Journal of Econometrics 157 (2010) 409–431

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Semiparametric estimation of a simultaneous game with incomplete information

Andres Aradillas-Lopez

Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, United States

ARTICLE INFO

Article history:
Received 28 August 2007
Received in revised form
15 March 2010
Accepted 16 March 2010
Available online 31 March 2010

JEL classification:
C14
C35
C62
C72

Keywords:
Semiparametric estimation
Trimming
Game-theoretic econometric models
Equilibrium beliefs

ABSTRACT

This paper studies the problem of estimating the normal-form payoff parameters of a simultaneous, discrete game where the realization of such payoffs is not common knowledge. The paper contributes to the existing literature in two ways. First, by making a comparison with the complete information case it formally describes a set of conditions under which allowing for private information in payoffs facilitates the identification of various features of the game. Second, focusing on the incomplete information case it presents an estimation procedure based on the equilibrium properties of the game that relies on weak semiparametric assumptions and relatively flexible information structures which allow players to condition their beliefs on signals whose exact distribution function is unknown to the researcher. The proposed estimators recover unobserved beliefs by solving a semiparametric sample analog of the population Bayesian–Nash equilibrium conditions. The asymptotic features of such estimators are studied for the case in which the distribution of unobserved shocks is known and the case in which it is unknown. In both instances equilibrium uniqueness is assumed to hold only in a neighborhood of the true parameter value and for a subset $\mathcal{Z}$ of realizations of the signals. Multiple equilibria are allowed elsewhere in the parameter space and no equilibrium selection theory is involved. Extensions to games where beliefs are conditioned on unobservables as well as general games with many players and actions are also discussed. An empirical application of a simple capital investment game is included.

1. Introduction

Recently, much attention has been devoted to the problem of econometric estimation of game-theoretic models with multiple equilibria. It has been argued Morris and Shin (2003) that multiple equilibria are often the result of assuming that agents have complete knowledge about the primitives of the model. The effect of these assumptions is to generate the type of coordinating behavior that results in multiple equilibria. The general issues that motivate this paper are addressed by revisiting a 2×2 game studied by Bresnahan and Reiss (1990), Bresnahan and Reiss (1991) and Tamer (2003). Such game-theoretic models were first considered in an econometric context in the pioneering work by Bjorn and Vuong (1984). If players have complete information and play pure strategies, their Nash equilibrium choices are described by a simultaneous discrete response model (see the thorough analysis by Heckman (1978)). We depart from a complete information structure and assume an incomplete information environment in which players observe a noisy signal of the game's normal-form payoffs before making their choices simultaneously. Some examples of estimation methods for *static* incomplete-information game-theoretic models include Brock and Durlauf (2001), Seim (2006), Sweeting (2010), Pesendorfer and Schmidt-Dengler (2008), and Bajari et al. (forthcoming).[1]

This paper contributes to the existing literature on estimation of simultaneous incomplete-information games in three ways. First, we assume a flexible information structure. Instead of confining the source of private information exclusively to the realization of unobservable shocks that are independent of all other variables in the model, we allow players to condition their beliefs on a vector of "signals" which are statistically related to the players' private information. We allow for some payoff covariates to be privately observed by the players when the game is played, but observable to the econometrician afterwards. Throughout, we allow interdependence between players' private information. Second, we propose estimation procedures that are based entirely on the Bayesian–Nash equilibrium (BNE) properties of the game. Finally, by comparing it to the complete information case, the paper illustrates the sense in which the presence of private information allows us to point-identify more features of the model, such as the conditional probabilities for each of the individual outcomes of the game.

[1] We focus exclusively on static games of incomplete information whose Bayesian–Nash equilibrium properties are directly comparable to those in this paper. We will not revisit, for example, the increasingly vast literature on the estimation dynamic games and auctions.

E-mail address: aaradill@ssc.wisc.edu.

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2010.03.043

Players here are not required to have exact knowledge about all parameters and distributional properties of the variables in the model. Instead, they are only assumed to be capable of predicting – in the spirit of Aumann (1987) – the distribution of *equilibrium* outcomes conditional on the realization of the vector of signals. Players' posterior beliefs are based on this conditional distribution. Players' beliefs are assumed to be unobserved by the econometrician, who also ignores the true distribution of the players' private information and the signals used. As a result, equilibrium beliefs have an unknown functional form and the estimation problem becomes semiparametric. The proposal will be to estimate equilibrium beliefs and payoff parameters simultaneously by solving a well-defined sample analog to the population equilibrium conditions and plugging the solution into a trimmed-likelihood function. We will characterize the asymptotic properties of such a procedure in two cases: (a) when the distribution of players' unobservable shocks is known up to a finite parameter value, and (b) when such a distribution is unknown. In both cases, we will enumerate conditions for $\sqrt{N}$-consistency and we will show that the efficiency of the proposed estimator is tied to the ability of the signals to explain the variation in the players' private information. To the best of our knowledge, this paper constitutes the first effort to simultaneously estimate beliefs and payoff parameters under explicitly weak semiparametric assumptions. It is also the first paper to illustrate explicitly the sense in which strategic interaction models with incomplete information are easier to identify than those with complete information.

The paper is organized as follows. Section 2 describes and analyzes the equilibrium properties of the game in question. Section 3 deals with estimation when the distribution of players' unobserved shocks is assumed to be known. The case in which such distribution is unknown is presented in Section 4. Section 5 discusses extensions to more general games. A simple empirical example of capital investment is included in Section 6. Concluding remarks are provided in Section 7. All proofs are included in the Appendix.

2. Properties of the game

Take $(X_1, \varepsilon_1) \in \mathbb{R}^{k_1} \times \mathbb{R}$ and $(X_2, \varepsilon_2) \in \mathbb{R}^{k_2} \times \mathbb{R}$, and denote $X \equiv (X_1, X_2), \varepsilon \equiv (\varepsilon_1, \varepsilon_2)$ and $k \equiv k_1+k_2$. Consider a simultaneous game with the following normal-form

		Player 2	
		$Y_2 = 1$	$Y_2 = 0$
PLAYER 1	$Y_1 = 1$	$X_1'\beta_1 - \varepsilon_1 + \alpha_1$, $X_2'\beta_2 - \varepsilon_2 + \alpha_2$	$X_1'\beta_1 - \varepsilon_1, 0$
	$Y_1 = 0$	$0, X_2'\beta_2 - \varepsilon_2$	$0, 0$

Upper case letters will be used to denote random variables, and lower case to denote particular realizations. $\mathbb{S}(U)$ will denote the support of a random variable U. Subscript $p \in \{1, 2\}$ will denote a particular player, and $-p$ will denote his opponent. We refer to $(\alpha_1, \alpha_2) \in \mathbb{R}^2$ as the strategic interaction parameters, which summarize the interaction effect between players' actions. Linearity of the index $X_p'\beta_p$ is not essential to our results. Identification with an alternative functional form would require Assumption A4(i) (full rank) below to be modified accordingly. The exclusion restriction between X_1 and X_2 in Assumption B1 would still need to be satisfied. The strategic interaction effect could also depend on (observable) covariates, and we explore this in the empirical example of Section 6. The Bayesian–Nash equilibrium analysis in Sections 2.2–2.4 hinges on the additive separability between X_p and ε_p. We leave the non-additively separable case for future research.

2.1. Players' information and optimal actions

We begin with the following assumption.

Assumption A1 (*Information Structure and Beliefs*).

(i) Player p observes the realization of (X_p, ε_p). The realization of ε_p is only privately observed by player p. We allow the possibility that the realization of some of the elements in X are only privately observed when the game is played, and we allow for the existence of a publicly observed vector of "signals" Z which is informative about X. If X is publicly observed then $Z = X$. Otherwise, $Z \neq X$.
(ii) Once Z is revealed, each player constructs a subjective assessment of $\Pr(Y_1, Y_2|Z)$, the joint distribution of optimal choices conditional on Z. As in Savage (1972), both players are assumed to be Bayesian expected utility maximizers given their beliefs, and this is assumed to be common knowledge.[2]

Our framework allows statistical interdependence between players' private information, and in particular between ε_1 and ε_2. However, players are not required or assumed to have exact knowledge about the distributions involved. Instead, as (A1(ii)) states, players are only assumed to be capable of predicting the distribution of *equilibrium* outcomes conditional on Z. As in Aumann (1987), *players do not condition their beliefs on their choices per-se, but on all the substantive information that leads them to make those choices*. The statistical interdependence between their private information is enclosed therein. In equilibrium, each player will choose a definite pure strategy by following a simple threshold-crossing decision rule. The apparent random nature of players' actions is a result of payoffs not being common knowledge.

Assumption A2 (*Distributional Properties of ε_1 and ε_2*).

(i) $(\varepsilon_1, \varepsilon_2)$ are jointly continuously distributed random variables with unbounded support conditional on each other. They are allowed to be correlated, but are assumed to be independent of all other variables in the model. The conditional support $\mathbb{S}(\varepsilon_p|\varepsilon_{-p})$, is assumed to be unbounded for $p = 1, 2$, for any possible realization of ε_{-p}.
(ii) The scale of ε_p is normalized to 1. From now on, the parameter vector (β_p, α_p) will be interpreted as relative to the scale of ε_p. We will denote the marginal distribution of ε_p by $G_p(\epsilon_p)$, with density $g_p(\epsilon_p)$. The joint distribution of $(\varepsilon_1, \varepsilon_2)$ is given by

$$G_{1,2}(\epsilon_1, \epsilon_2; \rho) = \mathcal{C}(G_1(\epsilon_1), G_2(\epsilon_2); \rho),$$

with corresponding joint density denoted by $g_{1,2}(\epsilon_1, \epsilon_2; \rho)$. By Sklar's Theorem (Sklar (1959) and Nelsen (2006) Chapter 2.3), $\mathcal{C}$ is interpreted as a copula function which depends on a finite-dimensional parameter ρ which summarizes the dependence between ε_1 and ε_2. We will focus on the case in which ρ is one-dimensional and denote $\theta = (\beta_1, \alpha_1, \beta_2, \alpha_2, \rho) \in \mathbb{R}^{k+3}$.

We normalize the scale of ε_p for identification purposes, since player's optimal actions will be given by threshold-crossing rules. Our results can potentially be extended to the case where ρ is vector-valued. Depending on the specific copula family a normalization of the vector ρ may be required.

2.2. Equilibrium beliefs and actions

Given the normal-form of the game and A1–A2, players' optimal actions are given by

[2] An event is considered common knowledge here if it is known to both players, it is known to both players that it is known to both players and so on ad infinitum.

$$Y_1 = \mathbb{1}\{X_1'\beta_1 + \alpha_1 \underbrace{\Pr_1(Y_2 = 1|Y_1 = 1, Z)}_{\text{Player 1's beliefs.}} - \varepsilon_1 \geq 0\}$$
$$Y_2 = \mathbb{1}\{X_2'\beta_2 + \alpha_2 \underbrace{\Pr_2(Y_1 = 1|Y_2 = 1, Z)}_{\text{Player 2's beliefs.}} - \varepsilon_2 \geq 0\}. \tag{1}$$

Each player uses a pure-strategy based on (1).[3] We will carefully examine conditions under which there exist a pair of *self-consistent* equilibrium beliefs that satisfy (1). Take any pair of scalars $\pi_1, \pi_2 \in \mathbb{R}^2$ and define

$$\varphi_1(\pi_1, \pi_2; \theta, Z) = \pi_1 - \frac{E\left[G_{1,2}(X_1'\beta_1 + \alpha_1\pi_2, X_2'\beta_2 + \alpha_2\pi_1; \rho)|Z\right]}{E\left[G_2(X_2'\beta_2 + \alpha_2\pi_1)|Z\right]},$$
$$\varphi_2(\pi_1, \pi_2; \theta, Z) = \pi_2 - \frac{E\left[G_{1,2}(X_1'\beta_1 + \alpha_1\pi_2, X_2'\beta_2 + \alpha_2\pi_1; \rho)|Z\right]}{E\left[G_1(X_1'\beta_1 + \alpha_1\pi_2)|Z\right]} \tag{2}$$
$$\varphi(\pi_1, \pi_2; \theta, Z) = (\varphi_1(\pi_1, \pi_2; \theta, Z), \varphi_2(\pi_1, \pi_2; \theta, Z)).$$

Proposition 1. *Fix a realization $z \in \mathbb{S}(Z)$ and a value of θ. For these fixed values we can view $\varphi(\pi_1, \pi_2; \theta, z)$ as a function of (π_1, π_2). If Assumptions A1 and A2 are satisfied, players' beliefs are deterministic given z and any pair of self-consistent beliefs that satisfy (1) must solve for (π_1, π_2) the system*

$$\varphi(\pi_1, \pi_2; \theta, z) = 0. \tag{3}$$

From now on, we will say that "there exists an equilibrium" for a given z and θ if there exists a pair (π_1, π_2) that solves (3). If there is more than one solution, we will say that there exist multiple equilibria. Conditions for existence and uniqueness of a solution to (3) will be crucial to our analysis. The following section addresses these issues in detail.

2.3. Existence of equilibria

Fix a realization $z \in \mathbb{S}(Z)$ and a value of θ. We will let $[0, 1]^2$ denote the unit-square in $\mathbb{R}^2$. For $(\pi_1, \pi_2) \in [0, 1]^2$, consider the mapping[4] $\Psi : [0, 1]^2 \to [0, 1]^2$ given by

$$\Psi(\pi_1, \pi_2; \theta, z) = \left(\frac{E\left[G_{1,2}(X_1'\beta_1 + \alpha_1\pi_2, X_2'\beta_2 + \alpha_2\pi_1; \rho)|Z = z\right]}{E\left[G_2(X_2'\beta_2 + \alpha_2\pi_1)|Z = z\right]}, \frac{E\left[G_{1,2}(X_1'\beta_1 + \alpha_1\pi_2, X_2'\beta_2 + \alpha_2\pi_1; \rho)|Z = z\right]}{E\left[G_1(X_1'\beta_1 + \alpha_1\pi_2)|Z = z\right]} \right)'. \tag{4}$$

Given A2, Ψ satisfies the requirements of Brouwer's Fixed Point Theorem[5] and there exists $(\pi_1^*, \pi_2^*) \equiv \pi^* \in (0, 1)^2$ such that $\pi^* = \Psi(\pi^*; \theta, z)$. By construction, π^* solves (3).

2.4. Cardinality of equilibria

2.4.1. Uniqueness via Jacobian restrictions

Denote the Jacobian matrix

$$\underset{2\times 2}{J(\pi_1, \pi_2; \theta, z)} = \nabla_\pi \varphi(\pi_1, \pi_2; \theta, z) = \begin{pmatrix} \nabla_{\pi_1}\varphi_1(\pi_1, \pi_2; \theta, z) & \nabla_{\pi_2}\varphi_1(\pi_1, \pi_2; \theta, z) \\ \nabla_{\pi_1}\varphi_2(\pi_1, \pi_2; \theta, z) & \nabla_{\pi_2}\varphi_2(\pi_1, \pi_2; \theta, z) \end{pmatrix}. \tag{5}$$

Definition 1. Fix $z \in \mathbb{S}(Z)$ and θ and suppose (π_1', π_2') solves (3). We will say that this solution is *locally unique* if there exists a neighborhood $\mathcal{N}$ around (π_1', π_2') such that no other $(\pi_1, \pi_2) \in \mathcal{N}$ solves (3). We will say it is *unique* if no other pair $(\pi_1, \pi_2) \in [0, 1]^2$ solves (3). We will say that it is *regular* if $J(\pi_1', \pi_2'; \theta, z)$ is invertible and *critical* otherwise.

Using Theorem 7 in Gale and Nikaido (1965) we obtain the following uniqueness result:

Proposition 2. *Fix $z \in \mathbb{S}(Z)$ and a value of θ. If none of the principal minors of $J(\pi_1, \pi_2; \theta, z)$ vanishes in the set $(\pi_1, \pi_2) \in [0, 1]^2$, the solution to (3) is unique.*

If Proposition 2 holds, $\varphi(\pi_1, \pi_2; \theta, z)$ is a one-to-one mapping everywhere[6] in $(\pi_1, \pi_2) \in [0, 1]^2$. Since $\mathbb{S}(\varepsilon) = \mathbb{R}^2$, the conditions of Proposition 2 are satisfied only if

$$\nabla_{\pi_1}\varphi_1(\pi_1, \pi_2; \theta, z) > 0, \quad \text{and} \quad \nabla_{\pi_2}\varphi_2(\pi_1, \pi_2; \theta, z) > 0 \quad \text{for all } (\pi_1, \pi_2) \in [0, 1]^2. \tag{6}$$

In Appendix A.4 of the Appendix we show that if (6) holds, we can represent the loci $\varphi_1(\pi_1, \pi_2; \theta, z) = 0$ and $\varphi_2(\pi_1, \pi_2; \theta, z) = 0$ as continuous, monotonic curves in $[0, 1]^2$. The signs of their slopes are equal to the signs of α_1 and α_2, respectively.

Fig. 1 presents four graphical examples. All points of intersection between the two curves are solutions to (3) and therefore candidate equilibrium beliefs. As (A) shows, uniqueness will follow whenever $\alpha_1\alpha_2 \leq 0$. (B) shows that uniqueness is also possible if $\alpha_1\alpha_2 > 0$. Critical equilibria occur at points of tangency between these curves. Let $\pi_{(s)}^*$ denote the equilibrium that has the smallest value for the π_1 component. If $\pi_{(s)}^*$ is regular, then $J(\pi_{(s)}^*; \theta, z)$ is positive-definite, otherwise there must be a second equilibrium. It follows that if the Jacobian is positive definite at all equilibria, there can only be one equilibrium.

Proposition 3. *Fix $z \in \mathbb{S}(Z)$ and θ. If (6) holds and if $J(\pi_1, \pi_2; \theta, z)$ is positive definite evaluated at all solutions to (3), then the solution to (3) is unique.*

If the support of $X_p'\beta_p$ is concentrated around large absolute values for certain realizations of Z, the conditions in Proposition 3 will be satisfied there. We explore this issue next.

2.4.2. Uniqueness of equilibrium when signals are informative

As we can see in Fig. 1, multiple solutions to (3) arise only if the curves $\varphi_1(\pi_1, \pi_2; \theta, z) = 0$ and $\varphi_2(\pi_1, \pi_2; \theta, z) = 0$ have enough variability in the interval $(\pi_1, \pi_2) \in [0, 1]^2$. In a number of cases, the conditional support of $X_p|Z$ may be rich enough that it includes regions of $\mathbb{S}(Z)$ where the variability needed to produce multiple equilibria is suppressed. Identification strategies involving some form of extreme support conditions have been widely used in microeconometrics. A few examples include Manski (1988), Heckman (1990), Heckman and Honoré (1990), Heckman and Honoré (1989), Matzkin (1992), Matzkin (1993), Lewbel (2000), Taber (2000), Carneiro et al. (2003) and, more

[3] Incomplete information makes it impossible for players to randomize in a way that makes their opponent *exactly* indifferent between $Y = 1$ and $Y = 0$.

[4] We ignore values of $\pi_1, \pi_2 \notin [0, 1]^2$ since all solutions to (3) belong inside $[0, 1]^2$.

[5] The unit square $[0, 1]^2$ satisfies the remaining requirements for Brouwer's Fixed Point Theorem.

[6] This result can be generalized to discrete games with more players or actions. Uniqueness would require all the principal minors of the Jacobian to be positive. Gale and Nikaido (1965) refer to such matrices as *P*-matrices.

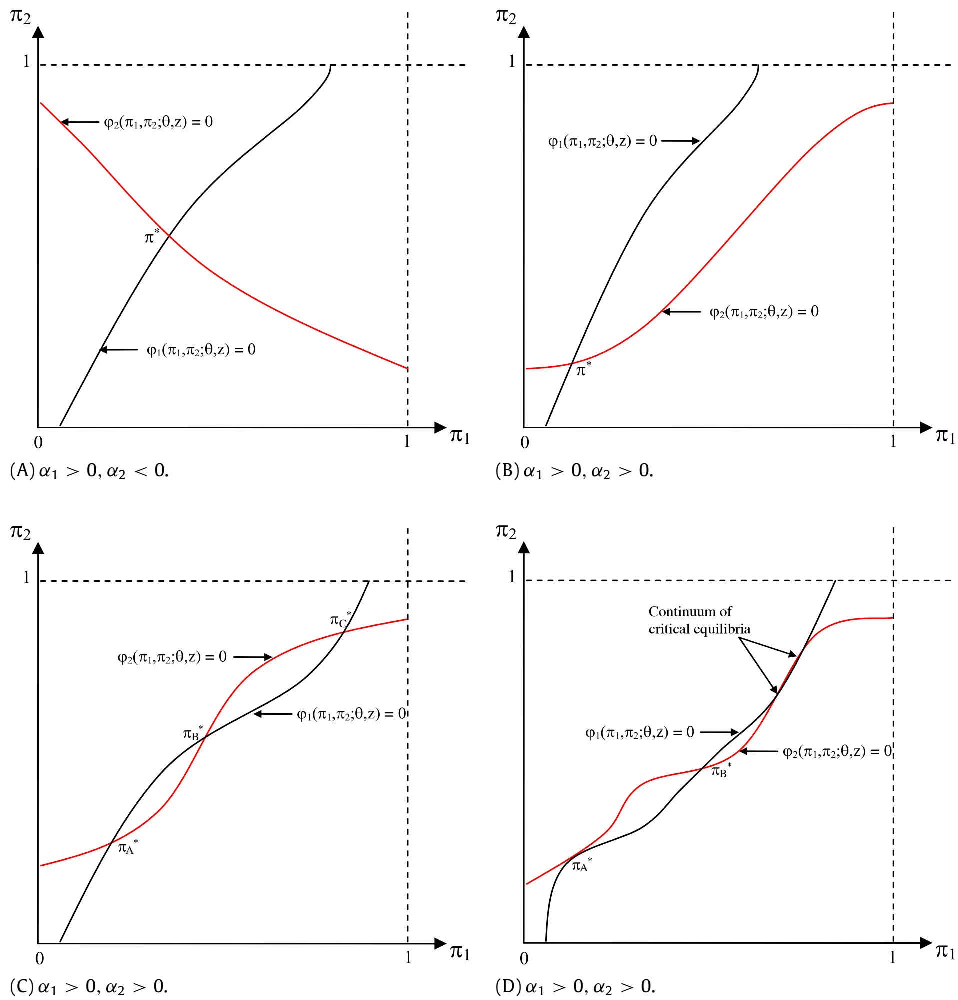

Fig. 1. Examples of the loci $\varphi_1(\pi_1, \pi_2; \theta, z) = 0$ and $\varphi_2(\pi_1, \pi_2; \theta, z) = 0$ assuming (6) holds. Each case corresponds to a fixed value of z and θ. Points of intersection are solutions to (3).

recently, Heckman and Navarro (2007). A thorough study of the implications of support conditions on identification and rates of convergence can be found in Khan and Tamer (2009). In the context of discrete choice models, extreme support assumptions imply the existence of realizations of observables where the agents' set of choices is essentially reduced. The general idea has been studied in psychological contexts, for example, in Thurstone (1959) and Falmagne (1985). In our case canceling out choices is not necessary to counteract the effect of multiple equilibria on identification. It would suffice if for each p there exists a region $\mathcal{Z} \subseteq \mathbb{S}(Z)$ such that the support of $X_p'\beta_p$ (but not of $X_{-p}'\beta_p$) is concentrated around large absolute values whenever $Z \in \mathcal{Z}$. In such regions, the Jacobian described in (5) would be approximately diagonal and the Gale–Nikaido conditions for a unique BNE would be satisfied.

3. Estimation when the distribution of $(\varepsilon_1, \varepsilon_2)$ is known

This section assumes that $G_{1,2}(\cdot, \cdot; \rho)$ is known up to the scalar parameter ρ.

3.1. Identification conditions

Henceforth, if there exists a *unique* solution to the equilibrium system (3) for a given z and θ, we will denote it by $\pi^*(\theta, z) \equiv (\pi_1^*(\theta, z), \pi_2^*(\theta, z))$.

Assumption A3. The parameter space $\Theta \subset \mathbb{R}^{k+3}$ is compact, Z is absolutely continuous with respect to Lebesgue measure and there exists a compact set $\mathcal{Z} \in \mathbb{S}(Z)$ such that $\inf z \in \mathcal{Z} f_Z(z) > \underline{f} > 0$. The set $\mathcal{Z}$ and the parameter space Θ are such that:

(i) (*Regularity of equilibria*) For every $\theta \in \Theta$ and for almost every $z \in \mathcal{Z}$, all solutions to (3) are regular. In particular, let $\pi_{(s)}^*(\theta, z) \equiv (\pi_{(s_1)}^*(\theta, z), \pi_{(s_2)}^*(\theta, z))$ be the solution to (3) that has the smallest value for the π_1 component. There exists $\overline{M} < \infty$ such that $\sup_{\substack{\theta\in\Theta \\ z\in\mathcal{Z}}} \|J(\pi_{(s)}^*(\theta, z); \theta, z)^{-1}\| < \overline{M}$, where $J(\pi; \theta, z)$ is the Jacobian described in (5).
(ii) (*Uniqueness of equilibrium*) Let $\theta_0 \in \Theta$ denote the true value of the parameter θ. There exists a neighborhood $\mathcal{N}_\theta \subseteq \Theta$ that contains θ_0 such that each $(\theta, z) \in \mathcal{N}_\theta \times \mathcal{Z}$ has a unique solution to the equilibrium conditions (3).

Take the Jacobian $\underset{2\times(k+5)}{D\varphi(\pi; \theta, z)} = \left(\underset{2\times 2}{\nabla_\pi \varphi(\pi; \theta, z)}, \underset{2\times(k+3)}{\nabla_\theta \varphi(\pi; \theta, z)}\right)$. Using the Transversality Theorem,[7] a sufficient condition for the almost-sure regularity condition in (i) to hold is if for every $(\theta, z) \in \Theta \times \mathcal{Z}$, the rank of $D\varphi(\pi; \theta, z)$ is equal to 2 when π is evaluated at any solution to (3).[8] Going back to Fig. 1, $\pi_{(s)}^*(\theta, z)$ defined in the second part of (i) is simply the first equilibrium along the horizontal axis. We assume all those equilibria to be regular, uniformly in $\Theta \times \mathcal{Z}$. Part (ii), asserts that *we will only assume*

7 See Mas-Collel et al. (1995), Theorem M.E.2.
8 See Assumption A4.

uniqueness of equilibrium at the true parameter value θ_0, and in a neighborhood $\mathcal{N}_\theta$ containing it. It would be sufficient to assume that Proposition 3 holds everywhere in $(\theta, z) \in \mathcal{N}_\theta \times \mathcal{Z}$. By construction, $\pi^*_{(s)}(\theta, z) \equiv \pi^*(\theta, z)$ everywhere in $\mathcal{N}_\theta \times \mathcal{Z}$. Our proposal will be to construct an estimator for players' beliefs that converges to $\pi^*_{(s)}(\theta, z)$ uniformly in $\Theta \times \mathcal{Z}$.

A candidate set $\mathcal{Z}$ could be found by relying on the relationship between Z and $(X_1'\beta_{1_0}, X_2'\beta_{2_0})$ predicted by economic theory. From our previous discussion, each $z \in \mathcal{Z}$ should be such that the support of either $X_1'\beta_{1_0}$ or $X_2'\beta_{2_0}$ (but not both) is concentrated around relatively large absolute values conditional on $Z = z$. This is feasible, for instance, when each X_p includes a publicly observed covariate whose coefficient has a known sign ex-ante, has rich support conditional on all other covariates and is excluded in X_{-p}. We can evaluate the validity of *any* candidate set $\mathcal{Z}$ by looking at nonparametric estimates of the Jacobian of

$$\Big(E\left[E\left[Y_1|X_1, E[Y_1|Z], E[Y_2|Z], Y_2 = 1\right]|Z\right], \quad E\left[E\left[Y_2|X_2, E[Y_1|Z], E[Y_2|Z], Y_1 = 1\right]|Z\right]\Big)'$$

with respect to $E[Y_1|Z]$ and $E[Y_2|Z]$. The principal minors of this Jacobian should not vanish in any set $\mathcal{Z}$ that satisfies Assumption A3.

Computing the equilibrium $\pi^*_{(s)}(\theta, z)$

For a given (θ, z), finding the equilibrium $\pi^*_{(s)}(\theta, z)$ described in Assumption A3 is computationally simple. Let φ_1, φ_2 and φ be as defined in Eq. (2) and consider the following iterative procedure

$$\begin{aligned}
&\text{Step 0.—Fix a value } \pi_2^0 \in [0, 1].\\
&\text{Step 1.—Take } \pi_2^0 \text{ from Step 0,}\\
&\quad \text{and let } \pi_1^1 \text{ be the solution to } \varphi_1(\pi_1, \pi_2^0; \theta, z) = 0.\\
&\text{Step 2.—Take } \pi_1^1 \text{ from Step 1,}\\
&\quad \text{and let } \pi_2^2 \text{ be the solution to } \varphi_2(\pi_1^1, \pi_2; \theta, z) = 0.\\
&\text{Step } k\text{.—Continue iteratively.}
\end{aligned} \tag{7}$$

For a given starting value, the solution in each one of the steps $2, \ldots$ is unique because both $\varphi_1(\pi_1, \pi_2; \theta, z) = 0$ and $\varphi_2(\pi_1, \pi_2; \theta, z) = 0$ are monotonic curves. This monotonicity is shown in Appendix A.4. If the BNE equilibrium is unique for z and θ, this procedure will converge to $\pi^*_{(s)}(\theta, z)$ (the unique equilibrium) regardless of the starting values. Otherwise, depending on the signs of α_1 and α_2, the appropriate starting values are: (i) $\pi_2^0 = 1$ if $\alpha_1 < 0$ and $\alpha_2 < 0$. (ii) $\pi_2^0 = 0$ if $\alpha_1 > 0$ and $\alpha_2 > 0$. (iii) Any $\pi_2^0 \in [0, 1]$ if $\alpha_1\alpha_2 \le 0$.

Given the previous conditions, the following assumption will be sufficient for identification.

Assumption A4. (i) (*Full rank*) X_p has full column rank with positive probability conditional on $Z \in \mathcal{Z}$.
(ii) (*Invertibility*) The joint distribution $G_{1,2}(\cdot; \rho)$ is an invertible function of ρ with positive probability in a region $\mathcal{R} \subseteq \mathbb{R}^2$ of realizations of $(\varepsilon_1, \varepsilon_2)$, and this region is such that

$$\Pr\left[\left(X_1'\beta_1 + \alpha_1\pi_2, X_2'\beta_2 + \alpha_2\pi_1\right) \in \mathcal{R} | Z \in \mathcal{Z}\right] > 0 \quad \forall \beta_1, \beta_2, \alpha_1, \alpha_2 \in \Theta, \ \forall \pi_1, \pi_2 \in [0, 1]^2.$$

Combining Assumptions A2 and A3 and A4(i), we have that for all $\theta \ne \theta_0$,

$$\begin{aligned}
&\Pr\left(X_1'\beta_1 + \alpha_1\pi^*_{(s_2)}(\theta, Z) \ne X_1'\beta_{1_0} + \alpha_{1_0}\pi_2^*(\theta_0, Z) | Z \in \mathcal{Z}\right) > 0\\
&\Pr\left(X_2'\beta_2 + \alpha_2\pi^*_{(s_1)}(\theta, Z) \ne X_2'\beta_{2_0} + \alpha_{2_0}\pi_1^*(\theta_0, Z) | Z \in \mathcal{Z}\right) > 0,
\end{aligned} \tag{8}$$

where $\pi^*_{(s)}(\cdot)$ is as defined in Assumption A3. We point out that (8) is not an artifact of the nonlinear nature[9] of $\pi^*_{(s)}(\theta, \cdot)$. Suppose $X = Z$ and $\varepsilon_p \sim U[-1, 1]$. Then (3) becomes a linear system in X. In this case (8) is satisfied if there exist $X_{1_\ell} \in X_1$ and $X_{2_\ell} \in X_2$ with nonzero coefficients and $\Pr(X_{1_\ell} \ne X_{2_\ell}) > 0$.

3.2. Estimation of equilibrium beliefs

As we mentioned before, we assume that players use self-consistent equilibrium beliefs according to (1). We will introduce the following assumption.

Assumption A5 (*Researcher*)**.** The econometrician has access to an i.i.d. sample of N games described by Assumptions A1–A4, and observes $(W_n)_{n=1}^N \equiv (Y_n, X_n, Z_n)_{n=1}^N$. The econometrician knows the functional form of $G_p(\cdot)$ and $G_{1,2}(\cdot, \cdot; \rho)$ up to the value of ρ. Henceforth, L will denote the number of elements in Z that are not in X.

For illustrative purposes, we will focus on the case in which all elements in X are privately observed, and $Z \cap X = \emptyset$. The case in which beliefs are conditioned on unobservables is addressed in Section 5.1. We will then explain how to adapt the methodology in the more general case. We will employ a kernel function $K : \mathbb{R}^L \to \mathbb{R}$ and a bandwidth sequence $h \to 0$. For any $\psi \in \mathbb{R}^L$, we denote $K_h(\psi) \equiv K(\psi/h)$. For a given $\theta \in \mathbb{R}^{k+3}$, take $z \in \mathbb{R}^L$ and let $\widehat{f}(z) = (Nh^L)^{-1}\sum_{n=1}^N K_h(Z_n - z)$. Now take $(\pi_1, \pi_2) \equiv \pi \in \mathbb{R}^2$ and define

$$\begin{aligned}
\widehat{\delta}_{1,2}(\pi; \theta, z) &= \left(Nh^L\widehat{f}(z)\right)^{-1}\sum_{n=1}^N G_{1,2}(X_{1n}'\beta_1 + \alpha_1\pi_2, X_{2n}'\beta_2 + \alpha_2\pi_1; \rho)K_h(Z_n - z)\\
\widehat{\delta}_p(\pi; \theta, z) &= \left(Nh^L\widehat{f}(z)\right)^{-1}\sum_{n=1}^N G_p(X_{p_n}'\beta_p + \alpha_p\pi_{-p})K_h(Z_n - z)\\
&\text{for } p = 1, 2,\\
\widehat{\varphi}_1(\pi; \theta, z) &= \pi_1 - \frac{\widehat{\delta}_{1,2}(\pi; \theta, z)}{\widehat{\delta}_2(\pi; \theta, z)},\\
\widehat{\varphi}_2(\pi; \theta, z) &= \pi_2 - \frac{\widehat{\delta}_{1,2}(\pi; \theta, z)}{\widehat{\delta}_1(\pi; \theta, z)}\\
\widehat{\varphi}(\pi; \theta, z), &= \left(\widehat{\varphi}_1(\pi; \theta, z), \widehat{\varphi}_2(\pi; \theta, z)\right)',\\
\widehat{Q}(\pi; \theta, z) &= -\widehat{\varphi}(\pi; \theta, z)'\widehat{\varphi}(\pi; \theta, z).
\end{aligned} \tag{9}$$

If X_{1_ℓ} were publicly observed and therefore included in Z, we would replace $X_{1_\ell n}$ with x_{1_ℓ}, its realization in z. If the entire vector X were publicly observed, we would have $\widehat{\delta}_{1,2}(\pi; \theta, z) = G_{1,2}(x_1'\beta_1 + \alpha_1\pi_2, x_2'\beta_2 + \alpha_2\pi_1; \rho)$ and $\widehat{\delta}_p(\pi; \theta, z) = G_p(x_p'\beta_p + \alpha_p\pi_{-p})$, and *the estimation problem would be entirely parametric*. Next, consider the following problem,

$$\underset{\pi \in [0,1]^2}{\text{Max}} \ \widehat{Q}(\pi; \theta, z), \tag{10}$$

$$\text{and let } \ \widehat{\pi}(\theta, z) = \text{Solution to (10) with the smallest value for the } \pi_1 \text{ component.} \tag{11}$$

We will use $\widehat{\pi}(\theta, z)$ as the estimator of players' beliefs. In the Appendix we carefully characterize the uniform (in $[0, 1]^2 \times \mathcal{Z} \times \Theta$) convergence properties of these objects defined in (9) to their population counterparts. Given this, the computational implementation of finding (11) would be undertaken using the iterative procedure described in Eq. (7), with the starting points mentioned there.

Assumption A6. There exists an $M \ge L + 1$ such that:
(i) (*Smoothness*) $G_p(\cdot)$ and $G_{1,2}(\cdot)$ are M-times differentiable with bounded derivatives. The conditional density $f_{X|Z}(x|z)$ is M-times differentiable with bounded derivatives with respect to all the elements in Z not included in X. In addition, $E[\|XX'\|^2|Z]$ is a bounded, continuous function of Z, and $E[\|XX'\|^4] < \infty$.

9 This nonlinearity is a consequence of the unbounded support assumption of ε_p.

(ii) (*Kernel and bandwidth*) The kernel $K(\cdot)$ is symmetric, Lipschitz-continuous, bounded and M-times differentiable with bounded derivatives. Denote $\Psi = (\psi_1, \ldots, \psi_L)$. The kernel $K(\cdot)$ is bias-reducing of order M: $\int K(\Psi)d\Psi = 1$, $\int \psi_1^{q_1} \cdots \psi_L^{q_L} K(\Psi)d\Psi = 0$ for all $(q_\ell)_{\ell=1}^L \in \mathbb{N}$ such that $1 \le q_1 + \cdots + q_L \le M-1$, and $\int \|\Psi\|^M |K(\Psi)|d\Psi < \infty$. As $N \to \infty$, the bandwidth h satisfies: $h \to 0$, $Nh^{L+2} \to \infty$, $Nh^{2M} \to 0$, and there exists $\overline{\delta} > 0$ such that $N^{1-\overline{\delta}}h^{2L} \to \infty$.

The existence-of-moments condition in part (i) could be replaced with the more restrictive assumption that $\mathbb{S}(X)$ is compact. It is shown in the Appendix that if Assumptions A3, A5 and A6 are satisfied, then $\sup_{\substack{\theta\in\Theta \\ z\in\mathcal{Z}}} \|\widehat{\pi}(\theta, z) - \pi^*_{(s)}(\theta, z)\| \xrightarrow{p} 0$. If Assumptions A1, A2 and A4 are also satisfied, then $\widehat{\pi}(\theta_0, z)$ converges to the players' equilibrium beliefs uniformly in $\mathcal{Z}$. Next, we show how we use $\widehat{\pi}(\theta, \cdot)$ to estimate θ.

3.3. Estimation of θ

Take any given value of the parameter vector θ. For two constants $(\pi_1, \pi_2) \equiv \pi \in \mathbb{R}^2$ define $t_1 = X_1'\beta_1 + \alpha_1\pi_2$ and $t_2 = X_2'\beta_2 + \alpha_2\pi_1$. Let[10]

$$
\begin{aligned}
P_{1,1}(X,\theta,\pi) &= \int_{-\infty}^{t_1}\int_{-\infty}^{t_2} g_{1,2}(\varepsilon_1,\varepsilon_2;\rho)d\varepsilon_2 d\varepsilon_1,\\
P_{1,0}(X,\theta,\pi) &= \int_{-\infty}^{t_1}\int_{t_2}^{\infty} g_{1,2}(\varepsilon_1,\varepsilon_2;\rho)d\varepsilon_2 d\varepsilon_1,\\
P_{0,1}(X,\theta,\pi) &= \int_{t_1}^{\infty}\int_{-\infty}^{t_2} g_{1,2}(\varepsilon_1,\varepsilon_2;\rho)d\varepsilon_2 d\varepsilon_1,\\
P_{0,0}(X,\theta,\pi) &= \int_{t_1}^{\infty}\int_{t_2}^{\infty} g_{1,2}(\varepsilon_1,\varepsilon_2;\rho)d\varepsilon_2 d\varepsilon_1.
\end{aligned} \tag{12}
$$

Denote $\ell_{1,1}(X,\theta,\pi) = \log P_{1,1}(X,\theta,\pi)$ and so on for the remaining probabilities. As before, let $W \equiv (Y, X, Z)$ and define

$$
\begin{aligned}
\ell(W,\theta,\pi) &= Y_1Y_2\ell_{1,1}(X,\theta,\pi) + (1-Y_1)Y_2\ell_{0,1}(X,\theta,\pi)\\
&\quad + Y_1(1-Y_2)\ell_{1,0}(X,\theta,\pi)\\
&\quad + (1-Y_1)(1-Y_2)\ell_{0,0}(X,\theta,\pi), \quad \text{and}\\
&\quad \ell_{\mathcal{Z}}(W,\theta,\pi) = \ell(W,\theta,\pi)\mathbb{1}\{Z\in\mathcal{Z}\}.
\end{aligned} \tag{13}
$$

If $Z \in \mathcal{Z}$, then $\ell_{\mathcal{Z}}(W, \theta_0, \pi^*(\theta_0, Z))$ is the conditional log-likelihood of Y given (X, Z). We will let $\nabla_\theta \ell_{\mathcal{Z}}(W, \theta, \pi)$ and $\nabla_\pi \ell_{\mathcal{Z}}(W, \theta, \pi)$ denote the vector of partial derivatives of $\ell_{\mathcal{Z}}(W, \theta, \pi)$ with respect to θ and π, respectively. For all $\theta \in \mathcal{N}_\theta$, we have

$$
\begin{aligned}
\frac{\partial \ell_{\mathcal{Z}}(W,\theta,\pi^*(\theta,Z))}{\partial\theta} &\equiv S_{\theta_{\mathcal{Z}}}(W,\theta)\\
&= \nabla_\theta \ell_{\mathcal{Z}}\left(W,\theta,\pi^*(\theta,Z)\right)\\
&\quad + \nabla_\theta \pi^*(\theta,Z)'\nabla_\pi \ell_{\mathcal{Z}}\left(W,\theta,\pi^*(\theta,Z)\right).
\end{aligned} \tag{14}
$$

The score function of $\ell_{\mathcal{Z}}(W, \theta_0, \pi^*(\theta_0, Z))$ is given by $S_{\theta_{\mathcal{Z}}}(W, \theta_0)$.

Assumption A7 (*Technical*)**.** θ_0 is an interior point of Θ. Let $P_{y_1,y_2}(X, Z, \theta, \pi)$ be as defined in (12). There exists a random variable U such that $E[U^2] < \infty$ and with probability one, $\sup_{\substack{\pi\in[0,1]^2 \\ \theta\in\Theta}} [P_{y_1,y_2}(X, Z, \theta, \pi)^{-1}] \le U$ for $(y_1, y_2) \in \{(1, 1), (1, 0), (0, 1), (0, 0)\}$. Let $S_{\theta_{\mathcal{Z}}}(W, \theta)$ be as defined in (14) and let $\Im_{\mathcal{Z}} = E\left[S_{\theta_{\mathcal{Z}}}(W, \theta_0)S_{\theta_{\mathcal{Z}}}(W, \theta_0)'\right]$, then $\Im_{\mathcal{Z}}$ is invertible.

The second part of A7 would be redundant if $\mathbb{S}(X)$ were compact. Coupled with our previous assumptions, it yields uniform convergence of objects like $\frac{1}{N}\sum_{n=1}^N \nabla_\theta \ell_{\mathcal{Z}}(W_n, \theta, \pi)$ and $\frac{1}{N}\sum_{n=1}^N \nabla_\pi \ell_{\mathcal{Z}}(W_n, \theta, \pi)$ in $\Theta \times [0, 1]^2$. In the Appendix, we show that an information-identity result holds and

$$
E\left[\frac{\partial^2 \ell_{\mathcal{Z}}(W,\theta,\pi^*(\theta,Z))}{\partial\theta\partial\theta'}\right]\Bigg|_{\theta=\theta_0} = -\Im_{\mathcal{Z}}.
$$

Finally, we define the terms in Eq. (15), shown in Box I, which we will require in the following.

We estimate θ by solving $\text{Max}_{\theta\in\Theta} \frac{1}{N}\sum_{n=1}^N \ell_{\mathcal{Z}}(W_n, \theta, \widehat{\pi}(\theta, Z_n))$. Theorem 1 presents the asymptotic properties of our estimator.

Theorem 1. *Let $\underset{(k+3)\times 2}{S_{\theta\pi_{\mathcal{Z}}}}(W_n, \theta_0)$ denote the partial derivative of $S_{\theta_{\mathcal{Z}}}(W_n, \theta_0)$ with respect to $\pi^*(\theta_0, Z_n)$ – see (14) –. Let $D(Z, \theta_0) = J(\pi^*(\theta_0, Z); \theta_0)^{-1}V(Z)$, where $J(\cdot)$ and $V(\cdot)$ are as defined in (5) and (15), respectively. Define $R_{\mathcal{Z}}(Z_n, \theta_0) = E[S_{\theta\pi_{\mathcal{Z}}}(W_n, \theta_0)|Z_n]D(Z_n, \theta_0)$ and*

$$
A_{\mathcal{Z}}(X_n, Z_n, \theta_0) = R_{\mathcal{Z}}(Z_n, \theta_0)\Big[E\left[B(Y_n)|X_n, Z_n\right] - E\left[B(Y_n)|Z_n\right]\Big],
$$

with $B(\cdot)$ as in (15). Let $\widehat{\pi}(\theta, z)$ be as defined in (11) and let $\widehat{\theta}$ be the solution to

$$
\underset{\theta\in\Theta}{\text{Max}} \frac{1}{N}\sum_{n=1}^N \ell_{\mathcal{Z}}\left(W_n, \theta, \widehat{\pi}(\theta, Z_n)\right),
$$

where $\ell_{\mathcal{Z}}(\cdot)$ is defined in (13). If Assumptions A1–A7 are satisfied, then

$$
\begin{aligned}
\widehat{\theta} - \theta_0 &= \Im_{\mathcal{Z}}^{-1} \times \frac{1}{N}\sum_{n=1}^N \Big[S_{\theta_{\mathcal{Z}}}(W_n, \theta_0) + A_{\mathcal{Z}}(X_n, Z_n, \theta_0)\Big]\\
&\quad + o_p(N^{-1/2})
\end{aligned} \tag{16}
$$

and therefore $\sqrt{N}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Im_{\mathcal{Z}}^{-1} + \Im_{\mathcal{Z}}^{-1}\Omega_{\mathcal{Z}}\Im_{\mathcal{Z}}^{-1})$, where

$$
\Omega_{\mathcal{Z}} = E\left[R_{\mathcal{Z}}(Z, \theta_0)\text{Var}\left[E[B(Y)|X, Z]|Z\right]R_{\mathcal{Z}}(Z, \theta_0)'\right].
$$

The structure of the asymptotic variance of $\widehat{\theta}$ resembles that of Ahn and Manski (1993), who estimate a discrete-choice model with nonparametric conditional expectations among the regressors but no strategic interaction. The matrix $\Im_{\mathcal{Z}}^{-1}$ is the semiparametric efficiency bound when $G_p(\cdot)$ and $G_{1,2}(\cdot)$ are known *and* players' beliefs are observed.[11] The term $\Omega_{\mathcal{Z}}$ reflects the loss in efficiency vis-a-vis the case in which $f_{x|z}(\cdot)$ is known and equilibrium beliefs can be exactly computed. The specific structure of $\Omega_{\mathcal{Z}}$ results from the asymptotically linear representation of $\widehat{\pi}(\cdot)$ – see Eq. (A.9) in the Appendix. Define $t(\theta_0) = (X_1'\beta_{1_0} + \alpha_{1_0}\pi_2^*(\theta_0, Z), X_2'\beta_{2_0} + \alpha_{2_0}\pi_1^*(\theta_0, Z))'$. Then $E[B(Y)|X, Z] = E[B(Y)|t(\theta_0)]$ and the magnitude of the efficiency loss would depend on the behavior of $E[B(Y)|t(\theta_0)] - E[B(Y)|Z]$. If Z is a perfect predictor for X, we would have $\Omega_{\mathcal{Z}} = 0$. In such a case, the only *effective* source of incomplete information is ε_p.[12] We also have $\Omega_{\mathcal{Z}} = 0$ if $\alpha_{p_0} = 0$ for $p = 1, 2$, since $S_{\theta\pi_{\mathcal{Z}}}(W_n, \theta_0) = 0$ in this case.

3.4. Incomplete information and coherency of the econometric model. A comparison with the complete-information case

There has been considerable effort devoted to the study of endogeneity in discrete and limited dependent variable models

[10] We take π_1, π_2 as constant here because we will eventually replace them with players' beliefs, which are deterministic conditional on Z, and Z is assumed to be independent of $(\varepsilon_1, \varepsilon_2)$.

[11] That is, the efficiency bound that corresponds to the trimmed MLE that uses $\mathcal{Z}$ as the trimming set.

[12] See the paragraph following Eq. (9).

$$\underset{2\times 3}{V(Z)} = \begin{pmatrix} \dfrac{1}{\Pr(Y_2=1|Z)} & 0 & -\dfrac{\Pr(Y_1=1|Y_2=1,Z)}{\Pr(Y_2=1|Z)} \\ \dfrac{1}{\Pr(Y_1=1|Z)} & -\dfrac{\Pr(Y_2=1|Y_1=1,Z)}{\Pr(Y_1=1|Z)} & 0 \end{pmatrix}, \qquad \underset{3\times 1}{B(Y)} = (Y_1Y_2, Y_1, Y_2)' \tag{15}$$

Box I.

in settings not necessarily motivated by game theory or strategic interaction. A partial list of examples includes Heckman (1978), Sickles and Schmidt (1978), Gourieroux et al. (1980), Blundell and Smith (1986), Blundell and Smith (1989), Sickles (1989), Blundell and Smith (1994), Dagenais (1999), Blundell and Powell (2004), Vytlacil and Yildiz (2007), Lewbel (2007), Chesher (2007), Abrevaya et al. (2007) and Klein and Vella (2009). Following the notation in Lewbel (2007), structural models of this type can be generically expressed as $Y = H(Y, W)$, where Y is a vector of endogenous variables and W includes observable and unobservable covariates and parameters. A model of this type is *coherent* (see Gourieroux et al. (1980)) if there exists a mapping $G(\cdot)$ such that we can write $Y = G(W)$ for each W. Pure-strategy Nash equilibrium behavior with complete information in the 2×2 game dictates optimal decision rules of the type

$$Y_p = \mathbb{1}\{X_p'\beta_p + \alpha_p \cdot Y_{-p} - \varepsilon_p \geq 0\} \quad \text{for } p = 1, 2.$$

Optimal choices depend on the actual ex-post action of the opponent. This is the source of endogeneity and it results from the definition of a complete-information pure strategy Nash equilibrium. If the support of the unobservable shock ε_p is assumed to be unbounded, the resulting model will be coherent and complete if and only if $\alpha_1 \cdot \alpha_2 \leq 0$ and mixed-strategies are allowed. If $\alpha_1 \cdot \alpha_2 > 0$ and only pure-strategies are considered, the model produces a unique prediction for $Y_1 + Y_2$ for any given realization of observables and a given value of payoff parameters (see, e.g., Bresnahan and Reiss (1991), Bresnahan and Reiss (1990), Berry (1992) and Tamer (2003)). In an incomplete information setting, for realizations of Z that produce a unique Bayesian–Nash equilibrium or, more generally, if the underlying equilibrium selection mechanism is degenerate, optimal decision rules are of the form

$$Y_p = \mathbb{1}\left\{X_p'\beta_p + \alpha_p \cdot \Pr(Y_{-p} = 1|Z, Y_p = 1) - \varepsilon_p \geq 0\right\}.$$

Thus, if Z is observable and independent of ε_p, incomplete information can naturally produce a coherent model.

4. Estimation when the distribution of ε is unknown

We now drop the assumption that $G_{1,2}(\cdot, \cdot; \rho)$ is known to the researcher.

4.1. Identification conditions

Under Assumptions A1–A2, the relationship between Y and (X, Z) is described by a double-index model. Thus, we can design an estimation procedure that relies only on this exclusion restriction. A few related examples include Han (1987), Powell et al. (1989), Ichimura and Lee (1991) and Stoker (1991).[13] In each instance, identification and estimation relies on some form of index exclusion restriction combined with either invertibility or smoothness assumptions concerning some functional (in our case, conditional choice probabilities). We will use a semiparametric likelihood-based procedure along the lines of Klein and Spady (1993) (KS). As before, we will rely on semiparametric analog BNE conditions.

Assumption B1 (*Exclusion Restriction and Parameter Normalization*). Normalize any intercept in β_p to zero for $p = 1, 2$. There exist two random variables $X_1^0 \in X_1$ and $X_2^0 \in X_2$ with nonzero coefficients such that both X_1^0 and X_2^0 are privately observed by each player, $\Pr(X_1^0 \neq X_2^0) > 0$, and (X_1^0, X_2^0, Z) is a continuously distributed random vector conditional on all other variables. We will partition $X_p = (X_p^0, X_p^*)$. Let β_p^0 be the coefficient of X_p^0 and let $\gamma_p \equiv \beta_p/\beta_p^0$, $\eta_p \equiv (1, \gamma_p)'$ and $\sigma_p \equiv \alpha_p/\beta_p^0$. We will normalize $X_p'\beta_p = X_p^0 + X_p^{*\prime}\gamma_p \equiv X_p'\eta_p$. Let $\gamma_p \equiv (\eta_p, \sigma_p)'$ and $\gamma \equiv (\gamma_1, \gamma_2)'$. The parameter of interest will now be $\gamma \in \mathbb{R}^k$ with parameter space $\Gamma \subset \mathbb{R}^k$, assumed to be compact. Note that the strategic interaction parameters are now denoted by σ_1, σ_2.

Exclusion restrictions similar to B1 are common in multiple-index and related models. We require X_1^0 and X_2^0 to be privately observed in order to make B1 compatible with Assumption B2 below. The parameter ρ is no longer identified, and the dimension of the identifiable parameter is now k instead of $k + 3$. When $G_p(\cdot)$ and $G_{1,2}(\cdot)$ were known, we were able to use this knowledge to estimate $\pi^*(\theta, z)$, the unique solution to (3) when $\theta, z \in \mathcal{N}_\theta \times \mathcal{Z}$. *This is no longer possible*. However, the self-consistent nature of players' equilibrium beliefs will allow us to work around this difficulty.[14] Define

$$\begin{aligned} t_1(\gamma_1) &= X_1'\eta_1 + \sigma_1 \Pr(Y_2 = 1|Y_1 = 1, Z), \\ t_2(\gamma_2) &= X_2'\eta_2 + \sigma_2 \Pr(Y_1 = 1|Y_2 = 1, Z), \end{aligned} \tag{17}$$

and let $t(\gamma) = (t_1(\gamma_1), t_2(\gamma_2))'$. We add the following assumption:

Assumption B2 (*Distributional Properties of $t(\gamma)$*). Let $\overline{M} > L + 1$ be the constant described in B4, below. Then:

(i) $\mathbb{S}(X, Z)$ is compact. $\exists\, \underline{f} > 0$ such that $f_{X,Z}(x, z) \geq \underline{f}$ for all $(x, z) \in \mathbb{S}(X, Z)$.

(ii) The random vector $(t(\gamma), Z)$ is jointly continuously distributed for all $\gamma \in \Gamma$. We will let $f_t(t; \gamma)$ and $f_{t,z}(t, z; \gamma)$ denote the marginal and joint densities of $t(\gamma)$ and Z evaluated at $t(\gamma) = t$ and $Z = z$. Both are $\overline{M}$ times differentiable with respect to t and z with bounded derivatives for all $\gamma \in \Gamma$. The conditional densities of $t(\gamma)$ given Z, and Z given $t(\gamma)$ will be denoted by $f_{t|z}(t|z; \gamma)$ and $f_{z|t}(t|z; \gamma)$, respectively. Define

$$\begin{aligned} \mu_{Y_1Y_2}(t, z; \gamma) &= E\left[Y_1Y_2|t(\gamma) = t, Z = z\right], \\ \mu_{Y_p}(t, z; \gamma) &= E\left[Y_p|t(\gamma) = t, Z = z\right]. \end{aligned} \tag{18}$$

Then $\mu_{Y_1Y_2}(t, z; \gamma)$ and $\mu_{Y_p}(t, z; \gamma)$ are $\overline{M}$ times differentiable with respect to t and z, with bounded derivatives for all $\gamma \in \Gamma$. Let $P_{\mathcal{Z}}(t; \gamma) = \Pr(Z \in \mathcal{Z}|t(\gamma) = t)$. Then $P_{\mathcal{Z}}(\cdot) > 0$ and is $\overline{M}$ times differentiable with bounded derivatives uniformly in $\Gamma \times \mathbb{S}(X, Z)$.

Estimation will now take place through nested semi-parametric procedures. Assumption B2 implies that the density $f_t(t; \gamma)$ is bounded away from zero uniformly over $\gamma \in \Gamma$ and $t \in \mathbb{S}(t(\gamma))$.

13 In the context of incomplete-information games, Aradillas-Lopez (2007) proposes a pairwise-difference estimation procedure based on a double-index model.

14 Self-consistency of beliefs implies that $\Pr_p(Y_{-p} = 1|Y_p = 1, Z) = \Pr(Y_{-p} = 1|Y_p = 1, Z)$ for $p = 1, 2$.

This avoids the need to do trimming in order to stay away from realizations of observables where the said density is arbitrarily close to zero. Trimming will now be done exclusively out of equilibrium considerations[15] to remain in the set $\mathcal{Z}$. Fix $\tau \in \mathbb{R}^2$ and define

$$P_{Y|t}(1,1|\tau;\gamma) = E[Y_1Y_2|t(\gamma)=\tau],\quad P_{Y_p|t}(1|\tau;\gamma) = E[Y_p|t(\gamma)=\tau],\quad \text{for } p=1,2. \tag{19}$$

Take two scalars $\pi_1, \pi_2 \in \mathbb{R}$ and let

$$\varphi_{1|t}(\pi_1,\pi_2;\gamma,Z) = \pi_1 - \frac{E\left[P_{Y|t}(1,1|X_1'\eta_1+\sigma_1\pi_2, X_2'\eta_2+\sigma_2\pi_1;\gamma)|Z\right]}{E\left[P_{Y_2|t}(1|X_1'\eta_1+\sigma_1\pi_2, X_2'\eta_2+\sigma_2\pi_1;\gamma)|Z\right]}$$

$$\varphi_{2|t}(\pi_1,\pi_2;\gamma,Z) = \pi_2 - \frac{E\left[P_{Y|t}(1,1|X_1'\eta_1+\sigma_1\pi_2, X_2'\eta_2+\sigma_2\pi_1;\gamma)|Z\right]}{E\left[P_{Y_1|t}(1|X_1'\eta_1+\sigma_1\pi_2, X_2'\eta_2+\sigma_2\pi_1;\gamma)|Z\right]},$$

and $\varphi_t(\pi_1,\pi_2;\gamma,Z) = (\varphi_{1|t}(\pi_1,\pi_2;\gamma,Z), \varphi_{2|t}(\pi_1,\pi_2;\gamma,Z))'$. For $z \in \mathbb{S}(Z)$ and $\gamma \in \Gamma$, consider a solution for π_1, π_2 to

$$\varphi_t(\pi_1,\pi_2;\gamma,z) = 0. \tag{20}$$

For $\gamma \neq \gamma_0$, a solution to (20) does *not* solve the equilibrium system (3). However, both solutions coincide when $\gamma = \gamma_0$ for any $z \in \mathcal{Z}$. We will exploit this property, and estimate players' beliefs using a sequence of solutions to (20). We explain the details in the following section.

4.2. Estimation of equilibrium beliefs

From now on, if (20) has a unique solution for the pair (γ, z), we will denote it by $\pi_t^*(\gamma, z)$. Note that $\pi_t^*(\gamma_0, z) = \pi^*(\gamma_0, z)$ (players' equilibrium beliefs) for all $z \in \mathcal{Z}$.

Assumption B3 (*Regularity of Solutions to* (20))**.** The uniform regularity conditions described in A3(i) hold for system (20), for the set $\mathcal{Z}$, the new parameter space Γ and a neighborhood $\mathcal{N}_\gamma \subseteq \Gamma$ that contains γ_0, the true parameter value. $\pi_{t_{(s)}}^*(\gamma, z)$ will be analogous to $\pi_{(s)}^*(\theta, z)$. Assumption A3(ii) is also maintained.

Let $H : \mathbb{R}^L \to \mathbb{R}$ be a kernel, and $b \to 0$ a bandwidth sequence. Let $H_b(\psi) \equiv H(\psi/b)$. Take $z \in \mathbb{R}^L$ and let $\overline{f}_Z(z) = (Nb^L)^{-1}\sum_{m=1}^N H_b(Z_m - z)$. Take the nonparametric estimators $\overline{P}(Y_1=1, Y_2=1|Z=z) = (Nb^L\overline{f}_Z(z))^{-1}\sum_{m=1}^N Y_{1m}Y_{2m}H_b(Z_m - z)$ and $\overline{P}(Y_p=1|Z=z) = (Nb^L\overline{f}_Z(z))^{-1}\sum_{m=1}^N Y_{p_m}H_b(Z_m - z)$, and define

$$\overline{\pi}_1(z) = \frac{\overline{P}(Y_1=1,Y_2=1|Z=z)}{\overline{P}(Y_2=1|Z=z)},\quad \overline{\pi}_2(z) = \frac{\overline{P}(Y_1=1,Y_2=1|Z=z)}{\overline{P}(Y_1=1|Z=z)}. \tag{21}$$

We will use the following sample analogs of (17) for the mth observation

$$\overline{t}_{1m}(\gamma_1) = X_{1m}'\eta_1 + \sigma_1\overline{\pi}_2(Z_m) \quad \text{and} \quad \overline{t}_{2m}(\gamma_2) = X_{2m}'\eta_2 + \sigma_2\overline{\pi}_1(Z_m),$$

with $\overline{t}_m(\gamma) = (\overline{t}_{1m}(\gamma_1), \overline{t}_{2m}(\gamma_2))'$. We will use a second set of kernel and bandwidth, given by $K : \mathbb{R}^2 \to \mathbb{R}$ and $h \to 0$, with $K_h(\psi) \equiv K(\psi/h)$. Fix a value $\tau \in \mathbb{R}^2$, let $\overline{f}_t(\tau;\gamma) = (Nh^2)^{-1}\sum_{m=1}^N K_h(\overline{t}_m(\gamma) - \tau)$ and define

$$\overline{P}_{Y|t}(1,1|\tau;\gamma) = \left(Nh^2\overline{f}_t(\tau;\gamma)\right)^{-1}\sum_{m=1}^N Y_{1m}Y_{2m}K_h\left(\overline{t}_m(\gamma)-\tau\right)$$

$$\overline{P}_{Y_p|t}(1|\tau;\gamma) = \left(Nh^2\overline{f}_t(\tau;\gamma)\right)^{-1}\sum_{m=1}^N Y_{p_m}K_h\left(\overline{t}_m(\gamma)-\tau\right)\quad \text{for } p=1,2. \tag{22}$$

We are finally ready to describe the semiparametric sample analog to (20). Take $z \in \mathbb{R}^L$ and let $\widehat{f}_Z(z) = \frac{1}{Nb^L}\sum_{n=1}^N H_b(Z_n - z)$. Now take $\pi \equiv (\pi_1, \pi_2) \in \mathbb{R}^2$ and define

$$\begin{aligned}
\overline{\delta}_{1,2|t}(\pi;\gamma,z) &= \left(Nb^L\widehat{f}_Z(z)\right)^{-1}\sum_{n=1}^N \overline{P}_{Y|t}\Big(1,1|X_{1n}'\eta_1 + \sigma_1\pi_2, X_{2n}'\eta_2+\sigma_2\pi_1;\gamma\Big)H_b(Z_n - z),\\
\overline{\delta}_{p|t}(\pi;\gamma,z) &= \left(Nb^L\widehat{f}_Z(z)\right)^{-1}\sum_{n=1}^N \overline{P}_{Y_p|t}\Big(1|X_{1n}'\eta_1 + \sigma_1\pi_2, X_{2n}'\eta_2+\sigma_2\pi_1;\gamma\Big)H_b(Z_n - z),\\
\overline{\varphi}_{1|t}(\pi;\gamma,z) &= \pi_1 - \frac{\overline{\delta}_{1,2|t}(\pi;\gamma,z)}{\overline{\delta}_{2|t}(\pi;\gamma,z)},\\
\overline{\varphi}_{2|t}(\pi;\gamma,z) &= \pi_2 - \frac{\overline{\delta}_{1,2|t}(\pi;\gamma,z)}{\overline{\delta}_{1|t}(\pi;\gamma,z)},\\
\overline{\varphi}_t(\pi;\gamma,z) &= \left(\overline{\varphi}_{1|t}(\pi;\gamma,z), \overline{\varphi}_{2|t}(\pi;\gamma,z)\right)',\\
\overline{Q}_t(\pi;\gamma,z) &= -\overline{\varphi}_t(\pi;\gamma,z)'\overline{\varphi}_t(\pi;\gamma,z).
\end{aligned} \tag{23}$$

Consider the following problem

$$\operatorname*{Max}_{\pi\in[0,1]^2} \overline{Q}_t(\pi;\gamma,z), \tag{24}$$

and define $\overline{\pi}_t(\gamma, z) =$ Solution to (24) with the smallest value for the π_1 component. (25)

In the Appendix we carefully characterize the uniform (in $[0,1]^2 \times \mathcal{Z} \times \Theta$) convergence properties of the objects in (23) to their population counterparts. Given this, and Assumption B3, the computational implementation of finding (25) would be undertaken using the iterative procedure described in Eq. (7), with the starting points mentioned there. We will use $\overline{\pi}_t(\gamma, z)$ to estimate players' beliefs. In Section 3.2, knowledge of $G_p(\cdot)$ and $G_{1,2}(\cdot)$ allowed us to estimate beliefs in one step. Now, the analogous procedure took us three steps and two sets of kernels and bandwidths, which must now satisfy Assumption B4, described below.

4.3. Estimation of γ

Let $\pi_{(s)}^*(\cdot)$ and $\pi_{t_{(s)}}^*(\cdot)$ be as defined in Assumptions A3 and B3. Define two new indices

$$t_1^*(\gamma) = X_1'\eta_1 + \sigma_1\pi_{2t_{(s)}}^*(\gamma, Z),\quad t_2^*(\gamma) = X_2'\eta_2 + \sigma_2\pi_{1t_{(s)}}^*(\gamma, Z),\qquad t^*(\gamma) = \left(t_1^*(\gamma), t_2^*(\gamma)\right). \tag{26}$$

We will exploit the fact that if $z \in \mathcal{Z}$, then $\left.\frac{\partial\pi_{(s)}^*(\gamma,z)}{\partial\gamma}\right|_{\gamma=\gamma_0} = \left.\frac{\partial\pi^*(\gamma,z)}{\partial\gamma}\right|_{\gamma=\gamma_0}$, where $\pi^*(\gamma_0, z)$ is pair of equilibrium beliefs that uniquely solve (3). For $y \in \{(0,0), (0,1), (1,0), (0,0)\}$ and

[15] Unknown densities were allowed to be arbitrarily close to zero in Section 3. Trimming in that case was done both to bound the density of Z uniformly away from zero, and to remain in the region $\mathcal{Z}$ where BNE equilibria was assumed to possess the features studied in Sections 2.3 and 2.4.

$\tau \equiv (\tau_1, \tau_2) \in \mathbb{R}^2$ define $P(y|\tau)$ in the following way: $P(1,1|\tau) = \Pr(\varepsilon_1 \le \tau_1, \varepsilon_2 \le \tau_2)$, $P(0,1|\tau) = \Pr(\varepsilon_1 > \tau_1, \varepsilon_2 \le \tau_2)$, $P(1,0|\tau) = \Pr(\varepsilon_1 \le \tau_1, \varepsilon_2 > \tau_2)$, and $P(0,0|\tau) = \Pr(\varepsilon_1 > \tau_1, \varepsilon_2 > \tau_2)$. Given our assumptions, if $Z \in \mathcal{Z}$, then $\Pr(Y = y|X,Z) = P(y|t^*(\gamma_0))$. Define

$$P_{Y|t^*_{\mathcal{Z}}}(y|\tau;\gamma) = \Pr(Y = y|t^*(\gamma) = \tau, Z \in \mathcal{Z}),$$

$$\frac{\partial P(y|t^*(\gamma_0))}{\partial\gamma} \equiv \left[\nabla_\tau P(y|t^*(\gamma))\frac{\partial t^*(\gamma)}{\partial\gamma}\right]_{\gamma=\gamma_0}.$$

Let $F_{t^*_{\mathcal{Z}}}(\cdot|\tau,\gamma)$ be the distribution of (X,Z) given $\{t^*(\gamma) = \tau, Z \in \mathcal{Z}\}$. Let $A(\gamma) = t^*(\gamma_0) - t^*(\gamma)$. Adding and subtracting $t^*(\gamma)$ and noting that $A(\gamma_0) = 0$, we get

$$\begin{aligned}
&\left.\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(y|t^*(\gamma);\gamma)}{\partial\gamma}\right|_{\gamma=\gamma_0}\\
&= \frac{\partial}{\partial\gamma}\left[\int P(y|A(\gamma)+t^*(\gamma))\mathrm{d}F_{t^*_{\mathcal{Z}}}(x,z|t^*(\gamma),\gamma)\right]_{\gamma=\gamma_0}\\
&= \frac{\partial}{\partial\gamma}\left\{\left[P(y|t^*(\gamma))\underbrace{\int \mathrm{d}F_{t^*_{\mathcal{Z}}}(x,z|t^*(\gamma),\gamma)}_{=1}\right]\right.\\
&\quad\left.+\left[\int P(y|A(\gamma)+t^*(\gamma_0))\mathrm{d}F_{t^*_{\mathcal{Z}}}(x,z|t^*(\gamma_0),\gamma_0)\right]\right\}_{\gamma=\gamma_0}\\
&= \frac{\partial P(y|t^*(\gamma_0))}{\partial\gamma} - E\left[\left.\frac{\partial P(y|t^*(\gamma_0))}{\partial\gamma}\right|t^*(\gamma_0), Z \in \mathcal{Z}\right]. \qquad (27)
\end{aligned}$$

This parallels the change-of-variable technique in Eqs. (40)–(42) in KS. As before, let $W \equiv (Y,X,Z)$ and define

$$\begin{aligned}
\ell^t_{\mathcal{Z}}(W,\gamma,\pi^*_{t_{(s)}}(\gamma,Z)) = \Big[&Y_1Y_2\log P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma);\gamma)\\
&+(1-Y_1)Y_2\log P_{Y|t^*_{\mathcal{Z}}}(1,0|t^*(\gamma);\gamma)\\
&+Y_1(1-Y_2)\log P_{Y|t^*_{\mathcal{Z}}}(1,0|t^*(\gamma);\gamma)\\
&+(1-Y_1)(1-Y_2)\log P_{Y|t^*_{\mathcal{Z}}}(0,0|t^*(\gamma);\gamma)\Big]\mathbb{1}\{Z \in \mathcal{Z}\},
\end{aligned} \qquad (28)$$

$$S_{\gamma\mathcal{Z}}(W,\gamma_0) = \left.\frac{\partial \ell^f_{\mathcal{Z}}(W,\gamma,\pi^*_t(\gamma,Z))}{\partial\gamma}\right|_{\gamma=\gamma_0}.$$

$S_{\gamma\mathcal{Z}}(W,\gamma_0)$ can be derived directly from (27), which also yields $E[S_{\gamma\mathcal{Z}}(W,\gamma_0)] = 0$. Take $\gamma \in \Gamma$ and let be $\overline{\pi}_t(\cdot)$ as in (25). Define

$$\widehat{t}_{1m}(\gamma_1) = X'_{1m}\eta_1 + \sigma_1\overline{\pi}_{2_t}(\gamma, Z_m),$$
$$\widehat{t}_{2m}(\gamma_2) = X'_{2m}\eta_2 + \sigma_2\overline{\pi}_{1_t}(\gamma, Z_m),$$
$$\widehat{t}_m(\gamma) = \left(\widehat{t}_{1m}(\gamma_1), \widehat{t}_{2m}(\gamma_2)\right).$$

Take $K(\cdot)$, h as in (22). Fix $\tau \in \mathbb{R}^2$, $\widehat{f}_{t_{\mathcal{Z}}}(\tau;\gamma) = (Nh^2)^{-1}\sum_{m=1}^N K_h(\widehat{t}_m(\gamma)-\tau)\mathbb{1}\{Z_m \in \mathcal{Z}\}$, and let

$$\begin{aligned}
\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\tau;\gamma) &= \left(Nh^2\widehat{f}_{t_{\mathcal{Z}}}(\tau;\gamma)\right)^{-1}\sum_{m=1}^N Y_{1m}Y_{2m}K_h\\
&\quad\times\left(\widehat{t}_m(\gamma)-\tau\right)\mathbb{1}\{Z_m \in \mathcal{Z}\}\\
\widehat{P}_{Y|t^*_{\mathcal{Z}}}(0,1|\tau;\gamma) &= \left(Nh^2\widehat{f}_{t_{\mathcal{Z}}}(\tau;\gamma)\right)^{-1}\sum_{m=1}^N(1-Y_{1m})Y_{2m}K_h\\
&\quad\times\left(\widehat{t}_m(\gamma)-\tau\right)\mathbb{1}\{Z_m \in \mathcal{Z}\}\\
\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,0|\tau;\gamma) &= \left(Nh^2\widehat{f}_{t_{\mathcal{Z}}}(\tau;\gamma)\right)^{-1}\sum_{m=1}^N Y_{1m}(1-Y_{2m})\\
&\quad\times K_h\left(\widehat{t}_m(\gamma)-\tau\right)\mathbb{1}\{Z_m \in \mathcal{Z}\}\\
\widehat{P}_{Y|t^*_{\mathcal{Z}}}(0,0|\tau;\gamma) &= \left(Nh^2\widehat{f}_{t_{\mathcal{Z}}}(\tau;\gamma)\right)^{-1}\sum_{m=1}^N(1-Y_{1m})(1-Y_{2m})\\
&\quad\times K_h\left(\widehat{t}_m(\gamma)-\tau\right)\mathbb{1}\{Z_m \in \mathcal{Z}\}.
\end{aligned}$$

If the logarithms exist for the nth observation, let

$$\begin{aligned}
\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n)) = \Big[&Y_{1n}Y_{2n}\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}_n(\gamma);\gamma)\\
&+(1-Y_{1n})Y_{2n}\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(0,1|\widehat{t}_n(\gamma);\gamma)\\
&+Y_{1n}(1-Y_{2n})\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,0|\widehat{t}_n(\gamma);\gamma)\\
&+(1-Y_{1n})(1-Y_{2n})\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(0,0|\widehat{t}_n(\gamma);\gamma)\Big]\mathbb{1}\{Z_n \in \mathcal{Z}\}, \qquad (29)
\end{aligned}$$

and let $\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n))$ be zero otherwise.

Assumption B4 (*Kernels and Bandwidths*). $N^{1/2-\widetilde{\delta}}b^Lh^6 \to \infty$, $N^{1/2}\left(b/h^4\right)^{\overline{M}} \to 0$, $N^{1/2}h^{\overline{M}}/b^L \to 0$. $K(\cdot)$ is $\overline{M}$-times differentiable with bounded derivatives. Both $K(\cdot)$ and $H(\cdot)$ are symmetric, Lipschitz-continuous, bias-reducing kernels of order $\overline{M}$.

Assumption B4 describes the relative difference in the rates of convergence to zero for the bandwidths b and h that will be required to achieve $\sqrt{N}$-consistency of the estimator $\widehat{\gamma}$. Going back to B2, $\overline{M}$ plays the same role as M in Assumption A6. They both indicate the degree of smoothness required from the distributions involved. It is clear from B4 that $\overline{M} > M$.

Assumption B5 (*Technical*). γ_0 is an interior point of Γ. The dominance condition in Assumption A7 is maintained. Let $S_{\gamma\mathcal{Z}}(W,\gamma_0)$ be as defined in Eq. (28). Let $\mathcal{I}_{\mathcal{Z}} = E[S_{\gamma\mathcal{Z}}(W,\gamma_0)S_{\gamma\mathcal{Z}}(W,\gamma_0)']$. Then $\mathcal{I}_{\mathcal{Z}}$ is invertible.

We estimate γ by solving $\mathrm{Max}_{\gamma\in\Gamma}\frac{1}{N}\sum_{n=1}^N\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n))$. Consistency will follow because $\sup_{\gamma\in\Gamma}|\frac{1}{N}\sum_{n=1}^N\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n))-E[\ell^t(W,\gamma,\pi^*_{t_{(s)}}(\gamma,Z))]| \xrightarrow{p} 0$, and $E[\ell^t(W,\gamma,\pi^*_{t_{(s)}}(\gamma,Z))]$ is uniquely maximized at $\gamma = \gamma_0$, when it is equal to $E[\ell(W,\gamma_0,\pi^*(\gamma_0,Z))]$, the expected value of the trimmed conditional log-likelihood of Y given X, Z. Asymptotic normality follows from a linear representation of $\frac{1}{N}\sum_{n=1}^N\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n))$ that holds uniformly up to a $o_p(N^{-1/2})$ term in $\mathcal{N}_\gamma$, and the usual Taylor approximation around γ_0.

Theorem 2. *Define $\Psi(Y,Z)$ as given in Box II.*
Let $\underset{k\times 2}{S_{\gamma\pi_{\mathcal{Z}}}}(W_n,\gamma_0)$ denote the partial derivative of $S_{\gamma\mathcal{Z}}(W_n,\gamma_0)$ with respect to $\pi^*(\gamma_0,Z_n)$. Let $D(Z,\gamma_0)$ be as in Theorem 1, with θ_0 replaced by γ_0. Let $B(\cdot)$ and $t^*(\gamma_0)$ be as defined in (15) and (26), respectively. Let $\underset{3\times 2}{\nabla_\pi}E[B(Y)|t^*(\gamma_0)]$ denote the partial derivative of $E[B(Y)|t^*(\gamma_0)]$ with respect to $\pi^*(\gamma_0,Z)$. Let $\mathcal{I}_{\mathcal{Z}}$ be as defined in Assumption B5. Denote:

$$\xi_{\mathcal{Z}}(W,\gamma_0) = E\left[S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)|t^*(\gamma_0)\right] - E\left[S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)|Z\right],$$

$$\begin{aligned}
\overline{A}_{\mathcal{Z}}(X_n,Z_n,\gamma_0) &= E\left[\xi_{\mathcal{Z}}(W,\gamma_0)|Z=Z_n\right]D(Z_n,\gamma_0)\\
&\quad\times\left[E\left[B(Y_n)|X_n,Z_n\right]-E\left[B(Y_n)|Z_n\right]\right],\\
\overline{B}_{\mathcal{Z}}(W_n,\gamma_0) &= E\left[E\left[\xi_{\mathcal{Z}}(W,\gamma_0)|Z\right]D(Z,\gamma_0)|t^*(\gamma_0)=t^*_n(\gamma_0)\right]\\
&\quad\times\left[E\left[B(Y_n)|X_n,Z_n\right]-B(Y_n)\right],\\
\overline{C}_{\mathcal{Z}}(W_n,\gamma_0) &= E\Big[E\left[E\left[\xi_{\mathcal{Z}}(W,\gamma_0)|Z\right]D(Z,\gamma_0)|t^*(\gamma_0)\right]\nabla_\pi\\
&\quad\times E\left[B(Y)|t^*(\gamma_0)\right]|Z=Z_n\Big]\Psi(Y_n,Z_n;\gamma_0).
\end{aligned}$$

Let $\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n))$ be as defined in (29) and let $\widehat{\gamma}$ be the solution to

$$\underset{\gamma\in\Gamma}{\mathrm{Max}}\frac{1}{N}\sum_{n=1}^N\widehat{\ell^t}_{\mathcal{Z}}(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n)).$$

$$\text{Define}\quad \Psi(Y,Z)=\begin{pmatrix}\dfrac{\Pr(Y_1=1|Y_2=1,Z)\,[Y_2-\Pr(Y_2=1|Z)]-[Y_1Y_2-\Pr(Y_1=1,Y_2=1|Z)]}{\Pr(Y_2=1|Z)}\\ \dfrac{\Pr(Y_2=1|Y_1=1,Z)\,[Y_1-\Pr(Y_1=1|Z)]-[Y_1Y_2-\Pr(Y_1=1,Y_2=1|Z)]}{\Pr(Y_1=1|Z)}\end{pmatrix}$$

Box II.

If Assumptions A1–A2 hold, and Assumptions A3–A7 are modified or replaced according to B1–B5, then

$$\widehat{\gamma}-\gamma_0=\mathcal{I}_{\mathcal{Z}}^{-1}\times\frac{1}{N}\sum_{n=1}^{N}\Big[S_{\gamma\mathcal{Z}}(W,\gamma_0)+\overline{A}_{\mathcal{Z}}(X_n,Z_n,\gamma_0)+\overline{B}_{\mathcal{Z}}(W_n,\gamma_0)+\overline{C}_{\mathcal{Z}}(W_n,\gamma_0)\Big]+o_p(N^{-1/2}).$$

As in Theorem 1, ignoring the exact functional form of equilibrium beliefs impacts the asymptotic variance of our estimator. If these beliefs were of known functional form, $\mathcal{I}_{\mathcal{Z}}^{-1}$ would be the semiparametric efficiency bound (see Chamberlain (1986), Cosslett (1987) and Newey (1990), Eq. (28) or KS, Eq. (50)). See Aradillas-Lopez (2007) for efficiency bounds and efficient influence functions in models with strategic interaction and incomplete information. The term $\overline{A}_{\mathcal{Z}}(X_n,Z_n,\gamma_0)$ in the influence function of $\widehat{\gamma}$ is analogous to $A_{\mathcal{Z}}(X_n,Z_n,\theta_0)$ in Eq. (16). The term $\overline{B}_{\mathcal{Z}}(W_n,\gamma_0)$ appears due to the lack of knowledge about $G_p(\cdot)$ and $G_{1,2}(\cdot)$ in the estimation of players' equilibrium beliefs. The use of the nonparametric estimators defined in (21) as plug-ins in (22) is the reason why the term $\overline{C}_{\mathcal{Z}}(W_n,\gamma_0)$ appears. Analogously to the case of $\widehat{\theta}$, the relative magnitude of these terms depends on the ability of signals Z to predict private information.

5. Extensions

5.1. Beliefs conditioned on unobservables

Suppose Player p conditions his beliefs on Z and ε_p, which is unobserved by the econometrician. In a recent paper, Grieco (2010) has described inferential methods for games where beliefs are conditioned on unobservables. However, the results there specialize to the case where $X=Z$. We will outline how to extend our methodology to the case where beliefs are conditioned on unobservables while allowing for $X\neq Z$. Let $g_{1|2}(\varepsilon_1|\varepsilon_2;\rho)$ denote the conditional pdf of ε_1 given ε_2 and denote its counterpart by $g_{2|1}(\varepsilon_2|\varepsilon_1;\rho)$. Players' optimal choices are now given by

$Y_1=\mathbb{1}\{X_1'\beta_1+\alpha_1\pi_2^*(Z,\varepsilon_1;\theta)-\varepsilon_1\geq 0\}$ and
$Y_2=\mathbb{1}\{X_2'\beta_2+\alpha_2\pi_1^*(Z,\varepsilon_2;\theta)-\varepsilon_2\geq 0\}$.

We maintain the assumption that X,Z are independent of ε_p. This yields

$$\begin{aligned}\pi_1^*(Z,\varepsilon_2;\theta)&=\int\left[\int\mathbb{1}\left\{X_1'\beta_1+\alpha_1\pi_2^*(Z,\varepsilon_1;\theta)-\varepsilon_1\geq 0\right\}\times g_{1|2}(\varepsilon_1|\varepsilon_2;\rho)\mathrm{d}\varepsilon_1\right]f(X_1|Z)\mathrm{d}X_1,\\ \pi_2^*(Z,\varepsilon_1;\theta)&=\int\left[\int\mathbb{1}\left\{X_2'\beta_2+\alpha_2\pi_1^*(Z,\varepsilon_2;\theta)-\varepsilon_2\geq 0\right\}\times g_{2|1}(\varepsilon_2|\varepsilon_1;\rho)\mathrm{d}\varepsilon_2\right]f(X_2|Z)\mathrm{d}X_2.\end{aligned}\tag{30}$$

The equilibrium analysis in Sections 2.3 and 2.4 can be extended to this case. Furthermore, a solution can be found through an iterative procedure along the lines of Eq. (7). For a pair of real-valued, measurable functions $\phi_A:\mathbb{S}(Z)\times\mathbb{S}(\varepsilon_1)\to[0,1]$ and $\phi_B:\mathbb{S}(Z)\times\mathbb{S}(\varepsilon_2)\to[0,1]$, define

$$\begin{aligned}\zeta_1\left(X_1,\phi_A(Z,\cdot),\varepsilon_2;\theta\right)&=\int\mathbb{1}\left\{X_1'\beta_1+\alpha_1\phi_A(Z,\varepsilon_1)-\varepsilon_1\geq 0\right\}\times g_{1|2}(\varepsilon_1|\varepsilon_2;\rho)\mathrm{d}\varepsilon_1,\\ \zeta_2\left(X_2,\phi_B(Z,\cdot),\varepsilon_1;\theta\right)&=\int\mathbb{1}\left\{X_2'\beta_2+\alpha_2\phi_B(Z,\varepsilon_2)-\varepsilon_2\geq 0\right\}\times g_{2|1}(\varepsilon_2|\varepsilon_1;\rho)\mathrm{d}\varepsilon_2.\end{aligned}$$

Choose a starting function $\pi_2^0:\mathbb{S}(Z)\times\mathbb{S}(\varepsilon_1)\to[0,1]$, and let

$$\begin{aligned}\pi_1^1(Z,\varepsilon_2;\theta)&=E\left[\zeta_1\left(X_1,\pi_2^0(Z,\cdot),\varepsilon_2;\theta\right)|Z,\varepsilon_2\right]\\ \pi_2^2(Z,\varepsilon_1;\theta)&=E\left[\zeta_2\left(X_2,\pi_1^1(Z,\cdot;\theta),\varepsilon_1;\theta\right)|Z,\varepsilon_1\right]\\ \pi_1^3(Z,\varepsilon_2;\theta)&=E\left[\zeta_1\left(X_1,\pi_2^2(Z,\cdot;\theta),\varepsilon_2;\theta\right)|Z,\varepsilon_2\right]\\ \pi_2^4(Z,\varepsilon_1;\theta)&=E\left[\zeta_2\left(X_2,\pi_1^3(Z,\cdot;\theta),\varepsilon_1;\theta\right)|Z,\varepsilon_1\right]\\ &\vdots\end{aligned}\tag{31}$$

This procedure converges to a solution to (30) for a given $(z,\epsilon_1,\epsilon_2)$. If we choose $\pi_2^0=0$, then the sequence $\{\pi_1^k(z,\epsilon_2),\pi_2^k(z,\epsilon_1)\}_k$ is monotonic for all possible values[16] of α_1 and α_2. It is also bounded in $[0,1]^2$ w.p.1 and consequently it has a limit. Any such limit is a solution to (30). This is also immediate to see for example if we choose $\pi_2^0=1$. Other starting values would also converge, but $\pi_2^0\in\{0,1\}$ simplify computations, and make the monotonicity argument transparent. Let $(H(\cdot),b)$ and $(K(\cdot),h)$ denote a pair of kernels and bandwidths, with $H_b(\psi)\equiv H(\psi/b)$ and $K_h(\psi)\equiv K(\psi/h)$. A sample analog to the iterative procedure in (31) can be done as follows

1. Draw an i.i.d. sample $(\varepsilon_{1s},\varepsilon_{2s})_{s=1}^S$ from the joint distribution $g_{1,2}(\cdot,\cdot;\rho)$.
2. For any realization z and a given value (ϵ_1,ϵ_2), define

$$\widehat{\zeta}_1\left(X_1,\phi_A(z,\cdot),\epsilon_2;\theta\right)=\sum_{s=1}^{S}\frac{\mathbb{1}\left\{X_1'\beta_1+\alpha_1\phi_A(z,\varepsilon_{1s})-\varepsilon_{1s}\geq 0\right\}H_b(\varepsilon_{2s}-\epsilon_2)}{\sum_{s=1}^{S}H_b(\varepsilon_{2s}-\epsilon_2)},$$

$$\widehat{\zeta}_2\left(X_2,\phi_B(z,\cdot),\epsilon_1;\theta\right)=\sum_{s=1}^{S}\frac{\mathbb{1}\left\{X_2'\beta_2+\alpha_2\phi_B(z,\varepsilon_{2s})-\varepsilon_{2s}\geq 0\right\}H_b(\varepsilon_{1s}-\epsilon_1)}{\sum_{s=1}^{S}H_b(\varepsilon_{1s}-\epsilon_1)}.$$

3. For any realization $\varepsilon_{1_r},\varepsilon_{2_r}$ in our simulated sample, a semiparametric sample analog to the iterative procedure in (31) would be

$$\widehat{\pi}_1^1(z,\varepsilon_{2r};\theta)=\sum_{n=1}^{N}\frac{\widehat{\zeta}_1\left(X_{1n},\pi_2^0(z,\cdot),\varepsilon_{2r};\theta\right)K_h(Z_n-z)}{\sum_{n=1}^{N}K_h(Z_n-z)}$$

$$\widehat{\pi}_2^2(z,\varepsilon_{1r};\theta)=\sum_{n=1}^{N}\frac{\widehat{\zeta}_2\left(X_{2n},\pi_1^1(z,\cdot;\theta),\varepsilon_{1r};\theta\right)K_h(Z_n-z)}{\sum_{n=1}^{N}K_h(Z_n-z)}$$

[16] As before, if $\alpha_1\alpha_2=0$, a solution to (30) is trivially unique.

$$\widehat{\pi}_1^3(z,\varepsilon_{2r};\theta)=\sum_{n=1}^{N}\frac{\widehat{\zeta}_1\left(X_{1n},\pi_2^2(z,\cdot;\theta),\varepsilon_{2r};\theta\right)K_h(Z_n-z)}{\sum_{n=1}^{N}K_h(Z_n-z)}$$

$$\vdots$$

Starting values could be determined as we described above. Let $\widehat{\pi}_1^*(Z_n,\varepsilon_{1s};\theta),\widehat{\pi}_1^*(Z_n,\varepsilon_{1s};\theta)$ denote the limit of this iterative procedure. Estimation of θ could rely on the simulated conditional probabilities

$$\begin{aligned}&\widehat{\Pr}(Y_1=1|X_{1n},Z_n;\theta)\\&\quad=\frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\left\{X_{1n}'\beta_1+\alpha_1\widehat{\pi}_2^*(Z_n,\varepsilon_{1s};\theta)-\varepsilon_{1s}\geq 0\right\}\\&\widehat{\Pr}(Y_2=1|X_{2n},Z_n;\theta)\\&\quad=\frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\left\{X_{2n}'\beta_1+\alpha_2\widehat{\pi}_1^*(Z_n,\varepsilon_{2s};\theta)-\varepsilon_{2s}\geq 0\right\}.\end{aligned}$$

The properties of an estimation procedure such as this one are currently under investigation.

5.2. Games with more players and actions

Consider a game played by a discrete set of players $\boldsymbol{\Omega}=\{1,\ldots,\mathcal{P}\}$, where player $p=1,\ldots,\mathcal{P}$ can choose among a set of S_p+1 possible actions, labeled $(a_p^0,a_p^1,\ldots,a_p^{S_p})$. Let A_p denote the action chosen by p and define

$$Y_p^\ell=\mathbb{1}\left\{A_p=a_p^\ell\right\}.$$

5.2.1. Action profiles

Let "$-p$" denote the collection of player p's opponents. Denote

$$\underset{\sim}{a}_{-p}=\left\{a_q^{\ell_q}\right\}_{q\neq p},$$

where $a_q^{\ell_q}$ denotes the action played by Player $q\neq p$.

The total number of action profiles by p's opponents is $\prod_{q\neq p}(S_q+1)\equiv\mathcal{M}_{-p}$. Denote the space of choice profiles by p's opponents by $\mathbb{A}_{-p}=\{\underset{\sim}{a}_{-p}^1,\ldots,\underset{\sim}{a}_{-p}^{\mathcal{M}_{-p}}\}$ and let $\underset{\sim}{\mathcal{A}}_{-p}\in\mathbb{A}_{-p}$ be the action profile selected by them. For each $j=1,\ldots,\mathcal{M}_{-p}$ define

$$\underset{\sim}{\mathcal{Y}}_{-p}^j=\mathbb{1}\left\{\underset{\sim}{A}_{-p}=\underset{\sim}{a}_{-p}^j\right\},\quad\text{and let}\quad\underset{\sim}{\mathcal{Y}}_{-p}=\left(\underset{\sim}{\mathcal{Y}}_{-p}^1,\ldots,\underset{\sim}{\mathcal{Y}}_{-p}^{\mathcal{M}_{-p}}\right)'.$$

5.2.2. Behavior

Let $V(a_p^\ell)$ denote the payoff to Player p of choosing the ℓth action. A generalization of the normal form payoffs in the 2×2 game is given by

$$\begin{aligned}&V_p(a_p^0)=0,\qquad V_p(a_p^\ell)=X_{p_\ell}'\beta_{p_\ell}+\alpha_{p_\ell}'\underset{\sim}{\mathcal{Y}}_{-p}-\varepsilon_{p_\ell}\\&\quad\text{for }\ell=1,\ldots,S_p,\ \text{and }\alpha_{p_\ell}\equiv\left(\alpha_{p_\ell}^1,\ldots,\alpha_{p_\ell}^{\mathcal{M}_{-p}}\right)'.\end{aligned}$$

Suppose the game has the same information structure as its 2×2 counterpart examined above. Player p's expected utility of choosing action a_p^ℓ is then

$$\begin{aligned}E_p\left[V_p(a_p^\ell)\right]&=X_{p_\ell}'\beta_{p_\ell}+\alpha_{p_\ell}'E_p\left[\underset{\sim}{\mathcal{Y}}_{-p}|Z\right]-\varepsilon_{p_\ell},\\&\qquad\ell=1,\ldots,S_p.\qquad E_p\left[V_p(a_p^0)\right]=0\\&\equiv X_{p_\ell}'\beta_{p_\ell}+\alpha_{p_\ell}'\underset{\sim}{\pi}_{-p}(Z)-\varepsilon_{p_\ell}\end{aligned}$$

where $\underset{\sim}{\pi}_{-p}(Z)$ denotes Player p's subjective expectation of $\underset{\sim}{\mathcal{Y}}_{-p}$ given Z. Players maximize their subjective expected utility which yields the decision rule

$$\begin{aligned}Y_p^\ell&=\mathbb{1}\left\{E_p\left[V_p(a_p^\ell)\right]>\max_{\kappa\neq\ell}\left\{E_p\left[V_p(a_\kappa^p)\right]\right\}\right\},\\&\quad\ell=0,\ldots,S_p.\end{aligned}$$

5.2.3. Equilibrium beliefs

Denote $\varepsilon_p=(\varepsilon_{p_1},\ldots,\varepsilon_{p_{S_p}})',X_p=(X_{p_1}',\ldots,X_{p_{S_p}}')',\theta_p=(\beta_{p_1}',\ldots,\beta_{p_{S_p}}',\alpha_{p_1}',\ldots,\alpha_{p_{S_p}}')'$ and $\theta=(\theta_1',\ldots,\theta_{\mathcal{P}}')'$. Fix an arbitrary vector of constants $\underset{\sim}{\pi}_{-p}=(\pi_{-p}^1,\ldots,\pi_{-p}^{\mathcal{M}_{-p}})'\in\mathbb{R}^{\mathcal{M}_{-p}}$ and let

$$\begin{aligned}&\mathbb{P}_p^\ell\left(X_p,Z;\underset{\sim}{\pi}_{-p},\theta_p\right)\\&=\begin{cases}\Pr\Bigg(X_{p_\ell}'\beta_{p_\ell}+\alpha_{p_\ell}'\underset{\sim}{\pi}_{-p}-\varepsilon_{p_\ell}\\\quad>\max\left\{0,\max\limits_{\substack{\kappa\neq\ell\\\kappa\geq 1}}\left(X_{p_\kappa}'\beta_{p_\kappa}+\alpha_{p_\kappa}'\underset{\sim}{\pi}_{-p}-\varepsilon_{p_\kappa}\right)\right\}\Bigg|X_p,Z\Bigg)\\\quad\text{if }\ell=1,\ldots,S_p\\\Pr\left(\max\limits_{\ell=1,\ldots,S_p}\left(X_{p_\ell}'\beta_{p_\ell}+\alpha_{p_\ell}'\underset{\sim}{\pi}_{-p}-\varepsilon_{p_\ell}\right)<0\Big|X_p,Z\right)\\\quad\text{if }\ell=0.\end{cases}\\&\mathbb{P}_p^\ell\left(Z;\underset{\sim}{\pi}_{-p},\theta_p\right)=E\left[\mathbb{P}_p^\ell\left(X_p,Z;\underset{\sim}{\pi}_{-p},\theta_p\right)|Z\right].\end{aligned}$$

Suppose the distribution of ε_p is assumed known up to a finite-dimensional parameter Σ_p. Extensions of Assumption A4(ii) to this multidimensional case would determine the identification features of Σ_p. Reexpress the probabilities in the previous equation as

$$\mathbb{P}_p^\ell\left(X_p,Z;\underset{\sim}{\pi}_{-p},\theta_p,\Sigma_p\right)\quad\text{and}\quad\mathbb{P}_p^\ell\left(Z;\underset{\sim}{\pi}_{-p},\theta_p,\Sigma_p\right)$$

to reflect their dependence on Σ_p. Suppose we maintain the assumption that ε_p is independent of (X_q,Z_q) for all q and p. Let $\overline{\mathcal{M}}\equiv\sum_{p=1}^{\mathcal{P}}\mathcal{M}_{-p}$ and fix a realization $z\in\mathbb{S}(Z)$. For such a realization, equilibrium beliefs would be any collection $\{\underset{\sim}{\pi}_{-p}\}_{p=1}^{\mathcal{P}}\in\mathbb{R}^{\overline{\mathcal{M}}}$ that solves the following $\overline{\mathcal{M}}\times\overline{\mathcal{M}}$ system

$$\begin{aligned}\pi_{-p}^j&=\prod_{\substack{q\neq p\\a_q^\ell\in\underset{\sim}{a}_{-p}^j}}\mathbb{P}_q^\ell\left(z;\underset{\sim}{\pi}_{-q},\theta_q,\Sigma_q\right),\\&\quad p=1,\ldots,\mathcal{P};\ j=1,\ldots,\mathcal{M}_{-p}.\end{aligned}$$

Existence and uniqueness of BNE can be studied in an analogous way to Sections 2.3 and 2.4. A sample analog to the BNE system described above would be of the form

$$\begin{aligned}\pi_{-p}^j&=\prod_{\substack{q\neq p\\a_q^\ell\in\underset{\sim}{a}_{-p}^j}}\left\{\sum_{n=1}^{N}\frac{\mathbb{P}_q^\ell\left(X_{p_n},z;\underset{\sim}{\pi}_{-q},\theta_q,\Sigma_q\right)K_h(Z_n-z)}{\sum_{n=1}^{N}K_h(Z_n-z)}\right\},\\&\quad p=1,\ldots,\mathcal{P};\ j=1,\ldots,\mathcal{M}_{-p}.\end{aligned}$$

The case in which the distribution of ε_p is unknown could be addressed as in Section 4 by using pseudo equilibrium conditions which would rely on the multiple-index nature of the game. The extension to beliefs conditioned on unobservables could be approached in a way analogous to the one outlined in Section 5.1.

6. A simple empirical example

Consider two firms, $p = 1, 2$ competing against each other in a given industry. At the end of year t, each firm commits to either an "aggressive" or a "passive" capital investment strategy for year $t + 1$. Let K_{p_t} and I_{p_t} denote firm p's net capital stock at the end of year t, and net capital investment during year t, respectively; and let $R_{p_t} = I_{p_t}/K_{p_{t-1}}$ denote firm p's *rate* of capital investment in year t. We say that firm p commits to an aggressive investment strategy for year $t + 1$ if $R_{p_{t+1}} > R_{p_t}$. We will let $Y_{p_t} = \mathbb{1}\{R_{p_{t+1}} > R_{p_t}\}$. In order to adhere to the structure of the game analyzed above, two features must be present. First, it must be a game played by two firms. Second, players must commit simultaneously to their choices in an incomplete-information environment.

All firm-level information was collected from Standard and Poor's Industrial Compustat-North America data set. Let S_{1_t} denote the total real net sales of the largest firm in the industry in year t and let S_{2_t} and S_{3_t} be the corresponding figures for the second and – if it exists – the third largest firms. For any given year t, we considered an industry at the NAICS level as a candidate for having two dominant players in year t if $S_{2_t} \geq 0.05 * S_{1_t}$ (otherwise it was considered to have only one dominant player), and if $S_{3_t} < 0.05 * S_{2_t}$ (otherwise it was considered to have at least three dominant players). From the resulting pool of candidate industries we kept only those that satisfied these criteria for years t, $t-1$ and $t-2$. Thus, the population of industries that we considered as having two dominant players in $t = 1990$ were those that satisfied the above criteria not only in 1990, but also in 1988 and 1989. We did this in order to eliminate those whose structure had shifted towards having two dominant firms only in the recent past.

For each industry, we labeled the firm with the largest market share as "Player 1", with the remaining firm labeled as "Player 2". Let Q_{p_t} denote player p's Tobin's Q at the end of year t. Denote $\Delta\%S_{p_t} = (S_{p_t} - S_{p_{t-1}})/S_{p_{t-1}}$ and $\Delta Q_{p_t} = Q_{p_t} - Q_{p_{t-1}}$. Let $X_{p_t} = (1, \Delta Q_{p_t}, \Delta\%S_{p_t}, R_{p_t})$. Tobin's Q compares the capitalized value of a marginal investment in real capital to its replacement cost. According to the net present value (NPV) theory of investment, Player p should adjust its investment decisions according to changes in Q_{p_t}. We include $\Delta\%S_{p_t}$ to account for the influence of short-term firm performance. Finally, a firm's propensity to act aggressively may depend on the investment rate itself. This is the reason why we also included R_{p_t} in X_{p_t}. Q_{p_t} was computed as in Jovanovic and Rousseau (2003), who also used the Compustat database. S_{p_t} was computed using firm p's net sales. To compute I_{p_t} we used capital expenditures in property, plant and equipment. K_{p_t} was measured as the net value of property, plant and equipment. Price deflators for the value of shipments and capital expenditures were taken from the NBER-CES Manufacturing Database.

Public knowledge of X_{p_t} is directly tied to the timing in the release of Player p's financial statements. Regulations in place during the period analyzed here (1986–1994) allowed firms to strategically delay such disclosure (see Chamley and Gale (1994) for a related model). This delay is the source of incomplete information in our example. At the end of year t, player p knows the exact realization of X_{p_t} but knows that of X_{-p_t} with a lag of at least one quarter. We will assume here that $Z_t = (X_{1_{t-1}}, X_{2_{t-1}})$.[17] Note that, after the fact, the researcher is able to observe X_{p_t} and Y_{p_t} along with Z_t for $p = 1, 2$. Players' unobserved shocks ε_{p_t} are assumed to have a logistic marginal distribution, and a joint distribution given by an FGM copula function

$$G_{1,2}(\epsilon_1, \epsilon_2; \rho) = G_1(\epsilon_1)G_2(\epsilon_2)\left[1 + \rho(1 - G_1(\epsilon_1))(1 - G_2(\epsilon_2))\right],$$

with $-1 \leq \rho \leq 1$. This yields $\mathrm{Corr}(\epsilon_1, \epsilon_2) = \rho$. In addition, both shocks are independent if and only if $\rho = 0$. These distributions are assumed to be time-invariant, which could be easily relaxed.

The years included in the sample were $t = \{1986, 1988, 1990, 1992, 1994\}$. All observations were pooled[18] together, resulting in a sample size of $N = 245$. We have $Z_t \in \mathbb{R}^4$, all of which are assumed continuously distributed. The following statistics summarize the data,

$$\widehat{E}[Y_1] = 0.5743, \quad \widehat{E}[Y_2] = 0.4729,$$
$$\widehat{E}[Y_1Y_2] = 0.2567, \quad \widehat{\mathrm{Corr}}(Y_1, Y_2) = -0.0602,$$
$$\widehat{E}[Y_1|Y_2 = 1] = 0.5428, \quad \widehat{E}[Y_1|Y_2 = 0] = 0.6025,$$
$$\widehat{E}[Y_2|Y_1 = 1] = 0.4470, \quad \widehat{E}[Y_2|Y_1 = 0] = 0.5079.$$

The choices observed in the data were relatively well spread across the four possible outcomes. The choice profile more frequently observed was (1, 0) and the least frequently observed was (0, 0). All these features remained relatively stable across the years in the sample.

The bias-reducing kernel used is the product of a polynomial and a standard normal density function. We chose a trimming set $\mathcal{Z}$ that avoids large, positive values of $\widehat{E}[Q_{p_t}|Z_t]$ for $p = 1, 2$. If mutual fighting hurts both players and the predictions of NPV investment theory hold, this trimming criterion would reduce the likelihood of multiple equilibria. Specifically, we dropped observations for which $\widehat{E}[Q_{p_t}|Z_t]$ was above the 98th percentile for $p = 1, 2$. This eliminated 8 observations. We estimated the model following the steps described in Section 3. Table 1 presents the results.[19] The sign for the coefficient of ΔQ_p is positive, as predicted by the NPV theory of investment. However, it is statistically significant only for Player 1 (the dominant firm). The coefficient of $\Delta\%S_p$ was statistically the same for both players, and this was also true for that of R_p. The composite hypothesis $H_0 : \beta_{\Delta\%S_1} = \beta_{\Delta\%S_2}, \beta_{R_1} = \beta_{R_2}$ could not be rejected.[20] Looking at the strategic interaction estimates, it appears that mutual aggression affects both players negatively. This effect is statistically significant only for the small player (Player 2). Finally, our estimation results provide evidence

Table 1
Estimation results. (Standard errors in parentheses.)

	Player 1		Player 2
Intercept	−1.6285*		−0.2799
	(0.7298)		(0.5804)
ΔQ_p	1.1452*		0.2826
	(0.4020)		(0.3199)
$\Delta\%S_p$	−2.8965*		−2.7802*
	(0.9805)		(0.6169)
R_p	7.4691*		5.5557*
	(1.6078)		(1.2537)
α_p	−0.1906		−1.4531*
	(0.7683)		(0.7027)
ρ		0.8093*	
		(0.3407)	

* Statistically significant at a 5% level.

17 Some of the variables used to compute X_p were not consistently reported for all firms in Compustat's quarterly database. However, all figures were consistently reported annually.

18 Implicit in our modeling choice of taking observations at least two years apart is the assumption that $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t})$ and $\varepsilon_s \equiv (\varepsilon_{1s}, \varepsilon_{2s})$ are jointly i.i.d. for all $|t - s| \geq 2$.

19 The bandwidth h and the kernel-parameter σ (in fact, the product $\sigma \times h$) were chosen doing a simple grid search over a set of candidate values and choosing the ones that minimized the difference (using the Frobenius matrix-norm) between the estimated asymptotic variance–covariance matrix described in Theorem 1 and a block-resampled counterpart. The latter was computed by eliminating blocks of ten observations at a time. These blocks were determined by ranking all observations according to $\|Z\|$ (after eliminating the observations in the trimming set $\mathcal{Z}$ described above). The values chosen were $h \times \sigma = 2/10$.

20 All tests mentioned here were conducted at a 5% significance level.

of a positive and statistically significant correlation coefficient ρ between the unobserved shocks, ε_1 and ε_2.

We estimated a modified version of the game where the degree of strategic interaction depends on players' relative "distance", measured simply by the sales ratio $S_2/S_1 \in (0, 1)$. Omitting the time subscript, players' optimal behavior is

$$Y_p = \mathbb{1}\Big\{X_p'\beta_p + \left[\delta_p + \gamma_p\,(S_2/S_1)\right] \times \underbrace{\Pr_p(Y_{-p} = 1|Y_p = 1, Z)}_{\text{Player p's beliefs.}} - \varepsilon_p \geq 0\Big\}, \quad \text{for } p = 1, 2, \qquad (32)$$

where the strategic interaction parameter α_p is now generalized to $\delta_p + \gamma_p(S_2/S_1)$. We re-estimated the model using the same signals, kernels, etc. as before. There was little change in the estimate results for β_p. Fig. 2 shows estimates and 95% confidence bands for the strategic interaction component. Our results suggest that mutually aggressive behavior becomes increasingly costly as firms become closer to each other (i.e., as $S_2/S_1 \to 1$). While the strategic interaction effect was statistically significant for player 2 at every value of S_2/S_1, we were still unable to reject the hypothesis that there is no strategic interaction effect for player 1 (the dominant firm).

7. Concluding remarks

The amount of information that agents are assumed to possess and their degree of rationality determine the equilibrium properties of any game-theoretic model. Such properties have an immediate impact on the identification features of the model. Under certain circumstances, private information in payoffs reduces the prevalence of multiple equilibria and this has significant identification implications. In such settings, conditional choice probabilities for each outcome of the game can be point-identified, and it is possible to design an estimation procedure that does not require prior knowledge of the sign of strategic among the players. This stands in contrast with the complete information version of the game. To illustrate these issues we studied a 2×2 game that has been previously analyzed in the context of complete information. We compared the equilibrium features of the game with and without complete information. Concentrating on the incomplete information environment, we proposed likelihood-based estimation methods that replicated the population equilibrium conditions using well-defined sample analogs. The estimation procedures were intrinsically tied to the equilibrium properties of the underlying game under weaker semiparametric assumptions than those in the existing literature. Separate procedures were described for the case in which the distribution of unobservables is assumed known, and the case in which it is unknown. In both cases, $\sqrt{N}$-consistency relied on equilibrium uniqueness in a neighborhood of the true parameter value and regularity of (possibly multiple) equilibria elsewhere in the parameter space. The efficiency features of the resulting estimators depended on the extent to which the signals explain the variation in players' private information. As we outlined, the principles and methods used here can be extended to games with more players and actions, or games with beliefs conditioned on unobservables. Ongoing research focuses on these extensions.

Acknowledgements

Financial support from the National Science Foundation Grants Nos. SES-0718409/0946894 and SES-0922062, the Gregory Chow Econometrics Research Program at Princeton University and a Shoemaker fellowship at the University of Wisconsin, Madison are gratefully acknowledged.

Appendix. Mathematical appendix

A.1. A uniform linear representation result

Linear representation theorems like the one we will present here have been known and used in the literature under assumptions very similar to ours. Closely related examples include Lemma 3 in Collomb and Hardle (1986) (henceforth CH) which was invoked for example by Stoker (1991) and by Ahn and Manski (1993). See also Theorem 1′ in Lewbel (1997) and the references cited there. A slight difference with respect to existing results is that our semiparametric estimators must have uniform convergence properties both over the parameter space Θ and over the trimming set $\mathcal{Z}$. The existing results cited consider uniform convergence over compact subsets of the support (e.g, over $\mathcal{Z}$ in our case). The specific conditions we enumerate in the result below are tailor-made to accommodate the technical assumptions made throughout the paper. The proof could rely on the type of combinatorial methods and maximal inequalities in Andrews (1994) and particularly, the results in Sherman (1994a,b). A direct, step-by-step proof using a chaining argument similar to Huber (1967) can be found in the author's website, at http://www.ssc.wisc.edu/~aaradill/spinfo-appendix-suppl.pdf and they are also readily available from the author. We omit the details here for space. Suppose $(X, Z) \in \mathbb{R}^P \times \mathbb{R}^L$ is a random vector with joint density $f_{X,Z}(x, z)$ and let $M \geq L + 1$. Assume an i.i.d. sample $\{X_n, Z_n\}_{n=1}^N$. Fix $\gamma \in \mathbb{R}^D$ and $z \in \mathbb{R}^L$, consider a function $\eta : \mathbb{R}^P \times \mathbb{R}^L \times \mathbb{R}^D \to \mathbb{R}$, a kernel $K : \mathbb{R}^L \to \mathbb{R}$ and a bandwidth $h_N \to 0$. Let $K_{h_N}(\psi) = K(\psi/h_N)$ and define $R_N(z, \gamma) = (Nh_N^L)^{-1}\sum_{n=1}^N \eta(X_n, z, \gamma)K_{h_N}(Z_n - z)$, $\widehat{f}_{Z_N}(z) = (Nh_N^L)^{-1}\sum_{n=1}^N K_{h_N}(Z_n - z)$ and $\mu_N(z, \gamma) = R_N(z, \gamma)/\widehat{f}_{Z_N}(z)$. For any $z \in \mathbb{S}(Z)$ let $\mu(z, \gamma) = E[\eta(X, z, \gamma)|Z = z]$.

Assumption S1. (A) Z is absolutely continuous w.r.t. Lebesgue measure. (B) $f_{X,Z}(x, z)$ and $f_Z(z)$ are bounded, M times differentiable with respect to z with bounded derivatives.

Assumption S2. There exist compact sets $\mathcal{Z} \subset \mathbb{S}(Z)$ with $\inf_{z\in\mathcal{Z}} f_Z(z) > 0$, and $\Gamma \subset \mathbb{R}^D$ such that: (A) $\mu(z, \gamma)$ is M times differentiable w.r.t. z and γ with bounded derivatives $\forall z \in \mathbb{S}(Z)$, $\gamma \in \Gamma$. (B) There exists $\overline{\eta} : \mathbb{R}^P \to \mathbb{R}_+$ such that $|\eta(X, z, \gamma)| \leq \overline{\eta}(X)$ w.p.1 for all $X \in \mathbb{S}(X), z \in \mathcal{Z}, \gamma \in \Gamma E[\overline{\eta}(X)^2|Z = z]$ is a continuous function of z for all $z \in \mathbb{S}(Z)$, and $E[\overline{\eta}(X)^4] < \infty$. (C) There exists $\overline{\eta}_1 : \mathbb{R}^P \to \mathbb{R}_+$, and $\varphi_1 > 0$ such that $|\eta(X, z, \gamma) - \eta(X, z', \gamma)| \leq \overline{\eta}_1(X)\|z - z'\|^{\varphi_1}$ w.p.1 for all $X \in \mathbb{S}(X)$, $z, z' \in \mathcal{Z}$, $\gamma \in \Gamma$, and $E[\overline{\eta}_1(X)] < \infty$. (D) There exists $\overline{\eta}_2 : \mathbb{R}^P \to \mathbb{R}_+$, and $\varphi_2 > 0$ such that $|\eta(X, z, \gamma) - \eta(X, z, \gamma')| \leq \overline{\eta}_2(X)\|\gamma - \gamma'\|^{\varphi_2}$ w.p.1 for all $X \in \mathbb{S}(X)$, $z \in \mathcal{Z}$, $\gamma, \gamma' \in \Gamma$, and $E[\overline{\eta}_2(X)] < \infty$.

Assumption S3. (A) The kernel $K(\cdot)$ has compact support, is Lipschitz-continuous, bounded and symmetric about zero. Denote $\psi = (\psi_1, \ldots, \psi_L)'$, then $\int K(\psi)\mathrm{d}\psi = 1$, $\int \|\psi\|^M|K(\psi)|\mathrm{d}\psi < \infty$ and $\int(\psi_1^{q_1}\cdots\psi_L^{q_L})K(\psi)\mathrm{d}\psi_1\cdots\mathrm{d}\psi_L = 0$ for all $0 < q_1+\cdots+q_L < M$. (B) $h_N \to 0$ satisfies: $Nh_N^{L+2} \to \infty Nh_N^{2L}/\log(N) \to \infty$ and $Nh_N^{2M} \to 0$.[21]

Theorem A.1. *If Assumptions S1–S3 are satisfied, then for any $z \in \mathcal{Z}, \gamma \in \Gamma$,*

$$\mu_N(z, \gamma) - \mu(z, \gamma) = \frac{1}{f_Z(z)}\frac{1}{Nh_N^L}\sum_{n=1}^N [\eta(X_n, z, \gamma) - \mu(z, \gamma)] \times K_{h_N}(Z_n - z) + \xi_N(z, \gamma)$$

where $\sup_{z\in\mathcal{Z}, \gamma\in\Gamma} |\xi_N(z, \gamma)| = O_p(N^{\delta-1}h_N^{-L})$ for any $\delta > 0$.

[21] If $L \geq 2$, $Nh_N^{2L}/\log(N) \to \infty$ implies $Nh_N^{L+2} \to \infty$.

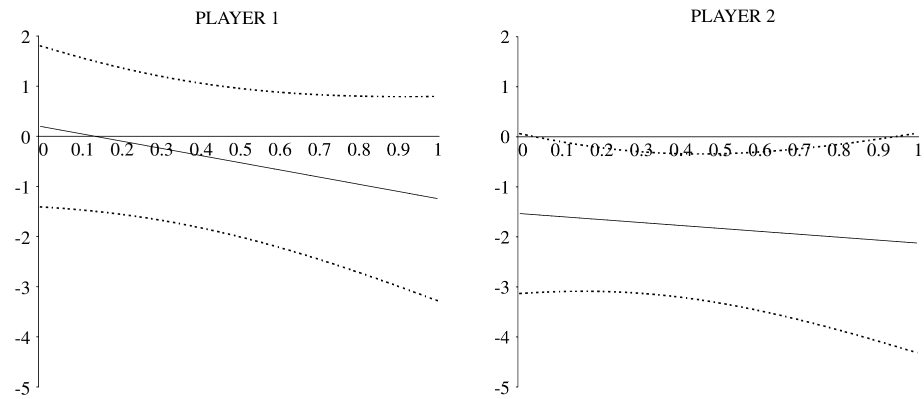

Fig. 2. 95% confidence intervals for $\delta_p+\gamma_p(S_2/S_1)$ (conditional on S_2/S_1) as defined in Eq. (32). Solid line depicts $\widehat{\delta}_p+\widehat{\gamma}_p(S_2/S_1)$. Horizontal axis represents S_2/S_1 in both panels.

Corollary 1. *If we strengthen the condition* $\log Nh_N^{-2L}=o(N)$ *to* $N^{\delta}h_N^{-2L}=o(N)$ *for some* $\delta>0$. *Let* $\xi_N(z,\gamma)$ *be as defined in Theorem A.1, then* $\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}|\xi_N(z,\gamma)|=o_p(N^{-1/2})$.

A.2. Theorem 1

A.2.1. Asymptotic properties of $\widehat{\pi}(\theta,z)$

Fix $(\theta,z)\in\Theta\times\mathcal{Z}$ and re-express the objects in Eq. (2) as

$$\frac{E\left[G_{1,2}(X_1'\beta_1+\alpha_1\pi_2,X_2'\beta_2+\alpha_2\pi_1;\rho)|Z=z\right]}{E\left[G_2(X_2'\beta_2+\alpha_2\pi_1)|Z=z\right]}\equiv\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_2(\pi;\theta,z)}\Rightarrow\varphi_1(\pi_1,\pi_2;\theta,z)=\pi_1-\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_2(\pi;\theta,z)},$$

$$\frac{E\left[G_{1,2}(X_1'\beta_1+\alpha_1\pi_2,X_2'\beta_2+\alpha_2\pi_1;\rho)|Z=z\right]}{E\left[G_1(X_1'\beta_1+\alpha_1\pi_2)|Z=z\right]}\equiv\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_1(\pi;\theta,z)}\Rightarrow\varphi_2(\pi_1,\pi_2;\theta,z)=\pi_2-\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_1(\pi;\theta,z)}.$$

Fix $\Delta>0$ and define the following two correspondences

$$\begin{aligned}\Upsilon_1(\Delta;\theta,z)&=\left\{\pi_1,\pi_2\in[0,1]^2:\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_2(\pi;\theta,z)}-\Delta\le\pi_1\le\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_2(\pi;\theta,z)}+\Delta\right\},\\ \Upsilon_2(\Delta;\theta,z)&=\left\{\pi_1,\pi_2\in[0,1]^2:\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_1(\pi;\theta,z)}-\Delta\le\pi_2\le\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_1(\pi;\theta,z)}+\Delta\right\}.\end{aligned}\tag{A.1}$$

$\Upsilon_p(\Delta;\theta,z)$ contains the locus $\varphi_p(\pi;\theta,z)=0$, for $p=1,2$. Let $\pi^*_{(s')}(\theta,z)$ be the solution to (3) that is closest to $\pi^*_{(s)}(\theta,z)$, with the latter as defined in Assumption S3(i). Uniform regularity of $\pi^*_{(s)}(\theta,z)$ as described in S3(i) yields $\inf_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\|\pi^*_{(s)}(\theta,z)-\pi^*_{(s')}(\theta,z)\|\ge\underline{\epsilon}>0$. Fix any ϵ: $0<\epsilon<\underline{\epsilon}$, and define $B(\epsilon;\theta,z)=\{\pi\in[0,1]^2:\|\pi-\pi^*_{(s)}(\theta,z)\|<\epsilon\}$. Then, for any such ϵ, $\exists\underline{\Delta}(\epsilon)>0$ such that $\forall(\theta,z)\in\Theta\times\mathcal{Z}$ the following holds: If we take *any* pair of continuous selections $\psi_p(\epsilon;\theta,z)\in\Upsilon_p(\underline{\Delta}(\epsilon);\theta,z)$, for $p=1,2$,[22] and we denote the set

$$\Pi_\psi(\epsilon;\theta,z)=\left\{\pi\in[0,1]^2:\pi\in\{\psi_1(\epsilon;\theta,z)\cap\psi_2(\epsilon;\theta,z)\}\right\},$$

and we let $\pi_{\psi_{(s)}}(\epsilon;\theta,z)$ be the element of $\Pi_\psi(\epsilon;\theta,z)$ with the smallest value for the π_1 component, then

$$\sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\left\|\pi_{\psi_{(s)}}(\epsilon;\theta,z)-\pi^*_{(s)}(\theta,z)\right\|\le\epsilon\quad\text{and}$$
$$\inf_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\left\|\pi_{\psi_{(s)}}(\epsilon;\theta,z)-\pi^*_{(s')}(\theta,z)\right\|>\underline{\epsilon}-\epsilon.\tag{A.2}$$

Fig. A.1 illustrates these objects for a case in which (3) has multiple (three) solutions.

Let $\widehat{\delta}_{1,2}(\pi;\theta,z)$ and $\widehat{\delta}_p(\pi;\theta,z)$, $p=1,2$ be as defined in Eq. (9). Given our assumptions, these semiparametric estimators satisfy the conditions of Theorem A.1 and in particular,

$$\sup_{\substack{\pi\in[0,1]^2\\ \theta\in\Theta\\ z\in\mathcal{Z}}}\left|\widehat{\delta}_{1,2}(\pi;\theta,z)-\delta_{1,2}(\pi;\theta,z)\right|\xrightarrow{p}0,$$
$$\sup_{\substack{\pi\in[0,1]^2\\ \theta\in\Theta\\ z\in\mathcal{Z}}}\left|\widehat{\delta}_p(\pi;\theta,z)-\delta_p(\pi;\theta,z)\right|\xrightarrow{p}0,\quad p=1,2.$$

Invoking the unbounded support property of ε_p, these results yield

$$\sup_{\substack{\pi\in[0,1]^2\\ \theta\in\Theta\\ z\in\mathcal{Z}}}\left\|\frac{\widehat{\delta}_{1,2}(\pi;\theta,z)}{\widehat{\delta}_p(\pi;\theta,z)}-\frac{\delta_{1,2}(\pi;\theta,z)}{\delta_p(\pi;\theta,z)}\right\|\xrightarrow{p}0,$$

and therefore,
$$\sup_{\substack{\pi\in[0,1]^2\\ \theta\in\Theta\\ z\in\mathcal{Z}}}\|\widehat{\varphi}(\pi;\theta,z)-\varphi(\pi;\theta,z)\|\xrightarrow{p}0.\tag{A.3}$$

Take any $\Delta>0$ and define the event

$$\mathcal{S}(\Delta;\theta,z)=\mathbb{1}\left\{\left\{\pi\in[0,1]^2:\widehat{\varphi}_p(\pi;\theta,z)=0\right\}\text{ is a continuous selection of }\Upsilon_p(\Delta;\theta,z)\text{ for }p=1,2\right\}.$$

Eqs. (A.1) and (A.3) yield $\sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}|\mathcal{S}(\Delta;\theta,z)-1|\xrightarrow{p}0$. Now take $\widehat{\pi}(\theta,z)$ as defined in Eq. (11) and define the event

$$\mathcal{T}(\theta,z)=\mathbb{1}\{\widehat{\pi}(\theta,z)\text{ is the solution to }\widehat{\varphi}(\pi;\theta,z)=0\text{ with the smallest value for the }\pi_1\text{ component}\}.$$

Since all solutions to (3) are assumed to be regular w.p.1 uniformly in $\Theta\times\mathcal{Z}$, we have $\sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}|\mathcal{T}(\theta,z)-1|\xrightarrow{p}0$. Now take any $0<\epsilon<\underline{\epsilon}$ and let $\underline{\Delta}(\epsilon)$ be as defined above, then these results and

[22] That is, any pair of continuous functions $\psi_1(\epsilon;\theta,z)$ and $\psi_2(\epsilon;\theta,z)$ such that $\psi_1(\epsilon;\theta,z)$ belongs to the correspondence $\Upsilon_1(\underline{\Delta}(\epsilon);\theta,z)$ and $\psi_2(\epsilon;\theta,z)$ belongs to the correspondence $\Upsilon_2(\underline{\Delta}(\epsilon);\theta,z)$.

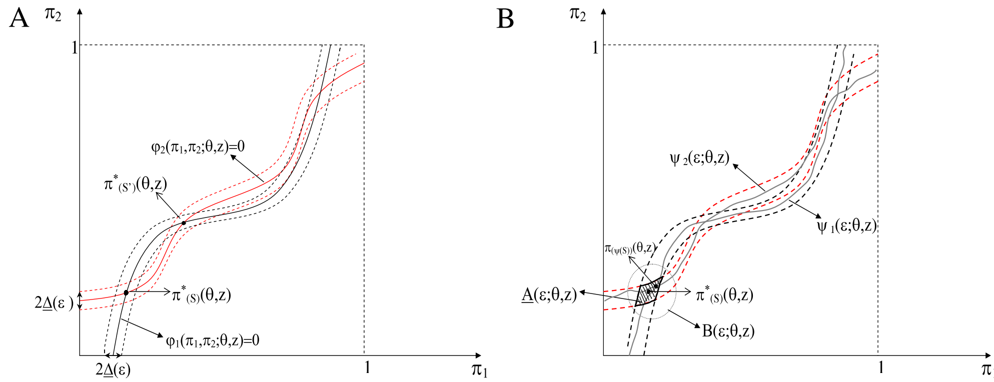

Fig. A.1. Example: The correspondence $\Upsilon_p(\underline{\Delta}(\epsilon);\theta,z)$ is given by the set of points between (and including) the dotted lines surrounding the locus $\varphi_p(\pi_1,\pi_2;\theta,z)=0$, for $p=1,2$. Panel (B) depicts two continuous selections $\psi_p(\epsilon;\theta,z)\in\Upsilon_1(\underline{\Delta}(\epsilon);\theta,z)$, for $p=1,2$. The first point of crossing along the π_1-axis of any such pair of continuous selections must be inside the set $\underline{A}(\epsilon;\theta,z)$, which is contained in $B(\epsilon;\theta,z)$.

Eq. (A.2) yield[23]

$$\sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\left\|\widehat{\pi}(\theta,z)-\pi^*_{(s)}(\theta,z)\right\|\xrightarrow{p}0\quad\text{and consequently,}$$

$$\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\left\|\widehat{\pi}(\theta,z)-\pi^*(\theta,z)\right\|\xrightarrow{p}0. \tag{A.4}$$

Next, we will look for a linear representation for $\widehat{\pi}(\theta,z)-\pi^*(\theta,z)$ in $\mathcal{N}_\theta\times\mathcal{Z}$. As in (5), let $\underset{2\times 2}{J(\pi;\theta,z)}=\nabla_\pi\varphi(\pi;\theta,z)$. Uniqueness implies regularity, and Assumption S3(ii) yields

$$\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\left\|J\left(\pi^*(\theta,z);\theta,z\right)^{-1}\right\|<\overline{M}\quad\text{for some }\overline{M}<\infty. \tag{A.5}$$

With probability approaching one (w.p.a.1) uniformly in $\mathcal{N}_\theta\times\mathcal{Z}$, $\widehat{\pi}(\theta,z)$ satisfies $\widehat{\varphi}(\widehat{\pi}(\theta,z);\theta,z)=0$. A first-order approximation around $\pi^*(\theta,z)$ yields

$$\begin{aligned}0=&\ \widehat{\varphi}\left(\pi^*(\theta,z);\theta,z\right)+\widehat{J}\left(\widetilde{\pi}(\theta,z);\theta,z\right)\\ &\times\left[\widehat{\pi}(\theta,z)-\pi^*(\theta,z)\right],\end{aligned} \tag{A.6}$$

where $\widehat{J}(\pi;\theta,z)=\nabla_\pi\widehat{\varphi}(\pi;\theta,z)$ and $\widetilde{\pi}(\theta,z)$ is between $\widehat{\pi}(\theta,z)$ and $\pi^*(\theta,z)$. Let $V(Z)$ and $B(Y)$ be as in Eq. (15), and let $V(Z,\theta)$ denote the value of $V(Z)$ for an arbitrary $\theta\in\mathcal{N}_\theta$. Our assumptions and the result in Theorem A.1 yield

$$\begin{aligned}&\widehat{\varphi}\left(\pi^*(\theta,z);\theta,z\right)\\ &\quad=-\frac{V(z,\theta)}{f_z(z)}\frac{1}{Nh^L}\sum_{m=1}^{N}\left[E\left[B(Y_m)|X_m,Z_m=z;\theta\right]\right.\\ &\qquad\left.-E\left[B(Y_m)|Z_m=z;\theta\right]\right]K_h\left(Z_m-z\right)+\widehat{\xi}_\varphi(z,\theta)\end{aligned} \tag{A.7}$$

where $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\widehat{\xi}_\varphi(z,\theta)\|=O_p(N^{\delta-1}h^{-L})$ for any $\delta>0$. Eqs. (A.4), (A.5) and Theorem A.1 yield $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\widehat{J}(\widetilde{\pi}(\theta,z);\theta,z)^{-1}\|=O_p(1)$. Combining this with Eqs. (A.6), (A.7) and the proof of Theorem A.1, we get

$$\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\left\|\widehat{\pi}(\theta,z)-\pi^*(\theta,z)\right\|=O_p\left(\sqrt{N^{\delta-1}h^{-L}}\right)\quad\text{for any }\delta>0. \tag{A.8}$$

Combining this result with Theorem A.1 and Eq. (A.5), a first order approximation yields $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\widehat{J}(\widetilde{\pi}(\theta,z);\theta,z)^{-1}-J(\pi^*(\theta,z);\theta,z)^{-1}\|=O_p(\sqrt{N^{\delta-1}h^{-L}})$ for any $\delta>0$. Let $D(z,\theta)=J(\pi^*(\theta,z);\theta,z)^{-1}V(z,\theta)$. Eqs. (A.6) and (A.7) yield[24]

$$\begin{aligned}\widehat{\pi}(\theta,z)-\pi^*(\theta,z)=&\ \frac{D(z,\theta)}{f_Z(z)}\frac{1}{Nh^L}\sum_{m=1}^{N}\left[E\left[B(Y_m)|X_m,Z_m=z;\theta\right]\right.\\ &\left.-E\left[B(Y_m)|Z_m=z;\theta\right]\right]K_h\left(Z_m-z\right)+\widehat{\xi}_\pi(z,\theta),\end{aligned} \tag{A.9}$$

where $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\widehat{\xi}_\pi(z,\theta)\|=O_p(N^{\delta-1}h^{-L})$ for any $\delta>0$. Now consider $\nabla_\theta\widehat{\pi}(\theta,z)$. w.p.a.1 uniformly in $\mathcal{N}_\theta\times\mathcal{Z}$, the Implicit Function Theorem (IFT) yields

$$\nabla_\theta\widehat{\pi}(\theta,z)=-\widehat{J}\left(\widehat{\pi}(\theta,z);\theta,z\right)^{-1}\nabla_\theta\widehat{\varphi}\left(\widehat{\pi}(\theta,z);\theta,z\right),$$

where $\nabla_\theta\widehat{\varphi}(\pi;\theta,z)$ denotes the vector of partial derivatives of $\widehat{\varphi}(\pi;\theta,z)$ with respect to θ, with π fixed. The IFT also yields $\nabla_\theta\pi^*(\theta,z)=-J(\pi^*(\theta,z);\theta,z)^{-1}\nabla_\theta\varphi(\pi^*(\theta,z);\theta,z)$. A detailed inspection into the components of $\nabla_\theta\widehat{\varphi}(\pi;\theta,z)$, along with our assumptions and the results from Theorem A.1 yield

$$\sup_{\substack{\pi\in[0,1]^2\\ \theta\in\Theta\\ z\in\mathcal{Z}}}\left\|\nabla_\theta\widehat{\varphi}\left(\pi;\theta,z\right)-\nabla_\theta\varphi\left(\pi;\theta,z\right)\right\|=O_p\left(\sqrt{N^{\delta-1}h^{-L}}\right)$$
$$\text{for any }\delta>0. \tag{A.10}$$

Using (A.9) and (A.10), component-wise Taylor approximations yield

$$\begin{aligned}\nabla_\theta\widehat{\pi}(\theta,z)-\nabla_\theta\pi^*(\theta,z)\ =&\ \frac{1}{Nh^L}\sum_{m=1}^{N}\Gamma(X_m,z;\theta)K_h(Z_m-z)\\ &+\widehat{\upsilon}(z,\theta),\end{aligned} \tag{A.11}$$

where

$$E\left[\Gamma(X_m,z;\theta)|Z_m=z\right]=0,$$

$$\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\left\|E\left[\frac{\Gamma(X_m,z;\theta)K_h(Z_m-z)}{h^L}\right]\right\|=O(h^M),$$

and $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\widehat{\upsilon}(z,\theta)\|=O_p(N^{\delta-1}h^{-L})$ for any $\delta>0$. M is as defined in Assumption A6. The exact expression for $\Gamma(X_m,z;\theta)$ is not relevant (see Eqs. (A.20) and (A.21)).

23 If all solutions to Eq. (10) are regular, then each one satisfies the first order conditions $\nabla_\pi\widehat{Q}(\pi;\theta,z)=0$. Since the Jacobian $\nabla_\pi\widehat{\varphi}(\pi;\theta,z)$ is invertible, this can be satisfied if and only if $\widehat{\varphi}(\pi;\theta,z)=0$.

24 Uniqueness of equilibrium along with our remaining assumptions imply that each one of the elements in $J(\pi^*(\theta,z);\theta,z)^{-1}$ is uniformly bounded away from zero everywhere in $\mathcal{N}_\theta\times\mathcal{Z}$. A first-order Taylor approximation of $\widehat{J}(\widetilde{\pi}(\theta,z);\theta,z)^{-1}-J(\pi^*(\theta,z);\theta,z)^{-1}$ component-wise along with our previous results and Theorem A.1 yield the result $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\widehat{J}(\widetilde{\pi}(\theta,z);\theta,z)^{-1}-J(\pi^*(\theta,z);\theta,z)^{-1}\|=O_p(\sqrt{N^{\delta-1}h^{-L}})$.

A.2.2. Proof of Theorem 1

Using our previous results and Assumption A7, we have

$$\sup_{\theta\in\Theta}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right)-\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\left(W_n,\theta,\pi^*_{(s)}(\theta,Z_n)\right)\right| \le \underbrace{\sup_{\substack{\theta\in\Theta\\ z\in\mathcal{Z}}}\left\|\widehat{\pi}(\theta,z)-\pi^*_{(s)}(\theta,z)\right\|}_{o_p(1)} \times \underbrace{\frac{1}{N}\sum_{n=1}^{N}\sup_{\substack{\theta\in\Theta\\ \pi\in[0,1]^2}}\left\|\nabla_\pi\ell_{\mathcal{Z}}\left(W_n,\theta,\pi\right)\right\|}_{O_p(1)} = o_p(1). \tag{A.12}$$

Using Assumption A7 and Lemma 2.4 in Newey and McFadden (1994), we obtain

$$\sum_{\theta\in\Theta}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\left(W_n,\theta,\pi^*_{(s)}(\theta,Z_n)\right) - E\left[\ell_{\mathcal{Z}}\left(W,\theta,\pi^*_{(s)}(\theta,Z)\right)\right]\right| \xrightarrow{p} 0.$$

Eq. (A.12) yields

$$\sum_{\theta\in\Theta}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right) - E\left[\ell_{\mathcal{Z}}\left(W,\theta,\pi^*_{(s)}(\theta,Z)\right)\right]\right| \xrightarrow{p} 0. \tag{A.13}$$

Let $L_{\mathcal{Z}}(W,\theta)=\exp\{\ell_{\mathcal{Z}}(W,\theta,\pi^*_{(s)}(\theta,Z))\}$, and note that $L_{\mathcal{Z}}(W,\theta_0)=\Pr(Y|X,Z)^{\mathbb{1}\{Z\in\mathcal{Z}\}}$. Take any $\theta\in\Theta$ such that $\theta\neq\theta_0$. By Assumption A4, it is not the case that the ratio $L_{\mathcal{Z}}(W,\theta)/L_{\mathcal{Z}}(W,\theta_0)$ is constant w.p.1. We have

$$\frac{L_{\mathcal{Z}}(W,\theta)}{L_{\mathcal{Z}}(W,\theta_0)} = \left\{\frac{\exp\left\{\ell\left(W,\theta,\pi^*_{(s)}(\theta,Z)\right)\right\}}{\Pr(Y|X,Z)}\ \text{if } Z\in\mathcal{Z} \text{ and 1 otherwise}\right\}.$$

This ratio is always positive for any θ. Therefore by Jensen's Inequality we have

$$-\log\left\{E\left[\frac{L_{\mathcal{Z}}(W,\theta)}{L_{\mathcal{Z}}(W,\theta_0)}\right]\right\} < E\left[-\log\left\{\frac{L_{\mathcal{Z}}(W,\theta)}{L_{\mathcal{Z}}(W,\theta_0)}\right\}\right]. \tag{A.14}$$

By construction, for any $(x,z)\in\mathbb{S}(X,Z)$ we have $\sum_y\exp\{\ell(y,x,z,\theta,\pi^*_{(s)}(\theta,z))\}=1$ when we sum over the four possible outcomes $y\in\{(1,1),(1,0),(0,1),(0,0)\}$.[25] Therefore

$$E\left[\frac{L_{\mathcal{Z}}(W,\theta)}{L_{\mathcal{Z}}(W,\theta_0)}\right] = \int_{(x,z):z\notin\mathcal{Z}} 1\cdot f_{X,Z}(x,z)\mathrm{d}x\mathrm{d}z + \int_{(x,z):z\in\mathcal{Z}}\left\{\sum_y \times\left[\frac{\exp\left\{\ell\left(y,x,z,\theta,\pi^*_{(s)}(\theta,Z)\right)\right\}}{\Pr(Y=y|X=x,Z=z)}\right] \times \Pr(Y=y|X=x,Z=z)\right\} f_{X,Z}(x,z)\mathrm{d}x\mathrm{d}z$$

$$= (1-\Pr(Z\in\mathcal{Z})) + \int_{(x,z):z\in\mathcal{Z}}\left\{\sum_y\exp\left\{\ell\left(y,x,z,\theta,\pi^*_{(s)}(\theta,z)\right)\right\}\right\} \times f_{X,Z}(x,z)\mathrm{d}x\mathrm{d}z = (1-\Pr(Z\in\mathcal{Z}))+\Pr(Z\in\mathcal{Z})=1.$$

Combining this with (A.14) we have

$$0 < E\left[-\log\left\{\frac{L_{\mathcal{Z}}(W,\theta)}{L_{\mathcal{Z}}(W,\theta_0)}\right\}\right] = E\left[\log L_{\mathcal{Z}}(W,\theta)\right] - E\left[\log L_{\mathcal{Z}}(W,\theta_0)\right] \quad \text{for any } \theta\in\Theta:\theta\neq\theta_0.$$

Since $L_{\mathcal{Z}}(W,\theta)=\exp\{\ell_{\mathcal{Z}}(W,\theta,\pi^*_{(s)}(\theta,Z))\}$, this implies that $E[\ell_{\mathcal{Z}}(W,\theta,\pi^*_{(s)}(\theta,Z))]$ is uniquely minimized at θ_0. Combining this with Eq. (A.13) and Theorem 2.1 in Newey and McFadden, we get

$$\widehat{\theta}\xrightarrow{p}\theta_0. \tag{A.15}$$

Let $S_{\theta_{\mathcal{Z}}}(W,\theta_0)$ be as defined in Eq. (14). Since θ_0 belongs in the interior of Θ, it satisfies $E[S_{\theta_{\mathcal{Z}}}(W,\theta_0)]=0$. Denote

$$\widehat{S}_{\theta_{\mathcal{Z}}}(W,\theta) \equiv \frac{\partial\ell_{\mathcal{Z}}(W,\theta,\widehat{\pi}(\theta,Z))}{\partial\theta} = \nabla_\theta\ell_{\mathcal{Z}}\left(W,\theta,\widehat{\pi}(\theta,Z)\right) + \nabla_\theta\widehat{\pi}(\theta,Z)'\nabla_\pi\ell_{\mathcal{Z}}\left(W,\theta,\widehat{\pi}(\theta,Z)\right).$$

w.p.a.1, $\widehat{\theta}$ satisfies $\frac{1}{N}\sum_{n=1}^{N}\widehat{S}_{\theta_{\mathcal{Z}}}(W_n,\widehat{\theta})=0$. We have

$$0=\frac{1}{N}\sum_{n=1}^{N}\widehat{S}_{\theta_{\mathcal{Z}}}\left(W_n,\theta_0\right)+\left[\frac{1}{N}\sum_{n=1}^{N}\frac{\partial\widehat{S}_{\theta_{\mathcal{Z}}}\left(W_n,\widetilde{\theta}\right)}{\partial\theta'}\right]\left(\widehat{\theta}-\theta_0\right), \tag{A.16}$$

with $\widetilde{\theta}$ between $\widehat{\theta}$ and θ_0. Eqs. (A.8), (A.10) and Assumption A7 yield

$$\frac{1}{N}\sum_{n=1}^{N}\left(\nabla_\theta\widehat{\pi}(\theta_0,Z_n)-\nabla_\theta\pi^*(\theta_0,Z_n)\right)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}} \times\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)\left(\widehat{\pi}(\theta_0,Z_n)-\pi^*(\theta_0,Z_n)\right) = O_p\left(N^{\delta-1}h^{-L}\right)$$

for any $\delta>0$. Therefore,

$$\frac{1}{N}\sum_{n=1}^{N}\widehat{S}_{\theta_{\mathcal{Z}}}\left(W_n,\theta_0\right)=\frac{1}{N}\sum_{n=1}^{N}\Big\{S_{\theta_{\mathcal{Z}}}\left(W_n,\theta_0\right) + \left[\nabla_\theta\pi^*(\theta_0,Z_n)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right) + \nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)\right]\left(\widehat{\pi}(\theta_0,Z_n)-\pi^*(\theta_0,Z_n)\right) + \left(\nabla_\theta\widehat{\pi}(\theta_0,Z_n)-\nabla_\theta\pi^*(\theta_0,Z_n)\right)' \times\ \nabla_\pi\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)\Big\}+O_p\left(N^{\delta-1}h^{-L}\right) \quad \text{for any } \delta>0. \tag{A.17}$$

Using (A.9), for any $\delta>0$,

$$\left[\nabla_\theta\pi^*(\theta_0,Z_n)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right) + \nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)\right]\left(\widehat{\pi}(\theta_0,Z_n)-\pi^*(\theta_0,Z_n)\right) = \frac{1}{N^2h^L}\sum_{m=1}^{N}\sum_{n=1}^{N}\left[\nabla_\theta\pi^*(\theta_0,Z_n)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}} \times\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)+\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)\right] \times\frac{D(Z_n,\theta_0)}{f_Z(Z_n)}\left[E\left[B(Y_m)|X_m,Z_m=Z_n\right] - E\left[B(Y_m)|Z_m=Z_n\right]\right]K_h\left(Z_m-Z_n\right)+O_p\left(N^{\delta-1}h^{-L}\right). \tag{A.18}$$

[25] See Eq. (12).

Let $S_{\theta\pi_{\mathcal{Z}}}(W_n,\theta_0)$ be as in Theorem 1. Note that $E[\nabla_{\pi_p}\ell_{\mathcal{Z}}(W_n,\theta_0,\pi^*(\theta_0,Z_n))|X_n,Z_n]=0$ for $p=1,2$. Therefore,

$$E\left[S_{\theta\pi_{\mathcal{Z}}}(W_n,\theta_0)|Z_n\right]=E\left[\nabla_\theta\pi^*(\theta_0,Z_n)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\right.\\ \times\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)+\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)|Z_n\left.\right].$$

The first term on the right-hand side of Eq. (A.18) is a U-statistic with projection

$$\frac{1}{N}\sum_{n=1}^{N}E\left[S_{\theta\pi_{\mathcal{Z}}}(W_n,\theta_0)|Z_n\right]D(Z_n,\theta_0)\left[E\left[B(Y_n)|X_n,Z_n\right]\right.\\ \left.-E\left[B(Y_n)|Z_n\right]\right]+\xi_N, \tag{A.19}$$

where $\xi_N=o_p(N^{-1/2})+O(h^M)$. Take the third term on the right-hand side of Eq. (A.17). Using (A.10) and (A.11),

$$\frac{1}{N}\sum_{n=1}^{N}\left(\nabla_\theta\widehat{\pi}(\theta_0,Z_n)-\nabla_\theta\pi^*(\theta_0,Z_n)\right)'\nabla_\pi\ell_{\mathcal{Z}}\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)\\ =\frac{1}{N^2h^L}\sum_{m=1}^{N}\sum_{n=1}^{N}\left(\Gamma(X_m,Z_n;\theta)K_h(Z_m-Z_n)\right)'\nabla_\pi\ell_{\mathcal{Z}}\\ \times\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)+O_p(N^{\delta-1}h^{-L}), \tag{A.20}$$

for any $\delta>0$. The first term on the right-hand side can be expressed as a symmetric, second-order U-statistic whose projection is given by a term of order $O(h^M)$. Therefore

$$\frac{1}{N^2h^L}\sum_{m=1}^{N}\sum_{n=1}^{N}\left(\Gamma(X_m,Z_n;\theta_0)K_h(Z_m-Z_n)\right)'\nabla_\pi\ell_{\mathcal{Z}}\\ \times\left(W_n,\theta_0,\pi^*(\theta_0,Z_n)\right)=o_p(N^{-1/2})+O(h^M). \tag{A.21}$$

Assumption A6(ii) yields $O_p(N^{\overline{\delta}-1}h^{-L})=o_p(N^{-1/2})$ for some $\overline{\delta}>0$ and $N^{1/2}h^M\to 0$. Combining this with the results in (A.19) and (A.21), Eq. (A.17) becomes

$$\frac{1}{N}\sum_{n=1}^{N}\widehat{S}_{\theta_{\mathcal{Z}}}(W_n,\theta_0)=\frac{1}{N}\sum_{n=1}^{N}\left[S_{\theta_{\mathcal{Z}}}(W_n,\theta_0)\right.\\ \left.+A_{\mathcal{Z}}(X_n,Z_n,\theta_0)\right]+o_p(N^{-1/2}) \tag{A.22}$$

where $A_{\mathcal{Z}}(X_n,Z_n,\theta_0)\equiv E[S_{\theta\pi_{\mathcal{Z}}}(W_n,\theta_0)|Z_n]D(Z_n,\theta_0)[E[B(Y_n)|X_n,Z_n]-E[B(Y_n)|Z_n]]$. Now, since $E[\nabla_\pi\ell_{\mathcal{Z}}(W_n,\theta_0,\pi^*(\theta_0,Z_n))|X_n,Z_n]=0$, we have

$$E\left[\frac{\partial S_{\theta_{\mathcal{Z}}}(W,\theta_0)}{\partial\theta'}\right]=E\left[\nabla_{\theta\theta'}\ell_{\mathcal{Z}}\left(W,\theta_0,\pi^*(\theta_0,Z)\right)\right.\\ +\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W,\theta_0,\pi^*(\theta_0,Z)\right)\nabla_\theta\pi^*(\theta_0,Z)\\ +\nabla_\theta\pi^*(\theta_0,Z)'\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W,\theta_0,\pi^*(\theta_0,Z)\right)'\\ \left.+\nabla_\theta\pi^*(\theta_0,Z)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\left(W,\theta_0,\pi^*(\theta_0,Z)\right)\nabla_\theta\pi^*(\theta_0,Z)\right]. \tag{A.23}$$

Take $a,b\in\{\theta,\pi\}$. It is easy to show that

$$E\left[\nabla_{a,b'}\ell_{\mathcal{Z}}\left(W,\theta_0,\pi^*(\theta_0,Z)\right)|X,Z\right]\\ =-E\left[\nabla_a\ell_{\mathcal{Z}}\left(W,\theta_0,\pi^*(\theta_0,Z)\right)\nabla_b\ell_{\mathcal{Z}}\right.\\ \left.\times\left(W,\theta_0,\pi^*(\theta_0,Z)\right)'|X,Z\right]. \tag{A.24}$$

Eqs. (A.23) and (A.24) yield

$$-E\left[\frac{\partial S_{\theta_{\mathcal{Z}}}(W,\theta_0)}{\partial\theta'}\right]=E\left[S_{\theta_{\mathcal{Z}}}(W,\theta_0)S_{\theta_{\mathcal{Z}}}(W,\theta_0)'\right]\equiv\Im_{\mathcal{Z}}, \tag{A.25}$$

an information-identity result for our trimmed log-likelihood function. We have

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial\widehat{S}_{\theta_{\mathcal{Z}}}(W_n,\theta)}{\partial\theta'}=\frac{1}{N}\sum_{n=1}^{N}\left[\nabla_{\theta\theta'}\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right)\right.\\ +\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right)\nabla_\theta\widehat{\pi}(\theta,Z_n)\\ +\nabla_{\theta\theta'}\widehat{\pi}(\theta,Z_n)'\left[\nabla_\pi\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right)\otimes I_{(k+3)}\right]\\ +\nabla_\theta\widehat{\pi}(\theta,Z_n)'\left\{\nabla_{\pi\theta'}\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right)\right.\\ \left.\left.+\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\left(W_n,\theta,\widehat{\pi}(\theta,Z_n)\right)\nabla_\theta\widehat{\pi}(\theta,Z_n)\right\}\right].$$

Using (A.5) and Theorem A.1, component-wise Taylor approximations yield $\sup_{\substack{\theta\in\mathcal{N}_\theta\\ z\in\mathcal{Z}}}\|\nabla_{\theta\theta'}\widehat{\pi}(\theta,z)-\nabla_{\theta\theta'}\pi^*(\theta,z)\|\xrightarrow{p}0$.Since $E[\nabla_{\pi_p}\ell_{\mathcal{Z}}(W,\theta_0,\pi^*(\theta_0,Z))|Z]=0$ for $p=1,2$ and $\widehat{\theta}\xrightarrow{p}\theta_0\in\mathcal{N}_\theta$. Therefore, $\frac{1}{N}\sum_{n=1}^{N}\nabla_{\theta\theta'}\widehat{\pi}(\widehat{\theta},Z_n)'[\nabla_\pi\ell_{\mathcal{Z}}(W_n,\widehat{\theta},\widehat{\pi}(\widehat{\theta},Z_n))\otimes I_{(k+3)}]\xrightarrow{p}0$. Going back to (A.16),

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial\widehat{S}_{\theta_{\mathcal{Z}}}\left(W_n,\widetilde{\theta}\right)}{\partial\theta'}\xrightarrow{p}-\Im_{\mathcal{Z}}.$$

Combining this with (A.16) and (A.22) yields

$$\widehat{\theta}-\theta_0=\Im_{\mathcal{Z}}^{-1}\times\frac{1}{N}\sum_{n=1}^{N}\left[S_{\theta_{\mathcal{Z}}}(W_n,\theta_0)+A_{\mathcal{Z}}(X_n,Z_n,\theta_0)\right]\\ +o_p(N^{-1/2}).$$

Using Eq. (A.25) and letting $R_{\mathcal{Z}}(Z,\theta_0)\equiv E[S_{\theta\pi_{\mathcal{Z}}}(W,\theta_0)|Z]D(Z,\theta_0)$,

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)\xrightarrow{d}\mathcal{N}\left(0,\Im_{\mathcal{Z}}^{-1}+\Im_{\mathcal{Z}}^{-1}\Omega_{\mathcal{Z}}\Im_{\mathcal{Z}}^{-1}\right),$$

where $\Omega_{\mathcal{Z}}=E[R_{\mathcal{Z}}(Z,\theta_0)\mathrm{Var}[E[B(Y)|X,Z]|Z]R_{\mathcal{Z}}(Z,\theta_0)']$, the result in Theorem 1. □

A.3. Theorem 2

A.3.1. Asymptotic properties of $\overline{\pi}_t(\gamma,z)$

Steps as those in Appendix A.2.1 yield the equivalent of Eq. (A.4) in this case,

$$\sup_{\substack{\gamma\in\Gamma\\ z\in\mathcal{Z}}}\left\|\overline{\pi}_t(\gamma,z)-\pi^*_{t_{(s)}}(\gamma,z)\right\|\xrightarrow{p}0,\\ \sup_{\substack{\gamma\in\mathcal{N}_\gamma\\ z\in\mathcal{Z}}}\left\|\overline{\pi}_t(\gamma,z)-\pi^*_t(\gamma,z)\right\|\xrightarrow{p}0, \tag{A.26}$$

with $\sup_{z\in\mathcal{Z}}\|\overline{\pi}_t(\gamma_0,z)-\pi^*(\gamma_0,z)\|\xrightarrow{p}0$ since $\pi^*(\gamma_0,z)=\pi^*_t(\gamma_0,z)$. Next, we look for a linear representation of $\overline{\pi}_t(\gamma,z)-\pi^*_t(\gamma,z)$ in $\mathcal{N}_\gamma\times\mathcal{Z}$. We start with a result analogous to (A.6),

$$0=\overline{\varphi}_t\left(\pi^*_t(\gamma,z);\gamma,z\right)+\nabla_\pi\overline{\varphi}_t\left(\widetilde{\pi}_t(\gamma,z);\gamma,z\right)\\ \times\left[\overline{\pi}_t(\gamma,z)-\pi^*_t(\gamma,z)\right], \tag{A.27}$$

with $\widetilde{\pi}_t(\gamma,z)$ between $\overline{\pi}_t(\gamma,z)$ and $\pi^*_t(\gamma,z)$. Now, let $t^*_{1m}(\gamma_1)=X'_{1m}\eta_1+\sigma_1\pi^*_{2_t}(\gamma,Z_m)$, $t^*_{2m}(\gamma_2)=X'_{2m}\eta_2+\sigma_2\pi^*_{1_t}(\gamma,Z_m)$, $t^*_{1m}(\gamma_1,z)=X'_{1m}\eta_1+\sigma_1\pi^*_{2_t}(\gamma,z)$ and $t^*_{2m}(\gamma_2,z)=X'_{2m}\eta_2+\sigma_2\pi^*_{1_t}(\gamma,z)$, with $t^*_m(\gamma)=(t^*_{1m}(\gamma_1),t^*_{2m}(\gamma_2))'$ and $t^*_m(\gamma,z)=(t^*_{1m}(\gamma_1,z),t^*_{2m}(\gamma_1,z))'$. Note that $t^*_m(\gamma_1,Z_m)=t^*_m(\gamma)$ and $t^*_m(\gamma_0)=t_m(\gamma_0)$. Take $\overline{P}_{Y|t}(1,1|\tau;\gamma)$ as in Eq. (22) and re-express it as $\overline{P}_{Y|t}(1,1|t^*_n(\gamma,z);\gamma)=\overline{R}_{Y|t}(1,1|t^*_n(\gamma,z);\gamma)/\overline{f}_t(t^*_n(\gamma,z);\gamma)$, where $\overline{R}_{Y|t}(1,1|t^*_n(\gamma,z);\gamma)\equiv\frac{1}{Nh^2}\sum_{m=1}^{N}Y_{1m}Y_{2m}K_h(\overline{t}_m(\gamma)-t^*_n(\gamma,z))$ and $\overline{f}_t(t^*_n$

$(\gamma, z); \gamma) = (Nh^2)^{-1} \sum_{m=1}^{N} K_h(\bar{t}_m(\gamma) - t_n^*(\gamma, z))$. Our assumptions yield

$$\begin{aligned}
\overline{R}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma) &= \frac{1}{Nh^2} \sum_{m=1}^{N} Y_{1m}Y_{2m}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&+ \frac{\sigma_1}{Nh^3} \sum_{m=1}^{N} Y_{1m}Y_{2m}\nabla_{t_1}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&\times \left[\overline{\pi}_2(Z_m) - \pi_2^*(\gamma_0, Z_m)\right] \\
&+ \frac{\sigma_2}{Nh^3} \sum_{m=1}^{N} Y_{1m}Y_{2m}\nabla_{t_2}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&\times \left[\overline{\pi}_1(Z_m) - \pi_1^*(\gamma_0, Z_m)\right] + \overline{\xi}^{A}_{11_N}(z, \gamma) \qquad (A.28)
\end{aligned}$$

where $\sup_{z\in\mathbb{S}(Z), \gamma\in\Gamma} |\overline{\xi}^{A}_{11_N}(z, \gamma)| = O_p(N^{\delta-1}b^{-L}h^{-4})$ for any $\delta > 0$. Invoking either Lemma 3 in CH or Theorem A.1, we have

$$\begin{aligned}
\overline{\pi}_1(z) - \pi_1^*(\gamma_0, z) &= \frac{1}{Nb^L} \sum_{m=1}^{N} \frac{Y_{1m}Y_{2m} - \Pr(Y_1 = 1, Y_2 = 1|z)}{\Pr(Y_2 = 1|z)f_Z(z)} \\
&\times H_b(Z_m - z) - \frac{\pi_1^*(\gamma_0, z)}{\Pr(Y_2 = 1|z)} \frac{1}{Nb^L} \sum_{m=1}^{N} \frac{Y_{2m} - \Pr(Y_2 = 1|z)}{f_Z(z)} \\
&\times H_b(Z_m - z) + \overline{\xi}_{1_N}(z) \\
\overline{\pi}_2(z) - \pi_2^*(\gamma_0, z) &= \frac{1}{Nb^L} \sum_{m=1}^{N} \frac{Y_{1m}Y_{2m} - \Pr(Y_1 = 1, Y_2 = 1|z)}{\Pr(Y_1 = 1|z)f_Z(z)} \\
&\times H_b(Z_m - z) - \frac{\pi_2^*(\gamma_0, z)}{\Pr(Y_1 = 1|z)} \frac{1}{Nb^L} \sum_{m=1}^{N} \frac{Y_{1m} - \Pr(Y_1 = 1|z)}{f_Z(z)} \\
&\times H_b(Z_m - z) + \overline{\xi}_{2_N}(z)
\end{aligned}$$

$\sup_{z\in\mathbb{S}(Z)} |\overline{\xi}_{p_N}(z)| = O_p(N^{\delta-1}b^{-L})$ for any $\delta > 0$. This representation result holds *uniformly over* $\mathbb{S}(Z)$, due to the conditions in Assumption B2(i). Using the previous result we obtain

$$\begin{aligned}
&\frac{1}{Nh^3} \sum_{m=1}^{N} Y_{1m}Y_{2m}\nabla_{t_1}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&\quad\times \left[\overline{\pi}_2(Z_m) - \pi_2^*(\gamma_0, Z_m)\right] \\
&= \frac{1}{N^2h^3b^L} \sum_{m=1}^{N}\sum_{\ell=1}^{N} Y_{1m}Y_{2m}\nabla_{t_1}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&\quad\times \left(\frac{Y_{1\ell}Y_{2\ell} - \Pr(Y_1 = 1, Y_2 = 1|Z_m)}{\Pr(Y_1 = 1|Z_m)f_Z(Z_m)} - \frac{\pi_2^*(\gamma_0, Z_m)}{\Pr(Y_1 = 1|Z_m)}\right. \\
&\quad\times \left.\frac{[Y_{1\ell} - \Pr(Y_1 = 1|Z_m)]}{f_Z(Z_m)}\right) H_b(Z_\ell - Z_m) + \overline{\xi}^{B}_{11_N}(z, \gamma)
\end{aligned}$$

where $\sup_{z\in\mathbb{S}(Z), \gamma\in\Gamma} |\overline{\xi}^{B}_{11_N}(z, \gamma)| = O_p(N^{\delta-1}b^{-L}h^{-3})$ for any $\delta > 0$. We can express the right-hand side as a symmetric, second-order U-statistic which satisfies all the assumptions of Lemma A.3 in Ahn and Powell (1993) uniformly in $\Gamma \times \mathbb{S}(Z)$. Let $\mu_{Y_1Y_2}(\cdot)$ and $f_{t|z}(\cdot)$ be as defined in Eq. (18). The projection of this U-statistic conditional on W_n is

$$\begin{aligned}
&\frac{1}{Nh^3} \sum_{m=1}^{N} Y_{1m}Y_{2m}\nabla_{t_1}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&\quad\times \left[\overline{\pi}_2(Z_m) - \pi_2^*(\gamma_0, Z_m)\right] \\
&= \frac{1}{N} \sum_{\ell=1}^{N} \frac{\nabla_{t_1}\left[\mu_{Y_1Y_2}\left(t_n^*(\gamma, z), Z_\ell; \gamma\right) \cdot f_{t|z}\left(t_n^*(\gamma, z)|Z_\ell; \gamma\right)\right]}{\Pr(Y_1 = 1|Z_\ell)} \\
&\quad\times \left(\pi_2^*(\gamma_0, Z_\ell)\left[Y_{1\ell} - \Pr(Y_1 = 1|Z_\ell)\right]\right. \\
&\quad\left.- \left[Y_{1\ell}Y_{2\ell} - \Pr(Y_1 = 1, Y_2 = 1|Z_\ell)\right]\right) + \overline{\xi}^{C}_{11_N}(z, \gamma)
\end{aligned}$$

where $\sup_{z\in\mathbb{S}(Z), \gamma\in\Gamma} |\overline{\xi}^{C}_{11_N}(z, \gamma)| = O_p(N^{\delta-1}b^{-L}h^{-3}) + O(b^{\overline{M}}/h^3) + o_p(N^{-1/2})$ for any $\delta > 0$. $\overline{M}$ is as described in Assumption B4. A parallel result holds for the third term on the right-hand side of Eq. (A.28). Combining both results (A.28) becomes

$$\begin{aligned}
\overline{R}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma) &= \frac{1}{Nh^2} \sum_{m=1}^{N} Y_{1m}Y_{2m}K_h\left(t_m(\gamma) - t_n^*(\gamma, z)\right) \\
&+ \frac{\sigma_1}{N} \sum_{\ell=1}^{N} \frac{\nabla_{t_1}\left[\mu_{Y_1Y_2}\left(t_n^*(\gamma, z), Z_\ell; \gamma\right) \cdot f_{t|z}\left(t_n^*(\gamma, z)|Z_\ell; \gamma\right)\right]}{\Pr(Y_1 = 1|Z_\ell)} \\
&\times \left(\pi_2^*(\gamma_0, Z_\ell)\left[Y_{1\ell} - \Pr(Y_1 = 1|Z_\ell)\right]\right. \\
&\left.- \left[Y_{1\ell}Y_{2\ell} - \Pr(Y_1 = 1, Y_2 = 1|Z_\ell)\right]\right) \\
&+ \frac{\sigma_2}{N} \sum_{\ell=1}^{N} \frac{\nabla_{t_2}\left[\mu_{Y_1Y_2}\left(t_n^*(\gamma, z), Z_\ell; \gamma\right) \cdot f_{t|z}\left(t_n^*(\gamma, z)|Z_\ell; \gamma\right)\right]}{\Pr(Y_2 = 1|Z_\ell)} \\
&\times \left(\pi_1^*(\gamma_0, Z_\ell)\left[Y_{2\ell} - \Pr(Y_2 = 1|Z_\ell)\right]\right. . \\
&\left.- \left[Y_{1\ell}Y_{2\ell} - \Pr(Y_1 = 1, Y_2 = 1|Z_\ell)\right]\right) + \overline{\xi}^{D}_{11_N}(z, \gamma) \qquad (A.29)
\end{aligned}$$

where $\overline{\xi}^{D}_{11_N}(z, \gamma)$ has the same properties as $\overline{\xi}^{C}_{11_N}(z, \gamma)$, described above. This is the linear representation for $\overline{R}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma)$ conditional on W_n. Take $\tau \in \mathbb{R}^2$ and let

$$\begin{aligned}
\widetilde{f}_t(\tau; \gamma) &= \frac{1}{Nh^2} \sum_{m=1}^{N} K_h\left(t_m(\gamma) - \tau\right), \\
\widetilde{P}_{Y|t}(1, 1|\tau; \gamma) &= \frac{1}{\widetilde{f}_t(\tau; \gamma)} \frac{1}{Nh^2} \sum_{m=1}^{N} Y_{1m}Y_{2m}K_h\left(t_m(\gamma) - \tau\right).
\end{aligned} \qquad (A.30)$$

$\widetilde{P}_{Y|t}(1, 1|\tau; \gamma)$ is the nonparametric estimator for $P_{Y|t}(1, 1|\tau; \gamma)$ *if we knew* $\pi^*(\gamma_0, \cdot)$. Define the term in Box III, where all the objects involved are as defined in Eqs. (18) and (19). After combining (A.29) with the corresponding representation of $\overline{f}_t(t_n^*(\gamma, z); \gamma)$ conditional on W_n, we obtain

$$\begin{aligned}
\overline{P}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma) &= \widetilde{P}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma) \\
&+ \frac{1}{N} \sum_{\ell=1}^{N} \frac{R_{11}(t_n^*(\gamma, z), Z_\ell; \gamma)'\Psi(Y_\ell, Z_\ell)}{f_t(t_n^*(\gamma, z); \gamma)} + \overline{\zeta}_{11_N}(z, \gamma), \qquad (A.31)
\end{aligned}$$

where $\sup_{z\in\mathbb{S}(Z), \gamma\in\Gamma} |\overline{\zeta}_{11_N}(z, \gamma)| = O_p(N^{\delta-1}b^{-L}h^{-4}) + O(b^{\overline{M}}/h^3) + o_p(N^{-1/2})$ for any $\delta > 0$. This is the linear representation for $\overline{P}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma)$ conditional on W_n. The equivalent representations for the remaining estimators described in Eq. (22) are (for $p = 1, 2$):

$$\begin{aligned}
\overline{P}_{Y_p|t}(1|t_n^*(\gamma, z); \gamma) &= \widetilde{P}_{Y_p|t}(1|t_n^*(\gamma, z); \gamma) \\
&+ \frac{1}{N} \sum_{\ell=1}^{N} \frac{R_p(t_n^*(\gamma, z), Z_\ell; \gamma)'\Psi(Y_\ell, Z_\ell)}{f_t(t_n^*(\gamma, z); \gamma)} + \overline{\zeta}_{p_N}(z, \gamma),
\end{aligned}$$

where $\widetilde{P}_{Y_p|t}(1|\tau; \gamma) = (\widetilde{f}_t(\tau; \gamma)Nh^2)^{-1} \sum_{m=1}^{N} Y_{pm}K_h(t_m(\gamma) - \tau)$, and $R_p(t_n^*(\gamma, z), Z_\ell; \gamma)$ is given by (A.32) in Box IV where the objects involved are defined in Eqs. (18) and (19). Using (23) and (A.31), we have

$$\begin{aligned}
&\overline{\delta}_{1,2|t}\left(\pi_t^*(\gamma, z); \gamma, z\right) \\
&= \frac{1}{\widehat{f}_Z(z)} \frac{1}{Nb^L} \sum_{n=1}^{N} \widetilde{P}_{Y|t}(1, 1|t_n^*(\gamma, z); \gamma)H_b(Z_n - z) + \frac{1}{\widehat{f}_Z(z)} \\
&\times \frac{1}{N^2b^L} \sum_{\ell=1}^{N}\sum_{n=1}^{N} \frac{R_{11}\left(t_n^*(\gamma, z), Z_\ell; \gamma\right)'\Psi(Y_\ell, Z_\ell)}{f_t(t_n^*(\gamma, z); \gamma)} \\
&\times H_b(Z_n - z) + \widehat{\xi}^{A}_{11_N}(\gamma, z), \qquad (A.33)
\end{aligned}$$

$$R_{11}(t_n^*(\gamma,z),Z_\ell;\gamma)=\begin{pmatrix}\sigma_2\nabla_{t_2}\left[\mu_{Y_1Y_2}\left(t_n^*(\gamma,z),Z_\ell;\gamma\right)\cdot f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\right]-\sigma_2 P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)\nabla_{t_2}f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\\ \sigma_1\nabla_{t_1}\left[\mu_{Y_1Y_2}\left(t_n^*(\gamma,z),Z_\ell;\gamma\right)\cdot f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\right]-\sigma_1 P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)\nabla_{t_1}f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\end{pmatrix}$$

Box III.

$$R_p(t_n^*(\gamma,z),Z_\ell;\gamma)=\begin{pmatrix}\sigma_2\nabla_{t_2}\left[\mu_{Y_p}\left(t_n^*(\gamma,z),Z_\ell;\gamma\right)\cdot f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\right]-\sigma_2 P_{Y_p|t}(1|t_n^*(\gamma,z);\gamma)\nabla_{t_2}f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\\ \sigma_1\nabla_{t_1}\left[\mu_{Y_p}\left(t_n^*(\gamma,z),Z_\ell;\gamma\right)\cdot f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\right]-\sigma_1 P_{Y_p|t}(1|t_n^*(\gamma,z);\gamma)\nabla_{t_1}f_{t|z}\left(t_n^*(\gamma,z)|Z_\ell;\gamma\right)\end{pmatrix}\quad (A.32)$$

Box IV.

where $\sup_{z\in\mathbb{S}(Z),\gamma\in\Gamma}|\widehat{\xi}^A_{11_N}(\gamma,z)|=O_p(N^{\delta-1}b^{-L}h^{-4})+O(b^{\overline{M}}/h^3)+o_p(N^{-1/2})$ for any $\delta>0$. Ignore for a moment the term $1/\widehat{f}_Z(z)$. The second term on the right hand side of (A.33) can be expressed as a symmetric, second-order U-statistic, with projection

$$\begin{aligned}&\frac{1}{N^2b^L}\sum_{\ell=1}^N\sum_{n=1}^N\frac{R_{11}\left(t_n^*(\gamma,z),Z_\ell;\gamma\right)'\Psi(Y_\ell,Z_\ell)}{f_t(t_n^*(\gamma,z);\gamma)}H_b(Z_n-z)\\&=\frac{1}{N}\sum_{n=1}^N\left(\int R_{11}\left(t,Z_n;\gamma\right)f_{z|t}(z|t;\gamma)\mathrm{d}t\right)'\Psi(Y_n,Z_n)\\&\quad+\widehat{\xi}^B_{11_N}(\gamma,z)\\&\equiv\frac{1}{N}\sum_{n=1}^N\phi_{11}\left(Z_n,z;\gamma\right)'\Psi(Y_n,Z_n)+\widehat{\xi}^B_{11_N}(\gamma,z),\end{aligned}$$

$$\text{with } \phi_{11}\left(Z_n,z;\gamma\right)\equiv\int R_{11}\left(t,Z_n;\gamma\right)f_{z|t}(z|t;\gamma)\mathrm{d}t,\quad (A.34)$$

where $\sup_{z\in\mathbb{S}(Z),\gamma\in\Gamma}|\widehat{\xi}^B_{11_N}(\gamma,z)|=O(b^{\overline{M}})$. The conditional density $f_{z|t}(\cdot)$ is as defined in Assumption B2(ii). Assumption B2 ensures that $\phi_{11}(Z_n,z;\gamma)$ is well-defined for all n.[26] Let $\widetilde{P}_{Y|t}(1,1|\tau;\gamma)$ be as defined in (A.30). Given our assumptions, either Lemma 3 in CH or Theorem A.1 yield

$$\begin{aligned}&\widetilde{P}_{Y|t}(1,1|\tau;\gamma)-P_{Y|t}(1,1|\tau;\gamma)\\&=\frac{1}{Nh^2}\sum_{m=1}^N\frac{Y_{1m}Y_{2m}-P_{Y|t}(1,1|\tau;\gamma)}{f_t(\tau;\gamma)}\\&\quad\times K_h(t_m(\gamma)-\tau)+\widetilde{\zeta}_{11_N}(\tau;\gamma),\end{aligned}$$

where $\sup_{\tau\in\mathbb{S}(t(\gamma)),\gamma\in\Gamma}|\widetilde{\zeta}_{11_N}(\tau;\gamma)|=O_p(N^{\delta-1}h^{-2})$ for any $\delta>0$.

$$\begin{aligned}&\frac{1}{Nb^L}\sum_{n=1}^N\widetilde{P}_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)H_b(Z_n-z)\\&=\frac{1}{Nb^L}\sum_{n=1}^N P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)H_b(Z_n-z)\\&\quad+\frac{1}{N^2h^2b^L}\sum_{n=1}^N\sum_{m=1}^N\frac{\left[Y_{1m}Y_{2m}-P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)\right]}{f_t\left(t_n^*(\gamma,z);\gamma\right)}\\&\quad\times K_h\left(t_m(\gamma)-t_n^*(\gamma,z)\right)H_b(Z_n-z)+\widehat{\xi}^C_{11_N}(\gamma,z),\end{aligned}\quad (A.35)$$

where $\sup_{z\in\mathbb{S}(Z),\gamma\in\Gamma}|\widehat{\xi}^C_{11_N}(\gamma,z)|=O_p(N^{\delta-1}b^{-L}h^{-4})+O(b^{\overline{M}}/h^3)+o_p(N^{-1/2})$ for any $\delta>0$. The second term on the right-hand side of Eq. (A.35) can be expressed as a second-order symmetric U-statistic, whose projection is given by

$$\begin{aligned}&\frac{1}{N^2h^2b^L}\sum_{n=1}^N\sum_{m=1}^N\frac{\left[Y_{1m}Y_{2m}-P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)\right]}{f_t\left(t_n^*(\gamma,z);\gamma\right)}\\&\quad\times K_h\left(t_m(\gamma)-t_n^*(\gamma,z)\right)H_b(Z_n-z)\\&=\frac{1}{N}\sum_{n=1}^N\left[Y_{1n}Y_{2n}-P_{Y|t}(1,1|t_n(\gamma);\gamma)\right]f_{z|t}\left(z|t_n(\gamma)\right)\\&\quad+\widehat{\xi}^D_{11_N}(\gamma,z),\end{aligned}$$

where $\sup_{z\in\mathbb{S}(Z),\gamma\in\Gamma}|\widehat{\xi}^D_{11_N}(\gamma,z)|=O(h^{\overline{M}}/b^L)+O((b/h^2)^{\overline{M}})+o_p(N^{-1/2})$. Throughout, we have used the fact that $t_n^*(\gamma,z)=t_n^*(\gamma)+(\sigma_1[\pi^*_{2_t}(\gamma,z)-\pi^*_{2_t}(\gamma,Z_n)],\sigma_2[\pi^*_{1_t}(\gamma,z)-\pi^*_{1_t}(\gamma,Z_n)])'$ to compute the relevant expectations. Combining Eqs. (A.34) and (A.35), Eq. (A.33) becomes

$$\begin{aligned}&\overline{\delta}_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right)=\frac{1}{\widehat{f}_Z(z)}\frac{1}{Nb^L}\sum_{n=1}^N P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)\\&\quad\times H_b(Z_n-z)+\frac{1}{N}\sum_{n=1}^N\frac{\left[Y_{1n}Y_{2n}-P_{Y|t}(1,1|t_n(\gamma);\gamma)\right]}{f_Z(z)}f_{z|t}\\&\quad\times\left(z|t_n(\gamma)\right)+\frac{1}{N}\sum_{n=1}^N\frac{\phi_{11}(Z_n,z;\gamma)'\Psi(Y_n,Z_n)}{f_Z(z)}+\widehat{\xi}^E_{11_N}(\gamma,z),\end{aligned}\quad (A.36)$$

where $\sup_{z\in\mathbb{S}(Z),\gamma\in\Gamma}|\widehat{\xi}^E_{11_N}(\gamma,z)|=\left(|\widehat{\xi}^C_{11_N}(\gamma,z)|+|\widehat{\xi}^D_{11_N}(\gamma,z)|\right)\times\left(1+O_p\left(\sqrt{N^{\delta-1}b^{-L}}\right)\right)+O_p\left(\sqrt{N^{\delta-2}b^{-L}}\right)$ for any $\delta>0$. For all $(\gamma,z)\in\mathcal{N}_\gamma\times\mathcal{Z}$, we have

$$\begin{aligned}&E\left[P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)|Z_n=z\right]=E\left[P_{Y|t}(1,1|t_n^*(\gamma);\gamma)|Z_n=z\right]\\&=\delta_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right).\end{aligned}$$

Applying Theorem A.1, the first term on the right-hand side of Eq. (A.35) satisfies

$$\begin{aligned}&\frac{1}{\widehat{f}_Z(z)}\frac{1}{Nb^L}\sum_{n=1}^N P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)H_b(Z_n-z)\\&=\delta_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right)\\&\quad+\frac{1}{N}\sum_{n=1}^N\frac{\left[P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)-\delta_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right)\right]}{f_Z(z)}\\&\quad\times H_b(Z_n-z)+\widehat{\xi}^F_{11_N}(\gamma,z),\end{aligned}$$

where $\sup_{z\in\mathcal{Z},\gamma\in\mathcal{N}_\gamma}|\widehat{\xi}^F_{11_N}(\gamma,z)|=O_p(N^{\delta-1}b^{-L})$ for any $\delta>0$. From above, (A.36) becomes

$$\begin{aligned}&\overline{\delta}_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right)=\delta_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right)\\&\quad+\frac{1}{N}\sum_{n=1}^N\frac{\left[P_{Y|t}(1,1|t_n^*(\gamma,z);\gamma)-\delta_{1,2|t}\left(\pi_t^*(\gamma,z);\gamma,z\right)\right]}{f_Z(z)}\end{aligned}$$

26 The condition $f_t(t;\gamma)>\underline{f}_t>0$ for all $t\in\mathbb{S}(t(\gamma))$ uniformly over Γ in Assumption B2 ensures that this integral is well-defined.

$$\times H_b(Z_n - z) + \frac{1}{N}\sum_{n=1}^{N}\frac{\left[Y_{1n}Y_{2n} - P_{Y|t}(1,1|t_n(\gamma);\gamma)\right]}{f_Z(z)}f_{z|t}\left(z|t_n(\gamma)\right) + \frac{1}{N}\sum_{n=1}^{N}\frac{\phi_{11}(Z_n,z;\gamma)'\Psi(Y_n,Z_n)}{f_Z(z)} + \nu_{1,2_N}(\gamma,z), \quad (A.37)$$

with $\nu_{1,2_N}(\gamma,z) = \widehat{\xi}^E_{11_N}(\gamma,z) + \widehat{\xi}^F_{11_N}(\gamma,z)$ for all $(\gamma,z)\in\mathcal{N}_\gamma\times\mathcal{Z}$. Parallel steps yield equivalent expressions for $\overline{\delta}_{p|t}(\pi^*_t(\gamma,z);\gamma,z)$, for $p=1,2$:

$$\overline{\delta}_{p|t}\left(\pi^*_t(\gamma,z);\gamma,z\right) = \delta_{p|t}\left(\pi^*_t(\gamma,z);\gamma,z\right) + \frac{1}{N}\sum_{n=1}^{N}\frac{\left[P_{Y_p|t}(1|t^*_n(\gamma,z);\gamma) - \delta_{p|t}\left(\pi^*_t(\gamma,z);\gamma,z\right)\right]}{f_Z(z)} \times H_b(Z_n-z) + \frac{1}{N}\sum_{n=1}^{N}\frac{\left[Y_{p_n} - P_{Y_p|t}(1|t_n(\gamma);\gamma)\right]}{f_Z(z)}f_{z|t}\left(z|t_n(\gamma)\right) + \frac{1}{N}\sum_{n=1}^{N}\frac{\phi_p(Z_n,z;\gamma)'\Psi(Y_n,Z_n)}{f_Z(z)} + \nu_{p_N}(\gamma,z), \quad (A.38)$$

where $\phi_p(Z_n,z;\gamma)\equiv\int R_p(t,Z_n;\gamma)f_{z|t}(z|t;\gamma)\mathrm{d}t$, and $R_p(t,Z_n;\gamma)$ as in Eq. (A.32). We are now ready to present the linear representation for $\overline{\pi}_t(\gamma,z)$ that holds uniformly in $\mathcal{N}_\gamma\times\mathcal{Z}$.

Lemma A.1. *Take any $(\gamma,z)\in\mathcal{N}_\gamma\times\mathcal{Z}$. Let $\phi_{11}(\cdot)$ and $\phi_p(\cdot)$ be as defined in Eqs. (A.34) and (A.38) respectively and define the terms shown in Box V.*
As in Theorem 1, let $\underset{3\times1}{B(Y_n)} = (Y_{1n}Y_{2n},Y_{1n},Y_{2n})'$. If our assumptions are satisfied, then

$$\overline{\pi}_t(\gamma,z) - \pi^*_t(\gamma,z) = D(z,\gamma) \times\frac{1}{N}\sum_{n=1}^{N}\left\{\left[\frac{E\left[B(Y_n)|t_n(\gamma)\right] - E\left[E\left[B(Y_n)|t_n(\gamma)\right]|Z_n=z\right]}{b^Lf_Z(z)}\right] \times H_b(Z_n-z) + \left[\frac{B(Y_n) - E\left[B(Y_n)|t_n(\gamma)\right]}{f_Z(z)}\right] \times f_{z|t}\left(z|t_n(\gamma);\gamma\right) + \frac{\phi\left(Z_n,z;\gamma\right)\Psi\left(Y_n,Z_n\right)}{f_Z(z)}\right\} + \overline{\xi}^{\pi}_N(\gamma,z),$$

where $\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\mathcal{N}_\gamma}}|\overline{\xi}^{\pi}_N(\gamma,z)| = [O_p(N^{\delta-1}b^{-L}h^{-4}) + O(b^{\overline{M}}/h^3) + O(h^{\overline{M}}/b^L) + O((b/h^2)^{\overline{M}}) + o_p(N^{-1/2})]\times[1+O_p(\sqrt{N^{\delta-1}b^{-L}})] + O_p(\sqrt{N^{\delta-2}b^{-L}}) + O_p(N^{\delta-1}b^{-L})$ *for any $\delta>0$.*

Proof. Note that $E[B(Y_n)|t_n(\gamma)] = (P_{Y|t}(1,1|t_n(\gamma);\gamma), P_{Y_1|t}(1|t_n(\gamma);\gamma), P_{Y_2|t}(1|t_n(\gamma);\gamma))'$, and $E[E[B(Y_n)|t_n(\gamma)]|Z_n=z] = (\delta_{1,2|t}(\pi^*_t(\gamma,z);\gamma,z), \delta_{1|t}(\pi^*_t(\gamma,z);\gamma,z), \delta_{2|t}(\pi^*_t(\gamma,z);\gamma,z))'$. The result follows from Eqs. (A.26), (A.27), a result equivalent to that of (A.5) – which follows from B3 –, and all the results that led to Eqs. (A.37) and (A.38). □

A.3.2. Proof of Theorem 2

Consistency of $\widehat{\gamma}$ follows from arguments parallel to those of Eqs. (A.12)–(A.15). Given the assumptions of Theorem 2, we can show a result equivalent to Eq. (A.13):

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{N}\sum_{n=1}^{N}\widehat{\ell^t}_{\mathcal{Z}}\left(W_n,\gamma,\overline{\pi}_t(\gamma,Z_n)\right) - E\left[\ell^t_{\mathcal{Z}}\left(W,\gamma,\pi^*_{t_{(s)}}(\gamma,Z)\right)\right]\right| \xrightarrow{p} 0.$$

Let $L^t_{\mathcal{Z}}(W,\gamma) = \exp\{\ell^t_{\mathcal{Z}}(W,\gamma,\pi^*_{t_{(s)}}(\gamma,Z))\}$. Then $\frac{L^t_{\mathcal{Z}}(W,\gamma)}{L^t_{\mathcal{Z}}(W,\gamma_0)} = \frac{\exp\{\ell^t_{\mathcal{Z}}(W,\gamma,\pi^*_{t_{(s)}}(\gamma,Z))\}}{\Pr(Y|X,Z)}$ if $Z\in\mathcal{Z}$ and 1 otherwise; note that $\sum_y\exp\{\ell^t_{\mathcal{Z}}(y,x,z,\gamma,\pi^*_{t_{(s)}}(\gamma,Z))\} = 1$. Assumption B1 implies that $\nexists\gamma\in\Gamma:\gamma\neq\gamma_0$ and $L^t_{\mathcal{Z}}(W,\gamma)/L^t_{\mathcal{Z}}(W,\gamma_0)$ is constant w.p.1. Using the same arguments as in Eqs. (A.14) and (A.15), we arrive at $\widehat{\gamma}\xrightarrow{p}\gamma_0$. Now let $\widehat{S^t}_{\gamma\mathcal{Z}}(W,\gamma) = \frac{\partial\widehat{\ell^t}_{\mathcal{Z}}(W,\gamma,\overline{\pi}_t(\gamma,Z))}{\partial\gamma}$. w.p.a.1, $\widehat{\gamma}$ satisfies $\frac{1}{N}\sum_{n=1}^{N}\widehat{S^t}_{\gamma\mathcal{Z}}(W_n,\widehat{\gamma}) = 0$. We have

$$0 = \frac{1}{N}\sum_{n=1}^{N}\widehat{S^t}_{\gamma\mathcal{Z}}(W_n,\gamma_0) + \left[\frac{1}{N}\sum_{n=1}^{N}\frac{\partial\widehat{S^t}_{\gamma\mathcal{Z}}(W_n,\widetilde{\gamma})}{\partial\gamma'}\right](\widehat{\gamma}-\gamma_0),$$

with $\widetilde{\gamma}$ between $\widehat{\gamma}$ and γ_0. (A.39)

We begin by analyzing $\frac{1}{N}\sum_{n=1}^{N}\widehat{S^t}_{\gamma\mathcal{Z}}(W_n,\gamma_0)$. Let $\phi(Z_n,z;\gamma_0)$ be as in Lemma A.1. Then $\phi(Z_n,z;\gamma_0) = \int R(t,Z_n;\gamma_0)f_{z|t}(z|t;\gamma_0)\mathrm{d}t$. Using Theorem A.1, Lemma A.1 and computing the projections of all the U-statistics involved, we obtain the following linear representation for $\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}_n(\gamma_0)) - P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*_n(\gamma_0))$ *conditional on W_n*:

$$\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}_n(\gamma_0)) - P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*_n(\gamma_0)) = \frac{1}{Nh^2}\sum_{\ell=1}^{N}\frac{\left[Y_{1\ell}Y_{2\ell} - P_{Y|t^*_{\mathcal{Z}}}(1,1|t_n(\gamma_0))\right]}{P_{\mathcal{Z}}(t^*_n(\gamma_0))f_t(t^*_n(\gamma_0))} \times K_h\left(t_\ell(\gamma_0) - t^*_n(\gamma_0)\right)\mathbb{1}\{Z_\ell\in\mathcal{Z}\} + \frac{\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(1,1|t^*_n(\gamma_0)\right)'D(Z_n,\gamma_0)}{f_Z(Z_n)} \times\frac{1}{Nb^L}\sum_{\ell=1}^{N}\Big[E\left[B(Y_\ell)|t_\ell(\gamma_0,Z_n)\right] - E\left[B(Y_\ell)|Z_\ell=Z_n\right]\Big]H_b(Z_\ell-Z_n) - \frac{\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(1,1|t^*_n(\gamma_0)\right)'}{f_t\left(t^*_n(\gamma_0)\right)P_{\mathcal{Z}}(t^*_n(\gamma_0))}\frac{1}{N}\sum_{\ell=1}^{N}\mathbb{1}\{Z_\ell\in\mathcal{Z}\} \times D(Z_\ell,\gamma_0)\Big[E[B(Y_\ell)|t_\ell(\gamma_0)] - E[B(Y_\ell)|Z_\ell]\Big]f_{t|z}\left(t^*_n(\gamma_0)|Z_\ell\right) + \frac{\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(1,1|t^*_n(\gamma_0)\right)'D(Z_n,\gamma_0)}{f_Z(Z_n)} \times\frac{1}{N}\sum_{\ell=1}^{N}\left[B(Y_\ell) - E[B(Y_\ell)|t_\ell(\gamma_0)]\right] \times f_{z|t}\left(Z_n|t_\ell(\gamma_0)\right) - \frac{\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(1,1|t^*_n(\gamma_0)\right)'}{f_t\left(t^*_n(\gamma_0)\right)P_{\mathcal{Z}}(t^*_n(\gamma_0))} \times\frac{1}{N}\sum_{\ell=1}^{N}E_Z\Big[\mathbb{1}\{Z\in\mathcal{Z}\} \times D(Z,\gamma_0)f_{t|z}(t^*_n(\gamma_0)|Z)|t(\gamma_0)=t_\ell(\gamma_0)\Big] \times\left[B(Y_\ell) - E[B(Y_\ell)|X_\ell,Z_\ell]\right] + \frac{\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(1,1|t^*_n(\gamma_0)\right)'D(Z_n,\gamma_0)}{f_Z(Z_n)} \times\frac{1}{N}\sum_{\ell=1}^{N}\phi(Z_\ell,Z_n;\gamma_0)\Psi(Y_\ell,Z_\ell) - \frac{\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(1,1|t^*_n(\gamma_0)\right)'}{f_t\left(t^*_n(\gamma_0)\right)P_{\mathcal{Z}}(t^*_n(\gamma_0))}$$

$$\underset{3\times 2}{\phi}(Z_n, z; \gamma) = (\phi_{11}(Z_n, z; \gamma), \phi_1(Z_n, z; \gamma), \phi_2(Z_n, z; \gamma))'$$

$$\underset{2\times 3}{V}(z,\gamma) = \begin{pmatrix} \dfrac{1}{\delta_{2|t}(\pi_t^*(\gamma, z); \gamma, z)} & 0 & -\dfrac{\delta_{1,2|t}(\pi_t^*(\gamma, z); \gamma, z)}{\delta_{2|t}(\pi_t^*(\gamma, z); \gamma, z)^2} \\ \dfrac{1}{\delta_{1|t}(\pi_t^*(\gamma, z); \gamma, z)} & -\dfrac{\delta_{1,2|t}(\pi_t^*(\gamma, z); \gamma, z)}{\delta_{1|t}(\pi_t^*(\gamma, z); \gamma, z)^2} & 0 \end{pmatrix}$$

$$\underset{2\times 3}{D}(z,\gamma) = \nabla_\pi \varphi_t\left(\pi_t^*(\gamma, z); \gamma, z\right)^{-1} V(z,\gamma)$$

Box V.

$$\times \frac{1}{N}\sum_{\ell=1}^{N}\left\{\int E_Z\left[\mathbb{1}\{Z \in \mathcal{Z}\}D(Z,\gamma_0)f\left(t_n^*(\gamma_0)|Z\right)|t(\gamma_0)\right]\right.$$
$$\left.\times R(t(\gamma_0), Z_\ell)\mathrm{d}t(\gamma_0)\right\}\Psi(Y_\ell, Z_\ell) + \widehat{\zeta}_{11_N}(W_n),$$

where

$$\sup_{W_n:Z_n\in\mathcal{Z}}\left|\widehat{\zeta}_{11_N}(W_n)\right| = o_p(N^{-1/2}) + O_p\left(N^{\delta-1}b^{-L}h^{-4}\right)$$
$$+ O_p\left(N^{\delta-1}h^{-2}\right) \quad \forall\delta > 0.$$

Equivalent expressions exist for $\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1, 0|\widehat{t}_n(\gamma_0)), \widehat{P}_{Y|t^*_{\mathcal{Z}}}(0, 1|\widehat{t}_n(\gamma_0))$ simply by replacing $P_{Y|t^*_{\mathcal{Z}}}(1, 1|\cdot)$ with $P_{Y|t^*_{\mathcal{Z}}}(1, 0|\cdot)$ and $P_{Y|t^*_{\mathcal{Z}}}(0, 1|\cdot)$ respectively. Let

$$\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma} = \left[\frac{1}{P_{Y|t^*_{\mathcal{Z}}}(1, 1|t^*(\gamma); \gamma)}\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(1, 1|t^*(\gamma); \gamma)}{\partial\gamma}\right]_{\gamma=\gamma_0}\mathbb{1}\{Z \in \mathcal{Z}\}$$

where $\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(1,1|\cdot)}{\partial\gamma}$ is as described in Eq. (27). Let

$$\frac{\partial L_{\mathcal{Z}}(1, 1|t_n^*(\gamma_0))}{\partial\gamma} = \left[\frac{1}{P_{Y|t^*_{\mathcal{Z}}}(1, 1|t_n^*(\gamma); \gamma)}\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(1, 1|t_n^*(\gamma); \gamma)}{\partial\gamma}\right]_{\gamma=\gamma_0}\mathbb{1}\{Z_n \in \mathcal{Z}\}.$$

Using our previous results, the fact that if $Z \in \mathcal{Z}$ then: $E[B(Y)|X, Z] = E[B(Y)|t(\gamma_0)] = E[B(Y)|t(\gamma_0)]$ and $t(\gamma_0) = t^*(\gamma_0)$, and computing the relevant U-statistic projections, algebraic manipulation yields

$$\frac{1}{N}\sum_{n=1}^{N}Y_{1n}Y_{2n}\frac{\partial\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1, 1|\widehat{t}_n(\gamma_0))}{\partial\gamma}\mathbb{1}\{Z_n \in \mathcal{Z}\}$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left[Y_{1n}Y_{2n}\frac{\partial L_{\mathcal{Z}}\left(1, 1|t_n^*(\gamma_0)\right)}{\partial\gamma}\right.$$
$$\left. - E\left[Y_1Y_2\,\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\right| t^*(\gamma_0) = t_n^*(\gamma_0)\right]\Bigg]$$
$$+ \frac{1}{N}\sum_{n=1}^{N}E\left[E\left[\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi\right.\right.$$
$$\left.\left.\times P_{Y|t^*_{\mathcal{Z}}}\left(1, 1|t^*(\gamma_0)\right)'|Z\right]D(Z,\gamma_0)|Z = Z_n\right]$$
$$\times\,[E[B(Y_n)|X_n, Z_n] - E[B(Y_n)|Z_n]]$$
$$- \frac{1}{N}\sum_{n=1}^{N}E\left[E\left[\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi\right.\right.$$
$$\left.\left.\times P_{Y|t^*_{\mathcal{Z}}}\left(1, 1|t^*(\gamma_0)\right)'|t^*(\gamma_0)\right]D(Z,\gamma_0)|Z = Z_n\right]$$
$$\times\,[E[B(Y_n)|X_n, Z_n] - E[B(Y_n)|Z_n]]$$
$$+ \frac{1}{N}\sum_{n=1}^{N}E\left[E\left[\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi\right.\right.$$
$$\left.\left.\times P_{Y|t^*_{\mathcal{Z}}}\left(1, 1|t^*(\gamma_0)\right)'|Z\right]D(Z,\gamma_0)|t^*(\gamma_0) = t_n^*(\gamma_0)\right]$$
$$\times\,[B(Y_n) - E[B(Y_n)|X_n, Z_n]]$$
$$- \frac{1}{N}\sum_{n=1}^{N}E\left[E\left[\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi\right.\right.$$
$$\left.\left.\times P_{Y|t^*_{\mathcal{Z}}}\left(1, 1|t^*(\gamma_0)\right)'|t^*(\gamma_0)\right]D(Z,\gamma_0)|t^*(\gamma_0) = t_n^*(\gamma_0)\right]$$
$$\times\,[B(Y_n) - E[B(Y_n)|X_n, Z_n]]$$
$$+ \frac{1}{N}\sum_{n=1}^{N}E\left[E\left[E\left[\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi\right.\right.\right.$$
$$\left.\times P_{Y|t^*_{\mathcal{Z}}}\left(1, 1|t^*(\gamma_0)\right)'|Z\right]$$
$$\left.\left.\times D(Z,\gamma_0)|t^*(\gamma_0)\right]\nabla_\pi E\left[B(Y)|t^*(\gamma_0)\right]|Z = Z_n\right]$$
$$\times\,\Psi(Y_n, Z_n) - \frac{1}{N}\sum_{n=1}^{N}E\left[E\left[E\left[\frac{\partial L_{\mathcal{Z}}(1, 1|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi\right.\right.\right.$$
$$\left.\left.\times P_{Y|t^*_{\mathcal{Z}}}\left(1, 1|t^*(\gamma_0)\right)'|t^*(\gamma_0)\right]D(Z,\gamma_0)|t^*(\gamma_0)\right]\nabla_\pi$$
$$\left.\times E\left[B(Y)|t^*(\gamma_0)\right]|Z = Z_n\right]\Psi(Y_n, Z_n) + o_p(N^{-1/2}). \tag{A.40}$$

Analogous results exist for $\frac{1}{N}\sum_{n=1}^{N}Y_{1n}(1 - Y_{2n})\frac{\partial\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,0|\widehat{t}_n(\gamma_0))}{\partial\gamma}\mathbb{1}\{Z_n \in \mathcal{Z}\}$, etc. by replacing $P_{Y|t^*_{\mathcal{Z}}}(1, 1|t(\gamma_0))$ and $L_{\mathcal{Z}}(1, 1|t(\gamma_0))$ with their counterparts in each case. Taking the sum over $y \in \{(1, 1), (1, 0), (0, 1), (0, 0)\}$ we have

$$\sum_y E\left[\frac{\partial L_{\mathcal{Z}}(y|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(y|t^*(\gamma_0)\right)|t^*(\gamma_0)\right]$$
$$= -E\left[S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)|t^*(\gamma_0)\right],$$
$$\sum_y E\left[\frac{\partial L_{\mathcal{Z}}(y|t^*(\gamma_0))}{\partial\gamma}\nabla_\pi P_{Y|t^*_{\mathcal{Z}}}\left(y|t^*(\gamma_0)\right)|Z\right]$$
$$= -E\left[S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)|Z\right],$$

with $S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)$ as in Theorem 2. Note that $E[Y_1Y_2|X,Z,Z\in\mathcal{Z}]=P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma_0))$. Thus,

$$E\left[Y_1Y_2\frac{\partial L_{\mathcal{Z}}(1,1|t^*(\gamma_0))}{\partial\gamma}\middle|t^*(\gamma_0),X,Z\right]=\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma_0))}{\partial\gamma}\mathbb{1}\{Z\in\mathcal{Z}\}.$$

The same result holds for $Y_1(1-Y_2)\frac{\partial L_{\mathcal{Z}}(1,0|t^*(\gamma_0))}{\partial\gamma}$, etc. By construction, $\sum_y\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(y|\tau)}{\partial\gamma}=0$. It follows by iterated expectations that

$$E\left[Y_1Y_2\frac{\partial L_{\mathcal{Z}}(1,1|t^*(\gamma_0))}{\partial\gamma}+(1-Y_1)Y_2\frac{\partial L_{\mathcal{Z}}(0,1|t^*(\gamma_0))}{\partial\gamma}+Y_1(1-Y_2)\frac{\partial L_{\mathcal{Z}}(1,0|t^*(\gamma_0))}{\partial\gamma}+(1-Y_1)(1-Y_2)\frac{\partial L_{\mathcal{Z}}(0,0|t^*(\gamma_0))}{\partial\gamma}\middle|t^*(\gamma_0)\right]=0 \tag{A.41}$$

Let

$$\xi_{\mathcal{Z}}(W,\gamma_0)=E\left[S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)|t^*(\gamma_0)\right]-E\left[S_{\gamma\pi_{\mathcal{Z}}}(W,\gamma_0)|Z\right]$$
$$\overline{A}_{\mathcal{Z}}(X_n,Z_n,\gamma_0)=E\left[\xi_{\mathcal{Z}}(W,\gamma_0)|Z=Z_n\right]D(Z_n,\gamma_0)\times\left[E\left[B(Y_n)|X_n,Z_n\right]-E\left[B(Y_n)|Z_n\right]\right]$$
$$\overline{B}_{\mathcal{Z}}(W_n,\gamma_0)=E\Big[E\left[\xi_{\mathcal{Z}}(W,\gamma_0)|Z\right]D(Z,\gamma_0)|t^*(\gamma_0)=t_n^*(\gamma_0)\Big]\times\left[E\left[B(Y_n)|X_n,Z_n\right]-B(Y_n)\right]$$
$$\overline{C}_{\mathcal{Z}}(W_n,\gamma_0)=E\Big[E\left[E\left[\xi_{\mathcal{Z}}(W,\gamma_0)|Z\right]D(Z,\gamma_0)|t^*(\gamma_0)\right]\nabla_\pi\times E\left[B(Y)|t^*(\gamma_0)\right]|Z=Z_n\Big]\Psi(Y_n,Z_n;\gamma_0).$$

Eqs. (A.40) and (A.41) yield

$$\frac{1}{N}\sum_{n=1}^N\widehat{S^t}_{\gamma_{\mathcal{Z}}}(W_n,\gamma_0)=\frac{1}{N}\sum_{n=1}^N\Big[S_{\gamma_{\mathcal{Z}}}(W_n,\gamma_0)+\overline{A}_{\mathcal{Z}}(X_n,Z_n,\gamma_0)+\overline{B}_{\mathcal{Z}}(W_n,\gamma_0)+\overline{C}_{\mathcal{Z}}(W_n,\gamma_0)\Big]+o_p(N^{-1/2}). \tag{A.42}$$

Now let $\widetilde{\gamma}$ be as described in Eq. (A.39). Denote $\mathbb{1}_{\mathcal{Z}_n}\equiv\mathbb{1}\{Z_n\in\mathcal{Z}\}$. The uniform convergence properties of our nonparametric estimators and consistency of $\widehat{\gamma}$ yield

$$\begin{aligned}&\frac{1}{N}\sum_{n=1}^N Y_{1n}Y_{2n}\frac{\partial^2\log\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}(\widetilde{\gamma}))}{\partial\gamma\partial\gamma'}\mathbb{1}\{Z_n\in\mathcal{Z}\}\\&=-\frac{1}{N}\sum_{n=1}^N\frac{Y_{1n}Y_{2n}\mathbb{1}_{\mathcal{Z}_n}}{\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}(\widetilde{\gamma}))^2}\times\frac{\partial\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}(\widetilde{\gamma}))}{\partial\gamma}\frac{\partial\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}(\widetilde{\gamma}))}{\partial\gamma}'\\&\quad+\frac{1}{N}\sum_{n=1}^N\frac{Y_{1n}Y_{2n}\mathbb{1}_{\mathcal{Z}_n}}{\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}(\widetilde{\gamma}))}\frac{\partial^2\widehat{P}_{Y|t^*_{\mathcal{Z}}}(1,1|\widehat{t}(\widetilde{\gamma}))}{\partial\gamma\partial\gamma'}\\&\xrightarrow{p}-E\left[\frac{\mathbb{1}\{Z\in\mathcal{Z}\}}{P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma_0))}\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma_0))}{\partial\gamma}\times\frac{\partial P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma_0))}{\partial\gamma}'\right]\\&\quad+E\left[\mathbb{1}\{Z\in\mathcal{Z}\}\frac{\partial^2P_{Y|t^*_{\mathcal{Z}}}(1,1|t^*(\gamma_0))}{\partial\gamma\partial\gamma'}\right].\end{aligned}$$

By construction, $\sum_y\frac{\partial^2P_{Y|t^*_{\mathcal{Z}}}(y|t^*(\gamma))}{\partial\gamma\partial\gamma'}=0$. The previous result and the definition of $\widehat{S^t}_{\gamma_{\mathcal{Z}}}(W_n,\widetilde{\gamma})$ in Eq. (A.39) yield

$$\frac{1}{N}\sum_{n=1}^N\frac{\partial\widehat{S^t}_{\gamma_{\mathcal{Z}}}(W_n,\widetilde{\gamma})}{\partial\gamma'}\xrightarrow{p}-E\left[S_{\gamma_{\mathcal{Z}}}(W,\gamma_0)S_{\gamma_{\mathcal{Z}}}(W,\gamma_0)'\right]\equiv-\mathcal{I}_{\mathcal{Z}}.$$

Using Assumption B5, (A.39), (A.42) and the above equation, we get

$$\widehat{\gamma}-\gamma_0=\mathcal{I}_{\mathcal{Z}}^{-1}\times\frac{1}{N}\sum_{n=1}^N\Big[S_{\gamma_{\mathcal{Z}}}(W_n,\gamma_0)+\overline{A}_{\mathcal{Z}}(X_n,Z_n,\gamma_0)+\overline{B}_{\mathcal{Z}}(W_n,\gamma_0)+\overline{C}_{\mathcal{Z}}(W_n,\gamma_0)\Big]+o_p(N^{-1/2}).$$

which concludes the proof. □

A.4. Monotonicity of the loci $\varphi_1(\pi_1,\pi_2;\theta,z)=0$, $\varphi_2(\pi_1,\pi_2;\theta,z)=0$

If (6) is satisfied, for each $\pi_2\in[0,1]$ there exists a unique $\pi_1\in[0,1]$ such that $\varphi_1(\pi_1,\pi_2;\theta,z)=0$. The same holds for $\varphi_2(\pi_1,\pi_2;\theta,z)=0$. Therefore, both loci can be depicted as continuous curves in $[0,1]^2$. Let $\mu_p(Z)=E[X_p|Z]$ and express $X_p=\mu_p(Z)+\nu_p$, where $E[\nu_p|Z]=0$. Let $\eta_p(\beta_p)=\varepsilon_p-\nu_p'\beta_p$. Take $(\eta_1,\eta_2)\in\mathbb{R}^2$ and let $H_{1|2}(\eta_1,\eta_2;\beta,z)=\Pr(\eta_1(\beta_1)\le\eta_1|\eta_2(\beta_2)\le\eta_2,Z=z)$ and $H_{2|1}(\eta_2,\eta_1;\beta,z)=\Pr(\eta_2(\beta_2)\le\eta_2|\eta_1(\beta_1)\le\eta_1,Z=z)$. We can re-express $\varphi_1(\pi_1,\pi_2;\theta,z)=\pi_1-H_{1|2}(\mu_1(z)'\beta_1+\alpha_1\pi_2,\mu_2(z)'\beta_2+\alpha_2\pi_1;\beta,z)$ and $\varphi_2(\pi_2,\pi_1;\theta,z)=\pi_2-H_{2|1}(\mu_2(z)'\beta_2+\alpha_2\pi_1,\mu_1(z)'\beta_1+\alpha_1\pi_2;\beta,z)$. Thus, the signs of the slopes of the loci $\varphi_1(\pi_1,\pi_2;\theta,z)=0$ and $\varphi_2(\pi_2,\pi_1;\theta,z)=0$ are given by the signs of $\alpha_1\nabla_{\eta_1}H_{1|2}(\mu_1(z)'\beta_1+\alpha_1\pi_2,\mu_2(z)'\beta_2+\alpha_2\pi_1;\beta,z)$ and $\alpha_2\nabla_{\eta_2}H_{2|1}(\mu_2(z)'\beta_2+\alpha_2\pi_1,\mu_1(z)'\beta_1+\alpha_1\pi_2;\beta,z)$, respectively. These signs are equal to those of α_1 and α_2 respectively, since $\nabla_{\eta_1}H_{1|2}(\eta_1,\eta_2;\beta,z)>0$ and $\nabla_{\eta_2}H_{2|1}(\eta_2,\eta_1;\beta,z)>0\,\forall(\eta_1,\eta_2)\in\mathbb{R}^2,z\in\mathbb{S}(Z),\theta\in\mathbb{R}^{k+3}$.

References

Abrevaya, J., Hausman, J., Khan, S., 2007. Testing for causal effects in a generalized regression model with endogenous regressors. Working Paper. Duke University.

Ahn, H., Manski, C., 1993. Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. Journal of Econometrics 56, 291–321.

Ahn, H., Powell, J., 1993. Semiparametric estimation of censored selection models. Journal of Econometrics 58, 3–29.

Andrews, D., 1994. Empirical process methods in econometrics. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North Holland.

Aradillas-Lopez, A., 2007. Semiparametric efficiency in models with rational expectations and strategic interactions. Working Paper. Princeton University.

Aradillas-Lopez, A., 2007. Pairwise difference estimation of incomplete information games. Working Paper. Princeton University.

Aumann, R., 1987. Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55, 1–18.

Bajari, P., Hong, H., Kreiner, J., Nekipelov, D., 2009. Estimating static models of strategic interaction. Journal of Business and Economic Statistics (forthcoming).

Berry, S., 1992. Estimation of a model of entry in the airline industry. Econometrica 60 (4), 889–917.

Bjorn, P.A., Vuong, Q., 1984. Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. SSWP No. 537. Caltech.

Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. Review of Economic Studies 71, 655–679.

Blundell, R., Smith, R., 1986. An exogeneity test for a simultaneous Tobit model. Econometrica 54, 679–685.

Blundell, R., Smith, R., 1989. Estimation in a class of simultaneous equation limited dependent variable models. Review of Economic Studies 56 (1), 37–57.

Blundell, R., Smith, R., 1994. Coherency and estimation in simultaneous models with censored or qualitative dependent variables. Journal of Econometrics 64, 355–373.

Bresnahan, T., Reiss, P., 1990. Entry in monopoly markets. Review of Economic Studies 57, 531–553.

Bresnahan, T., Reiss, P., 1991. Empirical models of discrete games. Journal of Econometrics 48, 57–81.

Brock, W., Durlauf, S., 2001. Discrete choice with social interactions. The Review of Economic Studies 68, 235–260.

Carneiro, P., Hansen, P., Heckman, J., 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. International Economic Review 44 (2), 273–301.

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

Chamley, C., Gale, D., 1994. Information revelation and strategic delay in a model of investment. Econometrica 62, 1065–1085.

Chesher, A., 2007. Endogeneity and discrete outcomes. CEMMAP Working Paper CWP05/07. The Institute for Fiscal Studies, University College London.

Collomb, G., Hardle, W., 1986. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression from dependent observations. Stochastic Processes and their Applications 23, 77–89.

Cosslett, S., 1987. Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. Econometrica 55 (3), 559–585.

Dagenais, M., 1999. A simultaneous probit model. Cahiers Economiques de Bruxelles 163, 325–346.

Falmagne, J., 1985. Elements of Psychophysical Theory. In: Oxford Psychology Series, vol. 6. Oxford University Press, New York.

Gale, D., Nikaido, H., 1965. The Jacobian matrix and global univalence of mappings. Mathematische Annalen 159, 81–93.

Gourieroux, C., Laffont, J., Montfort, A., 1980. Coherency conditions in simultaneous linear equation models with endogenous regimes. Econometrica 48 (3), 675–695.

Grieco, P., 2010. Discrete choice games with flexible information structure: an application to local grocery markets. Working Paper. Northwestern University.

Han, A., 1987. The the maximum rank correlation estimator. Journal of Econometrics 303–316.

Heckman, J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46, 931–960.

Heckman, J., 1990. Varieties of selection bias. American Economic Review 80 (2), 313–318.

Heckman, J., Honoré, B., 1989. The identifiability of the competing risks model. Biometrika 76 (2), 325–330.

Heckman, J., Honoré, B., 1990. The empirical content of the Roy model. Econometrica 58, 1121–1149.

Heckman, J., Navarro, S., 2007. Dynamic discrete choice and dynamic treatment effects. Journal of Econometrics 136 (2), 341–396.

Huber, P., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability, vol. 1. University of California Press, Berkeley, pp. 221–233.

Ichimura, H., Lee, L.-F., 1991. Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation. Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, p. 3.

Jovanovic, B., Rousseau, P., 2003. The Q-theory of mergers. American Economic Review: Papers and Proceedings 92, 198–204.

Khan, S., Tamer, E., 2009. Irregular identification, support conditions and inverse weight estimation. Working Paper. Northwestern University.

Klein, R., Spady, W., 1993. An efficient semiparametric estimator of binary response models. Econometrica 61, 387–421.

Klein, R., Vella, F., 2009. A semiparametric model for binary response and continuous outcomes under index heteroscedasticity. Journal of Applied Econometrics 24, 735–762.

Lewbel, A., 1997. Semiparametric estimation of location and other discrete choice moments. Econometric Theory 13 (1), 32–51.

Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. Journal of Econometrics 97, 145–177.

Lewbel, A., 2007. Coherency and completeness of structural models. International Economic Review 48 (4), 1379–1392.

Manski, C.F., 1988. Identification of binary response models. Journal of the American Statistical Association 83.

Mas-Collel, A., Whinston, M., Green, J., 1995. Microeconomic Theory. Oxford University Press, Oxford.

Matzkin, R., 1992. Nonparametric and districtuion-free estimation of the binary threshold crossing and the binary choice models. Econometrica 60 (2), 239–270.

Matzkin, R., 1993. Nonparametric identification and estimation of polychotomous choice models. Journal of Econometrics 58, 1339–1375.

Morris, S., Shin, H., 2003. Global games: theory and applications. In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), Advances in Economics and Econometrics. Proceedings of the Eighth World Congress of the Econometric Society. Cambridge University Press, pp. 56–114.

Nelsen, R., 2006. An Introduction to Copulas. Springer, New York.

Newey, W., 1990. Semiparametric efficiency bounds. Journal of Applied Econometrics 5 (2), 99–135.

Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North Holland.

Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic least squares estimators for dynamic games. Review of Economic Studies 75, 901–928.

Powell, J., Stock, J., Stoker, T., 1989. Semiparametric estimation of index coefficients. Econometrica 1403–1430.

Savage, L., 1972. The Foundations of Statistics. Dover Publications.

Seim, K., 2006. An empirical model of firm entry with endogenous product-type choices. Rand Journal of Economics 37 (3).

Sherman, R., 1994a. Maximal inequalities for degenerate U-processes with applications to optimization estimators. Annals of Statistics 22, 439–459.

Sherman, R., 1994b. U-processes in the analysis of a generalized semiparametric regression estimator. Econometric Theory 10, 372–395.

Sickles, R., 1989. An analysis of simultaneous limited dependent variable models and some nonstandard cases. Advances in Statistical Analysis and Statistical Computing 2, 85–122.

Sickles, R., Schmidt, P., 1978. Simultaneous equations models with truncated dependent variables. Journal of Economics and Business 30, 11–21.

Sklar, A., 1959. Fonctions de répartition á n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université de Paris 8, 229–231.

Stoker, T., 1991. Equivalence of direct and indirect estimators of average derivatives. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, pp. 99–118.

Sweeting, A., 2010. The strategic timing incentives of commercial radio stations: an empirical analysis using multiple equilibria. The RAND Journal of Economics 40 (4), 710–742.

Taber, C., 2000. Semiparametric identification and heterogeneity in discrete choice dynamic programming models. Journal of Econometrics 96 (2), 201–229.

Tamer, E.T., 2003. Incomplete bivariate discrete response model with multiple equilibria. Review of Economic Studies 70, 147–167.

Thurstone, L., 1959. The Measurement of Values. University of Chicago Press, Chicago.

Vytlacil, E., Yildiz, N., 2007. Dummy endogenous variables in weakly separable models. Econometrica 75, 757–779.