Using Rationalizable Bounds for Conditional Choice Probabilities as Control Variables

Andres Aradillas-Lopez*

December 2007 (Preliminary. Comments Welcome)

Abstract

This paper analyzes the problem of estimation and inference of the payoff parameters in a static game where players' beliefs are not necessarily correct in a Bayesian-Nash Equilibrium sense, but are consistent with a finite number of rounds of iterated deletion of dominated strategies. Assuming an upper bound on such number, we focus on a type of behavior that produces an exclusion restriction such that the influence of players' unobserved beliefs on their observed behavior is captured entirely by the bounds on rationalizable beliefs. Given our assumptions, these bounds are semiparametrically identified. We present a constructive identification result and analyze the asymptotic properties of an estimation procedure based on it.

1 Introduction

A vast number of econometric identification results for game theoretic models rely on the assumption of equilibrium behavior. In this context, much effort has been devoted to characterizing results which are robust to multiple equilibria, or to finding efficient algorithms capable of finding the set of equilibria in complicated games. Some examples of the econometric analysis of static games with complete information and Nash equilibrium include Bjorn and Vuong (1984), Bresnahan and Reiss (1991), Berry (1994), Tamer (2003), Andrews, Berry, and Jia (2004), Ciliberto and Tamer (2003), and Bajari, Hong, and Ryan (2005) among others. The game we will analyze here is played under incomplete information. Equilibrium games with different elements of incomplete information

*Department of Economics, Princeton University, Princeton NJ 08544. Financial support from the National Science Foundation Grant No. SES−0718409 and the Gregory Chow Econometrics Research Program at Princeton University are gratefully acknowledged.

are analyzed for example in Bajari, Hong, Kreiner, and Nekipelov (2006), Aradillas-Lopez (2006), Aguirregabiria and Mira (2007), Seim (2005), Pakes, Porter, Ho, and Ishii (2005), Berry and Tamer (2006) and McKelvey and Palfrey (1995) among others.

There exists substantial experimental evidence in the field of *behavioral game theory* which suggests that people's behavior often deviates systematically from equilibrium. Some examples include Stahl and Wilson (1994), Stahl and Wilson (1995), Nagel (1995), Ho, Camerer, and Weigelt (1998), Costa-Gomes, Crawford, and Broseta (2001), Costa-Gomes and Crawford (2006) and Crawford and Iriberri (2007). Theorists in this field have focused on *dominance-solvable games*, which are those that can be solved by iterated reasoning. From experimental studies, behavioral game theorists have found evidence which suggests that people tend to do only a few steps of iterated reasoning either because further steps are too complicated for them or because they believe they would be too complicated for their opponents. Such evidence suggests that people may fail to play an equilibrium strategy even if they are capable of understanding the iterated reasoning process that would eventually lead to it. These models depart from equilibrium behavior by dropping the assumption that each player has a perfect model of others' decisions and replacing it with the assumption that such subjective models survive k rounds of iterated elimination of dominated decisions. Thus, each player's subjective model about others' behavior is consistent with level-k interim rationalizability in the sense of Bernheim (1984). Their identification strategy is to assume the existence of a small number of types. One of those types, for example, would be a player who performs 2 rounds of iterated deletion of dominated actions, and best-responds to a uniform distribution over the space of actions that remain. Overwhelmingly, all these papers have rejected behavior that would be compatible with five or more rounds of iterated reasoning.

Partial and point-identification results for 2×2 discrete games, as well as first-price auctions based *exclusively* on the bounds implied by Level-k rationalizability are studied in Aradillas-Lopez and Tamer (2007). In this paper, we consider a type of behavior that encompasses the types analyzed in the aforementioned behavioral game studies. Here, we will refer to a player who maximizes his expected utility (given his information set) for any set of arbitrary beliefs as L_1-rational. We will refer to an L_k-rational player as one whose beliefs are consistent with $k-1$ mental rounds of deletion of dominated actions. We will assume a population of players who perform $k-1$ rounds of deletion of dominated actions (i.e, their beliefs correspond to those of an L_k-rational player) and best-respond to an (unknown) convex combination of the resulting upper and lower L_k-bounds for their opponents' expected actions. We will assume that the players in our population behave like this for some $k = 2, \ldots, k^*$ for a pre-specified k^*. The existing body of experimental evidence could be used to pick a conservative value of k^*. We develop an estimation methodology for an incomplete-information static game with $\mathcal{P} \geq 2$ players, each of which can take two actions. The parameters of interest will be those of the normal-form payoffs. We describe a set of semiparametric assumptions that allow us to identify the upper and lower L_k-bounds for opponents' expected behavior. We estimate these bounds semiparametrically through an iterative procedure, starting with $k = 2$. Using an exclusion restriction implied by the family of decision rules we consider, we present a constructive identification result and an accompanying estimation procedure. We show consistency of our estimator and outline its asymptotic distribution properties.

The paper proceeds as follows. Section 2 describes the game we will analyze, as well as some of the notation we will use throughout the paper. Section 3 studies the relevant implications of L_k-rationality in our game and characterizes the upper and lower bounds for opponents' expected behavior by any L_k-rational player. Section 4 defines $D_{k^*}-$rationality, the space of decision rules which we will assume describe the actions of the players in our population. An identification result based on $D_{k^*}-$rationality is also described there. Section 6 is devoted to estimation. First, it describes how the L_k-bounds will be estimated semiparametrically for any value of the parameter of interest. Then, it uses the constructive identification results from Section 4 to design an estimation procedure. The asymptotic properties of the estimator are described there. Section 7 concludes. Short proofs are included in the body of the paper. Longer derivations and proofs can be found in the mathematical appendix.

2 A $\mathcal{P} \times 2$ simultaneous game

2.1 Space of actions and payoffs

A group of $\mathcal{P}$ players, labeled $p = 1, \ldots, \mathcal{P}$ simultaneously choose among two available actions, labeled $Y_p \in \{0, 1\}$. Simultaneous choices are defined in the usual sense that each player must commit to an action before observing those actions chosen by any of his opponents. We will assume that each player p secures a payoff of zero with probability one if he chooses $Y_p = 0$. To be precise, we consider a parameterization of payoffs of the form

$$Y_p \times \left(X_p' \beta_p + \sum_{M=0}^{\mathcal{P}-1} \Delta_p^M \mathbb{Y}_{-p}^M - \varepsilon_p \right), \quad \text{where} \quad \mathbb{Y}_{-p}^M = \mathbb{1}\left\{ \sum_{q \neq p} Y_q = M \right\}, \tag{1}$$

where X_p denotes a vector of observable (to the econometrician) covariates, ε_p denotes an unobservable (to the econometrician) continuously distributed shock, and $(\beta_p, \Delta_p^0, \ldots, \Delta_p^{\mathcal{P}-1}) \equiv \theta_p$ denotes a vector of constant parameters. $\mathbb{Y}_{-p}^M$ is the indicator for the event "a total of M of player p's opponents chose $Y_q = 1$". The simultaneous nature of the game implies that $\mathbb{Y}_{-p}^M$ is unknown to player p at the moment of choosing Y_p. In Subsection 2.2 we will describe carefully how much information concerning opponents' payoff covariates and coefficients is known to each player. Notice that Equation (1) describes a strategic-interaction setting where player p is affected by his opponents' choices only through the total number of opponents who choose $Y_q = 1$, irrespective of their identities. We could consider alternative formulations where identities matter. The simplest example would be given by

$$Y_p \times \left(X_p' \beta_p + \sum_{q \neq p} \Delta_p^q Y_q - \varepsilon_p \right) \tag{1A}$$

The most general characterization would allow player p's payoffs to differ across the entire space of action profiles of his opponents. Denote the latter by $\mathscr{A}_{-p}$, and let

$$\mathbb{Y}_{-p}^{\mathbb{j}} = \mathbb{1}\left\{ \text{Player } p\text{'s opponents play action profile } a_{-p}^{\mathbb{j}} \in \mathscr{A}_{-p} \right\},$$

We could parameterize player p's payoffs as

$$Y_p \times \left(X_p' \beta_p + \sum_{\mathbb{j} \in \mathscr{A}_{-p}} \Delta_p^{\mathbb{j}} \mathbb{Y}_{-p}^{\mathbb{j}} - \varepsilon_p \right). \tag{1B}$$

Clearly, both (1) and ($1A$) are special cases of ($1B$), with implicit restrictions on the $\Delta_p^{\mathbb{j}}$'s. Without ex-ante strategic interaction restrictions of any kind, we would have a total of $2^{\mathcal{P}-1}$ unrestricted strategic-interaction parameters. As we shall see, this would have nontrivial implications for identification purposes. Unless explicitly stated otherwise, we will maintain the parameterization in (1). We will denote the payoff parameters by

$$\theta_p \equiv \left(\beta_p', \left\{ \Delta_p^M \right\}_{M=0}^{\mathcal{P}} \right), \quad \text{with} \quad \theta \equiv \left(\theta_p \right)_{p=1}^{\mathcal{P}}.$$

Our formulation is able to accommodate a number of special cases, each of which would imply further restrictions on the Δ_p^M's. For example,

(i) Only the choice made by the majority of player p's opponents matters to player p

$$\Delta_p^M = \begin{cases} \Delta_p^a & \text{if } \Big(M < \lfloor\frac{\mathcal{P}-1}{2}\rfloor + 1\Big)\mathbb{1}\big\{\mathcal{P} \text{ even}\big\} + \Big(M < \frac{\mathcal{P}-1}{2}\Big)\mathbb{1}\big\{\mathcal{P} \text{ odd}\big\} \\ \Delta_p^b & \text{if } \Big(M \geq \lfloor\frac{\mathcal{P}-1}{2}\rfloor + 1\Big)\mathbb{1}\big\{\mathcal{P} \text{ even}\big\} + \Big(M \geq \frac{\mathcal{P}-1}{2}\Big)\mathbb{1}\big\{\mathcal{P} \text{ odd}\big\}, \end{cases}$$

where $\lfloor x \rfloor$ denotes the smallest integer closest to x.

(ii) Payoffs depend linearly on M,

$$Y_p \times \left(X_p'\beta_p + \Delta_p \sum_{M=0}^{\mathcal{P}-1} M \cdot \mathbb{Y}_{-p}^M - \varepsilon_p \right) \Longrightarrow \Delta_p^M = \Delta_p \cdot M, \quad \text{for some scalar } \Delta_p.$$

We conclude by noting that in the special case where $\Delta_p^M = \Delta_p^{M'} \equiv \Delta_p$, the payoffs in (1) reduce to $Y_p \times \Big(X_p'\beta_p + \Delta_p - \varepsilon_p \Big)$. For the moment, we leave θ completely unrestricted. In Section 4 we will discuss in detail the identified features of θ given our behavioral assumptions. We describe the players' informational assumptions next.

2.2 Player's information, expected utility and beliefs.

We will make the following assumption

Assumption I1.- The expression for normal-form payoffs in Equation (1) and the true parameter value θ are common knowledge among the players. Each player p observes the realization of his own payoff covariates X_p and ε_p. The realization of ε_p is only privately observed by player p. We also allow (but not require) some elements in X_p to be privately observed when the game is played. Player p conditions his beliefs about his opponents' behavior on a vector of signals Z_p. Every covariate in $X \equiv \big(X_p\big)_{p=1}^{\mathcal{P}}$ that is publicly observed is included in Z_p, but the latter may include covariates which are not in X if some of its elements are private information. Z_p would also include covariates that are statistically interdependent with $\varepsilon_{-p} \equiv \big(\varepsilon_q\big)_{q \neq p}$. In particular, we could have $\varepsilon_p \in Z_p$ if ε_p is statistically interdependent with ε_{-p}. The identities of the covariates included in Z_p are common knowledge, but their exact realizations may be private information. The *true* (joint) distribution of $\big(X_p, Z_p, \varepsilon_p\big)_{p=1}^{\mathcal{P}}$ is known to all players. For each player p, the vector X_p is observed by the econometrician, but ε_p is not. For the moment we will say nothing about the observability of Z_p by the econometrician.

We will *maintain Assumption I1 for the rest of the paper.* Player p's beliefs are conditioned on Z_p. We will denote beliefs by

$$\pi_{-p}^{M}(Z_p) = E_p\big[\mathbb{Y}_{-p}^{M}\big|Z_p\big], \text{ where } E_p[\cdot] \text{ denotes player } p\text{'s subjective expectation.}$$

Player p's expected utility of choosing $Y_p = 1$ is given by

$$U_p(Y_p = 1) = X_p'\beta_p + \sum_{M=0}^{\mathcal{P}-1} \Delta_p^M \pi_{-p}^M(Z_p) - \varepsilon_p. \tag{2}$$

From above, we have $U_p(Y_p = 0) = 0$. Thus, conditional on the realization of Z_p, player p generates subjective expectations $\pi_{-p}^{M}(Z_p) = E_p\big[\mathbb{Y}_{-p}^{M}\big|Z_p\big]$ for $M = 0, \ldots, \mathcal{P}-1$, and he uses them to construct the expected utility in (2). *We will maintain expected utility maximization*, which yields optimal decision rules of the form

$$Y_p = \mathbb{1}\left\{X_p'\beta_p + \sum_{M=0}^{\mathcal{P}-1} \Delta_p^M \pi_{-p}^M(Z_p) - \varepsilon_p \geq 0\right\}. \tag{3}$$

This rule assumes that player p will choose $Y_p = 1$ when he is indifferent. Such ties will occur with probability zero given the continuously distributed nature of ε_p. Imperfect observation of opponents' payoffs hinders the ability of mixing in a way that makes opponents exactly indifferent between their two actions. Consequently, we focus on pure strategy strategies of the form (3).

Assumption I1 (continued).- It is common knowledge that players maximize their expected utility and follow an optimal decision rule of the form (3). However, we do not impose the assumption that the process of belief construction is common knowledge. In other words, the mapping $Z_p \longmapsto \sum_{M=0}^{\mathcal{P}-1} \Delta_p^M \pi_{-p}^M(Z_p)$ is not assumed to be known to player $q \neq p$. In particular, we do not impose the assumption that beliefs are "correct" in a Bayesian-Nash equilibrium sense.

Altogether, we assume a scenario where ε_p is only privately observed; the identities of the elements in X_p and Z_p are common knowledge, but their exact realizations may be only privately observed; the true joint distribution of $\big(X_p, Z_p, \varepsilon_p\big)_{p=1}^{\mathcal{P}}$ is common knowledge; expected utility maximization as in (3) is also common knowledge, but the process used to generate beliefs $\pi_{-p}^{M}(\cdot)$ may not be common knowledge. This departs from the restrictions of a Bayesian-Nash equilibrium, where *rationality is common knowledge.* We discuss this case in the following Subsection.

2.2.1 Bayesian-Nash equilibrium (BNE) beliefs

In this paper we depart from the notion of common knowledge of rationality and consequently, from the assumption that beliefs are "correct" in a Nayesian-Nash equilibrium (BNE) sense. By "correct" we refer to beliefs that match players' actual conditional choice probabilities. However, it is convenient to characterize the features of BNE beliefs. Let us begin by denoting player p's beliefs in vector form as

$$\pi_{-p}(Z_p) \equiv \Big(\pi^0_{-p}(Z_p), \ldots, \pi^{\mathcal{P}-1}_{-p}(Z_p)\Big)'.$$

As before, we let $a^{\mathbb{j}}_{-p}$ denote a particular action profile by player p's opponents. Let

$$\begin{aligned} &\mathbb{I}^{\mathbb{j}}_q \equiv \mathbb{1}\Big\{\text{Player } q \text{ is supposed to play } Y_q = 1 \text{ in profile } a^{\mathbb{j}}_{-p}\Big\}, \\ &\mathscr{A}^M_{-p} \equiv \Big\{a^{\mathbb{j}}_{-p} : \sum_{q\neq p} \mathbb{I}^{\mathbb{j}}_q = M\Big\} \quad \text{(collection of all action profiles where } M \text{ opponents choose } Y_q = 1). \end{aligned} \tag{4}$$

Next, let

$$\begin{aligned} &Y_q\big(a^{\mathbb{j}}_{-p}, \pi_{-q}\big) \\ &\quad = \mathbb{I}^{\mathbb{j}}_q \times \mathbb{1}\Big\{X_q'\beta_q + \sum_{M=0}^{\mathcal{P}-1} \Delta^M_q \pi^M_{-q}(Z_q) - \varepsilon_q \geq 0\Big\} + \big(1 - \mathbb{I}^{\mathbb{j}}_q\big) \times \mathbb{1}\Big\{X_q'\beta_q + \sum_{M=0}^{\mathcal{P}-1} \Delta^M_q \pi^M_{-q}(Z_q) - \varepsilon_q < 0\Big\}. \end{aligned}$$

$Y_q\big(a^{\mathbb{j}}_{-p}, \pi_{-q}\big)$ is the indicator function for the event "it is optimal for player q to play the action prescribed in profile $a^{\mathbb{j}}_{-p}$". Now let

$$\mathbb{B}^{\mathbb{j}}_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big) = \mathbb{1}\Big\{Y_q\big(a^{\mathbb{j}}_{-p}, \pi_{-q}\big) = 1 \quad \text{for all} \quad q \neq p.\Big\} = \prod_{q\neq p} Y_q\big(a^{\mathbb{j}}_{-p}, \pi_{-q}\big)$$

denote the indicator function for the event "it is optimal for player p's opponents to play the action profile $a^{\mathbb{j}}_{-p}$". Next, denote

$$\mathbb{Y}^M_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big) = \mathbb{1}\Big\{\mathbb{B}^{\mathbb{j}}_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big) = 1 \quad \text{for some } a^{\mathbb{j}}_{-p} \in \mathscr{A}^M_{-p}\Big\} = \sum_{\mathbb{j}\in\mathscr{A}^M_{-p}} \mathbb{B}^{\mathbb{j}}_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big).$$

$\mathbb{Y}^M_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big)$ is the indicator function for the event "it is optimal for exactly M of player p's opponents to choose $Y_q = 1$". A vector of BNE beliefs for this game is *any* collection of conditional choice probabilities $\big\{\pi_{-p}(\cdot)\big\}_{p=1}^{\mathcal{P}}$ that solve, for every realization of $Z_1, \ldots, Z_{\mathcal{P}}$, the BNE system

$$\pi^M_{-p}(Z_p) = E\Big[\mathbb{Y}^M_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big)\Big| Z_p\Big], \quad \text{for} \quad p = 1, \ldots, \mathcal{P}, \quad \text{and} \quad M = 0, \ldots, \mathcal{P}-1. \tag{5}$$

The foundation of the BNE system (5) is the requirement that equilibrium beliefs be self-consistent. Without ex-ante restrictions on the structure of the vector of strategic interaction parameters $\{\Delta_p^M\}_{M=0}^{\mathcal{P}-1}$, even for moderately large $\mathcal{P}$, the BNE system becomes an increasingly complex problem to solve for each individual player. This is true even in the case where all players condition on the same vector of signals ($Z_p = Z_q \equiv Z \ \forall \ p, q$).

3 Iterated dominance and the space of L_k−rational beliefs

3.1 Dominance and higher-order beliefs

We have stated previously that we wish to relax the assumption that players have perfect models about others and that rationality is common knowledge. We make the usual distinction between *first-order beliefs* as those that refer to other players' choices, and *higher-order beliefs* as those that refer to other players' beliefs. The first question we need to address has to do with the implications of higher order beliefs on players' behavior. The channel of this influence is through the effect of higher-order beliefs on their first-order counterparts. The type of behavior we analyze here is based on the premise that, even if players fail to have perfect models about others, it should be the case that first-degree beliefs satisfy the dominance features implied by their higher-order counterparts. We can illustrate this process using an iterative construction. We refer to player p's j^{th}−order beliefs as player p's assessment of his opponents' $(j-1)^{th}$−order beliefs. As we stated above, first-order beliefs refer to opponents' actions. Suppose player p is uncertain about the way player $q \neq p$ constructs his first-order beliefs $\pi_{-q}(Z_q)$ (see Equation 4), but he assumes that, w.p.1, for any realization of Z_q,[1]

$$\pi_{-q}(Z_q) \in \widetilde{\Pi}_{-q}^{(p)}(Z_q, \theta) \quad \text{where} \quad \widetilde{\Pi}_{-q}^{(p)}(Z_q, \theta) \subseteq \Delta^{\mathcal{P}} \quad \text{and} \quad \Delta^{\mathcal{P}} \equiv \text{probability simplex in } \mathbb{R}^{\mathcal{P}}.$$

In words, player p views $\widetilde{\Pi}_{-q}^{(p)}(Z_q, \theta)$ as the range of all possible values for player q's first-order beliefs for each given Z_q. For example, if we assumed BNE beliefs, the set $\widetilde{\Pi}_{-q}^{(p)}(Z_q, \theta)$ would be the collection of all solutions to the BNE system (5) for each given Z_q. The features of $\widetilde{\Pi}_{-q}^{(p)}(\cdot, \theta)$

[1] Note that this set is allowed to depend on Z_q and θ because player p knows that player q: (i) conditions his beliefs on Z_q, and (ii) knows the true value of θ.

summarize player p's *second-order beliefs* about player q. Let

$$\underline{\mu}_q\big(Z_q, \theta \mid \widetilde{\Pi}^{(p)}_{-q}\big) = \min_{(\pi^M)_{M=0}^{\mathcal{P}-1}} \left\{ \sum_{M=0}^{\mathcal{P}-1} \Delta_q^M \pi^M \right\} : \ \big(\pi^M\big)_{M=0}^{\mathcal{P}-1} \in \widetilde{\Pi}^{(p)}_{-q}(Z_q, \theta),$$

$$\overline{\mu}_q\big(Z_q, \theta \mid \widetilde{\Pi}^{(p)}_{-q}\big) = \max_{(\pi^M)_{M=0}^{\mathcal{P}-1}} \left\{ \sum_{M=0}^{\mathcal{P}-1} \Delta_q^M \pi^M \right\} : \ \big(\pi^M\big)_{M=0}^{\mathcal{P}-1} \in \widetilde{\Pi}^{(p)}_{-q}(Z_q, \theta).$$

According to player p's second-order beliefs, $\underline{\mu}_q(\cdot, \theta|\widetilde{\Pi}^{(p)}_{-q})$ and $\overline{\mu}_q(\cdot, \theta|\widetilde{\Pi}^{(p)}_{-q})$ represent the most pessimistic and the most optimistic assessments that player q can make for $\sum_{M=0}^{\mathcal{P}-1} \Delta_q^M \pi^M_{-q}(\cdot)$. If $\widetilde{\Pi}^{(p)}_{-q}(\cdot, \theta)$ is a connected subset of $\Delta^{\mathcal{P}}$ (e.g, the entire simplex $\Delta^{\mathcal{P}}$ itself), then any intermediate point between these bounds is a rationalizable in $\widetilde{\Pi}^{(p)}_{-q}$. For the moment, rationalizability of the entire interval is not essential —although this feature will hold in our formulation below—, for the current discussion only the bounds are important. For a given action profile $a^{\mathbb{j}}_{-p}$, let $\mathbb{I}^{\mathbb{j}}_q$ be as defined in (4), and let

$$\overline{Y}_q\big(a^{\mathbb{j}}_{-p}, \widetilde{\Pi}^{(p)}_{-q}\big) =$$
$$\mathbb{I}^{\mathbb{j}}_q \cdot \mathbb{1}\Big\{X_q'\beta_q + \overline{\mu}_q\big(Z_q, \theta \mid \widetilde{\Pi}^{(p)}_{-q}\big) - \varepsilon_q \geq 0\Big\} + \big(1 - \mathbb{I}^{\mathbb{j}}_q\big) \cdot \mathbb{1}\Big\{X_q'\beta_q + \underline{\mu}_q\big(Z_q, \theta \mid \widetilde{\Pi}^{(p)}_{-q}\big) - \varepsilon_q < 0\Big\},$$

$$\underline{Y}_q\big(a^{\mathbb{j}}_{-p}, \widetilde{\Pi}^{(p)}_{-q}\big) =$$
$$\mathbb{I}^{\mathbb{j}}_q \cdot \mathbb{1}\Big\{X_q'\beta_q + \underline{\mu}_q\big(Z_q, \theta \mid \widetilde{\Pi}^{(p)}_{-q}\big) - \varepsilon_q \geq 0\Big\} + \big(1 - \mathbb{I}^{\mathbb{j}}_q\big) \cdot \mathbb{1}\Big\{X_q'\beta_q + \overline{\mu}_q\big(Z_q, \theta \mid \widetilde{\Pi}^{(p)}_{-q}\big) - \varepsilon_q < 0\Big\}.$$

$\overline{Y}_q\big(a^{\mathbb{j}}_{-p}, \widetilde{\Pi}^{(p)}_{-q}\big)$ is the indicator function for the event "according to player p's second-order beliefs, the action that player q is supposed to play in profile $a^{\mathbb{j}}_{-p}$ *is not dominated* by his other available action". $\underline{Y}_q\big(a^{\mathbb{j}}_{-p}, \widetilde{\Pi}^{(p)}_{-q}\big)$ is the indicator for the event "according to player p's second-order beliefs, the action that player q is supposed to play in profile $a^{\mathbb{j}}_{-p}$ *dominates* his other available action". In this context, an action by player q is dominated —from the perspective of player p— if there do not exist beliefs in $\widetilde{\Pi}^{(p)}_{-q}$ that would rationalize it as a best response. Conversely, an action is not dominated if we can find such beliefs. Let us abbreviate player p's second-order beliefs for all his opponents by $\widetilde{\Pi}^{(p)} \equiv \big(\widetilde{\Pi}^{(p)}_{-q}\big)_{q \neq p}$. Define

$$\underline{\mathbb{B}}^{\mathbb{j}}_{-p}\big(\widetilde{\Pi}^{(p)}\big) = \mathbb{1}\Big\{\underline{Y}_q\big(a^{\mathbb{j}}_{-p}, \widetilde{\Pi}^{(p)}_{-q}\big) = 1 \ \forall\ q \neq p\Big\}, \quad \overline{\mathbb{B}}^{\mathbb{j}}_{-p}\big(\widetilde{\Pi}^{(p)}\big) = \mathbb{1}\Big\{\overline{Y}_q\big(a^{\mathbb{j}}_{-p}, \widetilde{\Pi}^{(p)}_{-q}\big) = 1 \ \forall\ q \neq p\Big\}.$$

Letting A_{-p} denote the profile of actions chosen by player p's opponents, it follows that *player p believes*

$$\underline{\mathbb{B}}^{\mathbb{j}}_{-p}\big(\widetilde{\Pi}^{(p)}\big) \leq \mathbb{1}\{A_{-p} = a^{\mathbb{j}}_{-p}\} \leq \overline{\mathbb{B}}^{\mathbb{j}}_{-p}\big(\widetilde{\Pi}^{(p)}\big) \quad \text{w.p.1.} \tag{6}$$

Equation (6) *summarizes the dominance features implied by player p's second-order beliefs* $\widetilde{\Pi}$. Now, according to our parameterization of payoffs, player p's expected utility depends on his expectation of $\big(\mathbb{Y}^{M}_{-p}\big)_{M=0}^{\mathcal{P}-1}$, the indicator functions for the event "exactly M opponents chose $Y_q = 1$". These subjective expectations constitute player p's first-order beliefs, and they must satisfy the dominance restrictions in (6). Let $\mathscr{A}^{M}_{-p}$ be as defined in (4) and let

$$\underline{\mathbb{Y}}^{M}_{-p}\big(\widetilde{\Pi}^{(p)}\big) = \mathbb{1}\Big\{\underline{\mathbb{B}}^{\mathbf{j}}_{-p}\big(\widetilde{\Pi}^{(p)}\big) = 1 \text{ for some } a^{\mathbf{j}}_{-p} \in \mathscr{A}^{M}_{-p}\Big\},$$
$$\overline{\mathbb{Y}}^{M}_{-p}\big(\widetilde{\Pi}^{(p)}\big) = \mathbb{1}\Big\{\overline{\mathbb{B}}^{\mathbf{j}}_{-p}\big(\widetilde{\Pi}^{(p)}\big) = 1 \text{ for some } a^{\mathbf{j}}_{-p} \in \mathscr{A}^{M}_{-p}\Big\}.$$

It follows from (6) that , with probability one, player p believes

$$\underline{\mathbb{Y}}^{M}_{-p}\big(\widetilde{\Pi}^{(p)}\big) \leq \mathbb{Y}^{M}_{-p} \leq \overline{\mathbb{Y}}^{M}_{-p}\big(\widetilde{\Pi}^{(p)}\big).$$

Recall that player p conditions his beliefs on Z_p. Let

$$\underline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|\widetilde{\Pi}^{(p)}\big) = E\big[\underline{\mathbb{Y}}^{M}_{-p}\big(\widetilde{\Pi}^{(p)}\big)\big|Z_p\big] \quad \text{and} \quad \overline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|\widetilde{\Pi}^{(p)}\big) = E\big[\overline{\mathbb{Y}}^{M}_{-p}\big(\widetilde{\Pi}^{(p)}\big)\big|Z_p\big].$$

These expectations are taken with respect to the true distribution of the covariates involved. Assumption (I1) asserts that this distribution is known to all players. The space of first-order beliefs for player p that satisfy the dominance restrictions in (6) is given by

$$\Pi_p\big(Z_p,\theta\big|\widetilde{\Pi}^{(p)}\big) = \left\{\big(\pi^{M}_{-p}\big)_{M=0}^{\mathcal{P}-1}:\ \underline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|\widetilde{\Pi}^{(p)}\big) \leq \pi^{M}_{-p} \leq \overline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|\widetilde{\Pi}^{(p)}\big) \text{ and } \sum_{M=0}^{\mathcal{P}-1}\pi^{M}_{-p} = 1\right\}. \tag{7}$$

With probability one, player p's first-order beliefs must satisfy $\pi_{-p}(Z_p) \in \Pi_p\big(Z_p,\theta\big|\widetilde{\Pi}^{(p)}\big)$. We can continue this construction iteratively by using player p's *third-order beliefs*, which would refer to his opponents' second order beliefs $\big(\widetilde{\Pi}^{(q)}\big)_{q\neq p}$. Using his third-order beliefs and (7), player p refine his second-order beliefs by eliminating the subset of $\widetilde{\Pi}^{(p)}$ that do not satisfy (7)[2]. Repeating the previous steps this would, in turn, lead player p to revise his first-order beliefs which would refine the set (7). Using higher-order beliefs we could repeat this process iteratively. Each round of this process of iterated thinking refines the set of lower-order beliefs using the dominance features described above. The following subsection describes this procedure starting with second-order beliefs that include the entire probability simplex. We will refer to the set of first-order beliefs (7) that results from the k^{th} step in this procedure as the set of L_k-rational beliefs.

[2] Recall that we defined player p's second-order beliefs $\widetilde{\Pi}^{(p)}$ as player p's assessment of the entire range of possible values for player q's first-order beliefs. Thus, player p would not consider first-order beliefs for his opponents that are not in $\widetilde{\Pi}^{(p)}$.

3.2 L_k−rational beliefs

For each player $p = 1, \ldots, \mathcal{P}$, let $\Pi_p(Z_p, \theta|L_k)$ be as defined in Equation (11), below. Let

$$\begin{aligned}\overline{\mu}_p(Z_p, \theta|L_k) &= \max_{(\pi^M_{-p})^{\mathcal{P}-1}_{M=0}} \left\{\sum_{M=0}^{\mathcal{P}-1} \Delta^M_p \pi^M_{-p}\right\} \text{ such that } (\pi^M_{-p})^{\mathcal{P}-1}_{M=0} \in \Pi_p(Z_p, \theta|L_k) \\ \underline{\mu}_p(Z_p, \theta|L_k) &= \min_{(\pi^M_{-p})^{\mathcal{P}-1}_{M=0}} \left\{\sum_{M=0}^{\mathcal{P}-1} \Delta^M_p \pi^M_{-p}\right\} \text{ such that } (\pi^M_{-p})^{\mathcal{P}-1}_{M=0} \in \Pi_p(Z_p, \theta|L_k).\end{aligned} \tag{8}$$

Take any action profile $a^{\mathbb{j}}_{-p} \in \mathscr{A}_{-p}$ and let $\mathbb{I}^{\mathbb{j}}_q$ be as defined in (4). Let

$$\begin{aligned}&\overline{Y}_q(a^{\mathbb{j}}_{-p}, L_k) = \mathbb{I}^{\mathbb{j}}_q \times \mathbb{1}\big\{X'_q\beta_q + \overline{\mu}_q(Z_q, \theta|L_k) - \varepsilon_q \geq 0\big\} + \big(1 - \mathbb{I}^{\mathbb{j}}_q\big) \times \mathbb{1}\big\{X'_q\beta_q + \underline{\mu}_q(Z_q, \theta|L_k) - \varepsilon_q < 0\big\}, \\ &\underline{Y}_q(a^{\mathbb{j}}_{-p}, L_k) = \mathbb{I}^{\mathbb{j}}_q \times \mathbb{1}\big\{X'_q\beta_q + \underline{\mu}_q(Z_q, \theta|L_k) - \varepsilon_q \geq 0\big\} + \big(1 - \mathbb{I}^{\mathbb{j}}_q\big) \times \mathbb{1}\big\{X'_q\beta_q + \overline{\mu}_q(Z_q, \theta|L_k) - \varepsilon_q < 0\big\}, \\ &\underline{\mathbb{B}}^{\mathbb{j}}_{-p}(L_k) = \mathbb{1}\Big\{\underline{Y}_q(a^{\mathbb{j}}_{-p}, L_k) = 1 \text{ for all } q \neq p\Big\} \quad \text{and} \quad \overline{\mathbb{B}}^{\mathbb{j}}_{-p}(L_k) = \mathbb{1}\Big\{\overline{Y}_q(a^{\mathbb{j}}_{-p}, L_k) = 1 \text{ for all } q \neq p\Big\}.\end{aligned} \tag{9}$$

For each $M = 0, \ldots, \mathcal{P} - 1$, let $\mathscr{A}^M_{-p}$ be as in (4) and define

$$\begin{aligned}&\underline{\mathbb{Y}}^M_{-p}(L_k) = \mathbb{1}\Big\{\underline{\mathbb{B}}^{\mathbb{j}}_{-p}(L_k) = 1 \text{ for some } a^{\mathbb{j}}_{-p} \in \mathscr{A}^M_{-p}\Big\}, \ \overline{\mathbb{Y}}^M_{-p}(L_k) = \mathbb{1}\Big\{\overline{\mathbb{B}}^{\mathbb{j}}_{-p}(L_k) = 1 \text{ for some } a^{\mathbb{j}}_{-p} \in \mathscr{A}^M_{-p}\Big\}, \\ &\underline{\mathbb{F}}^M_{-p}\big(Z_p, \theta\big|L_k\big) = E\big[\underline{\mathbb{Y}}^M_{-p}(L_k)\big|Z_p\big], \quad \text{and} \quad \overline{\mathbb{F}}^M_{-p}\big(Z_p, \theta\big|L_k\big) = E\big[\overline{\mathbb{Y}}^M_{-p}(L_k)\big|Z_p\big].\end{aligned} \tag{10}$$

These expectations are taken with respect to the true distribution of the covariates involved. From Assumption (I1), this distribution is known to all players. The set $\Pi_p(Z_p, \theta|L_k)$ is given by

For $k = 1$:

$$\Pi_p(Z_p, \theta|L_k) = \left\{(\pi^M_{-p})^{\mathcal{P}-1}_{M=0}: \quad 0 \leq \pi^M_{-p} \leq 1 \quad \text{and} \quad \sum_{M=0}^{\mathcal{P}-1} \pi^M_{-p} = 1\right\} \text{(the probability simplex in } \mathbb{R}^{\mathcal{P}}\text{)}.$$

For $k \geq 2$:

$$\Pi_p(Z_p, \theta|L_k) = \left\{(\pi^M_{-p})^{\mathcal{P}-1}_{M=0}: \quad \underline{\mathbb{F}}^M_{-p}\big(Z_p, \theta\big|L_{k-1}\big) \leq \pi^M_{-p} \leq \overline{\mathbb{F}}^M_{-p}\big(Z_p, \theta\big|L_{k-1}\big) \quad \text{and} \quad \sum_{M=0}^{\mathcal{P}-1} \pi^M_{-p} = 1\right\} \tag{11}$$

Thus, $\Pi_p(\cdot, \theta|L_1)$ is the entire probability simplex, the space of all possible, well-defined beliefs. We will show below that $\Pi_p(\cdot, \theta|L_k)$ is nonempty w.p.1 for any $k \geq 2$. First, we define the concept of L_k−rational beliefs and L_k−rationality.

Definition 1 (L_k−rationality) For each Z_p, we will refer to $\Pi_p(Z_p, \theta|L_k)$ as the *set of L_k−rational (first-order) beliefs* for player p. We will say that player p is L_k−rational if

$$X'_p\beta_p + \underline{\mu}_p(Z_p, \theta|L_k) - \varepsilon_p \leq U_p(Y_p = 1) \leq X'_p\beta_p + \overline{\mu}_p(Z_p, \theta|L_k) - \varepsilon_p \quad \text{w.p.1},$$

where, as before, $U_p(Y_p = 1)$ is player p's expected utility of choosing $Y_p = 1$.

The following remark summarizes the relationship between L_k-rationality and the iterated dominance construction discussed above in Subsection 3.1.

Remark 1 (L_k-rationality and iterated dominance) Any player who maximizes his expected utility using well-defined beliefs is L_1-rational. It also follows that

$$\underline{\mu}_p(Z_p, \theta|L_1) = \min\{\Delta_p^M : M = 0, \ldots, \mathcal{P}\}, \quad \overline{\mu}_p(Z_p, \theta|L_1) = \max\{\Delta_p^M : M = 0, \ldots, \mathcal{P}\}.$$

For $k \geq 2$, player p is L_k-rational if his beliefs are consistent with assuming that all his opponents are $L_{k-1}-$rational. A consequence of this is that, if player p is L_k-rational for $k \geq 2$, he is also $L_{k'}-$rational for all $1 \leq k' \leq k-1$. The assertion that player p is L_k-rational is a statement not only about his first-order beliefs, but also about his higher order beliefs up to order k. It is important to note that $\Pi_p(\cdot, \theta|L_k)$ is the entire set of first-order beliefs that survive $k-1$ steps of iterated dominance. This follows because $\Pi_p(\cdot, \theta|L_1)$ includes all possible beliefs and therefore $\Pi_p(\cdot, \theta|L_2)$ is the entire set of first-order beliefs that survive one round of dominance, and the assertion follows by induction for $k \geq 3$. Also, since $\Pi_p(\cdot, \theta|L_k)$ is the intersection between the probability simplex and a hypercube, it is convex and therefore connected for all $k \geq 1$ (we will show that it is nonempty below). This implies that all beliefs in $\Pi_p(\cdot, \theta|L_k)$ are rationalizable following $k-1$ steps of iterated dominance.

Proposition 1 *For any player p and all $k \geq 1$, the set of L_k-rational beliefs, $\Pi_p(Z_p, \theta|L_k)$, is nonempty for all $Z_p \in \mathbb{S}(Z_p)$ (the support of Z_p), and any θ.*

Proof: We will proceed by induction by showing that "if $\Pi_p(Z_p, \theta|L_k)$ is nonempty for all players $p = 1, \ldots, \mathcal{P}$, then $\Pi_p(Z_p, \theta|L_{k+1})$ is also nonempty". The result in the proposition will follow from here because, by construction, $\Pi_p(Z_p, \theta|L_k)$ is nonempty for $k = 1$ for all p and all θ. We focus on values $Z_p \in \mathbb{S}(Z_p)$ to ensure that the conditional expectations involved are well-defined. Suppose $\Pi_p(Z_p, \theta|L_k)$ is nonempty for every player p. Then $\underline{\mu}_p(Z_p, \theta|L_k)$ and $\overline{\mu}_p(Z_p, \theta|L_k)$ are well-defined and, by construction, satisfy $\underline{\mu}_p(Z_p, \theta|L_k) \leq \overline{\mu}_p(Z_p, \theta|L_k)$. Next, note that if $\underline{\mathbb{B}}^{\mathbb{j}}_{-p}(L_k) = 1$ for some action profile $a^{\mathbb{j}}_{-p}$, we must have $\underline{\mathbb{B}}^{\mathbb{i}}_{-p}(L_k) = 1$ for every other profile $a^{\mathbb{i}}_{-p} \neq a^{\mathbb{j}}_{-p}$. This follows because, if $a^{\mathbb{i}}_{-p} \neq a^{\mathbb{j}}_{-p}$, there must exist a player $q \neq p$ for whom $\underline{Y}_q(a^{\mathbb{i}}_{-p}, L_k) = 0$. This would be any player who is supposed to play different actions in profiles $a^{\mathbb{j}}_{-p}$ and $a^{\mathbb{i}}_{-p}$. Since both profiles are different, there must exist at least one such player. It follows that $\underline{\mathbb{Y}}^M_{-p}(L_k) = 1 \Longrightarrow \underline{\mathbb{Y}}^{M'}_{-p}(L_k) = 0$

for all $M' \neq M$. Therefore, $\mathbb{1}\{\underline{\mathbb{Y}}^{M}_{-p}(L_k)=1 \text{ for some } M=0,\ldots,\mathcal{P}-1\} = \sum_{M=0}^{\mathcal{P}-1}\underline{\mathbb{Y}}^{M}_{-p}(L_k)$. Using this, we get

$$
\begin{aligned}
&\Pr\Big(\underline{\mathbb{Y}}^{M}_{-p}(L_k)=1 \text{ for some } M=0,\ldots,\mathcal{P}-1 \;\Big|Z_p\Big)\\
&\quad=\sum_{M=0}^{\mathcal{P}-1}\underline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big) \leq \Pr\Big(\underline{\mathbb{B}}^{\mathfrak{j}}_{-p}(L_k)=1 \text{ for some } \; a^{\mathfrak{j}}_{-p}\in\mathscr{A}_{-p} \;\Big|Z_p\Big)\\
&\quad\leq \Pr\Big(X_q'\beta_q+\underline{\mu}_q(Z_q,\theta|L_k)-\varepsilon_q\geq 0 \;\text{ or }\; X_q'\beta_q+\overline{\mu}_q(Z_q,\theta|L_k)-\varepsilon_q<0 \;\text{ for some } q\neq p\;\Big|Z_p\Big)\\
&\quad=1-\Pr\Big(-\overline{\mu}_q(Z_q,\theta|L_k)\leq X_q'\beta_q-\varepsilon_q<-\underline{\mu}_q(Z_q,\theta|L_k)\;\text{ for all } q\neq p\;\Big|Z_p\Big)\leq 1.
\end{aligned}
$$

This yields $\sum_{M=0}^{\mathcal{P}-1}\underline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big)\leq 1$. Next, we will show that $\sum_{M=0}^{\mathcal{P}-1}\overline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big)\geq 1$. Note first that

$$
\Pr\Big(X_q'\beta_q+\overline{\mu}_q(Z_q,\theta|L_k)-\varepsilon_q\geq 0 \;\text{ or }\; X_q'\beta_q+\underline{\mu}_q(Z_q,\theta|L_k)-\varepsilon_q<0\;\Big|Z_p\Big)=1 \quad \forall\, q\neq p.
$$

This follows because it is impossible to have $X_q'\beta_q+\overline{\mu}_q(Z_q,\theta|L_k)-\varepsilon_q<0$ *and* $X_q'\beta_q+\underline{\mu}_q(Z_q,\theta|L_k)-\varepsilon_q\geq 0$ (this would imply that both of player q's available actions are dominated, which is impossible). Thus, we have $\Pr\Big(\overline{\mathbb{B}}^{\mathfrak{j}}_{-p}(L_k)=1 \text{ for some } a^{\mathfrak{j}}_{-p}\in\mathscr{A}_{-p}\;\Big|Z_p\Big)=1$. Since $\bigcup_M \mathscr{A}^{M}_{-p}=\mathscr{A}_{-p}$, it follows that $\Pr\Big(\overline{\mathbb{Y}}^{M}_{-p}(L_k)=1 \text{ for some } M=0,\ldots,\mathcal{P}-1\;\Big|Z_p\Big)=1$. A Bonferroni inequality yields $\sum_{M=0}^{\mathcal{P}-1}\overline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big)\geq 1$. Combining the previous results, we have

$$
0\leq 1-\sum_{M=0}^{\mathcal{P}-1}\underline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big)\leq 1\leq \sum_{M=0}^{\mathcal{P}-1}\overline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big).
$$

In fact, these inequalities are strict unless $\Delta_q^M=\Delta_q^{M'}$ for all M,M' and all $q\neq p$ (in this case, none of player p's opponents are affected by the actions of others). It follows that we can find a collection $(\pi^{M}_{-p})_{M=0}^{\mathcal{P}-1}$ that adds up to one and satisfies $\underline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big)\leq\pi^{M}_{-p}\leq\overline{\mathbb{F}}^{M}_{-p}\big(Z_p,\theta\big|L_k\big)$ for each M. This means that $\Pi_p(Z_p,\theta|L_{k+1})$ is nonempty. Thus, if $\Pi_p(Z_p,\theta|L_k)$ is nonempty for all $p=1,\ldots,\mathcal{P}$, then $\Pi_p(Z_p,\theta|L_{k+1})$ is also nonempty. The claim in the proposition follows because by construction $\Pi_p(Z_p,\theta|L_k)$ is nonempty for $k=1$ for all $p=1,\ldots,\mathcal{P}$ and all θ. □

The set of L_k-rational beliefs must contain all solutions to the BNE system (5) for all k. In other words, equilibrium behavior is a special case of L_k-rationality. This follows by the self-consistent nature of BNE beliefs.

Proposition 2 ***(L_k-rationality and BNE beliefs)*** *Suppose $\big(\pi^*_{-p}(Z_p)\big)_{p=1}^{\mathcal{P}}$ is a collection of beliefs that satisfy the Bayesian-Nash equilibrium conditions* (5) *described in Subsection 2.2.1 (i.e, a collection of BNE beliefs). Then, for each player p, we have $\pi^*_{-p}(Z_p)\in\Pi_p(Z_p,\theta|L_k)$ for every k.*

Proof: Throughout the proof we will use the notation introduced Subsection 2.2.1. We will proceed inductively, by showing that "if $\big(\pi^*_{-p}(Z_p)\big)_{p=1}^{\mathcal{P}}$ are BNE beliefs, then $\pi^*_{-p}(Z_p) \in \Pi_p(Z_p, \theta|L_k)$ implies $\pi^*_{-p}(Z_p) \in \Pi_p(Z_p, \theta|L_{k+1})$ for all p". The result in the proposition will follow because $\Pi_p(Z_p, \theta|L_1)$ contains all possible beliefs. Suppose $\pi^*_{-q}(\cdot) \in \Pi_q(\cdot, \theta|L_k)$ for all $q = 1, \ldots, \mathcal{P}$, but $\pi^*_{-p}(Z_p) \notin \Pi_p(Z_p, \theta|L_{k+1})$ for some p and some $Z_p \in \mathbb{S}(Z_p)$. This means that

$$\sum_{M=0}^{\mathcal{P}-1} \Delta_p^M \pi^*_{-p}(Z_p) < \underline{\mu}_p(Z_p, \theta|L_{k+1}) \quad \text{or} \quad \sum_{M=0}^{\mathcal{P}-1} \Delta_p^M \pi^*_{-p}(Z_p) > \overline{\mu}_p(Z_p, \theta|L_{k+1}).$$

By definition this, in turn, implies (using the notation from Subsection 2.2.1),

$$\pi_{-q}(\cdot) \in \Pi_{-q}(\cdot, \theta|L_k) \ \forall\ q \neq p \implies \pi^{*M}_{-p}(Z_p) \neq E\Big[\mathbb{Y}^M_{-p}\big(\{\pi_{-q}\}_{q\neq p}\big)\Big|Z_p\Big] \quad \text{for all } M.$$

(In fact, we only need this to hold for some M). Now, we have $\pi^*_{-q}(\cdot) \in \Pi_q(\cdot, \theta|L_k)\ \ \forall\ q \neq p$ by assumption. Thus, $\pi^*_{-p}(Z_p) \notin \Pi_p(Z_p, \theta|L_{k+1})$ implies $\pi^{*M}_{-p}(Z_p) \neq E\big[\mathbb{Y}^M_{-p}\big(\{\pi^*_{-q}\}_{q\neq p}\big)\big|Z_p\big]$. Therefore, $\big(\pi^*_{-p}(Z_p)\big)_{p=1}^{\mathcal{P}}$ do not satisfy the BNE system (5) which contradicts the assumption that they constitute BNE beliefs. Therefore, if $\big(\pi^*_{-p}(Z_p)\big)_{p=1}^{\mathcal{P}}$ are BNE beliefs, then $\pi^*_{-p}(Z_p) \in \Pi_p(Z_p, \theta|L_k)$ implies $\pi^*_{-p}(Z_p) \in \Pi_p(Z_p, \theta|L_{k+1})$ for all p. The result in the proposition follows because, for all players $p = 1, \ldots, \mathcal{P}$, the set $\Pi_p(Z_p, \theta|L_1)$ contains all possible beliefs including, in particular, any solution to the BNE system (5). □

Given the result in Proposition 2, an alternative proof of the claim that $\Pi_p(Z_p, \theta|L_k)$ is nonempty for all k would be to show that there exists at least one solution to the BNE system (5). Given the assumed continuous joint distribution of $\big(\varepsilon_p\big)_{p=1}^{\mathcal{P}}$, existence follows from Brouwer's fixed point theorem. Uniqueness of a solution is irrelevant to us, the result in Proposition 2 holds for all possible solutions to the BNE system.

Proposition 3 (Monotonicity of $\boldsymbol{L_k}$−rational bounds) For every player p, with probability one, we have

$$\underline{\mu}_p\big(Z_p, \theta\big|L_{k+1}\big) \geq \underline{\mu}_p\big(Z_p, \theta\big|L_k\big) \quad \text{and} \quad \overline{\mu}_p\big(Z_p, \theta\big|L_{k+1}\big) \leq \overline{\mu}_p\big(Z_p, \theta\big|L_k\big) \quad \forall\ k \geq 1. \tag{12}$$

Suppose that, for every player p and every opponent $q \neq p$,

$$0 < \Pr\Big[X_q'\beta_q + \max_M\big(\Delta_q^M\big) - \varepsilon_q \geq 0\Big|Z_p\Big] < 1; \quad 0 < \Pr\Big[X_q'\beta_q + \min_M\big(\Delta_p^M\big) - \varepsilon_q < 0\Big|Z_p\Big] < 1 \quad \forall\ Z_p \in \mathbb{S}(Z_p).$$

Then, for any player p who satisfies $\min_M\big(\Delta_p^M\big) \neq \max_M\big(\Delta_p^M\big)$, and who has a rival who also satisfies $\min_M\big(\Delta_q^M\big) \neq \max_M\big(\Delta_q^M\big)$, the inequalities in (12) are strict with positive probability one.

Proof: We begin with the trivial cases. Suppose $\min_M(\Delta_p^M) = \max_M(\Delta_p^M) \equiv \Delta_p$. Then $\sum_{M=0}^{\mathcal{P}-1}\Delta_p^M\pi_{-p}^M(Z_p) = \Delta_p$ for all well-defined beliefs, so $\underline{\mu}_p(Z_p,\theta|L_k) = \overline{\mu}_p(Z_p,\theta|L_k) = \Delta_p$ for all k and all $Z_p \in \mathbb{S}(Z_p)$. Next, suppose $\min_M(\Delta_q^M) = \max_M(\Delta_q^M) \equiv \Delta_q$ for all $q \neq p$. Let

$$Y_q(a_{-p}^{\mathbb{j}}) = \mathbb{I}_q^{\mathbb{j}} \times \mathbb{1}\Big\{X_q'\beta_q + \Delta_q - \varepsilon_q \geq 0\Big\} + \big(1-\mathbb{I}_q^{\mathbb{j}}\big) \times \mathbb{1}\Big\{X_q'\beta_q + \Delta_q - \varepsilon_q < 0\Big\},$$
$$\mathbb{B}_{-p}^{\mathbb{j}} = \prod_{q\neq p} Y_q(a_{-p}^{\mathbb{j}}), \quad \mathbb{Y}_{-p}^M = \sum_{\mathbb{j}\in\mathscr{A}_{-p}^M}\mathbb{B}_{-p}^{\mathbb{j}}, \quad \textit{and} \quad \mathbb{F}_{-p}^M(Z_p,\theta) = E\big[\mathbb{Y}_{-p}^M\big|Z_p\big].$$

Then, for all $k \geq 2$,

$$\Pi_p(Z_p,\theta|L_k) = \Big\{\mathbb{F}_{-p}^M(Z_p,\theta)\Big\}_{M=0}^{\mathcal{P}-1}, \quad \textit{and} \quad \underline{\mu}_p(Z_p,\theta|L_k) = \overline{\mu}_p(Z_p,\theta|L_k) = \sum_{M=0}^{\mathcal{P}-1}\Delta_p^M\mathbb{F}_{-p}^M(Z_p,\theta).$$

Thus, the result in the proposition follows with equality for all $k \geq 2$. For $k = 1$, note that $\sum_{M=0}^{\mathcal{P}-1}\Delta_p^M\mathbb{F}_{-p}^M(Z_p,\theta) = \overline{\mu}_p(Z_p,\theta|L_2) \leq \max_M(\Delta_p^M) = \overline{\mu}_p(Z_p,\theta|L_1)$ and $\sum_{M=0}^{\mathcal{P}-1}\Delta_p^M\mathbb{F}_{-p}^M(Z_p,\theta) = \underline{\mu}_p(Z_p,\theta|L_2) \geq \min_M(\Delta_p^M) = \underline{\mu}_p(Z_p,\theta|L_1)$. The only remaining case is when we have both $\min_M(\Delta_p^M) \neq \max_M(\Delta_p^M)$ *and* $\min_M(\Delta_q^M) \neq \max_M(\Delta_q^M)$ for some $q \neq p$. We proceed by induction, take the aforementioned player q and suppose $\underline{\mu}_q(Z_q,\theta|L_{k+1}) \geq \underline{\mu}_q(Z_q,\theta|L_k)$ and $\overline{\mu}_q(Z_q,\theta|L_{k+1}) \leq \overline{\mu}_q(Z_q,\theta|L_k)$ w.p.1. It follows from Equations (9) and (10) that $\underline{\mathbb{F}}_{-p}^M(Z_p,\theta|L_{k+1}) \geq \underline{\mathbb{F}}_{-p}^M(Z_p,\theta|L_k)$ and $\overline{\mathbb{F}}_{-p}^M(Z_p,\theta|L_{k+1}) \leq \overline{\mathbb{F}}_{-p}^M(Z_p,\theta|L_k)$ for each $M = 0,\ldots,\mathcal{P}-1$, w.p.1. It follows immediately that $\Pi_p(Z_p,\theta|L_{k+2}) \subseteq \Pi_p(Z_p,\theta|L_{k+1})$ w.p.1., and therefore by their definition in Equation (8), this implies $\underline{\mu}_p(Z_p,\theta|L_{k+2}) \geq \underline{\mu}_p(Z_p,\theta|L_{k+1})$ and $\overline{\mu}_p(Z_p,\theta|L_{k+2}) \leq \overline{\mu}_p(Z_p,\theta|L_{k+1})$ w.p.1. Thus, we have established that "$\underline{\mu}_p(Z_p,\theta|L_{k+1}) \geq \underline{\mu}_p(Z_p,\theta|L_k)$ and $\overline{\mu}_p(Z_p,\theta|L_{k+1}) \leq \overline{\mu}_p(Z_p,\theta|L_k)$ w.p.1 implies $\underline{\mu}_p(Z_p,\theta|L_{k+2}) \geq \underline{\mu}_p(Z_p,\theta|L_{k+1})$ and $\overline{\mu}_p(Z_p,\theta|L_{k+2}) \leq \overline{\mu}_p(Z_p,\theta|L_{k+1})$ w.p.1." If the condition below Equation (12) holds, then if the first set of inequalities are strict w.p.1, the second set of inequalities also hold as strict inequalities w.p.1. To conclude the proof, note that the condition below (12) implies, in fact, that $\underline{\mu}_p(Z_p,\theta|L_2) > \underline{\mu}_p(Z_p,\theta|L_1)$ and $\overline{\mu}_p(Z_p,\theta|L_2) < \overline{\mu}_p(Z_p,\theta|L_1)$ (strict inequalities) w.p.1. □

Remark 2 (L_k−rationality allows players to have higher order beliefs that differ across opponents) The model allows for player p to believe that the number of steps of iterated thinking differs across his opponents. The relevant value is the one that corresponds to the opponent q that player p believes performs the least number of steps of deletion. For example, suppose player p has three opponents and he believes that they perform 1, 4 and 8 steps of iterated thinking respectively. Then, according to our construction, player p is effectively L_2−rational: His beliefs are consistent with assuming that his opponents perform *at least* one round of deletion of dominated strategies. More generally, player p's higher order beliefs consist of probability distributions over the number of rounds of deletion of dominated strategies that his opponents compute, and player p could have different beliefs across his opponents. His behavior will be consistent with our definition of L_k-rationality for a given k if he believes that, with probability one, none of his opponents performs less than $k-1$ steps of deletion.

3.3 A compact expression for $\underline{\mu}_p(Z_p, \theta|L_k)$ and $\overline{\mu}_p(Z_p, \theta|L_k)$

Take any given θ and for any player p let

$$p^* \equiv \text{Number of coefficients in } (\Delta_p^M)_{M=0}^{\mathcal{P}-1} \text{ that are not zero.}$$

We will rank the nonzero Δ_p^M's according to their absolute value. Let $\mathcal{J}$ be any index set of the form

$$\mathcal{J} = \Big\{(1), (2), \ldots, (p^*)\Big\} \quad \text{with the property that} \quad \big|\Delta_p^{(j)}\big| \geq \big|\Delta_p^{(j+1)}\big| \quad \forall\ 1 \leq j \leq p^* - 1. \tag{13}$$

Without ties the index set $\mathcal{J}$ is unique, otherwise take any such index set. Using this set along with the L_k-bounds we will characterize $\overline{\mu}_p(Z_p, \theta|L_k)$. The features of $\underline{\mu}_p(Z_p, \theta|L_k)$ will be analogous after switching the signs of the Δ_p^M's. The problem of computing $\overline{\mu}_p(Z_p, \theta|L_k)$ is a simple linear programming problem consisting of assigning the weights $\big(\pi_{-p}^M(Z_p, \theta|L_k)\big)_{M=0}^{\mathcal{P}-1}$ optimally, according to (10). It is easy to see that these weights should be assigned sequentially according to the index set $\mathcal{J}$. Thus, we begin by assigning $\pi_{-p}^{(1)}(Z_p, \theta|L_k)$, followed by $\pi_{-p}^{(2)}(Z_p, \theta|L_k)$ and so forth until $\pi_{-p}^{(p^*)}(Z_p, \theta|L_k)$. To be precise, we would proceed sequentially by setting $\pi_{-p}^{(j)}(Z_p, \theta|L_k) = \overline{\mathbb{F}}_{-p}^{(j)}(Z_p, \theta|L_{k-1})$ if $\Delta_p^{(j)} > 0$, and $\pi_{-p}^{(j)}(Z_p, \theta|L_k) = \underline{\mathbb{F}}_{-p}^{(j)}(Z_p, \theta|L_{k-1})$ if $\Delta_p^{(j)} < 0$. As we proceed, we must verify that the constraints in (10) are satisfied. Let

$$\underline{\mathbb{P}}_{-p}^0\big(Z_p, \theta\big|L_k\big) = \sum_{M=0}^{\mathcal{P}-1} \underline{\mathbb{F}}_{-p}^M\big(Z_p, \theta\big|L_k\big)\mathbb{1}\big\{\Delta_p^M = 0\big\}; \quad \overline{\mathbb{P}}_{-p}^0\big(Z_p, \theta\big|L_k\big) = \sum_{M=0}^{\mathcal{P}-1} \overline{\mathbb{F}}_{-p}^M\big(Z_p, \theta\big|L_k\big)\mathbb{1}\big\{\Delta_p^M = 0\big\}.$$

The following objects indicate when the constraints in (10) become binding if we set $\pi_{-p}^{(j)}(Z_p,\theta|L_k)$ optimally in the sequential procedure we just described. Let

$$\overline{M}_k(Z_p,\theta) = \text{Max}\big\{1 \leq M \leq p^*\big\} \quad \text{such that:}$$

$$\sum_{j=1}^{M}\Big[\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Big]$$
$$\leq 1-\sum_{j=M+1}^{p^*}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)-\underline{\mathbb{P}}_{-p}^{0}\big(Z_p,\theta\big|L_{k-1}\big),$$

$$\overline{D}_k(Z_p,\theta) = \text{Max}\big\{1 \leq M \leq p^*\big\} \quad \text{such that:}$$

$$\sum_{j=1}^{M}\Big[\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Big]$$
$$\geq 1-\sum_{j=M+1}^{p^*}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)-\overline{\mathbb{P}}_{-p}^{0}\big(Z_p,\theta\big|L_{k-1}\big),$$

$$\text{and let}\quad \overline{M}_k^*(Z_p,\theta)=\text{Min}\Big\{\overline{M}_k(Z_p,\theta)\ ,\ \mathcal{P}-1\Big\},\quad \text{and}\quad \overline{D}_k^*(Z_p,\theta)=\text{Min}\Big\{\overline{D}_k(Z_p,\theta)\ ,\ \mathcal{P}-1\Big\}. \tag{14}$$

For notational simplicity we will denote $\overline{M}_k^*(Z_p,\theta) \equiv \overline{M}_k^*$ and $\overline{M}_k^*(Z_p,\theta) \equiv \overline{D}_k^*$, with the understanding that both objects depend on Z_p and θ. Ranking the action profiles $M=0,\ldots,\mathcal{P}-1$, the object $\overline{M}_k^*+1$ would be the action profile for which the L_k-bounds become binding because we would leave *too little* probability mass (according to the L_k-bounds) to distribute among the subsequent action profiles. $\overline{D}_k^*+1$ would be the action profile for which the L_k-bounds become binding because we would leave *too much* probability mass to distribute among the subsequent action profiles. We will let

$$\overline{U}_k^* = \text{Min}\{\overline{D}_k^*\ ,\ \overline{M}_k^*\}.$$

$\overline{U}_k^*+1$ indicates the strategy profile at which the probability-simplex restriction becomes binding in the construction of $\underline{\mu}_p\big(Z_p,\theta\big|L_k\big)$. For each $j=1,\ldots,\overline{U}_k^*$, we set $\pi_{-p}^{(j)}(Z_p,\theta|L_k)=\overline{\mathbb{F}}_{-p}^{(j)}(Z_p,\theta|L_{k-1})$ if $\Delta_p^{(j)}>0$, and $\pi_{-p}^{(j)}(Z_p,\theta|L_k)=\underline{\mathbb{F}}_{-p}^{(j)}(Z_p,\theta|L_{k-1})$ if $\Delta_p^{(j)}<0$. For $j\geq\overline{U}_k^*+1$ we have the following corner solution:

(A): If $\overline{M}_k^*\leq\overline{D}_k^*$,

$$\pi_{-p}^{(\overline{U}_k^*+1)}(Z_p,\theta|L_k)=\Bigg\{1-\sum_{j=1}^{\overline{U}_k^*}\Big[\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Big]$$
$$-\sum_{j=\overline{U}_k^*+2}^{p^*}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)-\underline{\mathbb{P}}_{-p}^{0}\big(Z_p,\theta\big|L_{k-1}\big)\Bigg\};\quad \pi_{-p}^{(j)}(Z_p,\theta|L_k)=\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\ \forall\ j\geq\overline{U}_k^*+2.$$

(B): If $\overline{M}_k^* > \overline{D}_k^*$,

$$\pi_{-p}^{(\overline{U}_k^*+1)}(Z_p,\theta|L_k) = \Bigg\{1 - \sum_{j=1}^{\overline{U}_k^*}\Big[\mathbb{1}\big\{\Delta_p^{(j)} > 0\big\}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big) + \mathbb{1}\big\{\Delta_p^{(j)} < 0\big\}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Big]$$
$$- \sum_{j=\overline{U}_k^*+2}^{p^*} \overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big) - \overline{\mathbb{P}}_{-p}^{0}\big(Z_p,\theta\big|L_{k-1}\big)\Bigg\}; \quad \pi_{-p}^{(j)}(Z_p,\theta|L_k) = \overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\ \forall\ j \geq \overline{U}_k^* + 2.$$

Note that[3] $\overline{M}_k^* = \overline{D}_k^*$ only if they are both equal to $\mathcal{P}-1$. An immediate consequence is that $\overline{M}_k^* = \overline{D}_k^*$ only if $p^* \geq \mathcal{P}-1$). We have

$$\overline{\mu}_p(Z_p,\theta|L_k) = \sum_{j=1}^{p^*}\Delta_p^{(j)}\pi_{-p}^{(j)}(Z_p,\theta|L_k),$$

where each $\pi_{-p}^{(j)}(Z_p,\theta|L_k)$ is as described above. This will yield a compact expression for $\overline{\mu}_p(Z_p,\theta|L_k)$ in terms of the bounds $\big(\underline{\mathbb{F}}_{-p}^{(j)}(Z_p,\theta|L_{k-1})\ ,\ \overline{\mathbb{F}}_{-p}^{(j)}(Z_p,\theta|L_{k-1})\big)_{j=1}^{p^*}$, $\underline{\mathbb{P}}_{-p}^{0}(Z_p,\theta|L_{k-1})$, and $\overline{\mathbb{P}}_{-p}^{0}(Z_p,\theta|L_{k-1})$. Denote

$$\eta_{p_j}^{a}\big(\Delta_p^{(j)},\overline{D}_k^*,\overline{M}_k^*\big) =$$
$$\mathbb{1}\big\{\overline{M}_k^* \leq \overline{D}_k^*\big\}\mathbb{1}\big\{\Delta_p^{(j)} > 0\big\}\mathbb{1}\big\{j \leq \overline{M}_k^*\big\} + \mathbb{1}\big\{\overline{M}_k^* > \overline{D}_k^*\big\}\Big[\mathbb{1}\big\{\Delta_p^{(j)} > 0\big\} + \mathbb{1}\big\{\Delta_p^{(j)} < 0\big\}\mathbb{1}\big\{j \geq \overline{D}_k^* + 1\big\}\Big],$$
$$\eta_{p_j}^{b}\big(\Delta_p^{(j)},\overline{D}_k^*,\overline{M}_k^*\big) =$$
$$\mathbb{1}\big\{\overline{M}_k^* > \overline{D}_k^*\big\}\mathbb{1}\big\{\Delta_p^{(j)} < 0\big\}\mathbb{1}\big\{j \leq \overline{D}_k^*\big\} + \mathbb{1}\big\{\overline{M}_k^* \leq \overline{D}_k^*\big\}\Big[\mathbb{1}\big\{\Delta_p^{(j)} < 0\big\} + \mathbb{1}\big\{\Delta_p^{(j)} > 0\big\}\mathbb{1}\big\{j \geq \overline{M}_k^* + 1\big\}\Big].$$

Then

$$\overline{\mu}_p\big(Z_p,\theta\big|L_k\big) = \Delta_p^{(\overline{U}_k^*+1)} \cdot \Bigg(1 - \Big[\mathbb{1}\big\{\overline{M}_k^* \leq \overline{D}_k^*\big\}\underline{\mathbb{P}}_{-p}^{0}(Z_p,\theta|L_{k-1}) + \mathbb{1}\big\{\overline{M}_k^* > \overline{D}_k^*\big\}\overline{\mathbb{P}}_{-p}^{0}(Z_p,\theta|L_{k-1})\Big]\Bigg)$$
$$+ \sum_{j=1}^{p^*}\Big(\Delta_p^{(j)} - \Delta_p^{(\overline{U}_k^*+1)}\Big) \cdot \Bigg(\eta_{p_j}^{a}\big(\Delta_p^{(j)},\overline{D}_k^*,\overline{M}_k^*\big)\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big) + \eta_{p_j}^{b}\big(\Delta_p^{(j)},\overline{D}_k^*,\overline{M}_k^*\big)\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Bigg). \tag{15}$$

[3] Otherwise there would exist a nonempty subset of actions $\mathcal{M}$ for which $\sum_{M\in\mathcal{M}}\underline{\mathbb{F}}_{-p}^{M}\big(Z_p,\theta\big|L_k\big) > \sum_{M\in\mathcal{M}}\overline{\mathbb{F}}_{-p}^{M}\big(Z_p,\theta\big|L_k\big)$, which is impossible even without strategic interaction across the players.

To derive the equivalent expression for $\underline{\mu}_p\big(Z_p,\theta\big|L_k\big)$ we proceed as above after switching the signs of the Δ_p^M's. Let

$$\underline{M}_k(Z_p,\theta) = \text{Max}\big\{1 \le M \le p^*\big\} \quad \text{such that:}$$

$$\sum_{j=1}^{M}\Big[\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Big]$$
$$\le 1-\sum_{j=M+1}^{p^*}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)-\underline{\mathbb{P}}_{-p}^{0}\big(Z_p,\theta\big|L_{k-1}\big),$$

$$\underline{D}_k(Z_p,\theta) = \text{Max}\big\{1 \le M \le p^*\big\} \quad \text{such that:}$$

$$\sum_{j=1}^{M}\Big[\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Big]$$
$$\ge 1-\sum_{j=M+1}^{p^*}\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)-\overline{\mathbb{P}}_{-p}^{0}\big(Z_p,\theta\big|L_{k-1}\big),$$

and let $\underline{M}_k^*(Z_p,\theta) = \text{Min}\Big\{\underline{M}_k(Z_p,\theta)\ ,\ \mathcal{P}-1\Big\}$, and $\overline{D}_k^*(Z_p,\theta) = \text{Min}\Big\{\underline{D}_k(Z_p,\theta)\ ,\ \mathcal{P}-1\Big\}$. (16)

and define $\underline{U}_k^* = \text{Min}\big\{\underline{D}_k^*\ ,\ \underline{M}_k^*\big\}$. Letting

$$\eta_{p_j}^c\big(\Delta_p^{(j)},\underline{D}_k^*,\underline{M}_k^*\big) =$$
$$\mathbb{1}\big\{\underline{M}_k^*\le\underline{D}_k^*\big\}\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\mathbb{1}\big\{j\le\underline{M}_k^*\big\}+\mathbb{1}\big\{\underline{M}_k^*>\underline{D}_k^*\big\}\Big[\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}+\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\mathbb{1}\big\{j\ge\underline{D}_k^*+1\big\}\Big]$$
$$\eta_{p_j}^d\big(\Delta_p^{(j)},\underline{D}_k^*,\underline{M}_k^*\big) =$$
$$\mathbb{1}\big\{\underline{M}_k^*>\underline{D}_k^*\big\}\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}\mathbb{1}\big\{j\le\underline{D}_k^*\big\}+\mathbb{1}\big\{\underline{M}_k^*\le\underline{D}_k^*\big\}\Big[\mathbb{1}\big\{\Delta_p^{(j)}>0\big\}+\mathbb{1}\big\{\Delta_p^{(j)}<0\big\}\mathbb{1}\big\{j\ge\underline{M}_k^*+1\big\}\Big],$$

we can express

$$\underline{\mu}_p\big(Z_p,\theta\big|L_k\big)=\Delta_p^{(\underline{U}_k^*+1)}\cdot\left(1-\Big[\mathbb{1}\big\{\underline{M}_k^*\le\underline{D}_k^*\big\}\underline{\mathbb{P}}_{-p}^{0}(Z_p,\theta|L_{k-1})+\mathbb{1}\big\{\underline{M}_k^*>\underline{D}_k^*\big\}\overline{\mathbb{P}}_{-p}^{0}(Z_p,\theta|L_{k-1})\Big]\right)$$
$$+\sum_{j=1}^{p^*}\Big(\Delta_p^{(j)}-\Delta_p^{(\underline{U}_k^*+1)}\Big)\cdot\left(\eta_{p_j}^c\big(\Delta_p^{(j)},\underline{D}_k^*,\underline{M}_k^*\big)\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\eta_{p_j}^d\big(\Delta_p^{(j)},\underline{D}_k^*,\underline{M}_k^*\big)\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\right). \quad (17)$$

Below, we will use the expressions in (15) and (17) to estimate the L_k−rational bounds. Both $\underline{\mu}_p(Z_p,\theta|L_k)$ and $\overline{\mu}_p(Z_p,\theta|L_k)$ are the solution to a simple, continuous linear programming problem. The next remark enumerates some of smoothness properties that result from this fact.

Remark 3 (Smoothness features of $\underline{\mu}_p(Z_p,\theta|L_k)$ and $\overline{\mu}_p(Z_p,\theta|L_k)$) As before, let $\varepsilon_{-p}\equiv\big(\varepsilon_q\big)_{q\neq p}$. Fix a realization $Z_p\in\mathbb{S}(Z_p)$ and let $G_{-p}(\cdot|Z_p)$ denote the joint conditional distribution of $\varepsilon_{-p}|Z_p$. If $G_{-p}(\cdot|Z_p)$ is a continuous, then $\underline{\mu}_p(Z_p,\cdot|L_k)$ and $\overline{\mu}_p(Z_p,\cdot|L_k)$ are continuous. Suppose $G_{-p}(\cdot|Z_p)$ is continuously differentiable. Then $\underline{\mu}_p(Z_p,\cdot|L_k)$ and $\overline{\mu}_p(Z_p,\cdot|L_k)$ are Lipschitz-continuous. Furthermore, they inherit *all* the differentiable properties of $G_{-p}(\cdot|Z_p)$ for all values of θ, except perhaps those where either of the inequalities in (14) holds with exact equality (for $\overline{\mu}_p(Z_p,\cdot|L_k)$) and those where either of the inequalities in (16) holds with exact equality (for $\underline{\mu}_p(Z_p,\cdot|L_k)$). However, as stated above, both $\underline{\mu}_p(Z_p,\cdot|L_k)$ and $\overline{\mu}_p(Z_p,\cdot|L_k)$ are still Lipschitz-continuous for such values.

4 "Dominance types" and $D_{k^*}-$rationality

Borrowing from the behavioral game-theory literature, we will refer to a dominance type of player as one who best-responds to some beliefs that survive a finite number of steps of iterated deletion of dominated strategies. According to our definitions, a dominance type who performs $k-1$ steps of iterated thinking is an L_k-rational player. Going back to our derivation, the set of L_k-rational beliefs $\Pi_p(Z_p,\theta|L_k)$ is convex (it is the intersection between a probability simplex and a hypercube). Thus, all rationalizable beliefs for $\sum_{M=0}^{\mathcal{P}-1}\Delta_p^M\pi_{-p}^M(Z_p)$ that are compatible with at least $k-1$ rounds of iterated deletion of dominated strategies can always be expressed as a convex combination of the lower and upper bounds $\underline{\mu}_p(Z_p,\theta|L_k)$ and $\overline{\mu}_p(Z_p,\theta|L_k)$. It follows that the best-response rule of any L_k-rational player can be expressed as[4]

$$Y_p = \mathbb{1}\Big\{X_p'\beta_p + \xi_p\underline{\mu}_p(Z_p,\theta|L_k) + (1-\xi_p)\overline{\mu}_p(Z_p,\theta|L_k) - \varepsilon_p \geq 0\Big\}, \quad \text{where} \ \ \xi_p \in [0,1].$$

This is the type of optimal decision rule that we will assume for what we will refer to as "dominance types". We will focus on a particular type of behavior where k is unknown, but assumed to be bounded above by some pre-specified k^*. The weights ξ_p are unobserved, but will be assumed to be deterministic conditional on what we will call a "rationality control function", observable. The result will be a generalization of dominance behavior, which we will refer to as $D_{k^*}-$rationality.

4.1 D_{k^*}-rationality

In their paper, Costa-Gomes and Crawford (2006) study dominance type of behavior, and define a D_k-rational player as one who performs performs k steps of iterated thinking and best-respond to

[4]See Remark 2.

a uniform distribution over the surviving beliefs. Using their definition, a D_k-player would follow the decision rule

$$Y_p = \mathbb{1}\Big\{X_p'\beta_p + \frac{1}{2}\Big[\underline{\mu}_p(Z_p,\theta\big|L_k) + \overline{\mu}_p(Z_p,\theta\big|L_k)\Big] - \varepsilon_p \geq 0\Big\}.$$

Here, we will define a $D_{k^*}-$rational decision rule as

$$Y_p = \mathbb{1}\Big\{X_p'\beta_p + \sum_{k=1}^{k^*}\Big[\underline{\xi}_p^k\cdot\underline{\mu}_p(Z_p,\theta\big|L_k) + \overline{\xi}_p^k\cdot\overline{\mu}_p(Z_p,\theta\big|L_k)\Big] - \varepsilon_p \geq 0\Big\},$$
$$\text{where } \underline{\xi}_p^k \in [0,1] \text{ and } \overline{\xi}_p^k \in [0,1] \text{ for each } k, \text{ and } \sum_{k=1}^{k^*}\Big(\underline{\xi}_p^k + \overline{\xi}_p^k\Big) = 1. \tag{18}$$

Thus, a $D_{k^*}-$rational player performs anything in between zero and k^* steps of iterated thinking, and best responds to some convex-combination of the rationalizable bounds. More generally, going back to Remark 2, a $D_{k^*}-$rational player may have higher order beliefs that differ across his opponents in the number of rounds of deletion of dominated strategies that each one of them computes. The implicit restriction in Equation (18) is that he believes that, with probability one, at least one of his opponents performs no more than k^*-1 rounds of deletion. It also follows from Proposition 2 that BNE beliefs are a special case of $D_{k^*}-$rationality.

Assumption D1 (Rationality control function).- For each player p, there exists a random variable r_p which we will call a rationality control function such that: (i) Each one of the weights $\underline{\xi}_p^k$ and $\overline{\xi}_p^k$ in (18) is deterministic conditional on r_p. (ii) $X_p \perp \varepsilon_p \big| r_p$ (X_p is independent of ε_p conditional on r_p). (iii) r_p is observable to the econometrician.

A special case of Assumption (D1) is when the weights $\underline{\xi}_p^k$ and $\overline{\xi}_p^k$ are unknown constants. From Assumption (D1), it follows that for a $D_{k^*}-$rational player,

$$E\Big[Y_p\Big|r_p, X_p, Z_p\Big] = E\Big[Y_p\Big|r_p, X_p'\beta, \Big(\underline{\mu}_p(Z_p,\theta\big|L_k)\ ,\ \overline{\mu}_p(Z_p,\theta\big|L_k)\Big)_{k=1}^{k^*}\Big] \tag{19}$$

Our identification strategy will be based on the exclusion restriction described in (19) plus a set of additional assumptions to be described below.

5 Identification of θ based on $D_{k^*}-$rationality

We present a set of conditions about the parameter space and the covariates involved under which identification is possible based only on the characterization of $D_{k^*}-$rationality under the conditions described above. As it is the case in the usual equilibrium models, *strategic interaction is a powerful identification mechanism.* From here on, we will let

$$\Theta \equiv \text{Parameter space, assumed to be compact, and } \ \theta_0 \equiv \text{True value of } \ \theta.$$

We will make the following assumption.

Assumption D2

1.– Z_p is observed by the econometrician. As before, we let $\varepsilon_{-p} \equiv \big(\varepsilon_q\big)_{q \neq p}$. Then, for every player p we have $\varepsilon_p \perp \varepsilon_{-p} \,\big| Z_p$ (conditional on Z_p, ε_p contains no information about ε_{-p}). As before, denote $X \equiv \underset{p}{\cup} X_p$, $Z \equiv \underset{p}{\cup} Z_p$ and $\varepsilon \equiv \big(\varepsilon_p\big)_{p=1}^{\mathcal{P}}$. The distribution of $\varepsilon \big| X, Z$ is assumed known by the econometrician, and the scale of each ε_p is normalized to one.

2.– For each player $q = 1, \ldots, \mathcal{P}$, the vector X_q has full column rank with positive probability, and there exists a publicly observed regressor $X_{q_\ell} \in X_q$ with $\beta_{q_{\ell_0}} \neq 0$, continuously distributed conditional on all other covariates in the model. Because it is publicly observed, it satisfies $X_{q_\ell} \in Z_p$ for all q and p. All the signals used by any player p are *informative*, in the sense that:

$$\nexists \ \widetilde{Z}_p \subset Z_p : \ \underline{\mu}_p(\widetilde{Z}_p, \theta_0 | L_k) = \underline{\mu}_p(Z_p, \theta_0 | L_k) \ \text{ and } \ \overline{\mu}_p(\widetilde{Z}_p, \theta_0 | L_k) = \overline{\mu}_p(Z_p, \theta_0 | L_k) \text{ w.p.1 for some } k \geq 2.$$

In addition, the dimensionality of Z_p satisfies $\dim(Z_p) \geq 2k^* - 1$ for every player p.

3.– For each player p, either the intercept in β_p is normalized to zero, or $\Delta_p^{\widetilde{M}} = 0$ for some $\widetilde{M}$. In addition, for each player p we have $\underline{\xi}_{p_1}^k(r_p) > 0$ or $\overline{\xi}_{p_1}^k(r_p) > 0$ w.p.1 for some $k \geq 2$. In words, this means that each player's behavior is compatible with assuming that, with strictly positive probability for all realizations of r_{p_1} and r_{p_2}, their opponents perform at least one round of deletion of dominated strategies. There exist (at least) two players, p_1 and p_2 for whom

$$\min_M \big(\Delta_{p_1}^M\big) \neq \max_M \big(\Delta_{p_1}^M\big) \quad \text{and} \quad \min_M \big(\Delta_{p_2}^M\big) \neq \max_M \big(\Delta_{p_2}^M\big).$$

4.– All players $p = 1, \ldots, \mathcal{P}$ are $D_{k^*}-$rational as described in Subsection 4.1 and the upper bound k^* is assumed known to the econometrician. Lastly, neither $X_p'\beta_0$ nor Z_p are deterministic conditional on r_p.

Part 1 of Assumption (D2) makes it possible to semiparametrically identify the bounds $\underline{\mu}_p(Z_p, \theta | L_k)$ and $\overline{\mu}_p(Z_p, \theta | L_k)$ for any given θ and k. Part 2 provides sufficient conditions to ensure that, for any $\theta \neq \theta_0$, there does not exist a deterministic mapping that recovers $\big(\underline{\mu}_p(Z_p, \theta_0 | L_k) \ , \ \overline{\mu}_p(Z_p, \theta_0 | L_k)\big)_{k=2}^{k^*}$

from $\left(\underline{\mu}_p(Z_p,\theta|L_k)\ ,\ \overline{\mu}_p(Z_p,\theta|L_k)\right)_{k=2}^{k^*}$ w.p.1. These sufficient conditions involve restrictions on the number of signals in Z_p and the fact that, evaluated at the true parameter value θ_0, not a single component of Z_p is redundant conditional on the others as a predictor of the L_k-rational bounds. The first condition in Part 2 is also meant to ensure that $X_p'\beta_0$ is not deterministic conditional on $\left(\underline{\mu}_p(Z_p,\theta|L_k)\ ,\ \overline{\mu}_p(Z_p,\theta|L_k)\right)_{k=2}^{k^*}$. Part 3 assumes that there exist at least two players whose payoffs are, in fact, affected by the actions of their opponents. In other words, strategic interaction is a relevant component of the model for at least two players. Moreover, these players' behavior is compatible with assuming that, with strictly positive probability for all realizations of r_{p_1} and r_{p_2}, their opponents perform at least one round of deletion of dominated strategies. This strategic interaction effect and the "rationality level" of these two players, combined with the other parts of Assumption (D2) will provide a powerful identification mechanism to detect impostor values of θ in $\left(\underline{\mu}_{p_1}(Z_{p_1},\theta|L_k)\ ,\ \overline{\mu}_{p_1}(Z_{p_1},\theta|L_k)\right)_{k=2}^{k^*}$ or $\left(\underline{\mu}_{p_2}(Z_{p_2},\theta|L_k)\ ,\ \overline{\mu}_{p_2}(Z_{p_2},\theta|L_k)\right)_{k=2}^{k^*}$. Combined with parts 1-3, last condition in part 4 is meant to ensure that neither $\left(\underline{\mu}_p(Z_p,\theta_0|L_k)\ ,\ \overline{\mu}_p(Z_p,\theta_0|L_k)\right)_{k=2}^{k^*}$, nor $X_p'\beta_0$ can be recovered in a deterministic way conditional on r_p for any player p.

Proposition 4 ***(Identification)*** *Let* $\mathcal{T}_p^*(\theta) \equiv \left(X_p'\beta_p, \left(\underline{\mu}_p\big(Z_p,\theta\big|L_k\big)\ ,\ \overline{\mu}_p\big(Z_p,\theta\big|L_k\big)\right)_{k=2}^{k^*}\right)'$. *Then, if Assumptions D1 and D2 holds,*

$$E\Big[Y_p\Big|X_p,Z_p,r_p\Big] = E\Big[Y_p\Big|\mathcal{T}_p^*(\theta),r_p\Big] \quad \textit{w.p.1 for each player } p=1,\ldots,\mathcal{P} \textit{ if and only if } \theta=\theta_0.$$

The proof of Proposition 4 is in the appendix. Aside from the features of $D_{k^*}-$rationality, a powerful mechanism for identification in this case is the strategic-interaction nature of the model. This effect would be absent , for example, if none of the players performed any round of deletion of dominated strategies, this is why we need the last part of Assumption (D2.3). Proposition 4 is a constructive result because $E\big[Y_p\big|X_p,Z_p,r_p\big]$ is nonparametrically identified and, given our assumptions, $\mathcal{T}_p^*(\theta)$ is also identified (i.e, it can be consistently estimated) for any *given* θ. The following section describes an estimation procedure based on this identification result.

6 Estimation

If X_p is publicly observed by all players and ε_p is the only source of private information for players' payoffs, knowing the true distribution of $\varepsilon_{-p}|Z_p$ (Assumption D2.1) implies that we can recover the exact expression for $\underline{\mu}_{-p}(Z_p,\theta|L_k)$ and $\overline{\mu}_{-p}(Z_p,\theta|L_k)$ for every Z_p, θ and k. In such a case, we would have $Z_p = X_{-p}$. The next subsection studies the case where some components of X_q are private information for some player q, so that $Z_p \neq X_{-p}$ for some p. In such a case, we will focus on the case where Z_p is a continuous random vector.

6.1 Estimation of $\underline{\boldsymbol{\mu}}_{-p}(Z_p, \theta | L_k)$ and $\overline{\boldsymbol{\mu}}_{-p}(Z_p, \theta | L_k)$ when $Z_p \neq X_{-p}$

As we mentioned above, given our previous assumptions, this subsection is irrelevant in its entirety for the case $Z_p = X_{-p}$ for all p (every component in X_q is public information for all q). In such a case, the reader can skip directly to Subsection 6.2 below. Here, we will describe the proposed estimator for the case when some elements in X_{-p} are not observed by Player p or, more generally, they are ignored by Player p in the construction of his beliefs (i.e, they are excluded from Z_p). The estimator we propose follows the analog principle by replacing unknown probabilities with semiparametric estimators in the construction described in Subsection 3.3. For a given θ and $Z_p \in \mathbb{S}(Z_p)$, let $\widehat{\underline{\mathbb{F}}}^M_{-p}(Z_p, \theta \big| L_k)$ and $\widehat{\overline{\mathbb{F}}}^M_{-p}(Z_p, \theta \big| L_k)$ be semiparametric estimators of

$$\underline{\mathbb{F}}^M_{-p}(Z_p, \theta \big| L_k) = E\big[\underline{\mathbb{Y}}^M_{-p}(L_k) \big| Z_p\big] \quad \text{and} \quad \overline{\mathbb{F}}^M_{-p}(Z_p, \theta \big| L_k) = E\big[\overline{\mathbb{Y}}^M_{-p}(L_k) \big| Z_p\big].$$

The construction of these semiparametric estimators is described in Subsection 6.1.1, below. As in Subsection 3.3 we will let p^* denote the number of nonzero elements in $\big(\Delta^M_p\big)^{\mathcal{P}-1}_{M=0}$ and we will let $\mathcal{J}$ be an index set as the one described in Equation (13). Using these estimators, we will use the following semiparametric analogs to $\overline{M}_k(Z_p, \theta)$ and $\overline{D}_k(Z_p, \theta)$ as defined in Equation (14),

$$\begin{aligned}
&\widehat{\overline{M}}_k(Z_p, \theta) = \text{Max}\big\{1 \le M \le p^*\big\} \quad \text{such that:} \\
&\sum_{j=1}^{M} \Big[\mathbb{1}\big\{\Delta^{(j)}_p > 0\big\} \widehat{\overline{\mathbb{F}}}^{(j)}_{-p}(Z_p, \theta \big| L_{k-1}) + \mathbb{1}\big\{\Delta^{(j)}_p < 0\big\} \widehat{\underline{\mathbb{F}}}^{(j)}_{-p}(Z_p, \theta \big| L_{k-1}) \Big] \\
&\qquad\qquad\qquad\qquad \le 1 - \sum_{j=M+1}^{p^*} \widehat{\underline{\mathbb{F}}}^{(j)}_{-p}(Z_p, \theta \big| L_{k-1}) - \widehat{\underline{\mathbb{P}}}^{0}_{-p}(Z_p, \theta \big| L_{k-1}), \\
&\widehat{\overline{D}}_k(Z_p, \theta) = \text{Max}\big\{1 \le M \le p^*\big\} \quad \text{such that:} \\
&\sum_{j=1}^{M} \Big[\mathbb{1}\big\{\Delta^{(j)}_p > 0\big\} \widehat{\overline{\mathbb{F}}}^{(j)}_{-p}(Z_p, \theta \big| L_{k-1}) + \mathbb{1}\big\{\Delta^{(j)}_p < 0\big\} \widehat{\underline{\mathbb{F}}}^{(j)}_{-p}(Z_p, \theta \big| L_{k-1}) \Big] \\
&\qquad\qquad\qquad\qquad \ge 1 - \sum_{j=M+1}^{p^*} \widehat{\overline{\mathbb{F}}}^{(j)}_{-p}(Z_p, \theta \big| L_{k-1}) - \widehat{\overline{\mathbb{P}}}^{0}_{-p}(Z_p, \theta \big| L_{k-1}), \\
&\text{and let} \quad \widehat{\overline{M}}^*_k(Z_p, \theta) = \text{Min}\Big\{\widehat{\overline{M}}_k(Z_p, \theta)\ ,\ \mathcal{P}-1\Big\}, \quad \text{and} \quad \widehat{\overline{D}}^*_k = \text{Min}\Big\{\overline{D}_k(Z_p, \theta)\ ,\ \mathcal{P}-1\Big\}.
\end{aligned} \tag{20}$$

If there exists no $1 \le M \le p^*$ that satisfies the conditions for $\widehat{\overline{M}}_k(Z_p,\theta)$, we will set

$$\widehat{\overline{M}}_k(Z_p,\theta) = 1,$$

and we will do the same with $\widehat{\overline{D}}_k(Z_p,\theta)$. For notational simplicity we will write $\widehat{\overline{M}}_k(Z_p,\theta) \equiv \widehat{\overline{M}}_k$ and $\widehat{\overline{D}}_k(Z_p,\theta) \equiv \widehat{\overline{D}}_k$, and will revisit the properties of these objects as functions of θ and Z_p when we characterize their asymptotic properties. Along the same lines, we estimate $\underline{M}_k(Z_p,\theta)$ and $\underline{D}_k(Z_p,\theta)$ with $\widehat{\underline{M}}_k(Z_p,\theta)$ and $\widehat{\underline{D}}_k(Z_p,\theta)$ respectively, by using sample analogs to Equation (16). We estimate $\overline{\mu}_p\big(Z_p,\theta\big|L_k\big)$ by using a sample analog to Equation (15),

$$\widehat{\overline{\mu}}_p\big(Z_p,\theta\big|L_k\big) = \Delta_p^{(\widehat{\overline{U}}^*_k+1)} \cdot \left(1 - \Big[\mathbb{1}\big\{\widehat{\overline{M}}^*_k \le \widehat{\overline{D}}^*_k\big\}\widehat{\underline{\mathbb{P}}}^0_{-p}(Z_p,\theta|L_{k-1}) + \mathbb{1}\big\{\widehat{\overline{M}}^*_k > \widehat{\overline{D}}^*_k\big\}\widehat{\overline{\mathbb{P}}}^0_{-p}(Z_p,\theta|L_{k-1})\Big]\right)$$
$$+ \sum_{j=1}^{p^*}\Big(\Delta_p^{(j)} - \Delta_p^{(\widehat{\overline{U}}^*_k+1)}\Big) \cdot \left(\eta^a_{p_j}\big(\Delta_p^{(j)},\widehat{\overline{D}}^*_k,\widehat{\overline{M}}^*_k\big)\widehat{\overline{\mathbb{F}}}^{(j)}_{-p}\big(Z_p,\theta\big|L_{k-1}\big) + \eta^b_{p_j}\big(\Delta_p^{(j)},\widehat{\overline{D}}^*_k,\widehat{\overline{M}}^*_k\big)\widehat{\underline{\mathbb{F}}}^{(j)}_{-p}\big(Z_p,\theta\big|L_{k-1}\big)\right). \tag{21}$$

Similarly, using Equation (17),

$$\widehat{\underline{\mu}}_p\big(Z_p,\theta\big|L_k\big) = \Delta_p^{(\widehat{\underline{U}}^*_k+1)} \cdot \left(1 - \Big[\mathbb{1}\big\{\widehat{\underline{M}}^*_k \le \widehat{\underline{D}}^*_k\big\}\widehat{\underline{\mathbb{P}}}^0_{-p}(Z_p,\theta|L_{k-1}) + \mathbb{1}\big\{\widehat{\underline{M}}^*_k > \widehat{\underline{D}}^*_k\big\}\widehat{\overline{\mathbb{P}}}^0_{-p}(Z_p,\theta|L_{k-1})\Big]\right)$$
$$+ \sum_{j=1}^{p^*}\Big(\Delta_p^{(j)} - \Delta_p^{(\widehat{\underline{U}}^*_k+1)}\Big) \cdot \left(\eta^c_{p_j}\big(\Delta_p^{(j)},\widehat{\underline{D}}^*_k,\widehat{\underline{M}}^*_k\big)\widehat{\overline{\mathbb{F}}}^{(j)}_{-p}\big(Z_p,\theta\big|L_{k-1}\big) + \eta^d_{p_j}\big(\Delta_p^{(j)},\widehat{\underline{D}}^*_k,\widehat{\underline{M}}^*_k\big)\widehat{\underline{\mathbb{F}}}^{(j)}_{-p}\big(Z_p,\theta\big|L_{k-1}\big)\right). \tag{22}$$

6.1.1 Estimation of $\underline{\mathbb{F}}^M_{-p}\big(Z_p,\theta\big|L_k\big)$ and $\overline{\mathbb{F}}^M_{-p}\big(Z_p,\theta\big|L_k\big)$.

We will use semiparametric estimators, obtained recursively according to sample analogs to Equations (8) – (11) in Subsection 3.2. Let us group

$$\overline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k) \equiv \Big(\overline{\mu}_q(Z_q,\theta|L_k)\Big)_{q\neq p}, \quad \underline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k) \equiv \Big(\underline{\mu}_q(Z_q,\theta|L_k)\Big)_{q\neq p}.$$

Using this notation, we will express

$$\begin{aligned} E\Big[\underline{\mathbb{Y}}^M_{-p}(L_k)\Big|X_{-p},Z_{-p},Z_p\Big] &= \underline{G}^M_{-p}\Big(\big(X'_q\beta_q\big)_{q\neq p}\ ,\ \overline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ,\ \underline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ;\ Z_p\Big),\\ E\Big[\overline{\mathbb{Y}}^M_{-p}(L_k)\Big|X_{-p},Z_{-p},Z_p\Big] &= \overline{G}^M_{-p}\Big(\big(X'_q\beta_q\big)_{q\neq p}\ ,\ \overline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ,\ \underline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ;\ Z_p\Big). \end{aligned} \tag{23}$$

The exact expressions for $\underline{G}^M_{-p}(\cdot,\cdot,\cdot)$ and $\overline{G}^M_{-p}(\cdot,\cdot,\cdot)$ are determined by the distribution of $\varepsilon_{-p}|X_{-p},Z_{-p},Z_p$ and by the events defined in Equations (9) – (10). Using iterated expectations,

$$\underline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_k) = E\Big[\underline{G}^M_{-p}\Big(\big(X'_q\beta_q\big)_{q\neq p}\ ,\ \overline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ,\ \underline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ;\ Z_p\Big)\ \Big|Z_p\Big],$$
$$\overline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_k) = E\Big[\overline{G}^M_{-p}\Big(\big(X'_q\beta_q\big)_{q\neq p}\ ,\ \overline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ,\ \underline{\mathbf{\Lambda}}_{-p}(Z_{-p},\theta|L_k)\ ;\ Z_p\Big)\ \Big|Z_p\Big].$$

We have assumed that the distribution of $\varepsilon_{-p}|X_{-p}, Z_{-p}, Z_p$ is known. For given θ and any k, this assumption will allow us to estimate $\underline{\mathbb{F}}^M_{-p}(Z_p, \theta|L_k)$ and $\overline{\mathbb{F}}^M_{-p}(Z_p, \theta|L_k)$ recursively, starting with $k = 1$. We proceed by replacing $\underline{\mathbf{\Lambda}}_{-p}(Z_{-p}, \theta|L_k)$ and $\overline{\mathbf{\Lambda}}_{-p}(Z_{-p}, \theta|L_k)$ with their corresponding semiparametric estimates, which are in turn computed recursively as described in Subsection 6.1. We employ a collection of bandwidth sequences

$$\big\{h_k\big\}_{k=2}^{k^*}, \quad \text{each of which converges to zero as } N \longrightarrow \infty.$$

We will be precise about their relative rates of convergence below. Let d_p denote the dimension of Z_p and let $K^p : \mathbb{R}^{d_p} \longrightarrow \mathbb{R}$ be a kernel function (we will be precise about its properties below). We will denote

$$K^p_{h_k}(\psi) \equiv K^p\bigg(\frac{\psi}{h_k}\bigg).$$

Beginning with $k = 1$, the estimators are

$$\widehat{\underline{\mathbb{F}}}^M_{-p}(Z_p, \theta|L_k) = \frac{1}{\widehat{f}_{Z_p}(Z_p)} \frac{1}{N h_k^{d_p}} \sum_{\ell=1}^{N} \underline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\big)_{q\neq p}\ , \widehat{\overline{\mathbf{\Lambda}}}_{-p}(Z_{-p\ell}, \theta|L_k)\ , \widehat{\underline{\mathbf{\Lambda}}}_{-p}(Z_{-p\ell}, \theta|L_k)\ ;\ Z_p\Big) K_{h_k}\big(Z_{p\ell} - Z_p\big),$$

$$\widehat{\overline{\mathbb{F}}}^M_{-p}(Z_p, \theta|L_k) = \frac{1}{\widehat{f}_{Z_p}(Z_p)} \frac{1}{N h_k^{d_p}} \sum_{\ell=1}^{N} \overline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\big)_{q\neq p}\ , \widehat{\overline{\mathbf{\Lambda}}}_{-p}(Z_{-p\ell}, \theta|L_k)\ , \widehat{\underline{\mathbf{\Lambda}}}_{-p}(Z_{-p\ell}, \theta|L_k)\ ;\ Z_p\Big) K_{h_k}\big(Z_{p\ell} - Z_p\big), \tag{24}$$

with

$$\widehat{\overline{\mathbf{\Lambda}}}_{-p}(Z_{-p}, \theta|L_k) \equiv \Big(\max_M\big(\Delta^M_q\big)\Big)_{q\neq p}, \quad \widehat{\underline{\mathbf{\Lambda}}}_{-p}(Z_{-p}, \theta|L_k) \equiv \Big(\min_M\big(\Delta^M_q\big)\Big)_{q\neq p} \quad \text{for } k = 1,$$
$$\widehat{\overline{\mathbf{\Lambda}}}_{-p}(Z_{-p}, \theta|L_k) \equiv \Big(\widehat{\overline{\mu}}_q(Z_q, \theta|L_k)\Big)_{q\neq p}, \quad \widehat{\underline{\mathbf{\Lambda}}}_{-p}(Z_{-p}, \theta|L_k) \equiv \Big(\widehat{\underline{\mu}}_q(Z_q, \theta|L_k)\Big)_{q\neq p} \quad \text{for } k \geq 2,$$

where $\widehat{\overline{\mu}}_q(Z_q, \theta|L_k)$ and $\widehat{\underline{\mu}}_q(Z_q, \theta|L_k)$ are computed as in Equations (21) and (22).

Assumption E1 As we mentioned at the beginning of this section, we maintain Assumptions D1 and D2, both of which led to Proposition 4. We introduce the following assumptions.

1.– We observe a random sample $\Big(\big(Y_{p_n}, X_{p_n}, Z_{p_n}\big)_{p=1}^{\mathcal{P}}\Big)_{n=1}^{N}$ from a population of players whose behavior satisfy the assumptions leading to Proposition 4.

2.− We will strengthen the first part of Assumption D2.3, and will assume that

$$\min_{M}\left(\Delta_p^M\right) \neq \max_{M}\left(\Delta_p^M\right) \quad \text{for all} \;\; p = 1, \ldots, \mathcal{P}.$$

3.− The signal vector Z_p is continuously distributed, with unknown density $f_{Z_p}(\cdot)$. Let $\underline{s}$ denote a pre-specified, strictly positive constant. We will assume that $\Pr\Big(f_{Z_p}(Z_p) = \underline{s}\Big) = 0$ for every p, and we will define

$$\mathcal{Z}_p = \Big\{z_p : \; f_{Z_p}(z) \geq \underline{s}\Big\}.$$

The features of the set $\mathcal{Z}_p$ will be important to us because we will ultimately use a trimming function $\mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\Big\}$ in the estimation of θ. We denote the dimension of the vector Z_p by d_p, and we let $\overline{d} = \max\big\{d_p : p = 1, \ldots, \mathcal{P}\big\}$. There exists an $\overline{M} \geq \overline{d} + 1$ such that, for every player p, the conditional density $f_{X_{-p}|Z_p}(x_{-p}|z_p)$ is $\overline{M}-$times differentiable with bounded derivatives with respect to all the elements in Z_p which are not included in X_{-p}. For each p, let $G_{\varepsilon_{-p}|Z_p}\big(\epsilon_{-p}|z_p\big)$ denote the (assumed to be known) distribution of ε_{-p} given Z_p. Then $G_{\varepsilon_{-p}|Z_p}(\cdot|\cdot)$ is $\overline{M}-$times differentiable with bounded derivatives (see parts 2 and 3 of Assumption D2). We assume that $E\big[\|XX'\|^2\big] < \infty$, where $X = \bigcup_{p=1}^{\mathcal{P}} X_p$.

4.− For each p, the kernel $K^p(\cdot)$ is bounded, symmetric, bias-reducing of order $\overline{M}$ and has compact support. It is also $\overline{M}$ times differentiable with bounded derivatives. Each one of the bandwidth sequences $\big(h_k\big)_{k=1}^{k^*}$ satisfies $N^{1-\overline{\epsilon}} h_k^{2d^*} \longrightarrow \infty$ and $N^{1/2}\big(h_k/h_{k+1}\big)^{\overline{M}} \longrightarrow 0$, where we defined above d^* as the maximum dimension of Z_p over all players.

Part 1 of Assumption (E1) will enable us to invoke the relevant laws of large numbers and central limit theorems in our asymptotic analysis. To illustrate what parts 2 and 3 are meant to accomplish, let $\mathbb{H}_{-p}^M\big(Z_p, \theta\big|L_k\big)$ denote either $\underline{\mathbb{F}}_{-p}^M\big(Z_p, \theta\big|L_k\big)$ or $\overline{\mathbb{F}}_{-p}^M\big(Z_p, \theta\big|L_k\big)$. Take any linear combination of the form $\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}_{-p}^M\big(Z_p, \theta\big|L_k\big)$. Parts 2 and 3 of Assumption (E1) are meant to ensure that

$$\Pr\left[\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}_{-p}^M\big(Z_p, \theta\big|L_k\big) = 1 \middle| Z_p \in \mathcal{Z}_p\right] = 0 \quad \text{for} \;\; k = 1, \ldots, k^* - 1. \tag{25}$$

This follows because, if part 2 holds,

$$\sum_{M=0}^{\mathcal{P}-1} \underline{\mathbb{F}}_{-p}^M\big(Z_p, \theta\big|L_1\big) < 1, \quad \text{and} \quad \sum_{M=0}^{\mathcal{P}-1} \overline{\mathbb{F}}_{-p}^M\big(Z_p, \theta\big|L_1\big) > 1 \quad \text{for almost all realizations of } Z_p.$$

Given the smoothness conditions in part 3, the relative rates of under/over smoothing described in part 4 will be necessary to achieve $\sqrt{N}-$consistency for our estimator of θ. Relative rates of convergence for the bandwidths involved must be chosen carefully, due to the iterative nature of $\left(\widehat{\underline{\mathbb{F}}}^{M}_{-p}(\cdot,\theta\big|L_k)\ ,\ \widehat{\overline{\mathbb{F}}}^{M}_{-p}(\cdot,\theta\big|L_k)\right)_{M=0}^{\mathcal{P}-1}$, where semiparametric estimates are plugged-in sequentially (see Equation 24), from $k=2,\ldots,k^*$.

6.2 Estimation of $\widehat{\boldsymbol{\theta}}$

We will maintain all the assumptions that led to Proposition 4. This includes of course, the features of the parameter space in Assumption D2. We introduce the following condition related to the control function described in Assumption D1.

Assumption E2

The rationality control function r_p can be expressed as a functional of the form

$$r_p = \zeta_p - E\big[T\big(U_p, W_p\big)\big|W_p\big],$$

where ζ_p, U_p and W_p are observable to the econometrician, and $T(\cdot,\cdot)$ has known functional form. We strengthen Assumption (E1.1) and assume that we observe a random sample $\left(\big(Y_{p_n}, X_{p_n}, Z_{p_n}, \zeta_{p_n}, U_{p_n}, W_{p_n}\big)_{p=1}^{\mathcal{P}}\right)_{n=1}^{N}$. The exclusion restrictions between these variables are those that are necessary for all the conditions in Assumption D2 to hold. From now on, to accommodate the case where r_p has to be estimated, we will denote it by

$$\widehat{r}\big(W_{p_n}\big) \equiv \widehat{r}_{p_n} \in \mathbb{R}^R.$$

Therefore, R denotes the dimensionality of $\widehat{r}_{p_n}$. We will assume that it admits a linear representation with the same features as those of $\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)$ and $\widehat{\underline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)$. That is,

$$\widehat{r}_{p_n} - r_{p_n} = \frac{1}{N\widetilde{h}^R}\sum_{m=1}^{N} R\big(U_{p_m}, W_{p_n}\big)\widetilde{K}\left(\frac{W_{p_m}-W_{p_n}}{\widetilde{h}}\right) + \upsilon_{p_N}\big(W_{p_n}\big). \tag{26}$$

If r_p is directly observed, then $R(\cdot,\cdot)=\upsilon_p(\cdot)=0$, and no bandwidths or kernels are involved. Otherwise, we focus on cases where (26) is such that $\max_{n=1,\ldots,N}\Big|\upsilon_{p_N}\big(W_{p_n}\big)\Big| = o_p(N^{-1/2})$, $\widetilde{K}(\cdot)$ is a bias-reducing kernel of order $\overline{M}$, where the latter is described in Assumption E1. The bandwidth sequence $\widetilde{h}$ is such that $N^{1/2-\epsilon}\widetilde{h}\longrightarrow\infty$ for some $\epsilon>0$ and $N^{1/2}\widetilde{h}^{\overline{M}}\longrightarrow 0$. In addition, for all $t\in\mathbb{R}^{\dim(W_p)}$, the conditional expectation $E\Big[R\big(U_p,t\big)\Big|W_p\Big]$ is $\overline{M}$-times differentiable with bounded derivatives with respect to W_p for every t. It also satisfies

$$E\Big[R\big(U_p,t\big)\Big|W_p=t\Big]=0 \quad \text{w.p.1.}$$

Nonparametric control variables with these features have been analyzed in Aradillas-Lopez, Honoré, and Powell (2007). This structure is also amenable to the control function studied by Imbens and Newey (2002). If there is no rationality control function, or if it is directly observable then $R(\cdot) = 0$ trivially. We assume that $f_{W_p}(\cdot)$, the density of W_{p_n} is bounded away from zero[5]

We will stack

$$\widehat{\overline{\mu}}_p^*(Z_{p_n}, \theta) = \begin{pmatrix} \widehat{\overline{\mu}}_p(Z_{p_n}, \theta | L_2) \\ \vdots \\ \widehat{\overline{\mu}}_p(Z_{p_n}, \theta | L_{k^*}) \end{pmatrix}, \quad \widehat{\underline{\mu}}_p^*(Z_{p_n}, \theta) = \begin{pmatrix} \widehat{\underline{\mu}}_p(Z_{p_n}, \theta | L_2) \\ \vdots \\ \widehat{\underline{\mu}}_p(Z_{p_n}, \theta | L_{k^*}) \end{pmatrix}, \quad \widehat{\mu}_p^*(Z_{p_n}, \theta) = \begin{pmatrix} \widehat{\underline{\mu}}_p^*(Z_{p_n}, \theta) \\ \widehat{\overline{\mu}}_p^*(Z_{p_n}, \theta) \end{pmatrix},$$

and

$$\widehat{t}_{p_n}^*(\theta) = \begin{pmatrix} X_{p_n}' \beta_p \\ \widehat{\mu}_p^*(Z_{p_n}, \theta) \\ \widehat{r}_{p_n} \end{pmatrix} \in \mathbb{R}^{2k^*+R-1}, \quad \text{henceforth, denote } d^* \equiv 2k^* + R - 1.$$

Therefore, the dimension of $\widehat{t}_{p_n}^*(\theta)$ is denoted by d^*. Let $H : \mathbb{R}^{2k^*+R-1} \longrightarrow \mathbb{R}$ be a kernel function, and let b denote a bandwidth sequence. As before, $H_b(\psi) \equiv H\left(\frac{\psi}{b}\right)$. We will define

$$\widehat{E}_p(t;\theta) = \frac{1}{\widehat{f}_{t^*}(t;\theta)} \frac{1}{Nb^{d^*}} \sum_{m=1}^{N} Y_{p_m} H_b\big(\widehat{t}_{p_m}^*(\theta) - t\big), \quad \text{where} \quad \widehat{f}_{t^*}(t;\theta) = \frac{1}{Nb^{d^*}} \sum_{m=1}^{N} H_b\big(\widehat{t}_{p_m}^*(\theta) - t\big). \tag{27}$$

Note that d^* is the same for all players by assumption (they are all assumed to be D_{k^*}−rational, and we use the same rationality control function for all of them).

Assumption E3

1.− The kernel $H(\cdot)$ is bias-reducing of order $\overline{M}$. The kernel $H(\psi)$ is also $\overline{M}$−times differentiable with bounded derivatives, where $\overline{M}$ is as described in Assumption E1. In addition, its first derivative, $H^{(1)}(\cdot)$ satisfies

$$\int H^{(1)}(\psi)d\psi = 0; \quad \int \psi H^{(1)}(\psi)d\psi = -1; \quad \int \psi^j H^{(1)}(\psi)d\psi = 0 \quad \text{for } j = 2, \ldots, \overline{M}+1.$$

In addition, the bandwidth b satisfies: $N^{1/2-\epsilon}b^{d^*} \longrightarrow \infty$ for some $\epsilon > 0$, $N^{1/2}b^{\overline{M}} \longrightarrow 0$, and

$$N^{1/2}h_k^{\overline{M}}/b^2 \longrightarrow 0 \text{ for each } k = 1, \ldots, k^* - 1, \text{ and } N^{1/2}\widetilde{h}^{\overline{M}}/b^2 \longrightarrow 0.$$

[5]We only need this to be true when $Z_{p_n} \in \mathcal{Z}_p$.

These conditions indicate the under/over-smoothing with respect to the other bandwidths employed.

2.– Let $t_p^*(\theta_0) = \Big(X_p'\beta_{p0} \quad \widehat{\mu}_p^*(Z_p,\theta_0) \quad r_p\Big)' \in \mathbb{R}^{d^*}$ and let $f_{Z_p|t_0}(z|t) \equiv f_{Z_p|t_0}(Z_p = z|t_p^*(\theta_0))$ denote the conditional density of Z_p given $t_p^*(\theta_0) = t$. Then $f_{Z_p|t_0}(Z_p|t)$ is $\overline{M}-$times differentiable with bounded derivatives with respect to Z_p. Note how this is consistent with the dimensionality and exclusion-restriction conditions in part 5 of Assumption D2. We also assume these smoothness conditions for $f_{W_p|t_0}(W_p|t)$, where W_p is as described in Equation (26). We also assume that $f_{t_p^*}(\cdot)$, the density function of $t_p^*(\theta_0)$ is bounded in the trimming set $\mathcal{Z}_p$.

6.2.1 An estimator based on $D_{k^*}-$rationality

As we did above, let us stack

$$\overline{\mu}_p^*(Z_{p_n},\theta) = \begin{pmatrix} \overline{\mu}_p(Z_{p_n},\theta|L_2) \\ \vdots \\ \overline{\mu}_p(Z_{p_n},\theta|L_{k^*}) \end{pmatrix}, \quad \underline{\mu}_p^*(Z_{p_n},\theta) = \begin{pmatrix} \underline{\mu}_p(Z_{p_n},\theta|L_2) \\ \vdots \\ \underline{\mu}_p(Z_{p_n},\theta|L_{k^*}) \end{pmatrix}, \quad \mu_p^*(Z_{p_n},\theta) = \begin{pmatrix} \underline{\mu}_p^*(Z_{p_n},\theta) \\ \overline{\mu}_p^*(Z_{p_n},\theta) \end{pmatrix},$$

and

$$t_{p_n}^*(\theta) = \begin{pmatrix} X_{p_n}'\beta_p \\ \widehat{\mu}_p^*(Z_{p_n},\theta) \\ r_{p_n} \end{pmatrix} \in \mathbb{R}^{d^*}.$$

Our general definition of D_k^*-rationality and the assumptions about the rationality control function r_{p_n} imply

$$E\big[Y_{p_n}|X_{p_n},Z_{p_n},r_{p_n}\big] = E\big[Y_{p_n}|t_{p_n}^*(\theta_0)\big] \quad \text{for each player } \; p = 1,\ldots,\mathcal{P}. \tag{28}$$

The proposed estimator is based on the exclusion restriction in (28). Let

$$\widehat{Q}_N^*(\theta) = \frac{1}{N}\sum_{n=1}^{N}\Bigg\{\sum_{p=1}^{\mathcal{P}}\omega_p\Big(Y_{p_n} - \widehat{E}_p\big(\widehat{t}_{p_n}^*(\theta)\big)\Big)^2 \mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\Big\}\Bigg\}, \tag{29}$$

where $\big(\omega_p\big)_{p=1}^{\mathcal{P}}$ are pre-specified, constant positive weights. We introduced the trimming function $\mathbb{1}\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\}$ previously. We could include a similar trimming function for X_p. Namely, $\mathbb{1}\{\widehat{f}_{X_p}(X_{p_n}) > \underline{q}\}$. We bypass this in view of the last part of Assumption E3. In general, Equation (28) implies that no additional trimming based on X_{p_n} or r_{p_n} would affect the consistency of our estimator (see Assumption E2 and Footnote 5). For simplicity, we introduced the last part of Assumption E3 in order to make $\mathbb{1}\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\}$ the only trimming sequence used.

We define our estimator as

$$\widehat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \widehat{Q}^*_N(\theta). \tag{30}$$

This estimator belongs to a more general version of the semiparametric multiple index models analyzed in Ichimura and Lee (1991), where the indices include semiparametrically estimated objects (each one of the L_k- bounds), and we condition also on an additional control variable, $\widehat{r}_p$. Each one of our control variables has a linear representation that holds uniformly in our sample for each θ in the parameter space. The asymptotic features of the estimator will depend on these representations.

6.2.2 Asymptotic features of $\widehat{\boldsymbol{\theta}}$.

This section will outline the asymptotic properties of the estimator described in Equation (30). Not surprisingly, the uniform convergence features of $\left(\widehat{\underline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)\ ,\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)\right)_{p=1}^{\mathcal{P}}$ will play a major role.

Consistency

In the appendix we show that given our assumptions about the continuously distributed nature of Z_p, we have

$$\Pr\Big(\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s} \quad \text{and} \quad f_{Z_p}(Z_{p_n}) \leq \underline{s} \ \text{ for some } n = 1\ldots,N\Big) \longrightarrow 0.$$

There, we also establish an asymptotically linear representation for $\widehat{\underline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big) - \underline{\mu}_p\big(Z_{p_n},\theta\big|L_k\big)$ and $\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big) - \overline{\mu}_p\big(Z_{p_n},\theta\big|L_k\big)$ for each player $p = 1,\ldots,\mathcal{P}$. Given our assumptions, these representations hold uniformly over our sample for each $\theta\in\Theta$ and $k = 2,\ldots,k^*$. In particular, for each $\theta\in\Theta$, the semiparametric estimators $\widehat{\underline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)$ and $\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)$ converge to their population counterparts uniformly over $\big(Z_{p_n}\big)_{n=1}^{N}$. Coupled with our construction of $\widehat{E}_p(\cdot)$ and assumptions E1-E3, we obtain[6]

$$\sum_{p=1}^{\mathcal{P}}\omega_p\Big(Y_{p_n} - \widehat{E}_p\big(\widehat{t}^*_{p_n}(\theta)\big)\Big)^2 \mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\Big\} \xrightarrow{p} \sum_{p=1}^{\mathcal{P}}\omega_p\Big(Y_{p_n} - E_p\big(t^*_{p_n}(\theta)\big)\Big)^2 \mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) > \underline{s}\Big\}$$

uniformly over $n = 1,\ldots,N$. This yields

$$\widehat{Q}^*_N(\theta) = \frac{1}{N}\sum_{n=1}^{N}\left\{\sum_{p=1}^{\mathcal{P}}\omega_p\Big(Y_{p_n} - E_p\big(t^*_{p_n}(\theta)\big)\Big)^2 \mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) > \underline{s}\Big\}\right\} + o_p(1)$$

[6] A mean-value approximation easily shows this, given our uniform convergence results for $\widehat{\underline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)$ and $\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_k\big)$.

for each $\theta \in \Theta$. By construction, each term

$$\sum_{p=1}^{\mathcal{P}} \omega_p \Big(Y_{p_n} - E_p\big(t^*_{p_n}(\theta)\big) \Big)^2 \mathbb{1}\Big\{ f_{Z_p}(Z_{p_n}) > \underline{s} \Big\} \Big\}$$

is trivially bounded w.p.1. They are also continuous functions of θ. Lemma 2.4 in Newey and McFadden (1994) yields

$$\sup_{\theta\in\Theta} \left\| \widehat{Q}_N(\theta) - \underbrace{E\Big[\sum_{p=1}^{\mathcal{P}} \omega_p \Big(Y_p - E_p\big(t^*_p(\theta)\big) \Big)^2 \mathbb{1}\Big\{ f_{Z_p}(Z_p) > \underline{s} \Big\} \Big]}_{\equiv Q(\theta)} \right\| \xrightarrow{p} 0. \tag{31}$$

From Proposition 4, $Q(\theta)$ is uniquely minimized in our parameter space Θ at its true value, θ_0. Using this fact and the uniform convergence feature in Equation (31), Theorem 2.1 in Newey and McFadden (1994) yields

$$\widehat{\theta} \xrightarrow{p} \theta_0.$$

We outline the asymptotic normality features of the estimator next.

Asymptotic normality

We begin with a brief overview of generalized gradients and some of their properties. Let $f : X \longrightarrow \mathbb{R}$ be a Lipschitz continuous function near a point x. The generalized directional derivative of f at x in the direction of v is given by

$$f^0(x; v) = \limsup_{\substack{y\to x \\ t\downarrow 0}} \frac{f(y+tv) - f(y)}{t}.$$

Suppose $X \subseteq \mathbb{R}^k$. The generalized gradient of f at the point x, denoted by $\partial f(x)$, is the subset of X given by

$$\partial f(x) = \Big\{ \zeta \in X : \; f^0(x; v) \geq \zeta' v \quad \text{for all } v \in X \Big\}.$$

If f is continuously differentiable at x, then $\partial f(x) = \big\{ \nabla_x f(x) \big\}$, the partial derivative of f with respect to x. In general, for the functions that will be relevant to us here, the set $\partial f(x)$ includes convex combinations of directional derivatives of f at x. If f is Lipschitz at x, then $\partial f(x)$ is a nonempty, convex set (Proposition 2.1.2 in Clarke (1990)). The following features of generalized gradients will be relevant to us.

(i) Chain Rule (Theorem 2.3.9 in Clarke (1990)): Let $h : X \longrightarrow \mathbb{R}^n$ and $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a pair of functions. Denote the component functions of h by h_i, for $i = 1, \ldots, n$. Suppose each h_i is Lipschitz near x, and g is Lipschitz near $h(x)$. Let $f = g \circ h$, then

$$\partial f(x) \subset \overline{\mathrm{co}}\left\{\sum_{i=1}^{n} \alpha_i \zeta_i : \ \zeta_i \in \partial h_i(x), \ \big(\alpha_1, \ldots, \alpha_n\big) \equiv \alpha \in \partial g\big(h(x)\big)\right\},$$

where $\overline{\mathrm{co}}$ denotes the closed convex hull. In particular, if g is a continuously differentiable function, we have

$$\partial f(x) \subset \left\{\sum_{i=1}^{n} \alpha_i \zeta_i : \ \zeta_i \in \partial h_i(x), \quad \alpha_i = \nabla_{h_i} g \ \text{ is the partial derivative of } g \text{ with respect to } h_i.\right\}.$$

(ii) Products, quotients and finite sums (Propositions 2.3.13, 2.3.14 and 2.3.2 in Clarke (1990)): Let f_1, f_2 be Lipschitz near x, then $f_1 f_2$ is Lipschitz near x and

$$\partial\big(f_1 f_2\big)(x) \subset f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x), \quad \text{and if } f_2(x) \neq 0, \ \partial\left(\frac{f_1}{f_2}\right) \subset \frac{f_2(x)\partial f_1(x) - f_1(x)\partial f_2(x)}{f_2^2(x)}.$$

If f_i, $i = 1, \ldots, n$ is a finite family of Lipschitz functions near x, the sum is also Lipschitz and

$$\partial\Big(\sum_{i=1}^{n} f_i\Big) \subset \sum_{i=1}^{n} \partial f_i(x).$$

(iii) Mean Value Theorem (Theorem 2.3.7 in Clarke (1990)): Let x and y be points in $X \subseteq \mathbb{R}^n$ and suppose f is Lipschitz on an open set containing the line segment $[x, y]$. Then there exists a point $u \in (x, y)$ such that

$$f(y) - f(x) \in \partial f(u)'(y - x).$$

(iv) Local Extrema (Proposition 2.3.2 in Clarke (1990)): If f attains a local minimum or maximum at x, then $0 \in \partial f(x)$.

Here, we will care about the generalized gradients of $\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big)$, $\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big)$ and their population counterparts, $\underline{\mu}_p\big(Z_{p_n}, \theta\big|L_k\big), \overline{\mu}_p\big(Z_{p_n}, \theta\big|L_k\big)$ *with respect to* θ, *given* Z_{p_n}. We will denote

$$\begin{aligned} \nabla^0 \widehat{\underline{\mu}}_p(Z_{p_n}, \theta; v|L_k) &= \limsup_{\substack{y \to \theta \\ t \downarrow 0}} \frac{\widehat{\underline{\mu}}_p(Z_{p_n}, y + tv|L_k) - \widehat{\underline{\mu}}_p(Z_{p_n}, y|L_k)}{t}, \\ \partial \widehat{\underline{\mu}}_p(Z_{p_n}, \theta|L_k) &= \Big\{\zeta \in \mathbb{R}^D : \ \nabla^0 \widehat{\underline{\mu}}_p(Z_{p_n}, \theta; v|L_k) \geq \zeta' v \quad \text{for all} \quad v \in \mathbb{R}^D.\Big\} \end{aligned} \tag{32}$$

We define $\partial \widehat{\overline{\mu}}_p(Z_{p_n}, \theta|L_k)$, $\partial \underline{\mu}_p(Z_{p_n}, \theta|L_k)$ and $\partial \overline{\mu}_p(Z_{p_n}, \theta|L_k)$ accordingly. As before, whenever they exist, let

$$\nabla_\theta \underline{\mu}_p(Z_{p_n}, \theta|L_k) \quad \text{and} \quad \nabla_\theta \overline{\mu}_p(Z_{p_n}, \theta|L_k)$$

denote the ordinary gradients (vector of partial derivatives). By Assumptions (D2), (E1) and the resulting asymptotic properties features of $\widehat{\underline{\mu}}_p(Z_{p_n}, \theta|L_k)$ and $\widehat{\overline{\mu}}_p(Z_{p_n}, \theta|L_k)$, with probability approaching one uniformly over $\left(Z_{p_n}\right)_{n=1}^N$ (all of which belong to the trimming set $\mathcal{Z}_p$), both estimators are Lipschitz-continuous for all $\theta \in \Theta$. Moreover, their generalized gradients $\partial\widehat{\overline{\mu}}_p(Z_{p_n}, \theta|L_k)$ and $\partial\widehat{\underline{\mu}}_p(Z_{p_n}, \theta|L_k)$ will differ from being simply their ordinary gradients $\nabla_\theta\widehat{\overline{\mu}}_p(Z_{p_n}, \theta|L_k)$ and $\nabla_\theta\widehat{\underline{\mu}}_p(Z_{p_n}, \theta|L_k)$ *only if* the realization of Z_{p_n} is such that the inequality in Equation (20) holds with exact equality for some $2 \leq k \leq k^*$ (or if the same is true for the equivalent equation for $\widehat{\underline{M}}_k(Z_p, \theta)$ or $\widehat{\underline{D}}_k(Z_p, \theta)$). Using Assumption (E1), Propositions A-6 and A-7 in the appendix and the results that lead to Proposition A-8, the probability that this happens goes to zero uniformly over $\left(Z_{p_n}\right)_{n=1}^N$ for each $\theta \in \Theta$.

As before, we stack θ as

$$\theta = \Big(\beta_1', \ldots, \beta_{\mathcal{P}}', \Delta_1', \ldots, \Delta_{\mathcal{P}}'\Big)', \quad \text{where} \quad \Delta_p = \Big(\Delta_p^0\ , \cdots, \Delta_p^{\mathcal{P}-1}\Big)'.$$

For each p, let $\nabla_{x_p\beta_p}\widehat{E}_p(\cdot)$ denote the derivative of $\widehat{E}_p(\cdot)$ with respect to $x_p'\beta_p$, let $\nabla_{\overline{\mu}_k}\widehat{E}_p(\cdot)$ denote the derivative with respect to $\widehat{\overline{\mu}}_p(\cdot|L_k)$ and so on for $\nabla_{\underline{\mu}_k}\widehat{E}_p(\cdot)$. It follows from Equation (27) and the properties of the kernel $\widetilde{K}(\cdot)$ that all these objects are the usual partial derivatives. Let

$$\widehat{T}_1\big(X_{1_n}, \widehat{t}^*_{1_n}(\theta)\big) = \begin{pmatrix} X_{1_n}\nabla_{x_1\beta_1}\widehat{E}_1\big(\widehat{t}^*_{1_n}(\theta)\big) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots\ ,\ \ldots\ , \widehat{T}_{\mathcal{P}}\big(X_{\mathcal{P}_n}, \widehat{t}_{\mathcal{P}_n}(\theta)\big) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ X_{\mathcal{P}_n}\nabla_{x_{\mathcal{P}}\beta_{\mathcal{P}}}\widehat{E}_{\mathcal{P}}\big(\widehat{t}_{\mathcal{P}_n}(\theta)\big) \end{pmatrix}.$$

Let 0_Δ denote a column vector of zeros in $\mathbb{R}^{\mathcal{P}^2}$ and define

$$\widehat{D}_p\big(X_{p_n}, \widehat{t}^*_{p_n}(\theta)\big) = \Big(\widehat{T}_p\big(X_{p_n}, \widehat{t}^*_{p_n}(\theta)\big)', 0_\Delta'\Big)'.$$

Let $\mathbb{1}^*_{p_n} = \mathbb{1}\big\{f_{Z_p}(Z_{p_n}) > \underline{s}\big\}$. Using the definition of $\widehat{\theta}$ and the properties of generalized gradients mentioned above, with probability approaching one uniformly in our sample, our estimator $\widehat{\theta}$ satisfies

$$\frac{1}{N}\sum_{n=1}^N\left\{\sum_{p=1}^{\mathcal{P}}\widehat{J}_{p_N}\left(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}(\widehat{\theta}), \widehat{\theta}\ \right)\Big(Y_{p_n} - \widehat{E}_p\big(\widehat{t}^*_{p_n}(\widehat{\theta})\big)\Big)\mathbb{1}^*_{p_n}\omega_p\right\} = 0,$$

where $\widehat{J}_{p_N}\Big(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}(\theta), \theta\ \Big)$ is of the form

$$\widehat{J}_{p_N}\Big(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}(\theta), \theta\ \Big) =$$
$$\widehat{D}_p\big(X_{p_n}, \widehat{t}^*_{p_n}(\theta)\big) + \sum_{k=2}^{k^*}\bigg(\widetilde{\nabla}_\theta\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big)\nabla_{\overline{\mu}_k}\widehat{E}_p\big(\widehat{t}^*_{p_n}(\theta)\big) + \widetilde{\nabla}_\theta\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big)\nabla_{\underline{\mu}_k}\widehat{E}_p\big(\widehat{t}^*_{p_n}(\theta)\big)\bigg), \tag{33}$$

where $\widetilde{\nabla}_\theta\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big) \in \partial\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big)$ and $\widetilde{\nabla}_\theta\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big) \in \partial\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta\big|L_k\big)$ and these generalized gradients are described in Equation (32). Using the Mean Value property described above (part iii) and the consistency of $\widehat{\theta}$, with probability approaching one we can express

$$\frac{1}{N}\sum_{n=1}^{N}\left\{\sum_{p=1}^{\mathcal{P}}\widehat{J}_{p_N}\Big(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}(\widetilde{\theta}), \widetilde{\theta}\ \Big)\widehat{J}_{p_N}\Big(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}(\widetilde{\theta}), \widetilde{\theta}\ \Big)'\mathbb{1}^*_{p_n}\omega_p\right\}\Big(\widehat{\theta} - \theta_0\Big)$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left\{\sum_{p=1}^{\mathcal{P}}\widehat{J}_{p_N}\Big(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}(\theta_0), \theta_0\ \Big)\Big(Y_{p_n} - \widehat{E}_p\big(\widehat{t}^*_{p_n}(\theta_0)\big)\Big)\mathbb{1}^*_{p_n}\omega_p\right\},$$

where $\widetilde{\theta} \in \big(\widehat{\theta}, \theta_0\big)$. Abbreviate $t^*_{p_n}(\theta_0) \equiv t^*_{p_n}$ and $\widehat{t}^*_{p_n}(\theta_0) \equiv \widehat{t}^*_{p_n}$. Let $J_p(X_{p_n}, Z_{p_n}, t^*_{p_n}(\theta), \theta\ )$ be the population counterpart of the object defined in Equation (33) and denote

$$J_p(X_{p_n}, Z_{p_n}, t^*_{p_n}(\theta_0), \theta_0\ ) \equiv J_p(X_{p_n}, Z_{p_n}, t^*_{p_n}).$$

Using the linear representation result in Proposition A-8 in the appendix, Assumptions E1-E3 and the resulting linear representation for $\widehat{E}_p\big(\widehat{t}^*_{p_n}\big) - E_p\big(t^*_{p_n}\big)$, we can express

$$\widehat{J}_{p_N}\Big(X_{p_n}, Z_{p_n}, \widehat{t}^*_{p_n}, \theta_0\ \Big)\Big(Y_{p_n} - \widehat{E}_p\big(\widehat{t}^*_{p_n}\big)\Big)\mathbb{1}^*_{p_n} - J_p\big(X_{p_n}, Z_{p_n}, t^*_{p_n}\big)\Big(Y_{p_n} - E_p\big(t^*_{p_n}\big)\Big)\mathbb{1}^*_{p_n}$$
$$= \frac{1}{N}\sum_{m=1}^{N}\left\{\sum_{k=1}^{k^*-1}\frac{1}{h_k^{d_p}}\frac{\psi^p_k\big(X_m, Z_m, t^*_{p_m}, Z_{p_n}\big)}{f_{Z_p}(Z_{p_n})}K^p_{h_k}\big(Z_{p_m} - Z_{p_n}\big) + \frac{1}{\widetilde{h}^R}\frac{\psi^p_r\big(X_m, Z_m, t^*_{p_m}, U_{p_m}, W_{p_n}\big)}{f_{W_p}(W_{p_n})}\widetilde{K}_{\widetilde{h}}\big(W_{p_m} - W_{p_n}\big)\right.$$
$$\left. + \gamma^p_{A}\big(X_m, Z_m, t^*_{p_m}\big)'\gamma^p_{B}(Z_{p_n}) + \nu^p_{A}\big(X_m, Z_m, t^*_{p_m}, U_{p_m}\big)'\nu^p_{B}(W_{p_n})\right\} + \eta_{p_N}(Z_{p_n}),$$

where $\max_{n=1,\ldots,N}\Big\|\eta_{p_N}(Z_{p_n})\Big\| = o_p(N^{-1/2})$. The remaining terms in this representation satisfy

$$E\Big[\psi^p_k\big(X_m, Z_m, t^*_{p_m}, Z_{p_n}\big)\Big|Z_{p_n}, Z_{p_m} = Z_{p_n}\Big] = 0\ \forall\ k = 2, \ldots, k^*,$$
$$E\Big[\psi^p_r\big(X_m, Z_m, t^*_{p_m}, U_{p_m}, W_{p_n}\big)\Big|W_{p_n}, W_{p_m} = W_{p_n}\Big] = 0$$
$$E\Big[\gamma^p_{A}\big(X_m, Z_m, t^*_{p_m}\big)\Big] = 0, \quad E\Big[\nu^p_{A}\big(X_m, Z_m, t^*_{p_m}, U_{p_m}\big)\Big] = 0.$$

We introduce the following assumption.

Assumption E4 Let $J_p(X_{p_n}, Z_{p_n}, t^*_{p_n}(\theta), \theta\ )$ denote denote the population counterpart of the object defined in Equation (33) and denote $J_p(X_{p_n}, Z_{p_n}, t^*_{p_n}(\theta_0), \theta_0\ ) \equiv J_p(X_{p_n}, Z_{p_n}, t^*_{p_n})$. Part 2 of Assumption E1 implies that with probability one, this object is given by

$$D_p\big(X_{p_n}, t^*_{p_n}\big) + \sum_{k=2}^{k^*}\Big(\nabla_\theta \overline{\mu}_p\big(Z_{p_n}, \theta_0 \big| L_k\big)\nabla_{\overline{\mu}_k} E_p\big(t^*_{p_n}\big) + \nabla_\theta \underline{\mu}_p\big(Z_{p_n}, \theta_0 \big| L_k\big)\nabla_{\underline{\mu}_k} E_p\big(t^*_{p_n}\big)\Big).$$

Let $\Sigma_p \equiv E\Big[J_p(X_{p_n}, Z_{p_n}, t^*_{p_n})J_p(X_{p_n}, Z_{p_n}, t^*_{p_n})'\mathbb{1}^*_{p_n}\Big]$. We will assume that $\sum_{p=1}^{\mathcal{P}} \Sigma_p \omega_p$ is an invertible matrix. We have the following result.

Theorem 1 *For each player $p = 1, \ldots, \mathcal{P}$, let*

$$\psi^p_\theta(X_n, Z_n, U_{p_n}, W_{p_n}) = \left(\sum_{p=1}^{\mathcal{P}} \Sigma_p \omega_p\right)^{-1} \times \Big\{ J_p\big(X_{p_n}, Z_{p_n}, t^*_{p_n}\big)\Big(Y_{p_n} - E_p\big(t^*_{p_n}\big)\Big)\mathbb{1}^*_{p_n} + \sum_{k=1}^{k^*-1} \psi^p_k\big(X_n, Z_n, t^*_{p_n}, Z_{p_n}\big)$$
$$+ \psi^p_r\big(X_n, Z_n, t^*_{p_n}, U_{p_n}, W_{p_n}\big) + \gamma^p_{A}\big(X_n, Z_n, t^*_{p_n}\big)'\overline{\gamma}^p_{B} + \nu^p_{A}\big(X_n, Z_n, t^*_{p_n}, U_{p_n}\big)'\overline{\nu}^p_{B}\Big\}$$

where this objects are as described above, $\overline{\gamma}^p_B = E\big[\gamma^p_B(Z_{p_n})\big]$, and $\overline{\nu}^p_B = E\big[\nu^p_B(Z_{p_n})\big]$. Note that $E\big[\psi^p_\theta(X_n, Z_n, U_{p_n}, W_{p_n})\big] = 0$. If Assumptions E1-E4 are satisfied,

$$\widehat{\theta} - \theta_0 = \left(\sum_{p=1}^{\mathcal{P}} \Sigma_p \omega_p\right)^{-1} \frac{1}{N}\sum_{n=1}^{N}\left\{\sum_{p=1}^{\mathcal{P}} \psi^p_\theta(X_n, Z_n, U_{p_n}, W_{p_n})\omega_p\right\} + o_p(N^{-1/2}),$$

where $\sum_{p=1}^{\mathcal{P}} \Sigma_p \omega_p$ is as described in Assumption E4.

If the only source of incomplete information is the unobserved payoff shock ε_p and $Z_p = X_{-p}$ for every player p, then L_k- bounds can be recovered exactly given that the distribution of ε is assumed known. In this case, none of the assumptions pertaining to the estimation of these bounds is necessary. If we retain the remaining assumptions, let

$$\widetilde{\psi}^p_\theta(X_n, Z_n, U_{p_n}, W_{p_n}) =$$
$$\left(\sum_{p=1}^{\mathcal{P}} \Sigma_p \omega_p\right)^{-1} \times \Big\{ J_p\big(X_{p_n}, Z_{p_n}, t^*_{p_n}\big)\Big(Y_{p_n} - E_p\big(t^*_{p_n}\big)\Big) + \psi^p_r\big(X_n, Z_n, t^*_{p_n}, U_{p_n}, W_{p_n}\big) + \nu^p_{A}\big(X_n, Z_n, t^*_{p_n}, U_{p_n}\big)'\overline{\nu}^p_{B}\Big\},$$

where these objects are as described above. In this case we would have

$$\widehat{\theta} - \theta_0 = \left(\sum_{p=1}^{\mathcal{P}} \Sigma_p \omega_p\right)^{-1} \frac{1}{N}\sum_{n=1}^{N}\left\{\sum_{p=1}^{\mathcal{P}} \widetilde{\psi_\theta}^p(X_n, Z_n, U_{p_n}, W_{p_n})\omega_p\right\} + o_p(N^{-1/2}).$$

The proof of this theorem relies heavily on the asymptotic representation result in Proposition A-8 and the fact that we assume the control variable $\widehat{r}_{p_n}$ to have an analogous representation. The only additional step is the analysis of $\widehat{E}_p\big(\widehat{t}^*_{p_n}\big) - E_p\big(t^*_{p_n}\big)$ and a characterization of the projections for all of the corresponding U-statistics. Note that the weights ω_p were not chosen in an optimal way. A natural way to choose them would come from the asymptotic result in Theorem 1.

7 Concluding remarks

The assumption of Nash equilibrium has been a powerful instrument for identification and inference in games. A growing body of experimental evidence suggests that in a number of strategic environments, people's behavior is the result of a few steps of iterated reasoning. In this paper we presented identification and inference results for a model in which players are assumed to best-respond to some element in the space of choices that survive $k-1$ rounds of deletion of dominated strategies. We described a set of assumptions which allowed us to semiparametrically identify the rationalizable bounds for any set of beliefs which survive this iterative process. We presented a constructive identification result based on the notion of D_k-rationality and studied the properties of an estimation procedure which is based on the identification result. We described a set of conditions under which the estimator has good asymptotic features. The convenient asymptotic properties of the semiparametric estimators for the L_k-rational bounds for beliefs suggest that one could device a specification test for competing models of behavior which depend on those bounds. This is a subject of ongoing research.

A Mathematical appendix

A.1 Proof of Proposition 4

Henceforth, we will abbreviate

$$\min_{M}\big(\Delta_p^M\big) \equiv \underline{\Delta}_p \quad \text{and} \quad \max_{M}\big(\Delta_p^M\big) \equiv \overline{\Delta}_p,$$

for each player $p = 1, \ldots, \mathcal{P}$.

Let players p_1 and p_2 be as described in Assumption (D2.3). Take player p_1 and consider any parameter value $\theta \in \Theta$ such that:

$$\text{either} \;\; \beta_q \neq \beta_{q0}, \;\; \underline{\Delta}_q \neq \underline{\Delta}_{q0}, \;\; \text{or} \;\; \overline{\Delta}_q \neq \overline{\Delta}_{q0} \;\; \text{for some } q \neq p_1.$$

From Assumption (D2), it will follow that for any such θ,

$$\Pr\Big(E\big[Y_{p_1}\big|\mathcal{T}^*_{p_1}(\theta), r_{p_1}\big] \neq E\big[Y_{p_1}\big|\mathcal{T}^*_{p_1}(\theta_0), r_{p_1}\big]\Big) > 0.$$

The same result will follow for player p_2 (this is why we need at least two players who satisfy Assumption D2.3). The effect of strategic interaction is behind this result, through the presence of the bounds $\left(\underline{\mu}_{p_1}(Z_{p_1},\theta|L_k)\ ,\ \overline{\mu}_{p_1}(Z_{p_1},\theta|L_k)\right)_{k=2}^{k^*}$ (and those of player p_2), all of which are identified for any given θ given Assumption (D2.1), and none of which depend on the rationality control functions r_{p_1} and r_{p_2}. The strategic interaction effect would not exist, for example, if none of the players performed any round of deletion of dominated strategies, this is why we need the last part of Assumption (D2.3). Thus, $\left(\beta_p,\underline{\Delta}_p,\overline{\Delta}_p\right)$ are identified for each player p. To see why the remaining strategic interaction parameters are also identified, note that Equations (15) and (17) yield

$$\sum_{k=2}^{k^*}\overline{\xi}_p^k(r_p)\overline{\mu}_p\big(Z_p,\theta\big|L_k\big)=$$
$$\sum_{k=2}^{k^*}\Bigg\{\overline{\xi}_p^k(r_p)\Delta_p^{(\overline{U}_k^*+1)}\cdot\Bigg(1-\Big[\mathbb{1}\{\overline{M}_k^*\le\overline{D}_k^*\}\underline{\mathbb{P}}_{-p}^0(Z_p,\theta|L_{k-1})+\mathbb{1}\{\overline{M}_k^*>\overline{D}_k^*\}\overline{\mathbb{P}}_{-p}^0(Z_p,\theta|L_{k-1})\Big]\Bigg)$$
$$+\overline{\xi}_p^k(r_p)\sum_{j=1}^{p^*}\Big(\Delta_p^{(j)}-\Delta_p^{(\overline{U}_k^*+1)}\Big)\cdot\Bigg(\eta_{p_j}^a\big(\Delta_p^{(j)},\overline{D}_k^*,\overline{M}_k^*\big)\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\eta_{p_j}^b\big(\Delta_p^{(j)},\overline{D}_k^*,\overline{M}_k^*\big)\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Bigg)\Bigg\},$$
$$\sum_{k=2}^{k^*}\underline{\xi}_p^k(r_p)\underline{\mu}_p\big(Z_p,\theta\big|L_k\big)=$$
$$\sum_{k=2}^{k^*}\Bigg\{\underline{\xi}_p^k(r_p)\Delta_p^{(\underline{U}_k^*+1)}\cdot\Bigg(1-\Big[\mathbb{1}\{\underline{M}_k^*\le\underline{D}_k^*\}\underline{\mathbb{P}}_{-p}^0(Z_p,\theta|L_{k-1})+\mathbb{1}\{\underline{M}_k^*>\underline{D}_k^*\}\overline{\mathbb{P}}_{-p}^0(Z_p,\theta|L_{k-1})\Big]\Bigg)$$
$$+\underline{\xi}_p^k(r_p)\sum_{j=1}^{p^*}(\Delta_p^{(j)}-\Delta_p^{(\underline{U}_k^*+1)})\Bigg(\eta_{p_j}^c\big(\Delta_p^{(j)},\underline{D}_k^*,\underline{M}_k^*\big)\overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)+\eta_{p_j}^d\big(\Delta_p^{(j)},\underline{D}_k^*,\overline{M}_k^*\big)\underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p,\theta\big|L_{k-1}\big)\Bigg)\Bigg\}.$$

Next, recall that, depending on the realization of Z_p, the value of U_k^* can be between 1 and $\mathcal{P}-1$. If the regressor X_{q_ℓ} described in Assumption (D2.2) has rich enough support, with strictly positive probability, conditional on r_p, we will have $\Delta_p^{(\underline{U}_k^*+1)}=\underline{\Delta}_p$, or $\Delta_p^{(\underline{U}_k^*+1)}=\overline{\Delta}_p$. This holds for each k, and it is also true for $\Delta_p^{(\overline{U}_k^*+1)}$. Since we showed that both $\underline{\Delta}_p$ and $\overline{\Delta}_p$ are identified, it will follow from the expressions above that for any $\theta\in\Theta$ such that $\Delta_p^M\neq\Delta_{p_0}^M$ for some M,

$$\Pr\Bigg(\sum_{k=2}^{k^*}\overline{\xi}_p^k(r_p)\overline{\mu}_p\big(Z_p,\theta\big|L_k\big)\neq\sum_{k=2}^{k^*}\overline{\xi}_p^k(r_p)\overline{\mu}_p\big(Z_p,\theta_0\big|L_k\big)$$
$$\text{or}\quad\sum_{k=2}^{k^*}\overline{\xi}_p^k(r_p)\overline{\mu}_p\big(Z_p,\theta\big|L_k\big)\neq\sum_{k=2}^{k^*}\overline{\xi}_p^k(r_p)\overline{\mu}_p\big(Z_p,\theta_0\big|L_k\big)\ \Bigg|\ r_p\Bigg)>0\quad\forall\ r_p\in\mathbb{S}(r_p).$$

It follows that for any such θ,

$$\Pr\Big(E\big[Y_p\big|\mathcal{T}_p^*(\theta),r_p\big]\neq E\big[Y_p\big|\mathcal{T}_p^*(\theta_0),r_p\big]\Big)>0.$$

Together, these results imply

$$\theta\neq\theta_0\Longrightarrow\Pr\Big(E\big[Y_p\big|\mathcal{T}_p^*(\theta),r_p\big]\neq E\big[Y_p\big|\mathcal{T}_p^*(\theta_0),r_p\big]\Big)>0\quad\text{for some}\ \ p=1,\ldots,\mathcal{P}.$$

The result in Proposition 4 follows from here because our assumptions imply that $E\big[Y_p\big|X_p,Z_p,r_p\big]=E\big[Y_p\big|\mathcal{T}_p^*(\theta_0),r_p\big]$ w.p.1. $\square$

A.2 A linear representation result for $\underline{\widehat{\mu}}_p(Z_p,\theta|L_k)$ and $\widehat{\overline{\mu}}_p(Z_p,\theta|L_k)$

Throughout the entire appendix, p and q will denote an arbitrary pair of players in $1,\ldots,\mathcal{P}$. We will proceed iteratively starting with L_1. Let

$$\overline{T}^M_{-p_N}(X_{-p\ell},Z_p,\theta|L_1)=\left(\frac{1}{f_{Z_p}(Z_p)}\right)\left[\overline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\ ,\ \overline{\Delta}_q\ ,\ \underline{\Delta}_q\big)_{q\neq p}\ ;\ Z_p\Big)-\overline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_1)\right]K^p_{h_1}(Z_{p\ell}-Z_p),$$
$$\underline{T}^M_{-p_N}(X_{-p\ell},Z_p,\theta|L_1)=\left(\frac{1}{f_{Z_p}(Z_p)}\right)\left[\underline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\ ,\ \overline{\Delta}_q\ ,\ \underline{\Delta}_q\big)_{q\neq p}\ ;\ Z_p\Big)-\underline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_1)\right]K^p_{h_1}(Z_{p\ell}-Z_p) \tag{A-1}$$

Let $\big(\alpha_M,\delta_M\big)_{M=0}^{\mathcal{P}-1}$ be an arbitrary collection of scalars, and let

$$H_p(Z_p,\theta|L_1)=\sum_{M=0}^{\mathcal{P}-1}\Big(\alpha_M\underline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_1)+\delta_M\overline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_1)\Big),$$
$$\widehat{H}_{p_N}(Z_p,\theta|L_1)=\sum_{M=0}^{\mathcal{P}-1}\Big(\alpha_M\widehat{\underline{\mathbb{F}}}^M_{-p}(Z_p,\theta|L_1)+\delta_M\widehat{\overline{\mathbb{F}}}^M_{-p}(Z_p,\theta|L_1)\Big). \tag{A-2}$$

For the same collection of scalars, let

$$S_{p_N}(X_{-p\ell},Z_p,\theta|L_1)=\sum_{M=0}^{\mathcal{P}-1}\Big(\alpha_M\underline{T}^M_{-p_N}(X_{-p\ell},Z_p,\theta|L_1)+\delta_M\overline{T}^M_{-p_N}(X_{-p\ell},Z_p,\theta|L_1)\Big). \tag{A-3}$$

All the objects involved in the estimation involve distribution functions, bounded with probability one between zero and one. Lemma 3 in Collomb and Hardle (1986) or Theorem A-1 in Aradillas-Lopez (2006) yield

$$\widehat{H}_{p_N}(Z_p,\theta|L_1)-H_p(Z_p,\theta|L_1)=\frac{1}{Nh_1^{d_p}}\sum_{\ell=1}^{N}S_{p_N}(X_{-p\ell},Z_p,\theta|L_1)+\xi_{p_N}(Z_p,\theta|L_1),$$
$$\text{where}\quad\sup_{\substack{z\in\mathcal{Z}_p\\ \theta\in\Theta}}\Big\|\xi_{p_N}(z,\theta|L_1)\Big\|=O_p\big(N^{\epsilon-1}h_N^{-d_p}\big)\quad\text{for any compact set }\Theta\text{ and any }\ \epsilon>0. \tag{A-4}$$

As before, we denote $\mathcal{Z}_p=\Big\{z:\ f_{Z_p}(z)\geq\underline{s}\Big\}$ (i.e, the population trimming set). In addition, due to bounded nature of $\underline{G}^M_{-p}(\cdot)$, $\overline{G}^M_{-p}(\cdot)$, $\underline{F}^M_{-p}(\cdot)$, $\overline{F}^M_{-p}(\cdot)$ and the kernel $K^p(\cdot)$, $h_1^{d_p}\xi_{p_N}(Z_p,\theta|L_1)$ is also bounded. Let $h_1^{d_p}\xi_{p_N}(Z_p,\theta|L_1)\equiv\psi_{p_N}(Z_p,\theta|L_1)$. We can reexpress Equation $(A-4)$ as

$$\widehat{H}_{p_N}(Z_p,\theta|L_1)-H_p(Z_p,\theta|L_1)=\frac{1}{N}\sum_{\ell=1}^{N}T_{p_N}(X_{-p\ell},Z_p,\theta|L_1),$$
$$\text{where}\quad T_{p_N}(X_{-p\ell},Z_p,\theta|L_1)=\left(\frac{1}{h_N^{d_p}}\right)\times\Big[S_{p_N}(X_{-p\ell},Z_p,\theta|L_1)+\psi_{p_N}(Z_p,\theta|L_1)\Big], \tag{A-5}$$

where $\sup_{\substack{z\in\mathcal{Z}_p\\ \theta\in\Theta}}\Big\|\psi_{p_N}(z,\theta|L_1)\Big\|=O_p\big(N^{\epsilon-1}\big)$ for any compact set Θ and any $\epsilon>0$. Note that

$$\Big|S_p(X_{-p\ell},Z_p,\theta|L_1)+\psi_{p_N}(Z_{p_n},\theta|L_1)\Big|\leq\overline{\mathcal{C}}\quad\text{for some constant }\overline{\mathcal{C}}.$$

For an arbitrary scalar b, let

$$\mathbb{I}_p\big(Z_p,\theta\big|L_1\big)=\mathbb{1}\Big\{H_p\big(Z_p,\theta\big|L_1\big)<b\Big\},\quad \text{and}\quad \widehat{\mathbb{I}}_{p_N}\big(Z_p,\theta\big|L_1\big)=\mathbb{1}\Big\{\widehat{H}_{p_N}\big(Z_p,\theta\big|L_1\big)<b\Big\}. \tag{A-6}$$

Suppose that for a particular value $\theta\in\Theta$ the following holds

$$\Pr\Big(H_p\big(Z_{p_n},\theta\big|L_1\big)=b\Big|Z_{p_n}\in\mathcal{Z}_p\Big)=0. \tag{A-7}$$

We also maintain the assumption that $\Pr\big(f_{Z_p}(Z_p)=\underline{s}\big)=0$. The following result will be useful.

Proposition A-5 *Given the above conditions and our previous assumptions,*

$$Pr\Big(\widehat{\mathbb{I}}_{p_N}\big(Z_{p_n},\theta\big|L_1\big)\neq\mathbb{I}_p\big(Z_{p_n},\theta\big|L_1\big)\ \textit{for some}\ \ n=1,\dots,N\ \Big)\longrightarrow 0.$$

Proof: Throughout our estimation procedure, we have used $\mathbb{1}\Big\{\widehat{f}_{Z_p}(z)>\underline{s}\Big\}$ as a trimming function, so $\mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n})>\underline{s}\Big\}=1$ for all n. Let

$$\mathbb{1}^*(Z_{p_n})=\mathbb{1}\Big\{H_p\big(Z_{p_n},\theta\big|L_1\big)=b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n})\geq\underline{s}\Big\}.$$

We proceed by noting that

$$\begin{aligned}&\mathbb{1}\Big\{\widehat{\mathbb{I}}_{p_N}\big(Z_{p_n},\theta\big|L_1\big)\neq\mathbb{I}_p\big(Z_{p_n},\theta\big|L_1\big)\ \text{for some}\ n=1,\dots,N\ \text{and some}\ \theta\in\Lambda_N.\Big\}\\ &\leq\sum_{n=1}^{N}\Bigg[\mathbb{1}\Big\{\widehat{H}_{p_N}\big(Z_{p_n},\theta\big|L_1\big)>b\Big\}\mathbb{1}\Big\{H_p\big(Z_{p_n},\theta\big|L_1\big)<b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n})\geq\underline{s}\Big\}\\ &+\mathbb{1}\Big\{\widehat{H}_{p_N}\big(Z_{p_n},\theta\big|L_1\big)<b\Big\}\mathbb{1}\Big\{H_p\big(Z_{p_n},\theta\big|L_1\big)>b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n})\geq\underline{s}\Big\}+\mathbb{1}^*(Z_{p_n}).\\ &+\mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n})>\underline{s}\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n})<\underline{s}\Big\}\Bigg].\end{aligned} \tag{A-8}$$

Let us begin with the term $\mathbb{1}\Big\{\widehat{H}_{p_N}\big(Z_{p_n},\theta\big|L_1\big)>b\Big\}\mathbb{1}\Big\{H_p\big(Z_{p_n},\theta\big|L_1\big)<b\Big\}\mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n})\geq\underline{s}\Big\}$. Let

$$\overline{T}_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)=T_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)-E\Big[T_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)\Big|Z_{p_n}\Big].$$

Using a triangle inequality,

$$\begin{aligned}&\left|\frac{1}{N}\sum_{\ell=1}^{N}T_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)\right|\leq\left|\frac{1}{N}\sum_{\ell=1}^{N}\overline{T}_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)\right|+\left|\frac{1}{N}\sum_{\ell=1}^{N}E\Big[T_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)\Big|Z_{p_n}\Big]\right|\\ &\leq\left|\frac{1}{N}\sum_{\ell=1}^{N}\overline{T}_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)\right|+\left|\frac{1}{N}\sum_{\ell=1}^{N}E\Big[S_{p_N}\big(X_{-p_\ell},Z_{p_n},\theta\big|L_1\big)\Big|Z_{p_n}\Big]\right|+\sup_{z\in\mathcal{Z}_p}\Big|\xi_{p_N}\big(z,\theta\big|L_1\big)\Big|\end{aligned} \tag{A-9}$$

Let $\bar{\epsilon}$ be the constant described in Assumption E1. Recall that $h_1^{d_p}\xi_{p_N}(z,\theta|L_1)$ is bounded w.p.1 and we also have $\sup_{z\in\mathcal{Z}_p}\Big|\xi_{p_N}(z,\theta|L_1)\Big| = O_p(N^{\epsilon-1}h_1^{d_p})$ for any $\epsilon > 0$. In particular, this holds for the aforementioned $\bar{\epsilon}$. This implies that[7]

$$\sup_{z\in\mathcal{Z}_p}\Big|\xi_{p_N}(Z_{p_n},\theta|L_1)\Big| \le \frac{N^{\bar{\epsilon}}}{Nh_1^{2d_p}} \times \overline{\mathcal{W}} \quad \text{w.p.1 for some scalar } \overline{\mathcal{W}}. \tag{A-10}$$

Let

$$\begin{aligned}\sigma_{p_N}(Z_{p_n}|L_1) &= \Big(b - H_{p_N}(Z_{p_n},\theta|L_1)\Big) - \max_{n=1,\dots,N}\Bigg|E\Big[S_{p_N}(X_{-p_\ell},Z_{p_n},\theta|L_1)\Big|Z_{p_n}\Big]\Bigg| - \frac{\overline{\mathcal{C}}K^p(0)}{Nh_1^{d_p}} - \frac{N^{\bar{\epsilon}}\cdot\overline{\mathcal{W}}}{Nh_1^{2d_p}}\\ &= \Big(b - H_{p_N}(Z_{p_n},\theta|L_1)\Big) + o(1).\end{aligned}$$

Equations $(A-8)-(A-10)$ yield

$$\begin{aligned}&\mathbb{1}\Big\{\widehat{H}_{p_N}(Z_{p_n},\theta|L_1) > b\Big\}\mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\}\\ &\le \mathbb{1}\Bigg\{\Bigg|\frac{1}{N}\sum_{\ell=1}^{N}E\Big[\overline{T}_{p_N}(X_{-p_\ell},Z_{p_n},\theta|L_1)\Big]\Bigg| > \sigma_{p_N}(Z_{p_n}|L_1)\Bigg\}\mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\}.\end{aligned} \tag{A-11}$$

For large enough N, we have

$$\begin{aligned}&\mathbb{1}\Big\{\sigma_{p_N}(Z_{p_n}|L_1) > 0\Big\}\mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\} =\\ &\qquad \mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\} \quad \text{w.p.1 for all } n = 1,\dots,N.\end{aligned}$$

Using Bernstein's inequality (which is pertinent here due to the boundedness of $S_{p_N}(X_{-p_\ell},Z_p,\theta|L_1)$, and the assumed independence across observations), this implies that for large enough N, there exist two constants $A > 0$ and $B > 0$ such that

$$\begin{aligned}&E\Bigg[\mathbb{1}\Big\{\widehat{H}_{p_N}(Z_{p_n},\theta|L_1) > b\Big\}\mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\}\Bigg|Z_{p_n}\Bigg]\\ &\qquad \le \exp\Bigg\{-\frac{Nh_1^{d_p}\sigma_{p_N}^2(Z_{p_n}|L_1)}{A + B\sigma_{p_N}(Z_{p_n}|L_1)}\Bigg\}\mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\} \quad \text{for each } (Z_{p_n})_{n=1}^{N}.\end{aligned} \tag{A-12}$$

Since $Nh_1^{d_p}/\log(N) \to \infty$, Lebesgue's convergence theorem[8] implies

$$E\Bigg[\max_{n=1,\dots,N}\Bigg(N \times \exp\Bigg\{-\frac{Nh_1^{d_p}\sigma_{p_N}^2(Z_{p_n}|L_1)}{A + B\sigma_{p_N}(Z_{p_n}|L_1)}\Bigg\}\mathbb{1}\Big\{H_p(Z_{p_n},\theta|L_1) < b\Big\}\mathbb{1}\Big\{f_{Z_p}(Z_{p_n}) \ge \underline{s}\Big\}\Bigg)\Bigg] \longrightarrow 0,$$

[7] For our purposes, the only relevant fact here is that h_1 has the property that $N^{1/2}h_1 \longrightarrow \infty$. . We will come back to this issue below, when we deal with L_k for $k \ge 2$.

[8] Note that, for large enough N, $\exp\Big\{\log(N) - \frac{Nh_1^{d_p}\sigma_{p_N}^2(Z_{p_n}|L_1)}{A+B\sigma_{p_N}(Z_{p_n}|L_1)}\Big\} \le 1$ w.p.1.

and therefore

$$\sum_{n=1}^{N} E\left[\mathbb{1}\left\{\widehat{H}_{p_N}(Z_{p_n},\theta|L_1) > b\right\}\mathbb{1}\left\{H_p(Z_{p_n},\theta|L_1) < b\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) \geq \underline{s}\right\}\right] \longrightarrow 0.$$

We can show that

$$\sum_{n=1}^{N} E\left[\mathbb{1}\left\{\widehat{H}_{p_N}(Z_{p_n},\theta|L_1) < b\right\}\mathbb{1}\left\{H_p(Z_{p_n},\theta|L_1) > b\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) \geq \underline{s}\right\}\right] \longrightarrow 0$$

by following the exact same steps if we start by redefining $\alpha^*_M \equiv -\alpha_M$ and $\delta^*_M \equiv -\delta_M$ in Equation $(A-2)$, and $b^* \equiv -b$ in Equation $(A-6)$. The term $\mathbb{1}\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\}\mathbb{1}\{f_{Z_p}(Z_{p_n}) < \underline{s}\}$ is simpler to analyze. Let

$$\begin{aligned}
\widetilde{\sigma}_{p_N}(Z_{p_n}) &= \underline{s} - f_{Z_p}(Z_{p_n}) - \max_{n=1,\ldots,N}\left(E\left[\frac{1}{h_1^{d_p}}K^p_{h_1}(Z_{p_\ell}-Z_{p_n})\Big|Z_{p_n}\right] - f_{Z_p}(Z_{p_n})\right) \\
&= \underline{s} - f_{Z_p}(Z_{p_n}) + o(1), \\
\widetilde{T}_{p_N}(Z_{p_\ell},Z_{p_n}) &= \frac{1}{h_1^{d_p}}K^p_{h_1}(Z_{p_\ell}-Z_{p_n}) - E\left[\frac{1}{h_1^{d_p}}K^p_{h_1}(Z_{p_\ell}-Z_{p_n})\Big|Z_{p_n}\right].
\end{aligned}$$

Then,

$$\mathbb{1}\left\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\} \leq \mathbb{1}\left\{\frac{1}{N}\sum_{\ell=1}^{N}\widetilde{T}_{p_N}(Z_{p_\ell},Z_{p_n}) > \widetilde{\sigma}_{p_N}(Z_{p_n})\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\}.$$

For large enough N, w.p.1 we have $\mathbb{1}\left\{\widetilde{\sigma}_{p_N}(Z_{p_n}) > 0\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\} = \mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\}$. For large enough N, Bernstein's inequality applies and we have

$$E\left[\mathbb{1}\left\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\}\Big|Z_{p_n}\right] \leq \exp\left\{-\frac{Nh_1^{d_p}\widetilde{\sigma}^2_{p_N}(Z_{p_n})}{\widetilde{A} + \widetilde{B}\cdot\widetilde{\sigma}_{p_N}(Z_{p_n})}\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\}. \tag{A-13}$$

for a pair of constants $\widetilde{A} > 0$ and $\widetilde{B} > 0$. Once again, since $Nh_N^{d_p}/\log(N) \to \infty$, Lebesgue's convergence theorem yields

$$E\left[\max_{n=1,\ldots,N}\left(N \times \exp\left\{-\frac{Nh_1^{d_p}\widetilde{\sigma}^2_{p_N}(Z_{p_n})}{\widetilde{A} + \widetilde{B}\cdot\widetilde{\sigma}_{p_N}(Z_{p_n})}\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\}\right)\right] \longrightarrow 0,$$

and therefore,

$$\sum_{n=1}^{N} E\left[\mathbb{1}\left\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\right\}\mathbb{1}\left\{f_{Z_p}(Z_{p_n}) < \underline{s}\right\}\right] \longrightarrow 0. \tag{A-14}$$

These steps we used to analyze the term $\mathbb{1}\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\}\mathbb{1}\{f_{Z_p}(Z_{p_n}) < \underline{s}\}$ are the same as those in Lemma 26 in Ichimura (2004). Finally, Equation $(A-7)$ yields

$$E\Big[\sum_{n=1}^{N}\mathbb{1}^*(Z_{p_n})\Big] = 0.$$

Together, these results prove the statement in Proposition A-5. □

Remark 4 Given the trimming function $\mathbb{1}\Big\{\widehat{f}_{Z_p}(Z_{p_n}) > \underline{s}\Big\}$ that we employ, Equation $(A-14)$ is equivalent to the statement

$$\Pr\Big(f_{Z_p}(Z_{p_n}) > \underline{s} \quad \text{for some} \quad n = 1, \ldots, N\Big) \longrightarrow 0. \tag{A-15}$$

Given our assumptions, this result holds for any one of the bandwidth sequences in the collection $h_1, \ldots, h_k$. This fact will be in the background of the remaining results.

Go back to the objects defined in Equations (14) and (16). We will be explicit about their dependence on Z_p and θ and will express them as $\underline{M}_k^*(Z_p, \theta)$, $\underline{D}_k^*(Z_p, \theta)$, $\overline{M}_k^*(Z_p, \theta)$ and $\overline{D}_k^*(Z_p, \theta)$. We will do the same with their semiparametric analog counterparts $\widehat{\underline{M}}_k^*(Z_p, \theta)$, $\widehat{\underline{D}}_k^*(Z_p, \theta)$, $\widehat{\overline{M}}_k^*(Z_p, \theta)$ and $\widehat{\overline{D}}_k^*(Z_p, \theta)$ respectively.

Proposition A-6 *Let* $\mathbb{H}_{-p}^M\big(Z_p, \theta \big| L_1\big)$ *denote either* $\underline{\mathbb{F}}_{-p}^M\big(Z_p, \theta \big| L_1\big)$ *or* $\overline{\mathbb{F}}_{-p}^M\big(Z_p, \theta \big| L_1\big)$. *Take any linear combination of the form* $\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}_{-p}^M\big(Z_p, \theta \big| L_1\big)$ *and suppose that for every* $\theta \in \Theta$ *such that* $\Delta_q^M \neq 0$ *for some* $q \neq p$ *and some* M *the following holds*

$$Pr\left[\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}_{-p}^M\big(Z_p, \theta \big| L_1\big) = 1 \middle| Z_p \in \mathcal{Z}_p\right] = 0 \tag{A-16}$$

for any such linear combination. Let $R_k^*(Z_p, \theta)$ *stand for* $\underline{M}_k^*(Z_p, \theta)$, $\underline{D}_k^*(Z_p, \theta)$, $\overline{M}_k^*(Z_p, \theta)$ *or* $\overline{D}_k^*(Z_p, \theta)$, *with* $\widehat{R}_k^*(Z_p, \theta)$ *being its semiparametric analog counterpart. Given our previous assumptions, if condition* $(A-16)$ *holds, then*

$$Pr\bigg(R_{k=2}^*(Z_{p_n}, \theta) \neq \widehat{R}_{k=2}^*(Z_{p_n}, \theta) \;\; \text{for some} \;\; n = 1, \ldots, N\bigg) \longrightarrow 0 \tag{A-17}$$

for all θ *that satisfies* $(A-16)$ *and any* θ *such that* $\Delta_p^M = 0$ *for all* M.

Proof: If θ is such that $\Delta_p^M = 0$ for all M, then Equations (14) and (20) immediately yield $R_k^*(Z_{p_n}, \theta) = \widehat{R}_k^*(Z_{p_n}, \theta) = \mathcal{P} - 2$ for all n and the result in Proposition A-6 follows. Otherwise the result follows from Proposition A-5 and the definitions of $R_k^*(Z_p, \theta)$ and $\widehat{R}_k^*(Z_p, \theta)$ in Equations (14) and (20). □

Remark 5 We have established before that, if $\underline{\Delta}_q \neq \overline{\Delta}_q$ for some $q \neq p$, we have

$$\sum_{M=0}^{\mathcal{P}-1} \underline{\mathbb{F}}_{-p}^M\big(Z_p, \theta \big| L_1\big) < 1, \quad \text{and} \quad \sum_{M=0}^{\mathcal{P}-1} \overline{\mathbb{F}}_{-p}^M\big(Z_p, \theta \big| L_1\big) > 1.$$

The condition in Equation $(A-16)$ assumes that the probability of finding a combination of $\underline{\mathbb{F}}_{-p}^M$'s and $\overline{\mathbb{F}}_{-p}^M$'s that is exactly equal to one is zero (recall also that we have assumed throughout that

Z_p is continuously distributed). If θ is such that $\Delta_q^M = 0$ for all $q \neq p$ and all M, then we have $\underline{\mathbb{F}}_{-p}^M(Z_p, \theta|L_1) = \overline{\mathbb{F}}_{-p}^M(Z_p, \theta|L_1) = \mathbb{F}_{-p}^M(Z_p, \theta)$ for all M and therefore, condition $(A-16)$ is violated.

Take any value of the parameter vector θ such that either $\Delta_p^M = 0$ for all M, or Equation $(A-16)$ is satisfied, and consider a corresponding index set $\mathcal{J}$ as described in Equation (13) (we stress once again the fact that the rankings in $\mathcal{J}$ depend exclusively on θ). Let $\overline{U}_k^* = \min\{\overline{M}_k^*, \overline{D}_k^*\}$ and $\underline{U}_k^* = \min\{\underline{M}_k^*, \underline{D}_k^*\}$. Proposition A-6 implies that with probability approaching one uniformly in $n = 1, \ldots, N$, we can express

$$
\begin{aligned}
&\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta \big| L_2\big) - \overline{\mu}_p\big(Z_{p_n}, \theta \big| L_2\big) = \\
&\sum_{j=1}^{p^*} \Bigg\{ \phi_{p_j}^a \left(\Delta_p^{(j)}, \Delta_p^{(\overline{U}_2^*)}, \overline{U}_2^*\right) \left(\widehat{\overline{\mathbb{F}}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big) - \overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big)\right) \\
&\qquad\qquad + \phi_{p_j}^b \left(\Delta_p^{(j)}, \Delta_p^{(\overline{U}_2^*)}, \overline{U}_2^*\right) \left(\widehat{\underline{\mathbb{F}}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big) - \underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big)\right) \Bigg\}, \\
&\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta \big| L_2\big) - \underline{\mu}_p\big(Z_{p_n}, \theta \big| L_2\big) = \\
&\sum_{j=1}^{p^*} \Bigg\{ \phi_{p_j}^c \left(\Delta_p^{(j)}, \Delta_p^{(\underline{U}_2^*)}, \underline{U}_2^*\right) \left(\widehat{\overline{\mathbb{F}}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big) - \overline{\mathbb{F}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big)\right) \\
&\qquad\qquad + \phi_{p_j}^d \left(\Delta_p^{(j)}, \Delta_p^{(\underline{U}_2^*)}, \underline{U}_2^*\right) \left(\widehat{\underline{\mathbb{F}}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big) - \underline{\mathbb{F}}_{-p}^{(j)}\big(Z_p, \theta \big| L_1\big)\right) \Bigg\},
\end{aligned}
\tag{A-18}
$$

where $\phi_{p_j}^a(\cdot)$, $\phi_{p_j}^b(\cdot)$, $\phi_{p_j}^c(\cdot)$ and $\phi_{p_j}^d(\cdot)$ would be given by Equation (15)

(15). Let $\overline{G}^M_{-p}(\cdot)$ and $\underline{G}^M_{-p}(\cdot)$ be as defined in Equation (23). As we have done throughout, we will denote $X_{-p} \equiv \big(X_q\big)_{q\neq p}$, $Z_{-p} \equiv \big(Z_q\big)_{q\neq p}$ and $Z = \big(Z_p\big)_{p=1}^{\mathcal{P}}$. Let

$$\overline{\varphi}_p\big(X_{-p_\ell}, Z_p, \theta\big|L_1\big) = \sum_{j=1}^{p^*}\Bigg\{\phi^a_{p_j}\Big(\Delta^{(j)}_p, \Delta^{(\overline{U}^*_2)}_p, \overline{U}^*_2\Big)\Big(\overline{G}^{(j)}_{-p}\Big(\big(X'_{q\ell}\beta_q, \overline{\Delta}_q, \underline{\Delta}_q\big)_{q\neq p}\ ;\ Z_p\Big) - \overline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta\big|L_1\big)\Big)$$
$$+ \phi^b_{p_j}\Big(\Delta^{(j)}_p, \Delta^{(\overline{U}^*_2)}_p, \overline{U}^*_2\Big)\Big(\underline{G}^{(j)}_{-p}\Big(\big(X'_{q\ell}\beta_q, \overline{\Delta}_q, \underline{\Delta}_q\big)_{q\neq p}\ ;\ Z_p\Big) - \underline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta\big|L_1\big)\Big)\Bigg\},$$

$$\underline{\varphi}_p\big(X_{-p_\ell}, Z_p, \theta\big|L_1\big) = \sum_{j=1}^{p^*}\Bigg\{\phi^c_{p_j}\Big(\Delta^{(j)}_p, \Delta^{(\underline{U}^*_2)}_p, \underline{U}^*_2\Big)\Big(\overline{G}^{(j)}_{-p}\Big(\big(X'_{q\ell}\beta_q, \overline{\Delta}_q, \underline{\Delta}_q\big)_{q\neq p}\ ;\ Z_p\Big) - \overline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta\big|L_1\big)\Big)$$
$$+ \phi^d_{p_j}\Big(\Delta^{(j)}_p, \Delta^{(\underline{U}^*_2)}_p, \underline{U}^*_2\Big)\Big(\underline{G}^{(j)}_{-p}\Big(\big(X'_{q\ell}\beta_q, \overline{\Delta}_q, \underline{\Delta}_q\big)_{q\neq p}\ ;\ Z_p\Big) - \underline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta\big|L_1\big)\Big)\Bigg\}. \tag{A-19}$$

Note that $E\Big[\overline{\varphi}_p\big(X_{-p_\ell}, z, \theta\big|L_1\big)\Big|Z_{p_\ell} = z\Big] = 0$, $E\Big[\underline{\varphi}_p\big(X_{-p_\ell}, z, \theta\big|L_1\big)\Big|Z_{p_\ell} = z\Big] = 0$. Given our assumptions, for each player p the following representation holds[9]:

$$\widehat{\overline{\mu}}_p\big(Z_{p_\ell}, \theta\big|L_2\big) - \overline{\mu}_p\big(Z_{p_\ell}, \theta\big|L_2\big) = \frac{1}{f_{Z_p}(Z_{p_\ell})}\frac{1}{Nh_1^{d_p}}\sum_{m=1}^{N}\overline{\varphi}_p\big(X_{-p_m}, Z_{p_\ell}, \theta\big|L_1\big)K^p_{h_1}\big(Z_{p_m} - Z_{p_\ell}\big) + \overline{\xi}_{p_N}\big(Z_{p_\ell}, \theta\big|L_2\big),$$
$$\widehat{\underline{\mu}}_p\big(Z_{p_\ell}, \theta\big|L_2\big) - \underline{\mu}_p\big(Z_{p_\ell}, \theta\big|L_2\big) = \frac{1}{f_{Z_p}(Z_{p_\ell})}\frac{1}{Nh_1^{d_p}}\sum_{m=1}^{N}\underline{\varphi}_p\big(X_{-p_m}, Z_{p_\ell}, \theta\big|L_1\big)K^p_{h_1}\big(Z_{p_m} - Z_{p_\ell}\big) + \underline{\xi}_{p_N}\big(Z_{p_\ell}, \theta\big|L_2\big), \tag{A-20}$$

where $\max_{\ell=1,\ldots,N}\big|\overline{\xi}_{p_N}(Z_{p_\ell}, \theta|L_2)\big| = o_p(N^{-1/2})$ and $\max_{\ell=1,\ldots,N}\big|\underline{\xi}_{p_N}(Z_{p_\ell}, \theta|L_2)\big| = o_p(N^{-1/2})$. For each $q \neq p$ we will let $\nabla_{\overline{\mu}_q}\overline{G}^M_{-p}(\cdot)$ and $\nabla_{\underline{\mu}_q}\overline{G}^M_{-p}(\cdot)$ denote the partial derivative of $\overline{G}^M_{-p}(\cdot)$ with respect to $\overline{\mu}_q(\cdot)$ and $\underline{\mu}_q(\cdot)$, respectively. Denote the second cross-partial derivatives as $\nabla_{\overline{\mu}_q\overline{\mu}_q}\overline{G}^M_{-p}(\cdot)$, $\nabla_{\underline{\mu}_q\underline{\mu}_q}\overline{G}^M_{-p}(\cdot)$ and $\nabla_{\overline{\mu}_q\underline{\mu}_q}\overline{G}^M_{-p}(\cdot)$. Likewise, we use $\nabla_{\overline{\mu}_q}\underline{G}^M_{-p}(\cdot)$, $\nabla_{\underline{\mu}_q}\underline{G}^M_{-p}(\cdot)$, $\nabla_{\overline{\mu}_q\overline{\mu}_q}\underline{G}^M_{-p}(\cdot)$, $\nabla_{\underline{\mu}_q\underline{\mu}_q}\underline{G}^M_{-p}(\cdot)$ and $\nabla_{\overline{\mu}_q\underline{\mu}_q}\underline{G}^M_{-p}(\cdot)$ to denote the same objects for $\underline{G}^M_{-p}(\cdot)$. Given our

[9] See Remark 4.

assumptions, all these objects are bounded w.p.1. For simplicity, we will abbreviate

$$\overline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\big)_{q\neq p}\ ,\ \overline{\mathbf{\Lambda}}_{-p}(Z_{-p_\ell},\theta|L_k)\ ,\ \underline{\mathbf{\Lambda}}_{-p}(Z_{-p_\ell},\theta|L_k)\ ;\ Z_{p_n}\Big) \equiv \overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_k\big),$$
$$\underline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\big)_{q\neq p}\ ,\ \overline{\mathbf{\Lambda}}_{-p}(Z_{-p_\ell},\theta|L_k)\ ,\ \underline{\mathbf{\Lambda}}_{-p}(Z_{-p_\ell},\theta|L_k)\ ;\ Z_{p_n}\Big) \equiv \underline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_k\big). \tag{A-21}$$

Given our assumptions, Equation $(A-20)$ and Lemma 3 in Collomb and Hardle (1986) (or Theorem A-1 in Aradillas-Lopez (2006)) yield

$$\frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{\ell=1}^{N}\nabla_{\overline{\mu}_q}\overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_2\big)K^p_{h_2}\big(Z_{p_\ell}-Z_{p_n}\big)\cdot\Big(\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_2\big)-\overline{\mu}_p\big(Z_{p_n},\theta\big|L_2\big)\Big)=$$
$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{\ell=1}^{N}\nabla_{\overline{\mu}_q}\overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_2\big)K^p_{h_2}\big(Z_{p_\ell}-Z_{p_n}\big)\cdot\Big(\widehat{\overline{\mu}}_p\big(Z_{p_n},\theta\big|L_2\big)-\overline{\mu}_p\big(Z_{p_n},\theta\big|L_2\big)\Big)$$
$$+\,\overline{\varrho}_{p_N}\big(Z_{p_n},\theta\big|L_2\big),\qquad \text{where}\quad \max_{n=1,\dots,N}\Big|\overline{\varrho}_{p_N}\big(Z_{p_n},\theta\big|L_2\big)\Big| = o_p(N^{-1/2}).$$

An equivalent result follows for

$$\frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{\ell=1}^{N}\nabla_{\underline{\mu}_q}\overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_2\big)K^{d_p}_{h_2}\big(Z_{p_\ell}-Z_{p_n}\big)\cdot\Big(\widehat{\underline{\mu}}_p\big(Z_{p_n},\theta\big|L_2\big)-\underline{\mu}_p\big(Z_{p_n},\theta\big|L_2\big)\Big).$$

Given Equation $(A-20)$, a second order approximation yields

$$\begin{aligned}
&\widehat{\overline{\mathbb{F}}}^M_p\big(Z_{p_n},\theta\big|L_2\big)\\
&=\frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{\ell=1}^{N}\overline{G}^M_{-p}\Big(\big(X'_{q\ell}\beta_q\big)_{q\neq p}\ ,\ \widehat{\overline{\mathbf{\Lambda}}}_{-p}(Z_{-p_\ell},\theta|L_2)\ ,\ \widehat{\underline{\mathbf{\Lambda}}}_{-p}(Z_{-p_\ell},\theta|L_2)\ ;\ Z_{p_n}\Big)K^p_{h_2}\big(Z_{p_\ell}-Z_{p_n}\big)\\
&=\frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{\ell=1}^{N}\overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_2\big)K^p_{h_2}\big(Z_{p_\ell}-Z_{p_n}\big)\\
&+\frac{1}{f_{Z_p}(Z_{p_n})N^2h_2^{d_p}}\sum_{m=1}^{N}\sum_{\ell=1}^{N}\sum_{q\neq p}\frac{1}{f_{Z_q}(Z_{q_\ell})}\Bigg\{\frac{1}{h_1^{d_q}}\nabla_{\overline{\mu}_q}\overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_2\big)\overline{\varphi}_q\big(X_{-q_m},Z_{q_\ell},\theta\big|L_1\big)K^q_{h_1}\big(Z_{q_m}-Z_{q_\ell}\big)\\
&+\frac{1}{h_1^{d_q}}\nabla_{\underline{\mu}_q}\overline{G}^M_{-p}\big(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta\big|L_2\big)\underline{\varphi}_q\big(X_{-q_m},Z_{q_\ell},\theta\big|L_1\big)K^q_{h_1}\big(Z_{q_m}-Z_{q_\ell}\big)\Bigg\}K^p_{h_2}\big(Z_{p_\ell}-Z_{p_n}\big)+\overline{\zeta}^M_{p_N}\big(Z_{p_n},\theta|L_2\big),
\end{aligned} \tag{A-22}$$

where $\max_{n=1,\ldots,N}\left|\overline{\zeta}^{M}_{p_N}(Z_{p_n},\theta|L_2)\right| = o_p(N^{-1/2})$. For each $q \neq p$ define[10]

$$\overline{D}^{M}_{-p,\overline{\mu}_q}(Z_{q_m},Z_{p_n},\theta|L_k) =$$
$$\begin{cases} E\Big[\nabla_{\overline{\mu}_q}\overline{G}^{M}_{-p}(X_{-p},Z_{-p},Z_p,\theta|L_k)\Big|Z_{p_n},Z_{q_m}\Big]\cdot f_{Z_p|Z_q}(Z_{p_n}|Z_{q_m}) & \text{if } Z_p \neq Z_q \text{ (players } p \text{ and } q \text{ use different signals)} \\ \frac{1}{h_k^{d_p}}E\Big[\nabla_{\overline{\mu}_q}\overline{G}^{M}_{-p}(X_{-p},Z_{-p},Z_q,\theta|L_k)\Big|Z_{p_n}\Big]K^{p}_{h_k}(Z_{p_m}-Z_{p_n}) & \text{if } Z_p = Z_q \text{ (players } p \text{ and } q \text{ use the same signals)} \end{cases},$$

$$\overline{D}^{M}_{-p,\underline{\mu}_q}(Z_{q_m},Z_{p_n},\theta|L_k) =$$
$$\begin{cases} E\Big[\nabla_{\underline{\mu}_q}\overline{G}^{M}_{-p}(X_{-p},Z_{-p},Z_p,\theta|L_k)\Big|Z_{p_n},Z_{q_m}\Big]\cdot f_{Z_p|Z_q}(Z_{p_n}|Z_{q_m}) & \text{if } Z_p \neq Z_q \text{ (players } p \text{ and } q \text{ use different signals)} \\ \frac{1}{h_k^{d_p}}E\Big[\nabla_{\underline{\mu}_q}\overline{G}^{M}_{-p}(X_{-p},Z_{-p},Z_p,\theta|L_k)\Big|Z_{p_n}\Big]K^{p}_{h_k}(Z_{p_m}-Z_{p_n}) & \text{if } Z_p = Z_q \text{ (players } p \text{ and } q \text{ use the same signals)} \end{cases} \quad \text{(A-23)}$$

Let us stack

$$\begin{aligned} \overline{D}^{M}_{-p,\mu_q}(Z_{q_m},Z_{p_n},\theta|L_k) &= \Big(\overline{D}^{M}_{-p,\underline{\mu}_q}(Z_{q_m},Z_{p_n},\theta|L_k)\ ,\ \underline{D}^{M}_{-p,\overline{\mu}_q}(Z_{q_m},Z_{p_n},\theta|L_k)\Big)', \\ \varphi_q(X_{-q_m},Z_{q_m},\theta|L_1) &= \Big(\underline{\varphi}_q(X_{-q_m},Z_{q_m},\theta|L_1)\ ,\ \overline{\varphi}_q(X_{-q_m},Z_{q_m},\theta|L_1)\Big)'. \end{aligned} \quad \text{(A-24)}$$

Given our assumptions, using the results in Ahn and Powell (1993) or Sherman (1994) the projection of the second-order U-statistic that appears in Equation $(A-22)$, conditional on Z_n, is given by

$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{m=1}^{N}\sum_{q\neq p}\overline{D}^{M}_{-p,\mu_q}(Z_{q_m},Z_{p_n},\theta|L_2)'\varphi_q(X_{-q_m},Z_{q_m},\theta|L_1)+\overline{\vartheta}_{p_N}(Z_{p_n},\theta|L_2),$$

where $\max_{n=1,\ldots,N}\left|\overline{\vartheta}_{p_N}(Z_{p_n},\theta|L_2)\right| = o_p(N^{-1/2})$. Notice that

$$E\Big[\overline{D}^{M}_{-p,\mu_q}(Z_{q_m},Z_{p_n},\theta|L_k)'\varphi_q(X_{-q_m},Z_{q_m},\theta|L_1)\Big|Z_{p_n}\Big]=0.$$

Now, from Lemma 3 in Collomb and Hardle (1986) or Theorem A-1 in Aradillas-Lopez (2006), we have

$$\begin{aligned} &\frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{m=1}^{N}\overline{G}^{M}_{-p}(X_{-p_m},Z_{-p_m},Z_{p_n},\theta|L_2)K^{p}_{h_2}(Z_{p_m}-Z_{p_n})-\overline{\mathbb{F}}^{M}_{-p}(Z_{p_n},\theta|L_2)= \\ &\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{Nh_2^{d_p}}\sum_{m=1}^{N}\Big[\overline{G}^{M}_{-p}(X_{-p_m},Z_{-p_m},Z_{p_n},\theta|L_2)-\overline{\mathbb{F}}^{M}_{-p}(Z_{p_n},\theta|L_2)\Big]K^{p}_{h_2}(Z_{p_m}-Z_{p_n})+\overline{\varsigma}_{p_N}(Z_{p_n},\theta|L_2), \end{aligned}$$

where $\max_{n=1,\ldots,N}\left|\overline{\varsigma}_{p_N}(Z_{p_n},\theta|L_2)\right| = o_p(N^{-1/2})$. Using the results above, we can extend Proposition A-6 to the case $k=3$.

[10]Recall that $Z\equiv(Z_p)_{p=1}^{\mathcal{P}}$, so $Z_n\equiv(Z_{p_n})_{p=1}^{\mathcal{P}}$.

Proposition A-7 *Suppose the conditions in Proposition A-6 are satisfied and let* $R^*_k(Z_p,\theta)$ *be as defined there. Let* $\mathbb{H}^M_{-p}(Z_p,\theta|L_2)$ *denote either* $\underline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_2)$ *or* $\overline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_2)$. *Suppose that for every* $\theta\in\Theta$ *such that* $\Delta^M_q\neq 0$ *for some* $q\neq p$ *and some* M *the following holds*

$$Pr\left[\sum_{M=0}^{\mathcal{P}-1}\mathbb{H}^M_{-p}(Z_p,\theta|L_2)=1\middle|Z_p\in\mathcal{Z}_p\right]=0. \tag{A-25}$$

Then

$$Pr\bigg(R^*_{k=3}(Z_{p_n},\theta)\neq\widehat{R}^*_{k=3}(Z_{p_n},\theta)\ \ \textit{for some}\ \ n=1,\ldots,N\bigg)\longrightarrow 0 \tag{A-26}$$

for all θ *that satisfies* $(A-25)$ *and any* θ *such that* $\Delta^M_p=0$ *for all* M.

Proof: It suffices to show that Proposition A-5 holds for $k=2$. The results preceding Proposition A-7 will be enough to establish that. Note that those results relied on the validity of Proposition A-6, which is why we maintain the assumption that such proposition holds. We begin by extending the definition in Equation $(A-1)$ to $k=2$. Let

$$\overline{T}^M_{-p_N}(X_{-p\ell},Z_p,\theta|L_2)=\left(\frac{1}{f_{Z_p}(Z_p)}\right)\left[\overline{G}^M_{-p}\Big(X_{-p\ell},Z_{-p\ell},Z_p,\theta\Big|L_2\Big)-\overline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_2)\right]K^p_{h_2}(Z_{p\ell}-Z_p),$$
$$\underline{T}^M_{-p_N}(X_{-p\ell},Z_p,\theta|L_2)=\left(\frac{1}{f_{Z_p}(Z_p)}\right)\left[\underline{G}^M_{-p}\Big(X_{-p\ell},Z_{-p\ell},Z_p,\theta\Big|L_2\Big)-\underline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_2)\right]K^p_{h_2}(Z_{p\ell}-Z_p) \tag{A-27}$$

The previous results yield the following linear representation for $\widehat{\overline{\mathbb{F}}}^M_{-p}(Z_{p_n},\theta|L_2)-\overline{\mathbb{F}}^M_{-p}(Z_{p_n},\theta|L_2)$,

$$\widehat{\overline{\mathbb{F}}}^M_{-p}(Z_{p_n},\theta|L_2)-\overline{\mathbb{F}}^M_{-p}(Z_{p_n},\theta|L_2)=$$
$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{\ell=1}^{N}\left[\frac{1}{h_2^{d_p}}\overline{T}^M_{-p_N}(X_{-p\ell},Z_{p_n},\theta|L_2)+\sum_{q\neq p}\overline{D}^M_{-p,\mu_q}(Z_{q\ell},Z_{p_n},\theta|L_2)'\varphi_q(X_{-q\ell},Z_{q\ell},\theta|L_1)\right]+\overline{\nu}^M_{pN}(Z_{p_n},\theta|L_2),$$

where $\max_{n=1,\ldots,N}\left|\overline{\nu}^M_{pN}(Z_{p_n},\theta|L_2)\right|=o_p(N^{-1/2})$. Let $\underline{D}^M_{-p,\mu_q}(\cdot)$ be the equivalent to the objects defined in Equation $(A-23)$ replacing $\overline{G}^M_{-p}(\cdot)$ with $\underline{G}^M_{-p}(\cdot)$. The same steps as above yield

$$\widehat{\underline{\mathbb{F}}}^M_{-p}(Z_{p_n},\theta|L_2)-\underline{\mathbb{F}}^M_{-p}(Z_{p_n},\theta|L_2)=$$
$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{\ell=1}^{N}\left[\frac{1}{h_2^{d_p}}\underline{T}^M_{-p_N}(X_{-p\ell},Z_{p_n},\theta|L_2)+\sum_{q\neq p}\underline{D}^M_{-p,\mu_q}(Z_{q\ell},Z_{p_n},\theta|L_2)'\varphi_q(X_{-q\ell},Z_{q\ell},\theta|L_1)\right]+\underline{\nu}^M_{pN}(Z_{p_n},\theta|L_2),$$

where $\max_{n=1,\ldots,N}\left|\underline{\nu}^M_{pN}(Z_{p_n},\theta|L_2)\right|=o_p(N^{-1/2})$. Let $(\alpha_M,\delta_M)_{M=0}^{\mathcal{P}-1}$ be an arbitrary collection of scalars, and let

$$H_p(Z_p,\theta|L_2)=\sum_{M=0}^{\mathcal{P}-1}\Big(\alpha_M\underline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_2)+\delta_M\overline{\mathbb{F}}^M_{-p}(Z_p,\theta|L_1)\Big),$$
$$\widehat{H}_{pN}(Z_p,\theta|L_2)=\sum_{M=0}^{\mathcal{P}-1}\Big(\alpha_M\widehat{\underline{\mathbb{F}}}^M_{-p}(Z_p,\theta|L_2)+\delta_M\widehat{\overline{\mathbb{F}}}^M_{-p}(Z_p,\theta|L_2)\Big). \tag{A-28}$$

For the same collection of scalars, let

$$S_{p_N}\left(X_{-p_\ell}, Z_p, \theta \middle| L_2\right) = \sum_{M=0}^{\mathcal{P}-1}\left(\alpha_M \underline{T}^M_{-p_N}\left(X_{-p_\ell}, Z_p, \theta \middle| L_2\right) + \delta_M \overline{T}^M_{-p_N}\left(X_{-p_\ell}, Z_p, \theta \middle| L_2\right)\right)$$

$$Q_{p_N}\left(X_{-q_\ell}, Z_{q_\ell}, Z_p, \theta \middle| L_2\right) = \frac{1}{f_{Z_p}(Z_p)}\sum_{M=0}^{\mathcal{P}-1}\sum_{q\neq p}\Big\{\alpha^M \underline{D}^M_{-p,\mu_q}\left(Z_{q_\ell}, Z_p, \theta \middle| L_2\right)'\varphi_q\left(X_{-q_\ell}, Z_{q_\ell}, \theta \middle| L_1\right)$$
$$+ \delta^M \overline{D}^M_{-p,\mu_q}\left(Z_{q_\ell}, Z_p, \theta \middle| L_2\right)'\varphi_q\left(X_{-q_\ell}, Z_{q_\ell}, \theta \middle| L_1\right)\Big\}.$$

Our previous results yield

$$\widehat{H}_{p_N}\left(Z_{p_n}, \theta \middle| L_2\right) - H_p\left(Z_{p_n}, \theta \middle| L_2\right) = \frac{1}{N}\sum_{\ell=1}^{N}\left[\frac{1}{h_2^{d_p}} S_{p_N}\left(X_{-p_\ell}, Z_{p_n}, \theta \middle| L_2\right) + Q_{p_N}\left(X_{-q_\ell}, Z_{q_\ell}, Z_{p_n}, \theta \middle| L_2\right)\right]$$
$$+ \xi_{p_N}\left(Z_{p_n}, \theta \middle| L_2\right), \tag{A-29}$$

where $\max_{n=1,\ldots,N}\left|\xi_{p_N}\left(Z_{p_n}, \theta \middle| L_2\right)\right| = o_p(N^{-1/2})$. The result in $(A-29)$ serves the same purpose as Equation $(A-4)$. From here, we can establish Proposition A-5 for L_2 by taking the same steps that followed Equation $(A-4)$. If we retrace the steps of the proof of Proposition A-5, we can see that the key facts were the following: $\xi_{p_N}\left(Z_p, \theta \middle| L_1\right)$ satisfied the following conditions: (i) $\max_{n=1\ldots,N}\left|E\left[h_1^{-d_p} S_{p_N}\left(X_{-p_\ell}, Z_{p_n}, \theta \middle| L_1\right)\middle| Z_{p_n}\right]\right| \longrightarrow 0$; (ii) $\max_{n=1,\ldots,N}\left|\xi_{p_N}\left(Z_{p_n}, \theta \middle| L_1\right)\right| = o_p(N^{-1/2})$ (see footnote 7); (iii) $h_1^{d_p}\xi_{p_N}\left(Z_{p_n}, \theta \middle| L_1\right)$ is bounded by some constant w.p.1 and (iv) $h_1^{d_p}N^{1/2} \longrightarrow \infty$. The equivalent results hold for Equation $(A-29)$: (ii) has already been established; (iv) follows from the bandwidth assumptions; (i) the same reasons that yielded $\max_{n=1\ldots,N}\left|E\left[h_1^{-d_p} S_{p_N}\left(X_{-p_\ell}, Z_{p_n}, \theta \middle| L_1\right)\middle| Z_{p_n}\right]\right| \longrightarrow 0$; (our smoothness assumptions, trimming and the bias-reducing nature of the kernels used) yield $\max_{n=1\ldots,N}\left|E\left[h_2^{-d_p} S_{p_N}\left(X_{-p_\ell}, Z_{p_n}, \theta \middle| L_2\right)\middle| Z_{p_n}\right]\right| \longrightarrow 0$. In addition, we showed above that $E\left[Q_{p_N}(X_{-q_\ell}, Z_{q_\ell}, Z_{p_n}, \theta \middle| L_2)\middle| Z_{p_n}\right] = 0$. Finally, to see why (iii) follows, note first that the bounded nature of probabilities and our trimming imply that $h_2^{d_p}\widehat{H}_{p_N}\left(Z_{p_n}, \theta \middle| L_2\right)$ is bounded w.p.1. Trivially, $h_2^{d_p}H_p\left(Z_{p_n}, \theta \middle| L_2\right)$ is also bounded w.p.1 since $H_p(\cdot)$ is some linear combination of probabilities; for the same reason (involving a linear combination of probabilities) $S_{p_N}\left(X_{-p_\ell}, Z_{p_n}, \theta \middle| L_2\right)$ is also bounded w.p.1. It follows immediately that $h_2^{d_p}\xi_{p_N}\left(Z_{p_n}, \theta \middle| L_2\right)$ is bounded w.p.1. All the bounds and results established after Equation $(A-4)$ can be established for $k=2$ following the same steps. From here, Proposition A-5 holds for $k=2$. Given the conditions in the statement of Proposition A-7, this is enough to prove its validity using the same arguments as in the proof of Proposition A-6. □

As before, we let $\overline{U}_k^* = \min\{\overline{M}_k^*, \overline{D}_k^*\}$ and $\underline{U}_k^* = \min\{\underline{M}_k^*, \underline{D}_k^*\}$. Proposition A-7 implies that with probability approaching one uniformly in $n = 1, \ldots, N$, we can express

$$
\begin{aligned}
&\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta \big| L_3\big) - \overline{\mu}_p\big(Z_{p_n}, \theta \big| L_3\big) = \\
&\sum_{j=1}^{p^*}\Bigg\{\phi^a_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\overline{U}_3^*)}, \overline{U}_3^*\right)\left(\widehat{\overline{\mathbb{F}}}{}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big) - \overline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big)\right) \\
&\qquad\qquad + \phi^b_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\overline{U}_3^*)}, \overline{U}_3^*\right)\left(\widehat{\underline{\mathbb{F}}}{}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big) - \underline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big)\right)\Bigg\}, \\
&\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta \big| L_3\big) - \underline{\mu}_p\big(Z_{p_n}, \theta \big| L_3\big) = \\
&\sum_{j=1}^{p^*}\Bigg\{\phi^c_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\underline{U}_3^*)}, \underline{U}_3^*\right)\left(\widehat{\overline{\mathbb{F}}}{}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big) - \overline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big)\right) \\
&\qquad\qquad + \phi^d_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\underline{U}_3^*)}, \underline{U}_3^*\right)\left(\widehat{\underline{\mathbb{F}}}{}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big) - \underline{\mathbb{F}}^{(j)}_{-p}\big(Z_p, \theta \big| L_2\big)\right)\Bigg\}.
\end{aligned}
\tag{A-30}
$$

As before, $\phi^a_{p_j}(\cdot)$, $\phi^b_{p_j}(\cdot)$, $\phi^c_{p_j}(\cdot)$ and $\phi^d_{p_j}(\cdot)$ are described in Equation (15). Using the notation from Equation $(A-21)$, let

$$
\begin{aligned}
&\overline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_2\big) = \\
&\qquad \sum_{j=1}^{p^*}\Bigg\{\phi^a_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\overline{U}_3^*)}, \overline{U}_3^*\right)\left(\overline{G}^{(j)}_{-p}\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_2\Big) - \overline{\mathbb{F}}^{(j)}_{-p}\big(Z_{p_n}, \theta \big| L_2\big)\right) \\
&\qquad\qquad + \phi^b_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\overline{U}_3^*)}, \overline{U}_3^*\right)\left(\underline{G}^{(j)}_{-p}\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_2\Big) - \underline{\mathbb{F}}^{(j)}_{-p}\big(Z_{p_n}, \theta \big| L_2\big)\right)\Bigg\}, \\
&\underline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_2\big) = \\
&\qquad \sum_{j=1}^{p^*}\Bigg\{\phi^c_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\underline{U}_3^*)}, \underline{U}_3^*\right)\left(\overline{G}^{(j)}_{-p}\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_2\Big) - \overline{\mathbb{F}}^{(j)}_{-p}\big(Z_{p_n}, \theta \big| L_2\big)\right) \\
&\qquad\qquad + \phi^d_{p_j}\left(\Delta_p^{(j)}, \Delta_p^{(\underline{U}_3^*)}, \underline{U}_3^*\right)\left(\underline{G}^{(j)}_{-p}\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_2\Big) - \underline{\mathbb{F}}^{(j)}_{-p}\big(Z_{p_n}, \theta \big| L_2\big)\right)\Bigg\}.
\end{aligned}
\tag{A-31}
$$

Let

$$\overline{\lambda}_p\big(X_m, Z_m, Z_{p_n},\theta\big|L_2\big) = \sum_{j=1}^{p^*}\Bigg\{\phi^a_{p_j}\left(\Delta^{(j)}_p, \Delta^{(\overline{U}^*_3)}_p, \overline{U}^*_3\right)\sum_{q\neq p}\overline{D}^{(j)}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta\big|L_2\big)'\varphi_q\big(X_{-q_m}, Z_{-q_m}, Z_{q_m}, \theta\big|L_1\big)$$
$$+ \phi^b_{p_j}\left(\Delta^{(j)}_p, \Delta^{(\overline{U}^*_3)}_p, \overline{U}^*_3\right)\sum_{q\neq p}\underline{D}^{(j)}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta\big|L_2\big)'\varphi_q\big(X_{-q_m}, Z_{-q_m}, Z_{q_m}, \theta\big|L_1\big)\Bigg\},$$

$$\underline{\lambda}_p\big(X_m, Z_m, Z_{p_n},\theta\big|L_2\big) = \sum_{j=1}^{p^*}\Bigg\{\phi^c_{p_j}\left(\Delta^{(j)}_p, \Delta^{(\underline{U}^*_3)}_p, \underline{U}^*_3\right)\sum_{q\neq p}\overline{D}^{(j)}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta\big|L_2\big)'\varphi_q\big(X_{-q_m}, Z_{-q_m}, Z_{q_m}, \theta\big|L_1\big)$$
$$+ \phi^d_{p_j}\left(\Delta^{(j)}_p, \Delta^{(\underline{U}^*_3)}_p, \underline{U}^*_3\right)\sum_{q\neq p}\underline{D}^{(j)}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta\big|L_2\big)'\varphi_q\big(X_{-q_m}, Z_{-q_m}, Z_{q_m}, \theta\big|L_1\big)\Bigg\}. \tag{A-32}$$

Given our previous results, the following representation holds

$$\widehat{\overline{\mu}}_p\big(Z_{p_\ell},\theta\big|L_3\big) - \overline{\mu}_p\big(Z_{p_\ell},\theta\big|L_3\big)$$
$$= \frac{1}{f_{Z_p}(Z_{p_\ell})}\frac{1}{N}\sum_{m=1}^{N}\left\{\frac{1}{h_2^{d_p}}\overline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_\ell},\theta\big|L_2\big)K^p_{h_2}\big(Z_{p_m} - Z_{p_\ell}\big) + \overline{\lambda}_p\big(X_m, Z_m, Z_{p_\ell},\theta\big|L_2\big)\right\} + \overline{\xi}_{p_N}(Z_{p_\ell},\theta|L_3),$$

$$\widehat{\underline{\mu}}_p\big(Z_{p_\ell},\theta\big|L_3\big) - \underline{\mu}_p\big(Z_{p_\ell},\theta\big|L_3\big)$$
$$= \frac{1}{f_{Z_p}(Z_{p_\ell})}\frac{1}{N}\sum_{m=1}^{N}\left\{\frac{1}{h_2^{d_p}}\underline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_\ell},\theta\big|L_2\big)K^p_{h_2}\big(Z_{p_m} - Z_{p_\ell}\big) + \underline{\lambda}_p\big(X_m, Z_m, Z_{p_\ell},\theta\big|L_2\big)\right\} + \underline{\xi}_{p_N}(Z_{p_\ell},\theta|L_3), \tag{A-33}$$

where $\max_{\ell=1,\ldots,N}\left|\overline{\xi}_{p_N}(Z_{p_\ell},\theta|L_3)\right| = o_p(N^{-1/2})$ and $\max_{\ell=1,\ldots,N}\left|\underline{\xi}_{p_N}(Z_{p_\ell},\theta|L_3)\right| = o_p(N^{-1/2})$. A second order approximation yields

$$\begin{aligned}
&\widehat{\overline{\mathbb{F}}}_p^M(Z_{p_n},\theta|L_3)\\
&= \frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_3^{d_p}}\sum_{\ell=1}^N \overline{G}_{-p}^M\Big((X'_{q_\ell}\beta_q)_{q\neq p}\ ,\widehat{\overline{\mathbf{\Lambda}}}_{-p}(Z_{-p_\ell},\theta|L_3)\ ,\underline{\widehat{\mathbf{\Lambda}}}_{-p}(Z_{-p_\ell},\theta|L_3)\ ;\ Z_{p_n}\Big)K_{h_3}^p(Z_{p_\ell}-Z_{p_n})\\
&= \frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_3^{d_p}}\sum_{\ell=1}^N \overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)K_{h_2}^p(Z_{p_\ell}-Z_{p_n})\\
&\overbrace{+\frac{1}{f_{Z_p}(Z_{p_n})N^2h_3^{d_p}}\sum_{m=1}^N\sum_{\ell=1}^N\sum_{q\neq p}\frac{1}{f_{Z_q}(Z_{q_\ell})}\Bigg\{\frac{1}{h_2^{d_q}}\nabla_{\overline{\mu}_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)\overline{\varphi}_q(X_{-q_m},Z_{-q_m},Z_{q_\ell},\theta|L_2)K_{h_2}^q(Z_{q_m}-Z_{q_\ell})}^{A}\\
&+\frac{1}{h_2^{d_q}}\nabla_{\underline{\mu}_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)\underline{\varphi}_q(X_{-q_m},Z_{-q_m},Z_{q_\ell},\theta|L_2)K_{h_2}^q(Z_{q_m}-Z_{q_\ell})\Bigg\}K_{h_3}^p(Z_{p_\ell}-Z_{p_n})\\
&\overbrace{+\frac{1}{f_{Z_p}(Z_{p_n})N^2h_3^{d_p}}\sum_{m=1}^N\sum_{\ell=1}^N\sum_{q\neq p}\frac{1}{f_{Z_q}(Z_{q_\ell})}\Bigg\{\nabla_{\overline{\mu}_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)\overline{\lambda}_q(X_m,Z_m,Z_{q_\ell},\theta|L_2)}^{B}\\
&+\nabla_{\underline{\mu}_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)\underline{\lambda}_q(X_m,Z_m,Z_{q_\ell},\theta|L_2)\Bigg\}K_{h_3}^p(Z_{p_\ell}-Z_{p_n})\\
&+\overline{\zeta}_{p_N}^M(Z_{p_n},\theta|L_3),
\end{aligned} \tag{A-34}$$

where $\max_{\ell=1,\ldots,N}\left|\overline{\zeta}_{p_N}(Z_{p_\ell},\theta|L_3)\right| = o_p(N^{-1/2})$. The term (A) is equivalent to the second term on the right-hand side of Equation $(A-22)$. Conditional on Z_n, we can the results in Ahn and Powell (1993) or Sherman (1994) and find that its projection is given by

$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{m=1}^N\sum_{q\neq p}\overline{D}_{-p,\mu_q}^M(Z_{q_m},Z_{p_n},\theta|L_3)'\varphi_q(X_{-q_m},Z_{-q_m},Z_{q_m},\theta|L_2)+\overline{\vartheta}_{p_N}(Z_{p_n},\theta|L_3),$$

where $\max_{n=1,\ldots,N}\left|\overline{\vartheta}_{p_N}(Z_{p_n},\theta|L_3)\right| = o_p(N^{-1/2})$, $\overline{D}_{-p,\mu_q}^M(Z_{q_m},Z_{p_n},\theta|L_k)$ is as defined in Equations $(A-23)$ and $(A-24)$ and

$$\varphi_q(X_{-q_m},Z_{-q_m},Z_{q_m},\theta|L_2) = \Big(\underline{\varphi}_q(X_{-q_m},Z_{-q_m},Z_{q_m},\theta|L_2)\ ,\overline{\varphi}_q(X_{-q_m},Z_{-q_m},Z_{q_m},\theta|L_2)\Big)',$$

where the last two are given in Equation $(A-31)$. Let us stack

$$\begin{aligned}
\nabla_{\mu_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3) &= \Big(\nabla_{\overline{\mu}_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)\ ,\nabla_{\underline{\mu}_q}\overline{G}_{-p}^M(X_{-p_\ell},Z_{-p_\ell},Z_{p_n},\theta|L_3)\Big)',\\
\lambda_q(X_m,Z_m,Z_{q_\ell},\theta|L_2) &= \Big(\overline{\lambda}_q(X_m,Z_m,Z_{q_\ell},\theta|L_2)\ ,\underline{\lambda}_q(X_m,Z_m,Z_{q_\ell},\theta|L_2)\Big)'.
\end{aligned}$$

Now let

$$\overline{J}_q^M(X_m, Z_m, Z_{p_n}, \theta|L_3) =$$

$$E\Bigg[\Bigg(E\Big[\nabla_{\mu_q}\overline{G}_{-p}^M(X_{-p}, Z_{-p}, Z_{p_n}, \theta|L_3)\Big|Z_p, Z_q, Z_{p_n}\Big]\Bigg)'\Bigg(\frac{\lambda_q(X_m, Z_m, Z_q, \theta|L_2)}{f_{Z_q}(Z_q)}\Bigg)\ \Bigg|\ Z_p = Z_{p_n}, X_m, Z_m\Bigg],$$

$$\overline{J}_{-p}^M(X_m, Z_m, Z_{p_n}, \theta|L_3) = \sum_{q\neq p}\overline{J}_q^M(X_m, Z_m, Z_{p_n}, Z_{p_n}, \theta|L_3).$$

The projection (conditional on Z_{p_n}) of the second order U-statistic in (B) is

$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{m=1}^{N}\overline{J}_{-p}^M(X_m, Z_m, Z_{p_n}, \theta|L_3) + \overline{\tau}_{p_N}(Z_{p_n}, \theta|L_3),$$

where $\max_{n=1,\ldots,N}\Big|\overline{\tau}_{p_N}(Z_{p_n}, \theta|L_3)\Big| = o_p(N^{-1/2})$. If we go back to the characterizations of the objects involved, it follows immediately that $E\Big[\overline{J}_{-p}^M(X_m, Z_m, Z_{p_n}, \theta|L_3)\Big|Z_{p_n}\Big] = 0$. As before, using Lemma 3 in Collomb and Hardle (1986) or Theorem A-1 in Aradillas-Lopez (2006), we get

$$\frac{1}{\widehat{f}_{Z_p}(Z_{p_n})}\frac{1}{Nh_3^{d_p}}\sum_{m=1}^{N}\overline{G}_{-p}^M(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta|L_3)K_{h_3}^p(Z_{p_m} - Z_{p_n}) - \overline{\mathbb{F}}_{-p}^M(Z_{p_n}, \theta|L_3) =$$

$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{Nh_3^{d_p}}\sum_{m=1}^{N}\Big[\overline{G}_{-p}^M(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta|L_3) - \overline{\mathbb{F}}_{-p}^M(Z_{p_n}, \theta|L_3)\Big]K_{h_3}^p(Z_{p_m} - Z_{p_n}) + \overline{\varsigma}_{p_N}(Z_{p_n}, \theta|L_3),$$

where $\max_{n=1,\ldots,N}\Big|\overline{\varsigma}_{p_N}(Z_{p_n}, \theta|L_3)\Big| = o_p(N^{-1/2})$. The previous results yield the following linear representation for $\widehat{\overline{\mathbb{F}}}_{-p}^M(Z_{p_n}, \theta|L_3) - \overline{\mathbb{F}}_{-p}^M(Z_{p_n}, \theta|L_3)$,

$$\widehat{\overline{\mathbb{F}}}_{-p}^M(Z_{p_n}, \theta|L_3) - \overline{\mathbb{F}}_{-p}^M(Z_{p_n}, \theta|L_3) =$$

$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{\ell=1}^{N}\Bigg[\frac{1}{h_3^{d_p}}\overline{T}_{-p_N}^M(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta|L_3) + \sum_{q\neq p}\overline{D}_{-p,\mu_q}^M(Z_{q_\ell}, Z_{p_n}, \theta|L_3)'\varphi_q(X_{-q_\ell}, Z_{-q_\ell}, Z_{q_\ell}, \theta|L_2)$$

$$+ \overline{J}_{-p}^M(X_\ell, Z_\ell, Z_{p_n}, \theta|L_3)\Bigg] + \overline{\nu}_{p_N}^M(Z_{p_n}, \theta|L_3),$$

where $\max_{n=1,\ldots,N}\Big|\overline{\nu}_{p_N}^M(Z_{p_n}, \theta|L_2)\Big| = o_p(N^{-1/2})$ and

$$\overline{T}_{-p_N}^M(X_{-p_\ell}, Z_{-p_\ell}, Z_p, \theta|L_3) = \left(\frac{1}{f_{Z_p}(Z_p)}\right)\Big[\overline{G}_{-p}^M\Big(X_{-p_\ell}, Z_{-p_\ell}, Z_p, \theta\Big|L_3\Big) - \overline{\mathbb{F}}_{-p}^M(Z_p, \theta|L_3)\Big]K_{h_3}^p(Z_{p_\ell} - Z_p).$$

As before, we can derive the equivalent result for $\widehat{\underline{\mathbb{F}}}_{-p}^M(Z_{p_n}, \theta|L_3) - \underline{\mathbb{F}}_{-p}^M(Z_{p_n}, \theta|L_3)$ as

$$\widehat{\underline{\mathbb{F}}}_{-p}^M(Z_{p_n}, \theta|L_3) - \underline{\mathbb{F}}_{-p}^M(Z_{p_n}, \theta|L_3) =$$

$$\frac{1}{f_{Z_p}(Z_{p_n})}\frac{1}{N}\sum_{\ell=1}^{N}\Bigg[\frac{1}{h_3^{d_p}}\underline{T}_{-p_N}^M(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta|L_3) + \sum_{q\neq p}\underline{D}_{-p,\mu_q}^M(Z_{q_\ell}, Z_{p_n}, \theta|L_3)'\varphi_q(X_{-q_\ell}, Z_{-q_\ell}, Z_{q_\ell}, \theta|L_2)$$

$$+ \underline{J}_{-p}^M(X_\ell, Z_\ell, Z_{p_n}, \theta|L_3)\Bigg] + \underline{\nu}_{p_N}^M(Z_{p_n}, \theta|L_3),$$

where $\underline{D}^M_{-p,\mu_q}\big(Z_{q_\ell}, Z_{p_n}, \theta \big| L_3\big)$ and $\underline{J}^M_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_3\big)$ are defined by replacing $\overline{G}^M_{-p}(\cdot)$ with $\underline{G}^M_{-p}(\cdot)$ in the definitions above, and

$$\underline{T}^M_{-p_N}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_p, \theta \big| L_3\big) = \left(\frac{1}{f_{Z_p}(Z_p)}\right)\left[\underline{G}^M_{-p}\Big(X_{-p_\ell}, Z_{-p_\ell}, Z_p, \theta \Big| L_3\Big) - \underline{\mathbb{F}}^M_{-p}\big(Z_p, \theta \big| L_3\big)\right] K^p_{h_3}\big(Z_{p_\ell} - Z_p\big).$$

If the conditions in Propositions A-6 and A-7 are satisfied and if the condition in Equation $(A-25)$ extends to L_3, namely

$$\Pr\left[\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}^M_{-p}\big(Z_p, \theta \big| L_3\big) = 1 \middle| Z_p \in \mathcal{Z}_p\right] = 0.$$

for all $\theta \in \Theta$ such that $\Delta^M_q \neq 0$ for some $q \neq p$ and some M, then we can follow the same steps as above to show that Proposition A-6 holds for $k = 4$, Namely

$$\Pr\Big(R^*_{k=4}(Z_{p_n}, \theta) \neq \widehat{R}^*_{k=4}(Z_{p_n}, \theta) \text{ for some } n = 1, \ldots, N\Big) \longrightarrow 0,$$

where $R^*_k(Z_p, \theta)$ is as defined in Proposition A-6. It follows from here that $\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta \big| L_4\big) - \overline{\mu}_p\big(Z_{p_n}, \theta \big| L_4\big)$ and $\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta \big| L_4\big) - \underline{\mu}_p\big(Z_{p_n}, \theta \big| L_4\big)$ can be expressed as in Equation $(A-30)$. Given this fact, their linear representations follow immediately from those of $\widehat{\underline{\mathbb{F}}}^M_{-p}\big(Z_{p_n}, \theta \big| L_3\big) - \underline{\mathbb{F}}^M_{-p}\big(Z_{p_n}, \theta \big| L_3\big)$ and $\widehat{\overline{\mathbb{F}}}^M_{-p}\big(Z_{p_n}, \theta \big| L_3\big) - \overline{\mathbb{F}}^M_{-p}\big(Z_{p_n}, \theta \big| L_3\big)$ as in the cases $k = 3, 2$ which we already established. If we continue this process iteratively for $k > 4$, a general result emerges. We present it next.

A.2.1 A general representation result

We start by denoting the coefficients described in Equation (15) compactly as $\phi^a_{p_j}\big(\Delta_p, \overline{U}^*_k\big)$, $\phi^b_{p_j}\big(\Delta_p, \overline{U}^*_k\big)$, $\phi^c_{p_j}\big(\Delta_p, \underline{U}^*_k\big)$ and $\phi^d_{p_j}\big(\Delta_p, \underline{U}^*_k\big)$. For $k \geq 1$ let

$$\begin{aligned}
\overline{R}^M_p\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big) &= \overline{G}^M_{-p}\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big) - \overline{\mathbb{F}}^M_{-p}\big(Z_{p_n}, \theta \big| L_k\big), \\
\underline{R}^M_p\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big) &= \underline{G}^M_{-p}\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big) - \underline{\mathbb{F}}^M_{-p}\big(Z_{p_n}, \theta \big| L_k\big).
\end{aligned} \tag{A-35}$$

For $k \geq 1$, let

$$\begin{aligned}
&\overline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_k\big) = \\
&\sum_{j=1}^{p^*}\left\{\phi^a_{p_j}(\Delta_p, \overline{U}^*_{k+1})\overline{R}^{(j)}_p\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big) + \phi^b_{p_j}(\Delta_p, \overline{U}^*_{k+1})\underline{R}^{(j)}_p\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big)\right\}, \\
&\underline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_k\big) = \\
&\sum_{j=1}^{p^*}\left\{\phi^c_{p_j}(\Delta_p, \underline{U}^*_{k+1})\overline{R}^{(j)}_p\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big) + \phi^d_{p_j}(\Delta_p, \underline{U}^*_{k+1})\underline{R}^{(j)}_p\Big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \Big| L_k\Big)\right\}, \\
&\varphi_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_k\big) = \Big(\overline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_k\big)\ ,\ \underline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_k\big)\Big)'.
\end{aligned} \tag{A-36}$$

Now let $\underline{D}^{(j)}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta \big| L_k\big)$ be as described in Equation $(A-23)$, and let $\underline{D}^{M}_{-p,\mu_q}(\cdot)$ be the objects that result from $\overline{G}^{M}_{-p}(\cdot)$ with $\underline{G}^{M}_{-p}(\cdot)$ there. For $k \geq 2$ we will let

$$\underline{E}^{M}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) = \sum_{q \neq p} \underline{D}^{M}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta \big| L_k\big)' \varphi_q\big(X_{-q_m}, Z_{-q_m}, Z_{q_m}, \theta \big| L_{k-1}\big),$$

$$\overline{E}^{M}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) = \sum_{q \neq p} \overline{D}^{M}_{-p,\mu_q}\big(Z_{q_m}, Z_{p_n}, \theta \big| L_k\big)' \varphi_q\big(X_{-q_m}, Z_{-q_m}, Z_{q_m}, \theta \big| L_{k-1}\big).$$

For $k = 1$, let the above objects be identically equal to zero. For $k \geq 2$, define

$$\begin{aligned}
\overline{\lambda}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) &= \sum_{j=1}^{p^*} \Big\{ \phi^a_{p_j}\big(\Delta_p, \overline{U}^*_{k+1}\big) \overline{E}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) + \phi^b_{p_j}\big(\Delta_p, \overline{U}^*_{k+1}\big) \underline{E}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) \Big\}, \\
\underline{\lambda}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) &= \sum_{j=1}^{p^*} \Big\{ \phi^c_{p_j}\big(\Delta_p, \underline{U}^*_{k+1}\big) \overline{E}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) + \phi^d_{p_j}\big(\Delta_p, \underline{U}^*_{k+1}\big) \underline{E}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) \Big\}, \\
\lambda_q\big(X_m, Z_m, Z_{q_\ell}, \theta \big| L_k\big) &= \Big(\overline{\lambda}_q\big(X_m, Z_m, Z_{q_\ell}, \theta \big| L_k\big) \ , \ \underline{\lambda}_q\big(X_m, Z_m, Z_{q_\ell}, \theta \big| L_k\big) \Big)'.
\end{aligned} \tag{A-37}$$

For $k = 1$, let the above objects be identically equal to zero. As before, for $k \geq 1$ we denote

$$\nabla_{\mu_q} \overline{G}^{M}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big) = \Big(\nabla_{\overline{\mu}_q} \overline{G}^{M}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big) \ , \ \nabla_{\underline{\mu}_q} \overline{G}^{M}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big) \Big)'.$$

For $k \geq 3$, define

$$\begin{aligned}
&\overline{J}^{M}_{q}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) = \\
&E\Bigg[\bigg(E\Big[\nabla_{\mu_q} \overline{G}^{M}_{-p}\big(X_{-p}, Z_{-p}, Z_{p_n}, \theta \big| L_k\big) \Big| Z_p, Z_q, Z_{p_n} \Big] \bigg)' \bigg(\frac{\lambda_q\big(X_m, Z_m, Z_q, \theta \big| L_{k-1}\big)}{f_{Z_q}(Z_q)} \bigg) \ \Bigg| \ Z_p = Z_{p_n}, X_m, Z_m \Bigg], \\
&\underline{J}^{M}_{q}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) = \\
&E\Bigg[\bigg(E\Big[\nabla_{\mu_q} \underline{G}^{M}_{-p}\big(X_{-p}, Z_{-p}, Z_{p_n}, \theta \big| L_k\big) \Big| Z_p, Z_q, Z_{p_n} \Big] \bigg)' \bigg(\frac{\lambda_q\big(X_m, Z_m, Z_q, \theta \big| L_{k-1}\big)}{f_{Z_q}(Z_q)} \bigg) \ \Bigg| \ Z_p = Z_{p_n}, X_m, Z_m \Bigg], \\
&\overline{J}^{M}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) = \sum_{q \neq p} \overline{J}^{M}_{q}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big), \\
&\underline{J}^{M}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) = \sum_{q \neq p} \underline{J}^{M}_{q}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big).
\end{aligned}$$

For $k \leq 2$, let the above objects be identically equal to zero. For $k \geq 3$ define

$$\begin{aligned}
\overline{\delta}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) &= \sum_{j=1}^{p^*} \Big\{ \phi^a_{p_j}\big(\Delta_p, \overline{U}^*_{k+1}\big) \overline{J}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) + \phi^b_{p_j}\big(\Delta_p, \overline{U}^*_{k+1}\big) \underline{J}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) \Big\}, \\
\underline{\delta}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) &= \sum_{j=1}^{p^*} \Big\{ \phi^c_{p_j}\big(\Delta_p, \underline{U}^*_{k+1}\big) \overline{J}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) + \phi^d_{p_j}\big(\Delta_p, \underline{U}^*_{k+1}\big) \underline{J}^{(j)}_{-p}\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) \Big\}, \\
\delta_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) &= \Big(\overline{\delta}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) \ , \ \underline{\delta}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) \Big)'.
\end{aligned} \tag{A-38}$$

For $k \leq 2$, let the above objects be identically equal to zero. We will define the following projections iteratively. For $k \geq 4$, let

$$\overline{\Gamma}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) =$$

$$\begin{aligned}
&\sum_{q \neq p} E\left[\frac{\nabla_{\mu_q}\overline{G}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big)'}{h_k^{d_p} f_{Z_q}(Z_{q_\ell})} \delta_q\Big(X_m, Z_m, Z_{q_\ell}, \theta \Big| L_{k-1}\Big) K_{h_k}\big(Z_{p_\ell} - Z_{p_n}\big) \,\Bigg|\, X_m, Z_m, Z_{p_n}\right] \text{ if } k = 4\\
&\sum_{q \neq p} E\left[\frac{\nabla_{\mu_q}\overline{G}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big)'}{h_k^{d_p} f_{Z_q}(Z_{q_\ell})} \Big\{\delta_q\Big(X_m, Z_m, Z_{q_\ell}, \theta \Big| L_{k-1}\Big)\right.\\
&\qquad\qquad \left. + \overline{\Gamma}_q\big(X_m, Z_m, Z_{q_\ell}, \theta \big| L_{k-1}\big)\Big\} K_{h_k}\big(Z_{p_\ell} - Z_{p_n}\big) \,\Bigg|\, X_m, Z_m, Z_{p_n}\right] \text{ if } k \geq 5,
\end{aligned}$$

and

$$\underline{\Gamma}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_k\big) =$$

$$\begin{aligned}
&\sum_{q \neq p} E\left[\frac{\nabla_{\mu_q}\underline{G}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big)'}{h_k^{d_p} f_{Z_q}(Z_{q_\ell})} \delta_q\Big(X_m, Z_m, Z_{q_\ell}, \theta \Big| L_{k-1}\Big) K_{h_k}\big(Z_{p_\ell} - Z_{p_n}\big) \,\Bigg|\, X_m, Z_m, Z_{p_n}\right] \text{ if } k = 4\\
&\sum_{q \neq p} E\left[\frac{\nabla_{\mu_q}\underline{G}_{-p}\big(X_{-p_\ell}, Z_{-p_\ell}, Z_{p_n}, \theta \big| L_k\big)'}{h_k^{d_p} f_{Z_q}(Z_{q_\ell})} \Big\{\delta_q\Big(X_m, Z_m, Z_{q_\ell}, \theta \Big| L_{k-1}\Big)\right.\\
&\qquad\qquad \left. + \underline{\Gamma}_q\big(X_m, Z_m, Z_{q_\ell}, \theta \big| L_{k-1}\big)\Big\} K_{h_k}\big(Z_{p_\ell} - Z_{p_n}\big) \,\Bigg|\, X_m, Z_m, Z_{p_n}\right] \text{ if } k \geq 5.
\end{aligned} \tag{A-39}$$

For $k \leq 3$, let the above objects be identically equal to zero. The following proposition summarizes the linear representation properties of our L_k-rational expectation bounds.

Proposition A-8 *Let* $\mathbb{H}^M_{-p}\big(Z_p, \theta \big| L_k\big)$ *denote either* $\underline{\mathbb{F}}^M_{-p}\big(Z_p, \theta \big| L_k\big)$ *or* $\overline{\mathbb{F}}^M_{-p}\big(Z_p, \theta \big| L_k\big)$. *Take any linear combination of the form* $\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}^M_{-p}\big(Z_p, \theta \big| L_k\big)$ *and suppose that for every* $\theta \in \Theta$ *such that* $\Delta^M_q \neq 0$ *for some* $q \neq p$ *and some* M *the following holds for any such linear combination:*

$$Pr\left[\sum_{M=0}^{\mathcal{P}-1} \mathbb{H}^M_{-p}\big(Z_p, \theta \big| L_k\big) = 1 \,\middle|\, Z_p \in \mathcal{Z}_p\right] = 0 \quad \textit{for} \;\; k = 1, \ldots, k^* - 1. \tag{A-40}$$

Then, given our assumptions the following linear representations hold for $k = 2, \ldots, k^*$ and each such parameter value θ,

$$
\begin{aligned}
&\widehat{\overline{\mu}}_p\big(Z_{p_n}, \theta \big| L_k\big) - \overline{\mu}_p\big(Z_{p_n}, \theta \big| L_k\big) \\
&= \frac{1}{f_{Z_p}(Z_{p_n})} \frac{1}{N} \sum_{m=1}^{N} \Bigg\{ \frac{1}{h_{k-1}^{d_p}} \overline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_{k-1}\big) K^p_{h_{k-1}}\big(Z_{p_m} - Z_{p_n}\big) + \overline{\lambda}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_{k-1}\big) \\
&\qquad + \overline{\delta}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_{k-1}\big) + \overline{\Gamma}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_{k-1}\big) \Bigg\} + \overline{\xi}_{p_N}(Z_{p_n}, \theta | L_k), \\
&\widehat{\underline{\mu}}_p\big(Z_{p_n}, \theta \big| L_k\big) - \underline{\mu}_p\big(Z_{p_n}, \theta \big| L_k\big) \\
&= \frac{1}{f_{Z_p}(Z_{p_n})} \frac{1}{N} \sum_{m=1}^{N} \Bigg\{ \frac{1}{h_{k-1}^{d_p}} \underline{\varphi}_p\big(X_{-p_m}, Z_{-p_m}, Z_{p_n}, \theta \big| L_{k-1}\big) K^p_{h_{k-1}}\big(Z_{p_m} - Z_{p_n}\big) + \underline{\lambda}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_{k-1}\big) \\
&\qquad + \underline{\delta}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_{k-1}\big) + \underline{\Gamma}_p\big(X_m, Z_m, Z_{p_n}, \theta \big| L_{k-1}\big) \Bigg\} + \underline{\xi}_{p_N}(Z_{p_n}, \theta | L_k).
\end{aligned}
\tag{A-41}
$$

The proof of this result consists of an iterative repetition of the arguments we described above. Given our smoothness, kernels and bandwidth assumptions, taking the corresponding projections of the successive U-statistics we obtain the general form for the expression in $(A - 41)$.

References

Aguirregabiria, V., and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–53.

Ahn, H., and J. Powell (1993): "Semiparametric Estimation of Censored Selection Models," *Journal of Econometrics*, 58, 3–29.

Andrews, D., S. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Working Paper, Yale University.

Aradillas-Lopez, A. (2006): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Working Paper, Princeton University.

Aradillas-Lopez, A., B. Honoré, and J. Powell (2007): "Pairwise Difference Estimation with Nonparametric Control Variables," *Forthcoming, International Economic Review.*

Aradillas-Lopez, A., and E. Tamer (2007): "The Identification Power of Equilibrium in Games," *Forthcoming, Journal of Business and Economic Statistics.*

Bajari, P., H. Hong, J. Kreiner, and D. Nekipelov (2006): "Estimating Static Models of Strategic Interaction," Working Paper, Duke University.

Bajari, P., H. Hong, and S. Ryan (2005): "Identification and Estimation of Discrete Games of Complete Information," Working Paper.

Bernheim, D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52(4), 1007–1028.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2), 242–262.

Berry, S., and E. Tamer (2006): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics. Proceedings of the Ninth world Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson., pp. 46–85. Cambridge University Press.

Bjorn, P. A., and Q. Vuong (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," SSWP No. 537, Caltech.

Bresnahan, T., and P. Reiss (1991): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99, 977–1009.

Ciliberto, F., and E. Tamer (2003): "Market Structure and Multiple Equilibria in Airline Markets," Working Paper.

Clarke, F. H. (1990): *Optimization and Nonstmooth Analysis*. SIAM Classics in Applied Mathematics.

Collomb, G., and W. Hardle (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.

Costa-Gomes, M., and V. Crawford (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96(5), 1737–1745.

Costa-Gomes, M., V. Crawford, and B. Broseta (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69(5), 1193–1235.

Crawford, V., and N. Iriberri (2007): "Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Forthcoming, Econometrica.*

Ho, T., C. Camerer, and K. Weigelt (1998): "Iterated Dominance and Iterated Best Response in Experimental 'p-Beauty Contests'," *American Economic Review*, 88(4), 947–969.

Ichimura, H. (2004): "Computation Of Asymptotic Distribution Semiparametric GMM Estimators," Working Paper. University College, London.

Ichimura, H., and L.-F. Lee (1991): "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," p. 3. Cambridge University Press.

Imbens, G., and W. Newey (2002): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper. Harvard University and MIT.

McKelvey, R., and T. Palfrey (1995): "Quantal Response Equilibria for Normal-Form Games," *Games and Economic Behavior*, 10, 6–38.

Nagel, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85(5), 1313–1326.

Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden. North Holland.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2005): "Moment Inequalities and Their Applications," Working Paper.

Seim, K. (2005): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *Forthcoming, RAND Journal of Economics.*

Sherman, R. (1994): "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator," *Econometric Theory*, 10, 372–395.

Stahl, D., and P. Wilson (1994): "Experimental Evidence on Players' Models of Other Players," *Journal of Economic Behavior and Organization*, 25(3), 309–327.

——— (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10(1), 218–254.

Tamer, E. T. (2003): "Incomplete Bivariate Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–167.