete information played by" bayesian" players," *Management science*, pp. 159–182, 320–334, 486–502.

HECKMAN, J. (1990): "Varieties of selection bias," *The American Economic Review*, pp. 313–318.

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

KIM, J., AND D. POLLARD (1990): "Cube root asymptotics," *The Annals of Statistics*, pp. 191–219.

LEWBEL, A., AND X. TANG (2012): "Identification and Estimation of Discrete Bayesian Games with Multiple Equilibria Using Excluded Regressors," Discussion paper, Boston College.

LIU, N., Q. VUONG, AND H. XU (2012): "Nonparametric Analysis of Binary Games of Incomplete Information," Working in progress.

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

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

——— (1988): *Analog estimation methods in econometrics*. Chapman and Hall New York.

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

NEWEY, W., AND D. MCFADDEN (1994): *Large sample estimation and hypothesis testing*. Elsevier.

NOLAN, D., AND D. POLLARD (1987): "U-Processes: Rates of Convergence," *The Annals of Statistics*, 15(2), 780–799.

——— (1988): "Functional Limit Theorems for U–Process," *The Annals of Probabilities*, 16(3), 1291–1298.

PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press.

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2003): "Identification and Estimation of Dynamic Games," Working Paper 9726, National Bureau of Economic Research.

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

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

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

VAN DER VAART, A., AND J. WELLNER (1996): *Weak convergence and empirical processes*. Springer Verlag.

WAN, Y., AND H. XU (2012): "Inferences in Semiparametric Binary Response Models with Interval Data," Working paper.

XU, H. (2011): "Estimation of Bayesian Nash Equilibria In Two-player Games With Correlated Private Signals," Working paper, Pennsylvania State University.

APPENDIX A. PROOFS

A.1. **Proof of Lemma 1.** We apply Athey (2001, Theorem 1) to show the existence of MSBE in our game. Note that the action space is finite in our setup. By Assumption A, SCC is satisfied, and the types have joint density with respect to Lebesgue measure. Thus it suffices to show that each player's interim payoff function is bounded in his type for all monotone pure strategies of other players. The payoff function in our model is a linear function in U_j, which thereafter is not bounded. We could, however, apply a monotone transformation on the payoff function:

$$\pi_j^*(y_{-j}, x_j, U_j) = \begin{cases} \alpha_j - \alpha_j y_j & \text{if } x_j'\beta_j - U_j > \alpha_j, \\ x_j'\beta_j - \alpha_j y_j - U_j & \text{if } -\alpha_j \leq x_j'\beta_j - U_j \leq \alpha_j, \\ -\alpha_j - \alpha_j y_j & \text{if } x_j'\beta_j - U_j < -\alpha_j. \end{cases}$$

to be the player j's payoff of choosing 1, given j's opponent chooses $y_{-j} \in \{0,1\}$. Note that for any x and realization of U, each player will make the same choice under both π_j and π_j^*. Hence, this transformation of payoff function doesn't change the equilibria solution set for each $x \in \mathscr{X}$. Using the new payoff functions π_j^*, it is then routine to verify all the conditions of Athey (2001, theorem 1), ensuring the existence of MSBE.

A.2. **Proof of Theorem 1.** When $\mathbb{P}(U_{-j} \leq t | X = x, U_j = u_j)$ is continuous in u_j for any $t \in \mathbb{R}$ and $x \in \mathscr{X}$, the result of Theorem 1 holds by the arguments in section 3. Without the continuity, the conclusion still holds. To see this, let

$$\nu_j^+(x) = \lim_{u_j \downarrow u_j^*(x)} \mathbb{P}\left(Y_{-j} = 1 | X = x, U_j = u_j\right), \quad \nu_j^-(x) = \lim_{u_j \uparrow u_j^*(x)} \mathbb{P}\left(Y_{-j} = 1 | X = x, U_j = u_j\right).$$

Under Assumption B, $\mathbb{P}\left(Y_{-j} = 1 | X = x, U_j = u_j\right)$ is a non–increasing function in u_j. Hence $\nu_j^+(x)$ and $\nu_j^-(x)$ are well defined and $\nu_j^+(x) \leq \nu_j^-(x)$. By the definition of MSBE, for all $u_j > u_j^*(x)$

$$x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1 | X = x, U_j = u_j\right) - u_j \leq 0.$$

Hence

$$\lim_{u_j \downarrow u_j^*(x)} \left\{x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1 | X = x, U_j = u_j\right) - u_j\right\} = x_j'\beta_j - \alpha_j \nu_j^+(x) - u_j^*(x) \leq 0.$$

Similarly, $x_j'\beta_j - \alpha_j \nu_j^-(x) - u_j^*(x) \geq 0$. It implies that $\nu_j^+(x) \geq \nu_j^-(x)$. So we have $\nu_j^+(x) = \nu_j^-(x)$. Hence

$$x_j'\beta_j - \alpha_j \mathbb{P}\left(Y_{-j} = 1 | X = x, U_j = u_j^*(x)\right) - u_j^*(x) = 0.$$

Thus

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

A.3. **Proof of Lemma 2.** We show the first equation in the statement of Lemma 2 holds. The second one holds by similar reasoning. Suppose that the first condition in Assumption F is satisfied, then,

$$\begin{aligned}\nu_1^0(x) &= \mathbb{P}\left(U_2 \leq u_2^*|X=x, Y_1=0\right) = \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2\nu_2|X=x, Y_1=0\right)\\ &\overset{(i)}{\geq} \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2|X=x, Y_1=0\right) = \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2|X=x, U_1 \geq u_1^*(x)\right)\\ &\overset{(ii)}{\geq} \mathbb{P}\left(U_2 \leq x_2'\beta_2 - \alpha_2|X=x, U_1 \geq x_1'\beta_1\right).\end{aligned}$$

Inequality (i) hold because $\nu_2 \leq 1$ and (ii) holds because $u_1^* \leq x_1'\beta_1$ and Assumption B. Hence, for any $\epsilon \in (0,1)$,

$$\begin{aligned}&\mathbb{P}\left(U_2 \leq t-\alpha_2|X_1=x_1, X_2'\beta_2=t, U_1 \geq x_1'\beta_1\right) \leq \mathbb{E}\left[\nu_1^0(X)|X_1=x_1, X_2'\beta_2=t\right]\\ &\leq (1-\epsilon)\mathbb{P}\left(\nu_1^0(X) \leq 1-\epsilon|X_j=x_j, X_2'\beta_2=t\right) + \mathbb{P}\left(\nu_1^0(X) > 1-\epsilon|X_j=x_j, X_2'\beta_2=t\right)\\ &= -\epsilon\mathbb{P}\left(\nu_1^0(X) \leq 1-\epsilon|X_j=x_j, X_2'\beta_2=t\right)+1.\end{aligned}$$

Let $t \to +\infty$, under Assumption F we have

$$\lim_{t\to+\infty}\mathbb{P}\left(\nu_1^0(X) \leq 1-\epsilon|X_j=x_j, X_2'\beta_2=t\right)=0.$$

A.4. **Proof of Theorem 2.** This proof simply follows from the proof of Manski (1985, Lemma 2).

Fix $\tilde{X}_1 = \tilde{x}_1$ and $\tilde{X}_2 = \tilde{x}_2$ be arbitrary values. It suffices to show that for any $(\tilde{b}_1, a_1) \neq (\tilde{\beta}_1, \alpha_1)$, either $\mathbb{P}(X_{1,1} + \tilde{x}_1'\tilde{b}_1 - a_1\nu_1^1 \geq 0 > X_{1,1} + \tilde{x}_1'\tilde{\beta}_1 - \alpha_1\nu_1^0|\tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0$ or $\mathbb{P}(X_{1,1} + \tilde{x}_1'\tilde{\beta}_1 - \alpha_1\nu_1^0 \geq 0 > X_{1,1} + \tilde{x}_1'\tilde{b}_1 - a_1\nu_1^1|\tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0$. Equivalently, we will show that, either

$$\mathbb{P}(\alpha_1\nu_1^0 - \tilde{x}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1\nu_1^1 - \tilde{x}_1'\tilde{b}_1|\tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0,$$

or

$$\mathbb{P}(a_1\nu_1^0 - \tilde{x}_1'\tilde{b}_1 > X_{1,1} \geq \alpha_1\nu_1^1 - \tilde{x}_1'\tilde{\beta}_1|\tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0.$$

Suppose $(\tilde{b}_1, a_1) \neq (\tilde{\beta}_1, \alpha_1)$. Consider the following two cases:

Case 1. $\tilde{b}_1 \neq \tilde{\beta}_1$. By Assumption D, there exists $\tilde{x}_1 \in \mathscr{X}_1$ such that $\tilde{x}_1'\tilde{b}_1 \neq \tilde{x}_1'\tilde{\beta}_1$. W.L.O.G., assume $\tilde{x}_1'\tilde{b}_1 > \tilde{x}_1'\tilde{\beta}_1$. Let $\epsilon \in \mathbb{R}^+$ such that $\tilde{x}_1'\tilde{b}_1 - \epsilon > \tilde{x}_1'\tilde{\beta}_1 + \epsilon$. By Lemma 2, for any

$x_{1,1} \in (-\tilde{x}_1'\tilde{b}_1 + \epsilon, -\tilde{x}_1'\tilde{\beta}_1 - \epsilon)$, there exists a $t > 0$ such that

$$\mathbb{P}\left(\max\left\{\alpha_1 \nu_1^0, a_1 \nu_1^1\right\} < \epsilon | X_1 = x_1, \tilde{X}_2 = \tilde{x}_2, X_{2,1} \leq -t\right) = 1,$$

where $x_1 = (x_{1,1}, \tilde{x}_1)$. Note that the value t depends on x_1. Then

$$\begin{aligned}
&\mathbb{P}(\alpha_1 \nu_1^0 (X) - \tilde{X}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1 \nu_1^1 (X) - \tilde{X}_1'\tilde{b}_1 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) \\
&\quad \geq \mathbb{P}(\alpha_1 \nu_1^0 (X) - \tilde{X}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1 \nu_1^1 (X) - \tilde{X}_1'\tilde{b}_1 ; X_{2,1} \leq -t(X_1) | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) \\
&\qquad \geq \mathbb{P}(X_{1,1} \in [-\tilde{x}_1'\tilde{b}_1 + \epsilon, -\tilde{x}_1'\tilde{\beta}_1 - \epsilon); X_{2,1} \leq -t(X_1) | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2) > 0,
\end{aligned}$$

where the last inequality holds by Assumption E.

Case 2. $\tilde{b}_1 = \tilde{\beta}_1$. Then $\alpha_1 \neq a_1$. W.L.O.G., assume that $\alpha_1 > a_1$. Let $\eta \in \mathbb{R}^+$ such that $\alpha_1 - \alpha_1\eta > a_1$. By Lemma 2, for any $x_{1,1} \in (-\tilde{x}_1'\tilde{b}_1 + a_1, -\tilde{x}_1'\tilde{\beta}_1 + \alpha_1 - \alpha_1\eta)$, there exists an $s > 0$ such that

$$\mathbb{P}\left(\min\left\{\nu_1^0, \nu_1^1\right\} > 1 - \eta | X_1 = x_1, \tilde{X}_2 = \tilde{x}_2, X_{2,1} \geq s\right) = 1.$$

Again, the value s depends on x_1. Then

$$\begin{aligned}
&\mathbb{P}\left(\alpha_1 \nu_1^0 (X) - \tilde{x}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1 \nu_1^1 (X) - \tilde{x}_1'\tilde{b}_1 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2\right) \\
&\quad \geq \mathbb{P}\left(\alpha_1 \nu_1^0 (X) - \tilde{x}_1'\tilde{\beta}_1 > X_{1,1} \geq a_1 \nu_1^1 (X) - \tilde{x}_1'\tilde{\beta}_1 ; X_{2,1} \geq s(X_1) | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2\right) \\
&\qquad \geq \mathbb{P}\left(X_{1,1} \in \left[a_1 - \tilde{x}_1'\tilde{\beta}_1, \alpha_1 - \alpha_1\eta - \tilde{x}_1'\tilde{\beta}_1\right); X_{2,1} \geq s(X_1) | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2\right) > 0.
\end{aligned}$$

The statement in the Lemma thus follows by combining case 1 and 2.

Appendix B. Proofs in Estimation

B.1. **Proof of Lemma 3.** It suffices to show that for $j = 1, 2$ and any $\theta_j \neq \theta_{0,j}$,

$$\mathbb{E} g_j\left(Z; \theta_{0,j}, \mathbf{v}_j, \delta_j, f_X\right) > \mathbb{E} g_j\left(Z; \theta_j, \mathbf{v}_j, \delta_j, f_X\right).$$

For the simplicity of notation, we only provide a proof for $j = 1$. Notice that the support $\mathscr{X}$ can be divided into three regions, A, B and C, as follows:

$$
\begin{aligned}
A &= \left\{x : x_1'\beta_1 - \alpha_1 \nu_1^0(x) < 0\right\}, \\
B &= \left\{x : x_1'\beta_1 - \alpha_1 \nu_1^1(x) \geq 0\right\}, \\
C &= \left\{x : x_1'\beta_1 - \alpha_1 \nu_1^0(x) \geq 0 > x_1'\beta_1 - \alpha_1 \nu_1^1(x)\right\}.
\end{aligned}
$$

For any $x \in A$, because $p_1(x) = \mathbb{P}\left(U_1 \leq X_1'\beta_1 - \nu_1(X) \,|X = x\right) \leq \mathbb{P}\left(U_1 \leq 0|X = x\right) = 1/2$, hence for any $x \in A$,

$$
\mathbb{E}\left\{g_1\left(Z;\theta_{0,1}, \mathbf{v}_1, \delta_1, f_X\right) \middle| X = x\right\} = \left(1 - 2p_1(x)\right) f_X(x) \times w_1(x) = \left|2p_1(x) - 1\right| f_X(x) \times w_1(x).
$$

Similarly, we have for any $x \in B$, $p_1(x) \geq 1/2$ and therefore

$$
\mathbb{E}\left\{g_1\left(Z;\theta_{0,1}, \mathbf{v}_1, \delta_1, f_X\right) \middle| X = x\right\} = \left(2p_1(x) - 1\right) f_X(x) \times w_1(x) = \left|2p_1(x) - 1\right| f_X(x) \times w_1(x).
$$

and for any $x \in C$

$$
\begin{aligned}
&\mathbb{E}\left\{g_1\left(Z, \alpha_1, \beta_1, \mathbf{v}, \delta, f_X\right) \middle| X = x\right\} = \left(2p_1(x) - 1\right) f_X(x) \times \left[2\delta_j(x) - 1\right] \times w_1(x) \\
&\quad = \left(2p_1(x) - 1\right) \times f_X(x) \times \operatorname{sgn}\left(2p_1(x) - 1\right) \times w_1(x) = \left|2p_1(x) - 1\right| f_X(x) \times w_1(x).
\end{aligned}
$$

From above discussion, we have that $\mathbb{E}g_1\left(Z;\theta_{0,1}, \mathbf{v}_1, \delta_1, f_X\right) = \int \left|2p_1(x) - 1\right| f_X^2(x) \times w_1(x)dx$. Now consider any $\theta_1 \neq \theta_{0,1}$. Similarly, we define three regions, $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ as follows:

$$
\begin{aligned}
\tilde{A} &= \left\{x : x_1'b_1 - a_1 \nu_1^0(x) < 0\right\}, \\
\tilde{B} &= \left\{x : x_1'b_1 - a_1 \nu_1^1(x) \geq 0\right\}, \\
\tilde{C} &= \left\{x : x_1'b_1 - a_1 \nu_1^0(x) \geq 0 > x_1'b_1 - a_1 \nu_1^1(x)\right\}.
\end{aligned}
$$

Note that $A \cap \tilde{B} \neq \emptyset$ and $B \cap \tilde{A} \neq \emptyset$.

For any $x \in \tilde{A}$,

$$
\mathbb{E}\left\{g_1\left(Z;\theta_1, \mathbf{v}_1, \delta_1, f_X\right) \middle| X = x\right\} = -\left(2p_1(x) - 1\right) f_X(x) \times w_1(x) \leq \left|2p_1(x) - 1\right| f_X(x) \times w_1(x).
$$

Similarly, $\mathbb{E}\left\{g_1\left(Z;\theta_1,\mathbf{v}_1,\delta_1,f_X\right)\middle|X=x\right\} \leq |2p_1(x)-1|\times f_X(x)w_1(x)$ for all $x\in\tilde{B}$ and $x\in\tilde{C}$. Thus we have

$$\mathbb{E}\left\{g_1\left(Z;\theta_1,\mathbf{v}_1,\delta_1,f_X\right)\middle|X=x\right\}\leq\mathbb{E}\left\{g_1\left(Z;\theta_{0,1},\mathbf{v}_1,\delta_1,f_X\right)\middle|X=x\right\}$$

for any x in the support.

Because for any $x\in A\cap\tilde{B}$, $p_1(x)<1/2$, and

$$\begin{aligned}\mathbb{E}\left\{g_1\left(Z;\theta_1,\mathbf{v}_1,\delta_1,f_X\right)\middle|X=x\right\}&=\left(2p_1(x)-1\right)f_X(x)\times w_1(x)\\&<|2p_1(x)-1|\,f_X(x)\times w_1(x)=\mathbb{E}\left\{g_1\left(Z;\theta_{0,1},\mathbf{v}_1,\delta_1,f_X\right)\middle|X=x\right\}.\end{aligned}$$

The inequality holds strictly as long as $f_X(x)>0$ and $w_1(x)>0$ for all $x\in A\cap\tilde{B}$. Therefore $\mathbb{E}\left\{g_1\left(Z;\theta_1,\mathbf{v}_1,\delta_1,f_X\right)\right\}<\mathbb{E}\left\{g_1\left(Z;\theta_{0,1},\mathbf{v}_1,\delta_1,f_X\right)\right\}$. □

B.2. **Proof of Theorem 3.** By Lemmas 5 and 6,

$$\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\hat{\lambda}\right)-L\left(\theta\right)\right|=o_p\left(1\right).$$

The consistency of the estimator then follows from Theorem 2.1 in Newey and McFadden (1994). □

Let $\hat{\xi}_j(x)=\left[2\hat{p}_j(x)-1\right]\hat{f}_X(x)$. Let further $\tilde{\xi}_j(x)=\hat{\xi}_j(x)+[\xi_j(x)-\hat{\xi}_j(x)]\times\mathbf{1}[|\hat{\xi}_j(x)-\xi_j(x)|>n^{-\gamma}]$ and $\tilde{\delta}_j(x)=\mathbf{1}[\tilde{\xi}_j(x)\geq 0]$. Then $\tilde{\xi}_j$ is an (infeasible) estimator of ξ_j such that $\sup_x|\tilde{\xi}_j(x)-\xi_j(x)|\leq n^{-\gamma}$ a.s..

Similarly, let $\tilde{v}^0_j(x)=\hat{v}^0_j(x)+[v^0_j(x)-\hat{v}^0_j(x)]\times\mathbf{1}\{|\hat{v}^0_j(x)-[v^0_j(x)-n^{-\gamma}]|>n^{-\gamma}\}$ and $\tilde{v}^1_j(x)=\hat{v}^1_j(x)+[v^1_j(x)-\hat{v}^1_j(x)]\times\mathbf{1}\{|\hat{v}^1_j(x)-[v^1_j(x)+n^{-\gamma}]|>n^{-\gamma}\}$. Moreover, we denote $\tilde{\lambda}\equiv(\tilde{\delta},\tilde{\mathbf{v}})$.

B.3. **Proof of Lemma 4.** Remember $\mathbf{v}_j(t)=(v^0_j(\cdot)-t,v^1_j(\cdot)+t)$. Let $\mathcal{V}_n(\theta,t)=\mathcal{U}_n(\theta;\delta,\mathbf{v}(t))$. First, we have

$$\begin{aligned}\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}})-\mathcal{U}_n(\theta;\delta,\mathbf{v}(0))&=\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}})-\mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}})\\&\quad+\mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}})-\mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}(0))+\mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}(0))-\mathcal{U}_n(\theta;\delta,\mathbf{v}(0))\\&\geq\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}})-\mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}})+\mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}(0))-\mathcal{U}_n(\theta;\delta,\mathbf{v}(0)),\end{aligned}$$

where the last step comes from Lemma 11. Then, by Lemmas 7 and 8, we have

$$\sup_{\theta\in\Theta} |\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}}) - \mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}})| \leq o_p(n^{-2/3}),$$

$$\sup_{\theta\in\Theta} |\mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}) - \mathcal{U}_n(\theta;\delta,\mathbf{v})| \leq o_p(n^{-2/3}).$$

Let $T_n \equiv \sup_{\theta\in\Theta}\left\{|\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}}) - \mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}})| + |\mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}) - \mathcal{U}_n(\theta;\delta,\mathbf{v})|\right\}$, which is $o_p(n^{-2/3})$. Thus

$$\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}}) - \mathcal{U}_n(\theta;\delta,\mathbf{v}(0)) \geq -T_n.$$

By similar arguments as above and let $\Gamma_n \equiv \sup_{\theta\in\Theta}\left\{|\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}}) - \mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}})| + |\mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}(2n^{-\gamma})) - \mathcal{U}_n(\theta;\delta,\mathbf{v}(2n^{-\gamma}))|\right\}$, we can show that

$$\mathcal{U}_n(\theta;\hat{\delta},\hat{\mathbf{v}}) - \mathcal{U}_n(\theta;\delta,\mathbf{v}(2n^{-\gamma})) \leq \Gamma_n. \quad \square$$

We define some notation before proceed the proof for Theorem 4. Define

$$\tilde{g}(Z_i, Z_\ell;\theta,\lambda) = \frac{1}{2}\sum_{j=1,2}\left[(2Y_{j\ell}-1)K_h(X_\ell,X_i)\vartheta_j(X_i;\theta,\lambda) + (2Y_{ji}-1)K_h(X_i,X_\ell)\vartheta_j(X_\ell;\theta,\lambda)\right], \quad (13)$$

and $\bar{g}(Z_i;\theta,\lambda) = \mathbb{E}[\tilde{g}(Z_i,Z_\ell;\theta,\lambda)|Z_i]$. Let $\overline{\mathscr{G}} = \{\bar{g}(\cdot;\theta,\lambda) : \theta\in\Theta\}$ be a class of functions indexed by θ. Let $\overline{G}$ be its envelope function. Note also that $\mathbb{E}\overline{G}^2 < \infty$.

To simply notation, let $\mathcal{V}_n(\theta,t) = \mathcal{U}_n(\theta;\delta,\mathbf{v}(t))$ and $\mathcal{V}(\theta;t) = \mathbb{E}\mathcal{V}_n(\theta;t)$. It then follows that $\mathcal{V}(\theta,0) = \mathbb{E}\mathcal{U}_n(\theta;\delta,\mathbf{v})$.

B.4. **Proof of Theorem 4.** The proof consists following two steps. In step 1, we establish the $\sqrt[3]{n}$–convergence of our estimator; in step 2, we show the weak convergence of the rescaled sample criterion function. Our proof is an extension of Nolan and Pollard (1988, Theorem 5).

Step 1. Recall that $\bar{g}(Z_i;\theta,\lambda) = \mathbb{E}[\tilde{g}(Z_i,Z_\ell;\theta,\lambda)|Z_i]$. By the symmetry of $\tilde{g}$,

$$\sqrt{n}\left(\frac{1}{n(n-1)}\sum_{i\neq\ell}\tilde{g}(Z_i,Z_\ell;\theta,\lambda) - \mathcal{V}(\theta,0)\right)$$
$$= 2\sqrt{n}\left(\frac{1}{n}\sum_i \bar{g}(Z_i;\theta,\lambda) - \mathcal{V}(\theta,0)\right) + \frac{\sqrt{n}}{n(n-1)}\sum_{i\neq\ell}\tilde{\tilde{g}}(Z_i,Z_\ell;\theta,\lambda), \quad (14)$$

where

$$\tilde{\tilde{g}}(Z_i,Z_\ell;\theta,\lambda) = \tilde{g}(Z_i,Z_\ell;\theta,\lambda) - \bar{g}(Z_i;\theta,\lambda) - \bar{g}(Z_\ell;\theta,\lambda) + \mathcal{V}(\theta,0).$$

Note that $\overline{\mathscr{G}}$ is actually a sum of two V–C class of functions and we can apply the maximal inequality of Kim and Pollard (1990, Section 3.1) to the first term on the right hand side of Equation (14), i.e. there exists a universal constant J such that

$$\sqrt{n}\mathbb{E}\sup_\theta\left|\frac{1}{n}\bar{g}(Z_i,\theta,\lambda) - \mathcal{V}(\theta,0)\right| < J\sqrt{\mathbb{E}\overline{G}^2}$$

The second term on the right hand side of Equation (14) is a degenerated $\mathcal{U}$–process. By Lemma 10, it is of order $O_p(1/\sqrt{nh^d})$ and hence is negligible. As a result, we can conclude that

$$\sqrt{n}\mathbb{E}\sup_\theta\left(\frac{1}{n(n-1)}\sum_{i\neq\ell}\tilde{g}(Z_i,Z_\ell;\theta,\lambda) - \mathcal{V}(\theta,0)\right) \le J\sqrt{\mathbb{E}\overline{G}^2}. \quad (15)$$

With Equation (15) in hand, and following the exact argument of Kim and Pollard (1990, Lemma 4.1), we know that for each $\epsilon > 0$, there exist random variables $\{M_n\}$ of order $O_p(1)$ independent with θ, such that

$$|[\mathcal{V}_n(\theta,t) - \mathcal{V}(\theta,t)] - [\mathcal{V}_n(\theta_0,0) - \mathcal{V}(\theta_0,0)]| \le \epsilon\|\theta-\theta_0\|^2 + \epsilon t^2 + n^{-\frac{2}{3}}M_n^2 \quad (16)$$

for all $(\theta,t)\in\Theta\times\mathbb{R}$, where $\mathcal{V}(\theta;t) = \mathbb{E}\mathcal{V}_n(\theta;t)$ and $\mathcal{V}_n(\theta,t) = \mathcal{U}_n(\theta;\delta,\mathbf{v}(t))$.

By Lemma 12, there exist $\epsilon_1,\epsilon_2>0$ such that $\mathcal{V}(\hat{\theta},2n^{-\gamma}) - \mathcal{V}(\theta_0,0) \le \epsilon_1 h^R(\|\theta-\theta_0\| + 2n^{-\gamma}) - 2\epsilon_2\left(\|\theta-\theta_0\|^2 + 4n^{-2\gamma}\right)$. Hence, if we choose $\epsilon = \epsilon_2$ and $(\theta,t) = (\hat{\theta},2n^{-\gamma})$ in Equation (16), we have

$$\mathcal{V}_n(\hat{\theta},2n^{-\gamma}) - \mathcal{V}_n(\theta_0,0) \le \epsilon_1 h^R\|\hat{\theta}-\theta_0\| - \epsilon_2\|\hat{\theta}-\theta_0\|^2 + 2\epsilon_1 n^{-\gamma}h^R - 4\epsilon_2 n^{-2\gamma} + n^{-\frac{2}{3}}M_n^2.$$

Moreover, by Lemma 4, we have

$$\mathcal{V}_n(\hat{\theta}, 2n^{-\gamma}) = \mathcal{U}_n(\hat{\theta}; \delta, \mathbf{v}(2n^{-\gamma})) \geq \mathcal{U}_n(\hat{\theta}; \hat{\delta}, \hat{\mathbf{v}}) - \Gamma_n$$
$$\geq \mathcal{U}_n(\theta_0; \hat{\delta}, \hat{\mathbf{v}}) - \Gamma_n \geq \mathcal{U}_n(\theta_0; \delta, \mathbf{v}(0)) - \Gamma_n - T_n = \mathcal{V}_n(\theta_0, 0) - \Gamma_n - T_n.$$

Note that $\epsilon_1 n^{-\gamma} h^R = o(n^{-2\gamma}) = o(n^{-2/3})$ by Assumption O. Hence

$$\epsilon_2 \left\|\hat{\theta} - \theta_0\right\|^2 - \epsilon_1 h^R \left\|\hat{\theta} - \theta_0\right\| \leq n^{-2/3} M_n^2 + \Gamma_n + T_n + 2\epsilon_1 n^{-\gamma} h^R = O_p(n^{-2/3}),$$

from which we obtain a $\sqrt[3]{n}$ rate of convergence for our estimator.

Step 2. Given we have established the $\sqrt[3]{n}$ convergence rate, we now focus on the $\sqrt[3]{n}$–neighborhood of θ_0. Let $Z_n(t) = n^{2/3}\left[\mathcal{U}_n(\theta_0 + tn^{-1/3}; \lambda) - \mathcal{U}_n(\theta_0; \lambda)\right]$. By Kim and Pollard (1990, Theorem 2.7), it suffices to show that

$$Z_n(t) \xrightarrow{d} W(t) - \frac{1}{2} t' V t.$$

Similarly as before, we define

$$f_{i,\ell,n}(t) = \frac{n^{\frac{1}{6}}}{2} \sum_{j=1,2} \Big((2Y_\ell - 1) K_h(X_\ell - X_i) \bar{\vartheta}_j(X_i; \theta_{0j} + t_j n^{-\frac{1}{3}})$$
$$+ (2Y_i - 1) K_h(X_i - X_\ell) \bar{\vartheta}_j(X_\ell; \theta_{0j} + t_j n^{-\frac{1}{3}}) \Big),$$

and a class of functions: $\mathscr{F}_n = \{f_{i,\ell,n} : t \in \mathbb{R}^d\}$. Let further $\widetilde{\mathscr{F}_n} = \{\tilde{f}_{i,\ell,n} : t \in \mathbb{R}^d\}$ where

$$\tilde{f}_{i,\ell,n}(t) = f_{i,\ell,n}(t) - \mathbb{E}[f_{i,\ell,n}(t)|X_i] - \mathbb{E}[f_{i,\ell,n}(t)|X_\ell] + \mathbb{E}[f_{i,\ell,n}(t)]. \tag{17}$$

Denote $\widetilde{F}_n$ the envelope function of $\widetilde{\mathscr{F}_n}$.

With these definitions, $Z_n(t) = \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq \ell} f_{i,\ell,n}(t)$ and

$$Z_n - \mathbb{E}[Z_n] = \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq \ell} \tilde{f}_{i,\ell,n} + 2\sqrt{n} \left(\frac{1}{n} \sum_i \mathbb{E}[\tilde{f}_{i,\ell,n}|X_i] - \mathbb{E}[Z_n] \right). \tag{18}$$

Applying Lemma 9 and by a similar argument as in Lemma 10, $\frac{\sqrt{n}}{n(n-1)} \mathbb{E} \sup_{\tilde{f} \in \widetilde{\mathscr{F}}} |\sum_{i \neq \ell} \tilde{f}_{i,\ell,n}|$ is of order $(nh^d)^{-1/2}$.[16] So the first term in Equation (18) is negligible. The distributional limit of $Z_n - \mathbb{E}[Z_n]$ is then determined by the second term in Equation (18). By Nolan and Pollard (1988,

[16]The extra $n^{1/6}$ re–norming appears in $f_{i,\ell,n}$ is taken care of by the $n^{-1/3}$ term inside the sign functions.

Theorem 5), it is a mean zero Guassian process with covariance kernel defined by

$$H(s,t) = \lim_{c\to\infty} 2 \sum_{j=1,2} \mathbb{E}\left\{\xi_j^2 \bar{\vartheta}_j(X;\theta_0 + t/c, \lambda_j)\bar{\vartheta}_j(X;\theta_0 + s/c, \lambda_j)\right\}.$$

Following the calculation in Kim and Pollard (1990, Example 6.4), we have $\mathbb{E}[Z_n] = -\frac{1}{2}t'Vt$ and

$$\begin{aligned} V &= 2n^{\frac{1}{2}} \left.\frac{\partial^2 \mathbb{E} f_{i,\ell,n}(t)}{\partial t^2}\right|_{t=0} \\ &= 2 \sum_{m=0,1}\sum_{j=1,2} \int \mathbf{1}\{x_j'\beta_j - \alpha_j v_j = 0\} \left[\dot{\kappa}_{m,j}(x_j, v_j)'\theta_{0,j}\right] f_{X_j, v_j(X)}(x_j, v_j)(x_j', v_j)'(x_j', v_j) d\sigma_j^m. \end{aligned}$$

This implies that $Z_n \xrightarrow{d} Z \equiv W - \frac{1}{2}t'Vt$. Given the convergence of Z_n and the $\sqrt[3]{n}$–convergence rate of $\hat{\theta}$, by Kim and Pollard (1990, Theorem 2.7), $n^{1/3}(\hat{\theta} - \theta_0)$ converges in distribution to the random vector that uniquely maximizes $Z(t)$. □

B.5. **Proof of Theorem 5.** The proof of Theorem 3 carries through for each $\theta \in \widehat{\Theta}$, which implies $\sup_{\theta\in\widehat{\Theta}} \rho\left(\theta, \Theta^I\right) \xrightarrow{p} 0$. On the other hand, following the proof of Manski and Tamer (2002, proposition 3.b), we have $\sup_{\theta\in\Theta^I} \rho\left(\theta, \widehat{\Theta}\right) \xrightarrow{p} 0$.

Appendix C. Auxiliary Lemmas

Lemma 5. *Suppose Assumptions G to K hold. Then*

$$\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\hat{\lambda}\right) - \mathcal{U}_n\left(\theta;\lambda\right)\right| = o_p\left(1\right).$$

Proof. It suffices to show that

$$\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\hat{\delta},\hat{\mathbf{v}}\right) - \mathcal{U}_n\left(\theta;\delta,\hat{\mathbf{v}}\right)\right| = o_p\left(1\right), \tag{19}$$

$$\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\delta,\hat{\mathbf{v}}\right) - \mathcal{U}_n\left(\theta;\delta,\mathbf{v}\right)\right| = o_p\left(1\right). \tag{20}$$

We show Equation (19) first. Note that $\mathrm{Sgn}(\cdot)$ only takes value $-1, 0$ or 1, hence

$$\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\hat{\delta},\hat{\mathbf{v}}\right) - \mathcal{U}_n\left(\theta;\delta,\hat{\mathbf{v}}\right)\right| \le \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{\ell\neq i}^{n}\sum_{j=1}^{2}\left|K_h(X_\ell - X_i)\right| \times \left|\hat{\delta}_j(X_i) - \delta_j(X_i)\right|,$$

then we have

$$\mathbb{E}\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\hat{\delta},\hat{\mathbf{v}}\right)-\mathcal{U}_n\left(\theta;\delta,\hat{\mathbf{v}}\right)\right|\leq 2\sum_{j=1}^{2}\mathbb{E}\left\{\left|K_h\left(X_\ell-X_i\right)\right|\times\left|\hat{\delta}_j(X_i)-\delta_j(X_i)\right|\right\},$$

where $\ell\neq i$. It suffices to show that for almost all $x\in\mathscr{X}$, there is

$$\mathbb{E}\left\{\left|K_h(X_\ell-x)\right|\times\left|\hat{\delta}_j(x)-\delta_j(x)\right|\right\}\to 0.$$

Note that

$$\mathbb{E}\left\{\left|K_h(X_\ell-x)\right|\times\left|\hat{\delta}_j(x)-\delta_j(x)\right|\right\}\leq\mathbb{E}\left|K_h(X_\ell-x)\right|\to f(x)\int|K(u)|du<\infty.$$

By dominant convergence theorem, it suffices to show that $\mathbb{E}\left(\left|\hat{\delta}_j(x)-\delta_j(x)\right|\,\middle|X_\ell\right)\to 0$. Because

$$\begin{aligned}\mathbb{E}\left(\left|\hat{\delta}_j(x)-\delta_j(x)\right|\,\middle|X_\ell\right)=\mathbb{P}&\left(\left[2\hat{p}_j(x)-1\right]\hat{f}_X(x)\geq 0>\left[2p_j(x)-1\right]f_X(x)\middle|X_\ell\right)\\&+\mathbb{P}\left(\left[2p_j(x)-1\right]f_X(x)\geq 0>\left[2\hat{p}_j(x)-1\right]\hat{f}_X(x)\middle|X_\ell\right),\end{aligned}$$

thus, by Assumptions I and K,

$$\begin{aligned}&\mathbb{P}\left(\left[2\hat{p}_j(x)-1\right]\hat{f}_X(x)\geq 0>\left[2p_j(x)-1\right]f_X(x)\middle|X_\ell\right)\\&=\mathbb{P}\left(\frac{1}{n-1}\sum_{q\neq i,\ell}^{n}\left\{(2Y_{jq}-1)\times K_h(X_q-x)\right\}\geq 0>\left[2p_j(x)-1\right]f_X(x)\right)+o_p(1)\to 0.\end{aligned}$$

Similarly,

$$\mathbb{P}\left(\left[2p_j(x)-1\right]f_X(x)\geq 0>\left[2\hat{p}_j(x)-1\right]\hat{f}_X(x)\middle|X_\ell\right)\to 0.$$

A similar argument holds for Equation (20), which concludes the proof. □

Lemma 6. *Suppose Assumptions G to K are satisfied, then*

$$\sup_{\theta\in\Theta}\left|\mathcal{U}_n\left(\theta;\lambda\right)-L\left(\theta\right)\right|=o_p(1).$$

Proof. Recall that $\mathcal{U}_n(\theta;\lambda)=\frac{1}{n(n-1)}\sum\sum_{i\neq\ell}\tilde{g}(Z_i,Z_\ell;\theta,\lambda)$ is a $\mathcal{U}$–process ($\tilde{g}$ is defined in Equation 13). It is easy to verify that the function class $\mathscr{G}=\{\tilde{g}:\theta\in\Theta\}$ has an integrable envelope function $\widetilde{G}(Z_i,Z_\ell;\theta,\lambda)$ such that $\mathbb{E}\widetilde{G}<\infty$. Since ϑ is a sum of indicator functions of θ, the VC index of $\mathscr{G}$ is bounded by a constant that only depends on the dimension of regressors d. Then by Van der Vaart and Wellner (1996, Theorem 2.6.7), its $\mathbb{L}_r$ covering number ($r \geq 1$) is bounded by a constant only depends on d. Then it follows from Theorem 7 of Nolan and Pollard (1987) that

$$\sup_{\theta \in \Theta} |\mathcal{U}_n(\theta; \lambda) - \mathbb{E}\mathcal{U}_n(\theta; \lambda)| = o_p(1).$$

To conclude the proof, note that θ only appears in ϑ, hence there exists some constant C such that

$$\sup_{\theta \in \Theta} |L(\theta) - \mathbb{E}\mathcal{U}_n(\theta; \lambda)| \leq C\mathbb{E}|\hat{f}(X_i) - f(X_i)| \xrightarrow{p} 0. \quad \square$$

Lemma 7. *Suppose that Assumptions G and L to O hold. Then we have*

$$\sup_{\theta \in \Theta} |\mathcal{U}_n(\theta; \hat{\delta}, \boldsymbol{\nu}) - \mathcal{U}_n(\theta; \tilde{\delta}, \boldsymbol{\nu})| \leq o_p(n^{-2/3}),$$
$$\sup_{\theta \in \Theta} |\mathcal{U}_n(\theta; \hat{\delta}, \hat{\boldsymbol{\nu}}) - \mathcal{U}_n(\theta; \hat{\delta}, \tilde{\boldsymbol{\nu}})| \leq o_p(n^{-2/3}).$$

Proof. Because of similarity of the proof, here we only show the first inequality. First, because

$$\sup_{\theta \in \Theta} |\mathcal{U}_n(\theta; \hat{\delta}, \mathbf{v}) - \mathcal{U}_n(\theta; \tilde{\delta}, \mathbf{v})| \leq \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\ell \neq i}^{n} \sum_{j=1}^{2} |K_h(X_\ell - X_i)| \times \mathbf{1}[|\hat{\xi}_j(X_i) - \xi_j(X_i)| > n^{-\gamma}].$$

By Assumption N,

$$\mathbb{E} \sup_{\theta \in \Theta} |\mathcal{U}_n(\theta; \hat{\delta}, \mathbf{v}) - \mathcal{U}_n(\theta; \tilde{\delta}, \mathbf{v})| \leq \frac{2\overline{K}}{h^d} \mathbb{E} \left\{ \sum_{j=1}^{2} \mathbf{1}[|\hat{\xi}_j(X_i) - \xi_j(X_i)| > n^{-\gamma}] \right\}.$$

Then it suffices to show that $n^{2/3}h^{-d}\mathbb{P}\left(|\hat{\xi}_j(X_i) - \xi_j(X_i)| > n^{-\gamma}\right) \to 0$, which follows from Lemma 14. $\square$

Lemma 8. *Suppose that Assumptions G and L to O hold, then*

$$\sup_{\theta \in \Theta; t \in \mathbb{R}} |\mathcal{U}_n(\theta; \tilde{\delta}, \boldsymbol{\nu}(t)) - \mathcal{U}_n(\theta; \delta, \boldsymbol{\nu}(t))| \leq o_p(n^{-2/3}).$$

Proof. Because

$$\sup_{\theta\in\Theta;t\in\mathbb{R}} |\mathcal{U}_n(\theta;\tilde{\delta},\mathbf{v}(t)) - \mathcal{U}_n(\theta;\delta,\mathbf{v}(t))|$$

$$\leq \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} \left|\hat{\xi}_j(X_i)\right| \times \left|\mathbf{1}[\tilde{\xi}_j(X_i) \geq 0] - \mathbf{1}[\xi_j(X_i) \geq 0]\right|$$

$$\leq \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} \left|\hat{\xi}_j(X_i)\right| \times \mathbf{1}[|\xi_j(X_i)| \leq n^{-\gamma}] \times \mathbf{1}[|\tilde{\xi}_j(X_i)| \leq n^{-\gamma}].$$

The last step comes from the fact $\sup_x |\tilde{\xi}_j(x) - \xi_j(x)| \leq n^{-\gamma}$.

Thus by Assumption M,

$$\mathbb{E} \sup_{\theta\in\Theta;t\in\mathbb{R}} |\mathcal{U}_n(\theta;\tilde{\delta},\mathbf{v}(t)) - \mathcal{U}_n(\theta;\delta,\mathbf{v}(t))| \leq 2n^{-\gamma}\sum_{j=1}^{2}\mathbb{P}\left(|\xi_j| \leq n^{-\gamma}\right) = O(n^{-2\gamma}).$$

The right hand side is $o(n^{-2/3})$ by Assumption O. This concludes the proof.

Lemma 9. *Let $\bar{\vartheta}_j(X;\theta_j) = \vartheta_j(X;\theta_j,\lambda_j) - \vartheta_j(X;\theta_{0,j},\lambda_j)$. Suppose that Assumptions L and M hold, then for any bounded functions $s(\cdot) : \mathscr{X} \to \mathbb{R}$ and any $t \in \mathbb{R}^d$,*

$$n^{1/3}\int s(X)\bar{\vartheta}^2(X,\theta_0 + t/n^{1/3})f(X)dX = O(\|t\|).$$

Proof. Denote $X_j^* = (X_j', v_j^0)$, $t = (t_1', t_2')'$, $t_j \in \mathbb{R}^{d_j}$. By the functional form of $\bar{\vartheta}$ and the assumption that $s(\cdot)$ is bounded, it is sufficient to verify that

$$n^{1/3}\int\left\{\mathbf{1}[X_j^*\theta_{0j} + X_j^*\frac{t_j}{\sqrt[3]{n}} \geq 0] - \mathbf{1}[X_j^*\theta_{0j} \geq 0]\right\}f(X)dX = O(\|t_j\|).$$

The integral is bounded by

$$n^{1/3}[\mathbb{P}(-X_j^*t_j/n^{1/3} \leq X_j^*\theta_{0j} \leq 0) + \mathbb{P}(-X_j^*t_j/n^{1/3} \geq X_j^*\theta_{0j} \geq 0)]$$

By Assumptions L and M, it is of order $O(\|t_j\|)$. □

Lemma 10. *Suppose that Assumptions G and L to N hold, then*

$$\mathbb{X}_n \equiv \frac{\sqrt{n}}{n(n-1)}\mathbb{E}\sup_{\theta\in\Theta}|\sum_{i\neq\ell}\tilde{\tilde{g}}_{i,\ell,n}| = O\left(\frac{1}{\sqrt{nh^d}}\right).$$

Proof. Write $\tilde{\tilde{g}}_{i,\ell}$ for $\tilde{\tilde{g}}_{i,\ell}(Z_i, Z_\ell; \theta, \lambda)$. Let $\widetilde{\mathscr{G}} = \{\tilde{\tilde{g}}_{i,\ell} : \theta \in \Theta\}$ be the class of functions of $\tilde{\tilde{g}}_{i,\ell}$ and $\tilde{\tilde{G}}$ be its envelop function. Note that $\frac{\sqrt{n}}{n(n-1)} \sum_{i \neq \ell} \tilde{\tilde{g}}_{i,\ell}$ is a degenerated $\mathcal{U}$–process. By Nolan and Pollard (1987, Theorem 6), $\mathbb{X}_n$ is bounded by $n^{-1/2} C_1 (\mathbb{E}\tilde{\tilde{G}}^2)^{1/2}[1 + (\mathbb{E}J_n^2(1))^{1/2}]$, where C_1 is some universal constant and

$$J_n(1) = \int_0^1 \log N_2(x, T_n, \widetilde{\mathscr{G}}, \tilde{\tilde{G}}) dx,$$

with $N_2(x, T_n, \widetilde{\mathscr{G}}, \tilde{\tilde{G}})$ being the $\mathbb{L}_2$ covering number and T_n being the empirical measure on all pairs (Z_i, Z_ℓ) in the definition of $g_{i,\ell}$. Hence it is sufficient to show that $\mathbb{E}\tilde{\tilde{G}}^2 = O\left(h^{-d}\right)$ and $J_n(1)$ is bounded.

Consider $\mathbb{E}\tilde{\tilde{G}}^2$ first. Remember that

$$\tilde{\tilde{g}}(Z_i, Z_\ell; \theta, \lambda) = \tilde{g}(Z_i, Z_\ell; \theta, \lambda) - \bar{g}(Z_i; \theta, \lambda) - \bar{g}(Z_\ell; \theta, \lambda) + \mathcal{V}(\theta, 0).$$

It is not hard to verify that there exists some positive constant C_2 such that

$$\begin{aligned} \mathbb{E}\tilde{\tilde{G}}^2 &\leq C_2 \mathbb{E}[K_h^2(X_\ell - X_i)\bar{\vartheta}^2(X_i, \theta)] \\ &\leq C_2 h^{-d} \int \int K^2(u) \bar{\vartheta}^2(X_i, \theta) f(X_i + uh) f(X_i) du dX_i \end{aligned} \tag{21}$$

The right hand side is $O(h^{-d})$ and since f is bounded and $\int |K(u)| du < \infty$.

Now we consider the term $J_n(1)$. By smoothness Assumptions L to N, we know that as an argument of $\tilde{\tilde{g}}_{i,\ell}$, θ either appears in an indicator function (through the term $\bar{\vartheta}$) or appears in a bounded and continuous function (through the expectations of $\bar{\vartheta}$). By Nolan and Pollard (1987, Lemma 16, 18, and 19), $\tilde{\tilde{\mathscr{G}}}$ is an Euclidean class and hence $J_n(1)$ is finite. □

Lemma 11. *Suppose that Assumptions A to G and L to O hold. Then* $\forall \theta \in \Theta$,

$$\mathcal{U}_n(\theta; \hat{\delta}, \boldsymbol{\nu}(0)) \leq \mathcal{U}_n(\theta; \hat{\delta}, \tilde{\boldsymbol{\nu}}) \leq \mathcal{U}_n(\theta; \hat{\delta}, \boldsymbol{\nu}(2n^{-\gamma})).$$

Proof. Write $\hat{\xi}_{ji}$ for $\hat{\xi}_j(X_i)$. Note that when $\tilde{v}_j^0 \le \tilde{v}_j^1$, we have

$$
\begin{aligned}
&\mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}}) = L_n(a,b;\hat{p},\hat{f}_X,\hat{\delta},\tilde{\mathbf{v}}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\hat{\xi}_j(X_i)\left\{\mathbf{1}[\hat{\xi}_j(X_i)\ge 0]\,\mathrm{sgn}(X_{ji}'b_j-\alpha_j\tilde{v}_j^0)+\mathbf{1}[\hat{\xi}_j(X_i)<0]\,\mathrm{sgn}(X_{ji}'b_j-\alpha_j\tilde{v}_j^1)\right\} \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\hat{\xi}_{ji}\,\mathrm{sgn}(\hat{\xi}_{ji}) - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\hat{\xi}_{ji}^{-}\times\left\{\mathbf{1}\left[X_{ji}'b_j-a_j\tilde{v}_j^1(X_i)\ge 0\right]\right\} \\
&\qquad\qquad - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\hat{\xi}_{ji}^{+}\times\left\{\mathbf{1}\left[X_{ji}'b_j-a_j\tilde{v}_j^0(X_i)\le 0\right]\right\},
\end{aligned}
$$

where $\hat{\xi}_{ji}^{-} = -\min\{\hat{\xi}_{ji},0\}$ and $\hat{\xi}_{ji}^{+} = \max\{\hat{\xi}_{ji},0\}$.

By construction, $\tilde{v}_j^0 \le v_j^0 \le v_j^1 \le \tilde{v}_j^1$. Thus

$$
\begin{aligned}
\mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}}) - \mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}) &= \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\hat{\xi}_i^{-}\times\left\{\mathbf{1}\left[a_j v_j^1(X_i)\le X_{ji}'b_j < a_j\tilde{v}_j^1(X_i)\right]\right\} \\
&\quad + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\hat{\xi}_i^{+}\times\left\{\mathbf{1}\left[a_j\tilde{v}_j^0(X_i) < X_{ji}'b_j \le a_j v_j^0(X_i)\right]\right\} \ge 0.
\end{aligned}
$$

Similarly, we can show $\mathcal{U}_n(\theta;\hat{\delta},\tilde{\mathbf{v}}) \le \mathcal{U}_n(\theta;\hat{\delta},\mathbf{v}(2n^{-\gamma}))$. □

Lemma 12. *Suppose that Assumptions A to G and L to O hold. Thus, there exist $\epsilon_1,\epsilon_2\in\mathbb{R}^+$ such that*

$$\mathcal{V}(\hat{\theta},2n^{-\gamma}) - \mathcal{V}(\theta_0,0) \le \epsilon_1 h^R(\|\hat{\theta}-\theta_0\| + 2n^{-\gamma}) - 2\epsilon_2\left(\|\hat{\theta}-\theta_0\|^2 + 4n^{-2\gamma}\right)$$

Proof. First, by Taylor expansion,

$$
\begin{aligned}
\mathcal{V}(\theta,t) - \mathcal{V}(\theta_0,0) &= \mathcal{V}_\theta(\theta_0,0)(\theta-\theta_0) + \mathcal{V}_t(\theta_0,0)\times t \\
&\quad + (\theta-\theta_0)'\mathcal{V}_{\theta\theta}(\theta^\dagger,t^\dagger)(\theta-\theta_0) + \mathcal{V}_{tt}(\theta^\dagger,t^\dagger)t^2 + 2(\theta-\theta_0)'\mathcal{V}_{\theta t}(\theta^\dagger,t^\dagger)t
\end{aligned}
$$

where $(\theta^\dagger, t^\dagger)$ between $(\theta_0, 0)$ and (θ, t). Because $\mathcal{V}_{\theta\theta}$, $\mathcal{V}_{\theta t}$ and $\mathcal{V}_{tt}$ are continuous in a neighborhood of $(\theta_0, 0)$ and strictly negative definite at $(\theta_0, 0)$, then

$$(\theta-\theta_0)'\mathcal{V}_{\theta\theta}(\theta^\dagger, t^\dagger)(\theta-\theta_0) + \mathcal{V}_{tt}(\theta^\dagger, t^\dagger)t^2 + 2(\theta-\theta_0)'\mathcal{V}_{\theta t}(\theta^\dagger, t^\dagger)t \le -2\epsilon_2\left(\|\theta-\theta_0\|^2 + t^2\right)$$

for some $\epsilon_2 > 0$ in a neighborhood of $(\theta_0, 0)$. Thus we have

$$\mathcal{V}(\hat{\theta}, 2n^{-\gamma}) - \mathcal{V}(\theta_0; 0) \le \mathcal{V}_\theta'(\theta_0, 0)(\hat{\theta}-\theta_0) + \mathcal{V}_t(\theta_0, 0) \times 2n^{-\gamma} - 2\epsilon_2\left(\left\|\hat{\theta}-\theta_0\right\|^2 + 4n^{-2\gamma}\right).$$

It remains to show that $\|\mathcal{V}_\theta(\theta_0, 0)\| \le \epsilon_1 h^R$ and $\|\mathcal{V}_t(\theta_0, 0)\| \le \epsilon_1 h^R$ for some $\epsilon_1 > 0$.

Let $\overline{\mathcal{V}}(\theta, t) \equiv \sum_{j=1,2}\mathbb{E}\left\{\xi_j(X) \times \vartheta_j(X; \theta, \delta, \mathbf{v}(t))\right\}$. Note that

$$\begin{aligned}\mathbb{E}\left\{\xi_j(X) \times \vartheta_j(X; \theta, \delta, \mathbf{v}(t))\right\} = -\mathbb{E}\xi_j(X) + 2\mathbb{E}\xi_j(X)\mathbf{1}(X_j'b_j - a_j v_j^0(X) + a_j t \ge 0)\mathbf{1}(\xi_j(X) \ge 0) \\ + 2\mathbb{E}\xi_j(X)\mathbf{1}(X_j'b_j - a_j v_j^1(X) - a_j t \ge 0)\mathbf{1}(\xi_j(X) < 0).\end{aligned}$$

Let $\bar{\xi}_j^+(x_j, v_j) = \mathbb{E}\left[\xi_j^+(X)|X_j = x_j, v_j^0(X) = v_j\right]$ and $\bar{\xi}_j^-(x_j, v_j) = \mathbb{E}\left[\xi_j^-(X)|X_j = x_j, v_j^1(X) = v_j\right]$. Thus

$$\begin{aligned}\overline{\mathcal{V}}_\theta(\theta_0, 0) = 2\sum_{j=1}^{2}\left(I + \|\theta_0\|^{-2}\theta_0\theta_0'\right)\int \mathbf{1}\left(x_j'b_j - a_j v_j = 0\right)\bar{\xi}_j^+(x_j, v_j) f_{X_j, v_j^0(X)}(x_j, v_j)(x_j', v_j)' d\sigma^0 \\ - 2\sum_{j=1}^{2}\left(I + \|\theta_0\|^{-2}\theta_0\theta_0'\right)\int \mathbf{1}\left(x_j'b_j - a_j v_j = 0\right)\bar{\xi}_j^-(x_j, v_j) f_{X_j, v_j^1(X)}(x_j, v_j)(x_j', v_j)' d\sigma^1 = 0\end{aligned}$$

where σ^0, σ^1 are surface measures on the corresponding lines and the last step comes from the fact that along the line $[X_j'b_j - a_j v_j^0(X) = 0]$, we have $\xi_j^+(X) = 0$; and along the other line $[X_j'b_j - a_j v_j^1(X) = 0]$, we have $\xi_j^-(X) = 0$. Similarly, $\mathcal{V}_t(\theta_0, 0) = 0$. So it suffices to show that $\|\mathcal{V}_\theta(\theta_0, 0) - \overline{\mathcal{V}}_\theta(\theta_0, 0)\| \le \epsilon_1 h^R$ and $\|\mathcal{V}_t(\theta_0, 0) - \overline{\mathcal{V}}_t(\theta_0, 0)\| \le \epsilon_1 h^R$ for some $\epsilon_1 > 0$.

Consider

$$\mathcal{V}(\theta, t) - \overline{\mathcal{V}}(\theta, t) = \sum_{j=1,2}\mathbb{E}\left\{\int_{\mathbb{R}^d}[\xi_j(X + hs) - \xi_j(X)]K(s)ds \times \vartheta_j(X; \theta, \delta, \mathbf{v}(t))\right\}.$$

Let $\phi_j(x) = \int_{\mathbb{R}^d} \xi_j(x+ht)K(t)dt - \xi_j(x)$. By Assumption N, $\forall x, n, j$, $\phi_j(x) \le C_1 \times h^R$ for some $C_1 \in \mathbb{R}^+$. Because

$$\frac{\partial \mathbb{E}\left\{\phi_j(X) \times \vartheta_j(X;\theta,\lambda)\right\}}{\partial\theta} = 2\frac{\partial \mathbb{E}\phi_j(X)\mathbf{1}(X_j'b_j - a_j\nu_j^0(X) \ge 0)\mathbf{1}(\xi_j(X) \ge 0)}{\partial\theta} + 2\frac{\partial \mathbb{E}\phi_j(X)\mathbf{1}(X_j'b_j - a_j\nu_j^1(X) \ge 0)\mathbf{1}(\xi_j(X) < 0)}{\partial\theta}.$$

Let further $\zeta_{n,j}^1(x_j, v_j) = \mathbb{E}\left[\phi_j(X)\mathbf{1}(\xi_j(X) \ge 0)|X_j = x_j, \nu_j^1(X) = v_j\right]$ and, similarly, $\zeta_{n,j}^0(x_j, v_j) = \mathbb{E}\left[\phi_j(X)\mathbf{1}(\xi_j(X) < 0)|X_j = x_j, \nu_j^0(X) = v_j\right]$. Then, by calculation,

$$\begin{aligned}
&\frac{\partial \mathbb{E}\phi_j(X)\mathbf{1}(X_j'b_j - a_j\nu_j^0(X) \ge 0)\mathbf{1}(\xi_j(X) \ge 0)}{\partial\theta}\Big|_{\theta=\theta_0} \\
&= \frac{\partial \int_{\mathbb{R}}\int_{\mathbb{R}^{d_j}} \zeta_{n,j}^0(x_j, v_j) f_{X_j,\nu_j^0(X)}(x_j, v_j)\mathbf{1}(x_j'b_j - a_j v_j(X) \ge 0)dx_j dv_j}{\partial\theta}\Big|_{\theta=\theta_0} \\
&= (I + |\theta_0|^{-2}\theta_0\theta_0') \int_{\mathbb{R}^{d_j+1}} \mathbf{1}(x_j'\beta_j - \alpha_j v_j(x) = 0)\zeta_{n,j}^0(x_j, v_j) f_{X_j,\nu_j^0(X)}(x_j, v_j)(x_j', v_j)' d\sigma^0.
\end{aligned}$$

Since $\zeta_{n,j}^0(x_j, v_j) \le C_1 h^R$ for all x_j, v_j and n. Then there exists a constant $C \in \mathbb{R}^+$ such that

$$\left\|\frac{\partial \mathbb{E}\phi_j(X)\mathbf{1}(X_j'b_j - a_j\nu_j^0(X) \ge 0)\mathbf{1}(\xi_j(X) \ge 0)}{\partial\theta}\Big|_{\theta=\theta_0}\right\| \le Ch^R.$$

Therefore, we can choose $\epsilon_1 > 0$ such that $\left\|\mathcal{V}_\theta(\theta_0, 0) - \overline{\mathcal{V}}_\theta(\theta_0, 0)\right\| \le \epsilon_1 h^R$. By a similar argument, we also have $\left\|\mathcal{V}_t(\theta_0, 0) - \overline{\mathcal{V}}_t(\theta_0, 0)\right\| \le \epsilon_1 h^R$. □

Lemma 13 (Bernstein's tail inequality)**.** *Let $X_1, \cdots, X_n$ be independent real-valued random variables with zero mean, such that $\forall$ i, $|X_i| \le M$ a.s. Defining $\sigma^2 = n^{-1}\sum_{i=1}^n Var(X_i)$ and $S_n = \sum_{i=1}^n X_i$. Then for any $\epsilon > 0$, we have*

$$\mathbb{P}\left(\frac{1}{n}|S_n| > \epsilon\right) \le 2\exp\left(-\frac{n\epsilon^2}{2\sigma^2 + \frac{2}{3}M\epsilon}\right).$$

Lemma 14. *Suppose that assumptions G, L to O hold. Then for any $k > 0$*

$$n^k \mathbb{P}\left\{|\hat{\xi}_j(X_i) - \xi_j(X_i)| > n^{-\gamma}\right\} \to 0.$$

Proof. By Dominated Convergence theorem, it suffices to show $n^k \mathbb{P}\left\{|\hat{\xi}_j(x) - \xi_j(x)| > n^{-\gamma}\right\} \to 0$ for any $x \in \mathscr{X}$.

Because

$$\mathbb{P}\left\{|\hat{\xi}_j(x) - \xi_j(x)| > n^{-\gamma}\right\} \le \mathbb{P}\left\{|\hat{\xi}_j(x) - \mathbb{E}\hat{\xi}_j(x)| + |\mathbb{E}\hat{\xi}_j(x) - \xi_j(x)| > n^{-\gamma}\right\}$$
$$\le \mathbb{P}\left\{|\hat{\xi}_j(x) - \mathbb{E}\hat{\xi}_j(x)| > n^{-\gamma} - |\mathbb{E}\hat{\xi}_j(x) - \xi_j(x)|\right\} = \mathbb{P}\left\{\frac{1}{n}\left|\sum_{\ell=1}^{n}(w_{j\ell} - \mathbb{E}w_{j\ell})\right| > \tau_{jn}\right\},$$

where $w_{j\ell} = (2y_{j\ell} - 1) \times K\left(\frac{X_\ell - x}{h}\right)$ and $\tau_{jn} = h^d\left[n^{-\gamma} - |\mathbb{E}\hat{\xi}_j(x) - \xi_j(x)|\right]$. Thus, by Bernstein's tail inequality (Lemma 13),

$$\mathbb{P}\left\{\frac{1}{n}\left|\sum_{\ell=1}^{n}(w_{j\ell} - \mathbb{E}w_{j\ell})\right| > \tau_n\right\} \le 2\exp\left(-\frac{n\tau_n^2}{2\mathrm{Var}(w_{j\ell}) + \frac{2}{3}\overline{K}\tau_n}\right).$$

Note that by Assumption M and N, we have $\mathbb{E}\hat{\xi}_j(x) - \xi_j(x) = O_p(h^R)$. Then under Assumption O, for sufficient large n, we have $0.5h^d n^{-\gamma} \le \tau_{jn} \le h^d n^{-\gamma}$. It should also be noted that

$$\mathrm{Var}(w_{j\ell}) \le \mathbb{E}w_{j\ell}^2 \le \mathbb{E}K^2\left(\frac{X - x}{h}\right) \le C_0 h^d,$$

where $C_0 = \overline{K}^2 \sup_x f_X(x) < \infty$. Hence, we have

$$\mathbb{P}\left\{\frac{1}{n}\left|\sum_{\ell=1}^{n}(w_{j\ell} - \mathbb{E}w_{j\ell})\right| > \tau_n\right\} \le 2\exp\left(-\frac{\frac{1}{4}nh^{2p}n^{-2\gamma}}{2C_0h^d + \frac{2}{3}\overline{K}h^d n^{-\gamma}}\right) = 2\exp\left(-\frac{\frac{1}{4}nh^d n^{-2\gamma}}{2C_0 + \frac{2}{3}\overline{K}n^{-\gamma}}\right).$$

For sufficient large n, we have $\frac{2}{3}\overline{K}n^{-\gamma} \le 1$ and $nh^d n^{-2\gamma} > n^{2r_0}$ (by Assumption O). Hence,

$$\mathbb{P}\left\{\frac{1}{n}\left|\sum_{\ell=1}^{n}(w_{j\ell} - \mathbb{E}w_{j\ell})\right| > \tau_n\right\} \le 2\exp\left(-\frac{n^{2r_0}}{2C_0 + 1}\right).$$

Therefore, given arbitrary $k > 0$

$$n^k \mathbb{P}\left\{|\hat{\xi}_j(x) - \xi_j(x)| > n^{-\gamma}\right\} \le n^k \exp\left(-\frac{n^{2r_0}}{2C_0 + 1}\right) \to 0. \qquad \square$$

Appendix D. Tables in Section 6

Table 1. Average bias and std of mis–specified MLE (Figure 1)

	$n = 500$	$n = 1000$	$n = 3000$
$r = 0$	0.074(0.426)	0.085(0.291)	0.060(0.167)
$r = 0.1$	0.110(0.418)	0.114(0.285)	0.093(0.159)
$r = 0.2$	0.146(0.414)	0.140(0.284)	0.141(0.161)
$r = 0.3$	0.186(0.390)	0.200(0.278)	0.188(0.153)
$r = 0.4$	0.275(0.398)	0.285(0.278)	0.272(0.149)
$r = 0.5$	1.085(0.397)	0.382(0.280)	0.367(0.147)
$r = 0.6$	0.461(0.373)	0.475(0.263)	0.461(0.145)
$r = 0.7$	0.583(0.379)	0.587(0.252)	0.570(0.144)

Standard deviation reported in the parentheses.

Table 2. Average bias and std of our estimator (Figure 1)

	$n = 500$	$n = 1000$	$n = 3000$
$r = 0$	0.018(1.108)	0.003(1.002)	0.042(0.801)
$r = 0.1$	0.128(1.068)	0.007(0.992)	0.070(0.742)
$r = 0.2$	0.030(1.042)	0.051(0.950)	0.042(0.770)
$r = 0.3$	0.045(1.072)	0.042(0.911)	0.043(0.781)
$r = 0.4$	0.086(1.047)	0.063(0.965)	0.120(0.802)
$r = 0.5$	0.095(1.043)	0.125(0.953)	0.071(0.732)
$r = 0.6$	0.076(1.063)	0.067(0.900)	0.046(0.747)
$r = 0.7$	0.065(1.020)	0.080(0.946)	0.061(0.734)

Standard deviation reported in the parentheses.

Table 3. Average bias and std under different DGPs (Figure 3)

DGP	Estimator	$n = 1000$	$n = 3000$
eq. (1)	Ours	0.045(0.771)	0.029(0.527)
	L–Z's	−0.284(0.109)	−0.272(0.071)
eq. (11)	Ours	0.057(0.742)	0.070(0.589)
	L–Z's	0.004(0.111)	0.001(0.074)

Standard deviation reported in the parentheses.