Semiparametric Estimation of a Simultaneous Game with Incomplete Information

Andres Aradillas-Lopez[1]

THIS VERSION: 12/30/05

Abstract

We study a simultaneous game with a parameterized normal-form representation that has been the focus of previous work. By relying exclusively on the assumption of Nash equilibrium we show that if players observe only a noisy signal of the payoff realization then, if the amount of noise satisfies some conditions it is possible to identify more features than in a complete information environment. In particular, the model produces a well-defined likelihood function for symmetric versions of the game only if there is incomplete information. If players condition their beliefs on observable signals whose exact distribution is unknown to the econometrician, the resulting likelihood function is only partially known. Assuming a Bayesian Nash equilibrium (BNE) is being played, we approach this problem by estimating beliefs based on sample analog BNE conditions using semiparametric methods for conditional moment estimation. The vector of payoff parameters is estimated by maximizing a trimmed log-likelihood function in which unobserved BNE beliefs are replaced by their analog estimators. The trimming set is an interior subset of the support of the signals where each BNE is unique. The resulting estimator is $\sqrt{N}$-consistent and its asymptotic variance depends partially on the degree to which the signals used by the players account for the variation of the players' private information. A simple specification test is proposed by comparing estimated equilibrium beliefs against nonparametric choice probabilities.

Keywords: Semiparametric estimation; Incomplete econometric model; Incomplete information game; Bayesian-Nash equilibria; Trimming.

JEL Classification: C13, C14, C35, C62, C73.

[1]Special thanks to Jim Powell, Guido Imbens and Bo Honore. I also thank seminar participants at Berkeley, Princeton, Caltech, Carnegie-Mellon, El Colegio de Mexico, Harvard-MIT, Nuffield College Oxford, NYU, UCSD, UCLA and UPenn for their valuable comments. Correspondence: Andres Aradillas-Lopez, Department of Economics, 312 Fisher Hall, Princeton, NJ 08544. email: aaradill@princeton.edu.

1 Introduction

In the recent past, much attention has been devoted to the problem of econometric estimation of game-theoretic models with multiple equilibria. This is a potentially difficult task because a particular realization of observables and a particular value of the parameter of interest(in a parameterized model) may be consistent with different outcomes of the model. This may be true even if agents' actions and beliefs are logically coherent and consistent with some equilibrium concept (e.g, Nash equilibrium). It has been argued (Morris and Shin (2001)) that multiple equilibria is often the result of assuming that agents have complete knowledge about the primitives of the model: economic fundamentals as well as other agents' behavior are assumed to be common knowledge. The effect of these assumptions is to generate the type of coordinating behavior that results in multiple equilibria.

In this paper we revisit the 2×2 game studied by Bresnahan and Reiss (1990, 1991) and Tamer (2003). Assuming that players have complete information and ruling out mixed-strategies, these authors showed that this game had multiple or nonexistent equilibria with positive probability unless the payoffs of (at least) one of the players were independent of the actions of his opponent. They showed that point-identification of the payoff parameters was still possible even if players' actions have a mutual effect on their opponent's payoffs and suggested different ways to obtain $\sqrt{N}$-consistent estimators. However, other important features of the game are not point-identified under these assumptions. In particular, it is not possible to point-estimate outcome probabilities for a subset of the four outcomes of the game without introducing an ad-hoc equilibrium selection mechanism. Payoff parameters are point-identified, but choice probabilities for a subset of outcomes are not.

In this paper we depart from the complete-information structure and assume an incomplete-information environment in which players observe a noisy signal of the game's realized payoff structure. We assume that players condition their beliefs on this signal and maximize their expected utility given these beliefs. The resulting Bayesian-Nash equilibrium (BNE) conditions can be expressed as the solution to a vector of conditional moment restrictions. Relying only on the assumption of a Nash equilibrium (without a theory of equilibrium selection) we show that if the amount of noise in the signals is rich enough in a well-defined sense, the game has more point-identified features if it is played with incomplete information. The model generates a well-defined likelihood function for the four outcomes of the simultaneous game under conditions that are generically weaker if players have incomplete information. In particular, the model produces a well-defined likelihood function for symmetric versions of the game only if there is incomplete information. This result holds with or without mixed strategies.

Next, we focus on the incomplete-information case and analyze the problem of estimating the payoff parameters when players' beliefs are unobserved but a BNE is assumed. In addition, the exact distribution of the signals is assumed to be known only by the players, but not the researcher. Consequently, even if payoffs are assumed to have a known parametric representation and if the distribution of unobservables is assumed to be known, BNE beliefs turn out to be unknown functions. Therefore, the exact functional form of the likelihood function is also unknown. BNE beliefs are the solution to a system of conditional moment restrictions, which contain information about the parameter of interest. The proposal is to estimate players' beliefs by solving a semiparametric analog to the population BNE conditions. Such estimated beliefs also depend on the parameter to be estimated. We then proceed to maximize a trimmed version of the game's log-likelihood function replacing the unknown beliefs with their estimates. The trimming set is an interior subset of the support of the signals where each realization has a unique BNE. The game's likelihood function is well-defined everywhere in that set.

We show that the resulting estimator is $\sqrt{N}$-consistent and characterize its asymptotic distribution. Its efficiency will be partially determined by the predictive power of the signals observed by the players. A simple specification test is also suggested based a test-statistic that measures the difference between the estimated BNE beliefs (a byproduct of the estimation procedure) and the fully nonparametric choice probabilities. The procedure is computationally simple to implement and –unlike the estimation procedures required for the complete-information case– does not rely on prior knowledge about the signs of the strategic-interaction parameters. The paper proceeds as follows: the properties of the game that will be relevant for estimation purposes are presented in Section 2, we make a detailed comparison between the complete and incomplete-information cases. We focus on the incomplete information version thereafter. Section 3 outlines the proposed estimation procedure and describes the asymptotic properties of the resulting estimators for BNE beliefs and for the payoff parameter vector as well as a simple specification test based on a comparison between estimated equilibrium beliefs and nonparametric choice-probability estimators. Section 4 briefly discusses extensions and Section 5 presents some concluding remarks. All the proofs are included in the appendix.

2 Properties of the game

We focus on a 2×2 simultaneous game with the following normal-form representation,

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

This normal-form game was studied by Bresnahan and Reiss (1990, 1991) and Tamer(2003). Let $\varepsilon = (\varepsilon_1, \varepsilon_2)$ and $X = X_1 \cup X_2$, where $X_p \in \mathbb{R}^{k_p}$ for $p = 1, 2$. We will denote $\theta_p = (\beta_p, \alpha_p) \in \mathbb{R}^{k_p+1}$ and $\theta = (\theta_1, \theta_2) \in \mathbb{R}^{k+2}$ with $k \equiv k_1 + k_2$. Following the aforementioned authors, we will assume throughout that the researcher observes the realization of X but does not observe that of ε. Here, we will assume ε to be a vector of mutually independent idiosyncratic shocks which are also independent of all other variables in the model. For $p = 1, 2$ we will assume ε_p to be continuously distributed with unbounded support ($\mathbb{S}(\varepsilon_p) = \mathbb{R}$) and distribution function denoted by $G_p(\epsilon_p)$ –which does not depend on θ–, with corresponding density function $g_p(\epsilon_p)$. The strategic interaction of the game is summarized by (α_1, α_2). We will say that the game is *symmetric* if $\alpha_1\alpha_2 > 0$, *asymmetric* if $\alpha_1\alpha_2 < 0$ and *not jointly strategic* if $\alpha_1\alpha_2 = 0$.

2.1 Incomplete information assumptions

The following assumptions describe the way in which we will incorporate incomplete information into the game:

Assumption I1. (A) The functional form for the game's payoffs and the true value of θ are common knowledge to both players. The realization of (X_p, ε_p) is perfectly observed by player p. (B) The realization of ε_p is observed only by player p, and is independent of any other variable in the model. (C) We allow the possibility that some elements of X_p are only privately observed by player p when the game is played.

Assumption I2. (A) There exists a vector of publicly observable variables $Z_p \in \mathbb{R}^{L_p}$, statistically related to X_p; we will denote $Z \equiv Z_1 \cup Z_2$. All publicly observable elements of X are included in Z, but we can have $Z \neq X$. The distribution functions of X and Z do not depend on θ. (B) Players condition their beliefs about their opponent's expected actions exclusively on the realization of Z.

Assumption I3. (A) Both players have perfect knowledge of the distribution of ε and (X, Z). (B) Players' actions constitute a Bayesian Nash Equilibrium (BNE).

Payoff functional form and the rationality of both players are assumed to be common knowledge. The only source of incomplete information for each player is the inability to observe the realization of his opponent's payoffs. We will refer to Z as "signals", although they are not necessarily strategic signals sent by the players in a signalling game. Signals are used only to extract information about the privately observed elements in X. Thus, if the latter were publicly observed, we would have $X = Z$. The results and methods discussed in this paper can be adapted to the case in which players use different sets of signals. The only critical assumption would be that the conditioning signals be observable by the researcher. This is a particular case of a more general result by Manski (1991), who showed that a discrete choice model with uncertainty is estimable only if expectations are fulfilled and are conditioned only on variables observed by the researcher.

This setup resembles that of a global game, which are games of incomplete information in which players observe a noisy signal of the underlying state. Unlike the general setup of a global game, here we assume that each player has perfect knowledge of his own realized payoffs[1] Attributing the source of incomplete information exclusively to the unobserved idiosyncratic shocks ε would result in beliefs represented by unconditional expectations. This is the case in Seim (2002) and Sweeting (2003). The existence of informative signals implies that in a BNE, players' beliefs are conditional expectations. This complicates the estimation problem if beliefs are unobserved by the researcher and if the conditioning variables have continuous support. This is exactly the scenario this paper deals with.

2.2 Equilibrium with incomplete information

Players condition their beliefs exclusively on the realization of Z, so conditional on Z beliefs are completely deterministic. Denote player p's equilibrium belief as $\pi^*_{-p}(Z,\theta) \equiv E_p[Y_{-p} | \; Z,\theta]$ where $E_p[\cdot]$ denotes player p's subjective expectation and $-p$ denotes player p's opponent. BNE optimal actions are therefore given by

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

Fix a parameter vector $\theta \in \mathbb{R}^{k+2}$ and a realization $z \in \mathbb{S}(Z)$. For this pair (z,θ), define the following objects as functions of $\pi \equiv (\pi_1,\pi_2) \in \mathbb{R}^2$:

$$\varphi_1(\pi_2| \; z,\theta_1) = E\big[G_1\big(X_1'\beta_1 + \alpha_1\pi_2\big)\big| \; Z = z\big] \text{ and } \varphi_2(\pi_1| \; z,\theta_2) = E\big[G_2\big(X_2'\beta_2 + \alpha_2\pi_1\big)\big| \; Z = z\big],$$

and denote $\varphi(\pi| \; z,\theta) = \big(\varphi_1(\pi_2| \; z,\theta_1) \; , \; \varphi_2(\pi_1| \; z,\theta_2)\big)'$. Then $\pi^*(z,\theta) \equiv \big(\pi_1^*(z,\theta),\pi_2^*(z,\theta)\big)$ is the solution for π to the system

$$\pi - \varphi(\pi| \; z,\theta) = 0. \tag{2}$$

Together, (1) and (2) summarize players' behavior and beliefs in a BNE. (2) characterizes BNE beliefs as the solution to a system of conditional moment restrictions. It will play a

[1] Global games were first studied by Carlsson and van Damme (1993). Their properties are thoroughly analyzed by Morris and Shin (2001).

crucial role throughout the paper. As we will see, the incomplete-information model will be self-contained for (z, θ) whenever (2) has a unique solution. The next result establishes existence of a BNE:

Lemma 1 ***(BNE existence)*** *A solution to (2) exists for each $z \in \mathbb{S}(Z)$ and each $\theta \in \mathbb{R}^{k+2}$. Every solution $\pi^*(z, \theta)$ is strictly inside the unit-square $[0,1]^2$.*

By the properties of $G_1(\cdot)$ and $G_2(\cdot)$, $\varphi(\pi|\, z, \theta)$ is continuous function of π and strictly bounded inside the set $[0,1]^2$ for each $z \in \mathbb{S}(Z)$ and each $\theta \in \mathbb{R}^{k+2}$. Existence of a solution to (2) follows from Brouwer's Fixed Point Theorem. This result does not rely on assuming unbounded support for ε_p. The features of $\mathbb{S}(\varepsilon_p)$ will become relevant for BNE uniqueness.

Lemma 2 ***(BNE uniqueness)*** *Consider a particular realization $z \in \mathbb{S}(Z)$ and a given $\theta \in \mathbb{R}^{k+2}$. The solution to (2) is unique for this particular z and θ if and only if all solutions $\big(\pi_1^*(\theta, z), \pi_2^*(\theta, z)\big)$ satisfy*

$$1 - \alpha_1\alpha_2 E\big[g_1\big(X_1'\beta_1 + \alpha_1\pi_2^*(\theta, z)\big)\big|\; Z = z\big] E\big[g_2\big(X_2'\beta_2 + \alpha_2\pi_1^*(\theta, z)\big)\big|\; Z = z\big] > 0 \qquad (3)$$

A sufficient –but not necessary– condition for the requirement of Lemma 2 to hold is

$$1-\alpha_1\alpha_2 E\big[g_1\big(X_1'\beta_1+\alpha_1\pi_2\big)\big|\; Z = z\big] E\big[g_2\big(X_2'\beta_2+\alpha_2\pi_1\big)\big|\; Z = z\big] > 0 \;\; \forall\; (\pi_1, \pi_2) \in [0,1]^2. \quad (3')$$

The object on the left-hand side of (3) is the determinant of the Jacobian of the BNE system (2). If condition (3′) is satisfied, then the system is a contraction in $[0,1]^2$. The condition of Lemma 2 is related to the notion of "regular" equilibrium, which would require the left-hand side of (3) to be different from zero for all BNE solutions. All regular equilibria are locally unique. Global uniqueness in our case requires that the determinant be strictly positive; we could think of (3) as a "strong regularity" condition. The following result presents a simple, sufficient condition for BNE uniqueness with probability one in $\mathbb{S}(Z)$. It will also illustrate that multiplicity of equilibria is possible only in symmetric games.

Corollary 1 Let $\overline{g}_p = \max_{\epsilon \in \mathbb{R}} g_p(\epsilon)$. Suppose $\alpha_1\alpha_2 < 1/(\overline{g}_1\overline{g}_2)$, then each $z \in \mathbb{S}(Z)$ has a unique solution to (2) for any $\beta_1 \in \mathbb{R}^{k_1}$, $\beta_2 \in \mathbb{R}^{k_2}$.

If the condition of Corollary 1 is satisfied, then (3′) is true for all $\pi \in \mathbb{R}^2$ and all $z \in \mathbb{S}(Z)$. This very simple condition is satisfied by all asymmetric games, but it is also satisfied by a subset of symmetric games. This illustrates an important result: under our assumptions, multiple BNE can occur only if the game is symmetric. Next, we summarize the equilibrium properties of the game with complete information.

2.3 Comparison with the complete information case

Players have complete information if they have perfect knowledge about each other's payoff realization. If the game is played with complete information and mixed-strategies are ruled out, players' Nash equilibrium actions are described by the simultaneous system $Y_1 = \mathbb{1}\big\{X_1'\beta_1 + \alpha_1 Y_2 - \varepsilon_1 \geq 0\big\}$ and $Y_2 = \mathbb{1}\big\{X_2'\beta_2 + \alpha_2 Y_1 - \varepsilon_2 \geq 0\big\}$. The properties of this and other nonlinear systems were carefully studied by Heckman (1978). Under these circumstances, the assumption of Nash equilibrium alone will yield a well-defined likelihood function for the four outcomes of the game if and only if $\alpha_1\alpha_2 = 0$. This is the setting of Tamer (2003) and Bresnahan and Reiss (1990, 1991).

If mixed-strategies are allowed, the nature of the game changes nontrivially. In a Nash equilibrium, players choose the optimal randomization of their two available actions. Let $\pi_p(X, \varepsilon; \theta) = \Pr(Y_p = 1|X, \varepsilon; \theta)$ and consider a strategy profile $\big(\pi_1^*(X, \varepsilon; \theta), \pi_2^*(X, \varepsilon; \theta)\big)$ expressed compactly as (π_1^*, π_2^*). This profile is a Nash equilibrium if and only if

$$\pi_1^* \in \underset{\pi_1 \in [0,1]}{\operatorname{argmax}}\ \pi_1 \cdot \big(X_1'\beta_1 - \varepsilon_1 + \pi_2^*\alpha_1\big) \quad \text{and} \quad \pi_2^* \in \underset{\pi_2 \in [0,1]}{\operatorname{argmax}}\ \pi_2 \cdot \big(X_2'\beta_2 - \varepsilon_2 + \pi_1^*\alpha_2\big)$$

Complete information allows each player to randomize in such a way that makes his opponent *exactly* indifferent between his two actions. This is not possible with incomplete information. Let $\mathcal{M}(X, \theta) = \big\{\varepsilon :\ \text{Min}\{X_p'\beta_p, X_p'\beta_p + \alpha_p\} \leq \varepsilon_p \leq \text{Max}\{X_p'\beta_p, X_p'\beta_p + \alpha_p\} \text{ for } p = 1, 2\ \big\}$, where as before $\varepsilon \equiv (\varepsilon_1, \varepsilon_2)$. If $\varepsilon \in \mathcal{M}(X, \theta)$, the game has a unique equilibrium in mixed-strategies if $\alpha_1\alpha_2 < 0$ and it has three equilibria (two pure, one mixed) if $\alpha_1\alpha_2 > 0$. Equilibrium is unique w.p.1 whenever $\varepsilon \notin \mathcal{M}(X, \theta)$ or if $\alpha_1\alpha_2 = 0$. Next, we describe the concept of a self-contained model.

Consider a model with a finite set of outcomes $\mathcal{Y}$ and agents $\mathcal{P}$. Each agent $p \in \mathcal{P}$ chooses an action $S_p \in \mathcal{J}_p$, where $\mathcal{J}_p$ could be finite or not. Let $V = (W_p, \nu_p)_{p \in \mathcal{P}}$ be a vector of covariates, where W_p is observed by the econometrician and ν_p is not and let $\xi = (\xi_p)_{p\in\mathcal{P}}$ be a vector of parameters. Suppose players' actions –according to the model– are described by a binary relation $S_p : \mathbb{S}(V) \times \Xi \longrightarrow \mathcal{J}_p$, where Ξ is the parameter space for ξ. We will say that the model is *self-contained* if it produces a well-defined likelihood function for the entire space of outcomes $\mathcal{Y}$ conditional on $W = (W_p)_{p\in\mathcal{P}}$ and ξ. A model is self-contained if it generates a function $\mathbb{P} : 2^{\mathcal{Y}} \times \mathbb{S}(W) \times \Xi \longrightarrow [0, 1]$ such that $\mathbb{P}(A, w, \xi) = \Pr\big(A \big| W = w; \xi\big)$ for all $A \in 2^{\mathcal{Y}}$. and $\mathbb{P}(\cdot|w, \xi)$ satisfies all the axioms of probability in $(\mathcal{Y}, 2^{\mathcal{Y}})$. This property may hold only for a subset of $\mathbb{S}(W)$. We will say that the model is self-contained w.p.1 if it is self-contained for all $w \in \mathbb{S}(W)$.

This concept is closely related to that of a coherent model introduced by Gourieroux, Laffont and Monfort (1980) and more generally, that of a *complete* econometric model[2]. The model described above will be complete if the binary relation $S_p(\cdot)$ is a well-defined function. The only reason why we do not use the term "incomplete" here is to avoid cumbersome statements such as "all symmetric games are incomplete if players have complete information". Self-containment focuses on the properties of the resulting likelihood function instead of those of agents' decision rules. Self-contained models are of interest because they are potentially amenable to Maximum Likelihood-based estimation. They are also capable of generating predictions for any subset of outcomes conditional on observables.

Assuming players behave optimally, the game analyzed here can be characterized as the model described above. In the incomplete-information case we have $W = (X, Z)$ and $\mathcal{J}_p = Y_p \in \{0, 1\}$ –mixed strategy equilibria occur with probability zero with incomplete information, so we can limit the space of actions to $Y_p \in \{0, 1\}$. With complete information we have $W = X$ and $\mathcal{J}_p = \pi_p \in [0, 1]$ -mixed strategy equilibria may arise with complete information-; with $\nu_p = \varepsilon_p$ and $\Xi = \Theta$ in both cases. The next result compares the conditions under which each version of the game is self-contained if we assume a Nash equilibrium.

[2]Tamer makes the distinction between an incomplete and a (truly) incoherent model. A model is incomplete if it has multiple solutions and incoherent if it has no solutions. See Lewbel (2005) for coherency conditions of models with dummy endogenous variables.

Lemma 3 *Suppose* $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$. *Assuming that players' actions constitute a Nash equilibrium conditional on the information they possess, then:* (1) *Not-jointly strategic games are self-contained w.p.1 for all* $(\beta_1, \beta_2) \in \mathbb{R}^k$. (2) *If players have incomplete information according to I1-I3, every asymmetric game is self-contained w.p.1 for all* $(\beta_1, \beta_2) \in \mathbb{R}^k$. *This is true with complete information if and only if mixed-strategies are allowed.* (3) *Symmetric games are self-contained only if players have incomplete information.*

If players have complete information and mixed-strategies are ruled out, the resulting model is not self-contained unless $\alpha_1\alpha_2 = 0$ due to multiple equilibria (if $\alpha_1\alpha_2 > 0$) and non-existence of equilibrium (if $\alpha_1\alpha_2 < 0$). The assumption $\mathbb{S}(\varepsilon_p) = \mathbb{R}$ for $p = 1, 2$ is sufficient for $\Pr\big(\varepsilon \in \mathcal{M}(X, \theta)\big) > 0$ to be true. Using Lemma 2 and Corollary 1, we conclude that symmetric games are self-contained only if players have incomplete information. Asymmetric games are always self-contained with incomplete information and also with complete information as long as mixed-strategies are allowed. Symmetric games are never self-contained if players have perfect knowledge of the fundamentals of the game. If players have incomplete information and observe the realization of payoffs with the right amount of noise, symmetric games become self-contained. Symmetric games are amenable to MLE estimation only with incomplete information. The remainder of the paper deals with the problem of estimating θ when the game is played with incomplete information.

3 Estimation

Looking at (1), players' choices depend on their beliefs which are assumed to be unobserved. Any estimation procedure must involve a strategy to estimate these beliefs. If the outcome of the game is a BNE, players' beliefs must satisfy (2). Lack of knowledge about the conditional distributions involved implies that equilibrium beliefs are unknown functions of θ and Z. We will focus on estimation procedures which recover players' unobserved beliefs by using well-defined sample analogs to the BNE conditions (2). Let L denote the dimension of Z, with $M \equiv L + 1$. We will make the following stochastic assumptions:

Assumption S1. (A) $G_1(\cdot)$ and $G_2(\cdot)$ are $M+2$ times differentiable with bounded derivatives everywhere. (B) Z is absolutely continuous with respect to Lebesgue measure. (C) $f_{X,Z}(x,z)$ and $f_Z(z)$ are bounded, M times differentiable with respect to z with bounded derivatives everywhere. (D) $E\big[\|XX'\|^2 \big| Z\big]$ is a continuous function of Z everywhere in $\mathbb{S}(Z)$, and $E\big[\|XX'\|^4\big] < \infty$.

Assumption S2. (A) The parameter space $\Theta \subset \mathbb{R}^k$ is compact. (B) There exists a compact set $\mathcal{Z} \subset \text{interior}\{\mathbb{S}(Z)\}$ with $\Pr(Z \in \mathcal{Z}) > 0$ and $\inf_{z\in\mathcal{Z}} f_z(z) > 0$, such that Condition (3) of Lemma 2 is satisfied for each $(z,\theta) \in \mathcal{Z} \times \Theta$.

We will assume the econometrician knows $G_p(\cdot)$ possibly up to a finite parameter vector. This is not a crucial assumption, it could be dropped and we would still be able to obtain a uniformly consistent and asymptotically normal estimator[3]. Moreover, knowing $G_p(\cdot)$ does not affect the fact that equilibrium beliefs are still unknown functions because the distribution of Z is not known. Assumptions S1(B),(C) and (D) have to be satisfied only by the elements of Z not included in X –recall that all publicly observed elements of X are included in Z, and L would denote their dimensionality. Assumption S1(B) could be relaxed and we would simply replace kernels with indicator functions for all variables in Z with discrete support and all the results presented here would follow. If $\alpha_1\alpha_2 < 1/\overline{g}_1\overline{g}_2$, any arbitrary set $\mathcal{Z} \in \text{int}\mathbb{S}(Z)$ satisfies S2(B). This however is far more restrictive than what we require, but it illustrates that we do not need to have prior knowledge of the individual signs of α_p. S1(D) and the uniformly bounded nature of $G_p(\cdot)$ will result in a "in probability" Lipschitz condition that will be useful in establishing the uniform results in Lemmas 4 and 5[4]. Next, we describe two procedures to estimate equilibrium beliefs based on (2).

[3]See for example Ahn (1995).

[4]See Assumption A2 of Theorem 2 in the appendix. See also Lemma 2.9 in Newey and McFadden.

3.1 Estimation of equilibrium beliefs

Take $z \in \mathbb{S}(Z)$, $\pi \in \mathbb{R}^2$, $\theta \in \mathbb{R}^{k+2}$ and let $Q(\pi|z,\theta) = -\big(\pi - \varphi(\pi|z,\theta)\big)'\big(\pi - \varphi(\pi|z,\theta)\big)$. Given S2(B), we can re-interpret $\pi^*(\theta, z)$ as an optimization solution:

$$\pi^*(\theta, z) = \underset{\pi \in [0,1]^2}{\operatorname{argmax}}\ Q(\pi|z,\theta). \tag{4}$$

We will present two estimators of $\pi^*(\cdot)$ based on this representation. The first estimator solves a semiparametric sample analog to 4. The second is a two-step estimator based on a linearization of this condition. Let $K(\cdot)$ and h_N be a kernel function and a bandwidth sequence respectively, with $K_h(t) \equiv K(t/h)$. We will assume the following:

Assumption S3. (A) Define $\mathbb{S}(K(\cdot)) \equiv \big\{\psi \in \mathbb{R}^L : K(\psi) \neq 0\big\}$, then $\mathbb{S}(K(\cdot))$ is compact. $K(\cdot)$ is bounded and symmetric about zero, with $\int K(\psi)d\psi = 1$. Denote $\psi = (\psi_1, \ldots, \psi_L)'$, then $\int \big\|\psi\big\|^M \big|K(\psi)\big| d\psi < \infty$ and $\int \big(\psi_1^{q_1} \cdots \psi_L^{q_L}\big) K(\psi) d\psi_1 \ldots d\psi_L = 0$ for all $0 < q_1 + \cdots + q_L < M$. There exists $c_{_K}$ such that $\big\|K(\psi) - K(\psi')\big\| \leq c_{_K} \big\|\psi - \psi'\big\|$ $\forall$ $\psi, \psi' \in \mathbb{R}^L$. (B) $h_{_N} \to 0$ satisfies: $Nh_{_N}^{L+2} \to \infty$; $Nh_{_N}^{2M} \to 0$ and $N^{1-\sigma}h_{_N}^{2L} \to \infty$ for some $\sigma > 0$. (C) We have a sample of N games, and observe their respective outcomes $\{Y_n\}_{n=1}^N$ and covariates $\{(X_n, Z_n)\}_{n=1}^N$, where the latter is an i.i.d. sample.

Compactness of $\mathbb{S}(K(\cdot))$ can be easily dispensed with, but we maintain this assumption to avoid making assumptions about the size of $\mathbb{S}(Z)$. Some of the elements of X_n are private information at the time the game is played but are assumed to be observed by the econometrician. This is compatible with a situation where some covariates are only privately observed when choices are made, but become publicly disclosed afterwards (e.g, financial statements which are made public after the fact with a lag). Alternatively, the researcher may have access to information which was prohibitively costly for the players to collect when they played the game. Clearly, experimental data sets could be designed to fit exactly our description. In each case, the econometrician is assumed to observe all covariates that were publicly observed by the players, but is able to observe some pieces of private information some time after the game is played and the outcome is known. We now present the first estimator of beliefs.

3.1.1 Analog estimator

Take $z \in \mathbb{R}^L$, $\pi \in \mathbb{R}^2$ and $\theta \in \mathbb{R}^{k+2}$. Let $\widehat{f}_{Z_N}(z) = (Nh_N^L)^{-1}\sum_{n=1}^N K_h(Z_n - z)$, and for $p = 1,2$ define $\widehat{\varphi}_{p_N}(\pi_{\text{-}p} \mid z, \theta_p) = \big(Nh_N^L \cdot \widehat{f}_{Z_N}(z)\big)^{-1}\sum_{n=1}^N G_p\big(X'_{p_n}\beta_p + \alpha_p\pi_{\text{-}p}\big)K_h(Z_n - z)$, with $\widehat{\varphi}_N(\pi \mid z, \theta) = \big(\ \widehat{\varphi}_{1_N}(\pi_2 \mid z, \theta_1)\ ,\ \widehat{\varphi}_{2_N}(\pi_1 \mid z, \theta_2)\ \big)'$. Define the corresponding quadratic function $\widehat{Q}_N(\pi \mid z, \theta) = -\big(\pi - \widehat{\varphi}_N(\pi \mid z, \theta)\big)'\big(\pi - \widehat{\varphi}_N(\pi \mid z, \theta)\big)$. Our semiparametric analog estimator will be defined as

$$\widehat{\pi^*_N}(\theta, z) = \underset{\pi\in[0,1]^2}{\operatorname{argmax}}\ \widehat{Q}_N(\pi \mid z, \theta),$$

which is a semiparametric sample analog to 4. The next result summarizes the properties of $\widehat{\pi^*_N}(\cdot)$ that will be relevant here.

Lemma 4 *Let $\mathcal{Z}$ be as defined in Assumption S2(B). Take $z \in \mathbb{R}^L$ and $\theta \in \mathbb{R}^k$ and let $\widehat{\pi^*_N}(\theta, z) = \underset{\pi\in[0,1]^2}{\operatorname{argmax}}\ \widehat{Q}_N(\pi \mid z, \theta)$. If Assumptions (S1)-(S3) are satisfied, then for any $\delta > 0$,*

$$\text{(A)} \quad \underset{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}}{sup}\ \big(N^{1-\delta}h_N^L\big)^{1/2}\Big\|\widehat{\pi^*_N}(\theta, z) - \pi^*(\theta, z)\Big\| = O_p(1),$$

$$\text{(B)} \quad \underset{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}}{sup}\ \big(N^{1-\delta}h_N^L\big)^{1/2}\Big\|\nabla_\theta\widehat{\pi^*_N}(\theta, z) - \nabla_\theta\pi^*(\theta, z)\Big\| = O_p(1),$$

$$\underset{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}}{sup}\ \big(N^{1-\delta}h_N^L\big)^{1/2}\Big\|\nabla_{\theta\theta'}\widehat{\pi^*_N}(\theta, z) - \nabla_{\theta\theta'}\pi^*(\theta, z)\Big\| = O_p(1),$$

where for each $z \in \mathcal{Z}$ and $\theta \in \Theta$, $\pi^*(\theta, z)$ is the unique solution to the BNE conditions (2).

The key ingredients for this result are a set of in-probability Lipschitz conditions that follow from Assumption S1, and a uniform bound for the inverse-Jacobian of the BNE system (2), which follows from the strong regularity condition S2(B). The last part of assumption S3(B) implies that the uniform rates of convergence in the previous lemma are in fact $o_p(N^{-1/4})$, which is the rate we need for $\sqrt{N}$-consistency of the estimators for θ which we will describe below. The proof of Lemma 4 relies on a refinement –for the i.i.d case– to an existing result by Collomb and Hardle (1986) which is useful to establish rates of uniform convergence over compact sets for nonparametric estimators. Using their results would have required us to assume compactness of $\mathbb{S}(X)$. The results of Collomb and Hardle have been used previously by Stoker (1991) and Ahn and Manski (1993).

3.1.2 Linearized, two-step estimator

Let $J\big(\pi\big|z,\theta\big)$ denote the 2×2 Jacobian $\nabla_\pi\big(\pi-\varphi(\pi|z,\theta)\big)$. By S2(B), $J\big(\pi^*(\theta,z)\big|z,\theta\big)$ is invertible everywhere in $\mathcal{Z}\times\Theta$. Now denote $\widehat{J}_N\big(\pi\big|z,\theta\big)=\nabla_\pi\big(\pi-\widehat{\varphi}_N(\pi|z,\theta)\big)$ and let $\overline{\pi}_{p_N}(z)=\big(Nh_N^L\cdot\widehat{f}_{Z_N}(z)\big)^{-1}\sum_{n=1}^N Y_{p_n}K_h(Z_n-z)$ be the usual kernel-weighed estimator of $E[Y_p|Z=z]$, which does not exploit the BNE conditions. Finally, let $\overline{\overline{\pi}}_{p_N}(z)=\text{Max}\big\{0,\text{Min}\{\overline{\pi}_{p_N}(z),1\}\big\}$ and $\overline{\overline{\pi}}_N(z)\equiv\big(\overline{\pi}_{1_N}(z),\overline{\pi}_{2_N}(z)\big)'$. The proposed linearized estimator of $\pi^*(\cdot)$ is given by $\widetilde{\pi_N^*}(\theta,z)=\overline{\overline{\pi}}_N(z)+\widehat{J}_N\big(\overline{\overline{\pi}}_N(z)\big|z,\theta\big)^{-1}\big[\widehat{\varphi}_N\big(\overline{\overline{\pi}}_N(z)\big|z,\theta\big)-\overline{\overline{\pi}}_N(z)\big]$, which is essentially a linear approximation of (4) around $\overline{\overline{\pi}}_N(z)$. We force the first-step estimator inside $[0,1]^2$ because all BNE are contained inside that set. It also allows us to take advantage of the uniform convergence results over compact sets which are established in the Appendix. Define $\rho(\theta,Z)=\pi^*(\theta_0,z)+J\big(\pi^*(\theta_0,z)\big|z,\theta\big)^{-1}\big[\varphi\big(\pi^*(\theta_0,z)\big|z,\theta\big)-\pi^*(\theta_0,z)\big]$. The next result summarizes the properties of $\widetilde{\pi_N^*}(\cdot)$ that will be relevant here.

Lemma 5 *Let $\mathcal{Z}$ be as defined in Assumption S2(B). Take $z\in\mathbb{R}^L$ and $\theta\in\mathbb{R}^k$, and let $\widetilde{\pi_N^*}(\theta,z)$ and $\rho(\theta,Z)$ be as described above. If Assumptions I1-I3 and S1-S3 are satisfied, then for any $\delta>0$,*

$$\text{(A)}\quad \sup_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}}\big(N^{1-\delta}h_N^L\big)^{1/2}\Big\|\widetilde{\pi_N^*}(\theta,z)-\rho(\theta,z)\Big\|=O_p(1),$$

$$\text{(B)}\quad \sup_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}}\big(N^{1-\delta}h_N^L\big)^{1/2}\Big\|\nabla_\theta\widetilde{\pi_N^*}(\theta,z)-\nabla_\theta\rho(\theta,z)\Big\|=O_p(1),$$

$$\sup_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}}\big(N^{1-\delta}h_N^L\big)^{1/2}\Big\|\nabla_{\theta\theta'}\widetilde{\pi_N^*}(\theta,z)-\nabla_{\theta\theta'}\rho(\theta,z)\Big\|=O_p(1).$$

Note that $\rho(\theta_0,Z)=\pi^*(\theta_0,z)$, $\nabla_\theta\rho(\theta_0,z)=\nabla_\theta\pi^*(\theta_0,z)$ but –in general– $\nabla_{\theta\theta'}\rho(\theta_0,z)\neq\nabla_{\theta\theta'}\pi^*(\theta_0,z)$. To first order of approximation this will not make a difference between the asymptotic properties of the estimator for θ that uses $\widetilde{\pi_N^*}(\cdot)$ as the the estimator of $\pi^*(\cdot)$ and the one that uses $\widehat{\pi_N^*}(\cdot)$. We describe the proposed estimation procedure for θ next.

3.2 Estimation of the payoff parameter vector

Given S2(B), the model is self-contained everywhere in $\mathcal{Z}\times\Theta$. The proposal will be to maximize a trimmed log-likelihood function where the unknown equilibrium beliefs are replaced with either of the estimators studied in sections 3.1.1 and 3.1.2. The trimming set is $\mathcal{Z}$, where the model is self-contained. Since the true functional form of the likelihood function is not known (because beliefs are unknown functions), we call this a trimmed-quasi maximum likelihood procedure. We begin by addressing the issue of identification of θ.

3.2.1 Identification

Assuming that the game is in a BNE, and that each BNE is unique (and therefore the model is self-contained) in $\mathcal{Z} \times \Theta$ is not essential for identification of θ. The fact that all beliefs –even nonequilibrium beliefs– must always be in $[0, 1]$ implies that conditional choice probabilities for $p = 1, 2$ must satisfy $G_p(X_p'\beta_p + \underline{\Delta}_p) \leq \Pr(Y_p = 1|X, Z, \theta) \leq G_p(X_p'\beta_p + \overline{\Delta}_p)$, where $\underline{\Delta}_p = \text{Min}\{\alpha_p, 0\}$ and $\overline{\Delta}_p = \text{Max}\{\alpha_p, 0\}$. Given a scale-normalization for ε_p and provided that X_p has full rank for $p = 1, 2$, point-identification of θ would be viable if for all $\theta \neq \theta_0$, there exist $\mathcal{X}, \mathcal{X}' \subset \mathbb{S}(X)$ such that $G_p(X_p'\beta_p + \underline{\Delta}_p) > G_p(X_p'\beta_{p0} + \overline{\Delta}_{p0})$ for all $X_p \in \mathcal{X}$ and $G_p(X_p'\beta_p + \overline{\Delta}_p) < G_p(X_p'\beta_{p0} + \underline{\Delta}_{p0})$ for all $X_p \in \mathcal{X}'$. This would be possible for example, if there exist a continuously distributed regressor $X_{p_\ell} \in X_p$ with $\beta_{p_\ell} \neq 0$ such that its conditional support $\mathbb{S}(X_{p_\ell}|Z, X_{p_j}, j \neq \ell)$ is unbounded for all $Z \in \mathcal{Z}$ –see for example Theorem 2 in Tamer (2003)–. Such a condition would be perfectly compatible with our assumptions, in particular S1(D). Estimation could be carried out using some of the recently developed estimation methods for semi-identified models or some variation thereof. See for example Manski and Tamer (2002), Chernozhukov et al. (2004), Andrews et al. (2004) or Pakes et al. (2005).

Multiple equilibria does come at a cost however. First of all, any estimation technique would require exact prior knowledge of the signs of α_1 and α_2. Most importantly, since the model would fail to be self-contained, point-estimation of all four outcome probabilities conditional on observables would be impossible[5] Given Lemma 3 and Equation (1), if assumption S2(B) is satisfied then the model is self-contained, outcome probabilities are identified and the parameter vector θ would be identified too –modulo a scale normalization for ε_p– if a simple full-rank condition is satisfied –see for example McFadden (1981)–. We

[5]The set of observable implications (identified features) is smaller. See Jovanovic (1989) for a general analysis of observable implications in models with multiple equilibria.

will add the following assumption.

Assumption S4. If $\theta \neq \theta_0$ then $\Pr\Big(\beta_p' X_p + \alpha_1 \pi_{-p}^*(\theta, Z) \neq \beta_{p0}' X_p + \alpha_{p0} \pi_{-p}^*(\theta_0, Z) \;\Big|\, Z \in \mathcal{Z}\Big) > 0$ for $p = 1, 2$ and all $\theta \in \Theta$.

This full-rank condition does not depend crucially on the assumption $\mathbb{S}(\varepsilon_p) = \mathbb{R}$ and the resulting nonlinear nature of $\pi^*(\cdot)$ –remember that some elements in X may be included in Z–. Suppose momentarily that $\varepsilon_p \sim U[-1, 1]$, $\alpha_1\alpha_2 < 4$ and $X_p'\beta_p + \alpha_p\pi \in (-1, 1)$ w.p.1 for all $\pi \in [0, 1]$. Suppose now that X is publicly observable so that $X = Z$. The resulting equilibrium beliefs are in fact linear functions of X. In this case, Assumption S4 would be satisfied if there exist $X_{1,\ell_1} \in X_1$ and $X_{2,\ell_2} \in X_2$ such that $\Pr(X_{1,\ell_1} \neq X_{2,\ell_2}) > 0$ and $\beta_{1,\ell_1} \neq \beta_{2,\ell_2}$, which is a simple exclusion restriction. Now let us define $F(X_p, Y_p, \theta_p, \pi) = G_p\big(X_p'\beta_p + \alpha_p\pi\big)^{Y_p}\big[1 - G_p\big(X_p'\beta_p + \alpha_p\pi\big)\big]^{1-Y_p}$ and $W \equiv (Y', X', Z')'$. Then $L\big(W, \theta, \pi^*(Z, \theta)\big) = F(X_1, Y_1, \theta_1, \pi_2^*(Z, \theta)) \times F(X_2, Y_2, \theta_2, \pi_1^*(Z, \theta))$ is the likelihood function of Y conditional on (X, Z) and a value of θ. Let $L_{\mathcal{Z}}(W, \theta, \pi) = L(W, \theta, \pi)^{\mathbf{1}\{Z \in \mathcal{Z}\}}$ and $\ell_{\mathcal{Z}}(W, \theta, \pi) = \ln L_{\mathcal{Z}}(W, \theta, \pi)$. Then, $\ell_{\mathcal{Z}}\big(W, \theta, \pi^*(\theta, Z)\big)$ denotes the trimmed log-likelihood function which is always well-defined since the model is self-contained in $\mathcal{Z}$. The next result establishes an information-inequality result for $\ell_{\mathcal{Z}}\big(W, \theta, \pi^*(\theta, Z)\big)$.

Lemma 6 *Suppose Assumptions I1-I3, S1, S2 and S4 are satisfied, then*

$$E\big[\ell_{\mathcal{Z}}\big(W, \theta, \pi^*(\theta, Z)\big)\big] < E\big[\ell_{\mathcal{Z}}\big(W, \theta_0, \pi^*(\theta_0, Z)\big)\big] \quad \forall\ \theta \neq \theta_0,\ \ \theta \in \Theta.$$

Lemma 6 establishes identification of θ. In order to exploit this inequality to estimate θ consistently we need to estimate the unobserved equilibrium beliefs first. Next, we study the properties of procedures which use the the estimators examined in sections 3.1.1 and 3.1.2.

3.2.2 Trimmed quasi-maximum likelihood estimation

One of the attractive properties of self-contained models is that they are amenable to likelihood-based estimation. Choosing $\mathcal{Z}$ as the trimming set allows us to exploit that feature. This also allows us to take advantage of Lemmas 4 and 5 which will ultimately yield $\sqrt{N}$-consistency of our estimator of θ. Before proceeding, let $\nabla_\theta \ell_{\mathcal{Z}}(w, \theta, \pi)$ and $\nabla_\pi \ell_{\mathcal{Z}}(w, \theta, \pi)$

denote the partial derivative of $\ell_{\mathcal{Z}}$ with respect to θ and π respectively, leaving all other arguments constant. The total first partial derivative of our trimmed log-likelihood function with respect to θ is given by

$$\frac{\partial \ell_{\mathcal{Z}}\big(w,\theta,\pi^*(\theta,z)\big)}{\partial\theta} = \nabla_\theta \ell_{\mathcal{Z}}\big(w,\theta,\pi^*(\theta,z)\big) + \nabla_\theta \pi^*(\theta,z)'\nabla_\pi \ell_{\mathcal{Z}}\big(w,\theta,\pi^*(\theta,z)\big),$$

which depends on $\nabla_\theta\pi^*(\theta,z)$ unless $\alpha_1=\alpha_2=0$. This illustrates why it is important to estimate beliefs also as an (unknown) function of θ. Define

$$\overline{D}_{\mathcal{Z}}(Z) = E\left[\frac{\partial^2 \ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)}{\partial\theta\partial\pi'}\,\middle|\, Z\right],$$

where $\partial^2\ell_{\mathcal{Z}}(w,\theta,\pi)\big/\partial\theta\partial\pi'$ denotes the partial derivative of the score with respect to π. As before, let $J(\pi|Z,\theta) = \nabla_\pi\big(\pi - \varphi(\pi|Z,\theta)\big)$ denote the Jacobian of the BNE conditions. We will denote $J_0(Z) = J\big(\pi^*(\theta_0,Z)|Z,\theta_0\big)$ and $B_{\mathcal{Z}}(Z) = \overline{D}_{\mathcal{Z}}(Z)J_0(Z)^{-1}$, which is well-defined by assumption S2(B)[6]. We will add the following assumption:

Assumption S5. (A) $\theta_0 \in \operatorname{int}\Theta$. (B) $E\left[\sup_{\substack{\theta\in\Theta\\ \pi\in[0,1]^2}} \left|\ell_{\mathcal{Z}}\big(W,\theta,\pi\big)\right|\right] < \infty$. Now, define $\Upsilon_p\big(X_p,Z,\theta,\pi\big) = \mathbb{1}\{Z\in\mathcal{Z}\}\times\Big[G_p\big(X_p'\beta_p+\alpha_p\pi\big)\big[1-G_p\big(X_p'\beta_p+\alpha_p\pi\big)\big]\Big]^{-1}$. Then for $p=1,2$ we have: $E\left[\sup_{\substack{\theta\in\Theta\\ \pi\in[0,1]}} \Upsilon_p\big(X_p,Z,\theta,\pi\big)^2\right] < \infty$; $E\left[\sup_{\substack{\theta\in\Theta\\ \pi\in[0,1]}} \left\|X_p\cdot\Upsilon_p\big(X_p,Z,\theta,\pi\big)\right\|^2\right] < \infty$ and $E\left[\sup_{\substack{\theta\in\Theta\\ \pi\in[0,1]}} \left\|\Big(X_pX_p'\Big)\cdot\Upsilon_p\big(X_p,Z,\theta,\pi\big)\right\|^2\right] < \infty$. (C) Let $\Im_{\mathcal{Z}}$ denote the trimmed information matrix $\Im_{\mathcal{Z}} = E\Big[\Big(\partial\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)/\partial\theta\Big)\Big(\partial\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)/\partial\theta\Big)'\Big]$. Then $\Im_{\mathcal{Z}}$ is invertible.

Assumption S5 includes the usual regularity conditions for maximum likelihood estimation. The following result establishes $\sqrt{N}$-consistency of the proposed estimator, we also show that its asymptotic properties (to first order of approximation) are the same if we use the analog estimator $\widehat{\pi}(\cdot)$ or the two-step linearized one $\widetilde{\pi}(\cdot)$.

[6] In fact, all we would have to assume for $B_{\mathcal{Z}}(Z)$ to be well-defined is local-uniqueness of BNE, a much weaker condition than global uniqueness.

Theorem 1 *Let $\widehat{\pi_N^*}(\theta,z)$ and $\widetilde{\pi_N^*}(\theta,z)$ be as defined in Sections 3.1.1 and 3.1.2, and define*

$$\widehat{\theta} = \underset{\theta\in\Theta}{\text{argmax}}\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(W_n,\theta,\widehat{\pi_N^*}(\theta,Z_n)\big) \quad \text{and} \quad \widetilde{\theta} = \underset{\theta\in\Theta}{\text{argmax}}\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(W_n,\theta,\widetilde{\pi_N^*}(\theta,Z_n)\big).$$

If I1-I3 and S1-S5 are satisfied: $\widetilde{\theta} = \widehat{\theta} + o_p(N^{-1/2})$ *and* $\widehat{\theta} - \theta_0 = \frac{1}{N}\sum_{n=1}^{N}\psi_n + o_p(N^{-1/2})$, *where*

$$\psi_n = \Im_{\mathcal{Z}}^{-1}\left[\frac{\partial \ell_{\mathcal{Z}}\big(W_n,\theta_0,\pi^*(\theta_0,Z_n)\big)}{\partial\theta} + B_{\mathcal{Z}}(Z_n)\Big(E\big[Y\big|\ X_n,Z_n\big] - E\big[Y\big|\ Z_n\big]\Big)\right].$$

Consequently, $\sqrt{N}\big(\widehat{\theta}-\theta_0\big) \xrightarrow{d} \mathcal{N}\big(0,\Im_{\mathcal{Z}}^{-1} + \Im_{\mathcal{Z}}^{-1}\Omega_{\mathcal{Z}}\Im_{\mathcal{Z}}^{-1}\big)$ *and* $\sqrt{N}\big(\widehat{\theta}-\widetilde{\theta}\big) \xrightarrow{p} 0$, *where* $\Omega_{\mathcal{Z}} = E\Big[B_{\mathcal{Z}}(Z)\text{Var}\Big[E\big[Y\big|X,Z\big]\Big|Z\Big]B_{\mathcal{Z}}(Z)'\Big]$.

If we had exact knowledge about $f_{X,Z}(\cdot)$, we could find $\pi^*(\theta,z_n)$ by solving (2) –perhaps numerically– and perform MLE[7]. The asymptotic variance of the resulting estimator would simply be $\Im_{\mathcal{Z}}^{-1}$. In the case of the estimators in Theorem 1, lack of knowledge about about $f_{X,Z}(\cdot)$ results in an increase of the variance by $\Im_{\mathcal{Z}}^{-1}\Omega_{\mathcal{Z}}\Im_{\mathcal{Z}}^{-1}$. The correction matrix $\Omega_{\mathcal{Z}}$ depends on $\overline{D}_{\mathcal{Z}}(Z)$, $J_0(Z)^{-1}$ and $\text{Var}\big[E\big[Y\big|X,Z\big]\big|Z\big]$. The term $\overline{D}_{\mathcal{Z}}(Z)$ is a measure of the interdependence between the problems of estimating θ and $\pi^*(\theta,z)$. Such interdependence comes from the BNE conditions unless $\alpha_1 = \alpha_2 = 0$, in which case the estimation of θ is adaptive. $J_0(Z)^{-1}$ appears because both $\widehat{\pi_N^*}(\theta,z)$ and $\widetilde{\pi_N^*}(\theta,z)$ have an asymptotically linear representation up to a term of order $o_p(N^{-1/2})$ and this expression comes from a linear approximation to the BNE conditions.

The asymptotic linear representation of $\widehat{\pi_N^*}(\theta,z)$ and $\widetilde{\pi_N^*}(\theta,z)$ also causes the term $\text{Var}\big[E\big[Y\big|X,Z\big]\big|Z\big]$ to appear because the BNE conditions (2) are simply $E\big[E[Y|X,Z] - E[Y|Z]\ \big|\ Z\big] = 0$ when $\theta = \theta_0$. The magnitude of this term is related to the explanatory power of the signals Z. Suppose for example that $\text{Var}\big[\phi(Z)\cdot E[Y|X,Z]\big] = \text{Var}\big[\phi(Z)\cdot E[Y|Z]\big]$ for all functions $\phi(\cdot)$ with bounded variance, so that all the variability in $E[Y|X,Z]$ is explained by its best predictor based on Z (i.e, $E[Y|Z]$). Then, by iterated expectations we would have $E\big[\phi(Z)\text{Var}\big[\cdot E[Y|X,Z]\big]\phi(Z)'\big] = 0$ and therefore $\Omega_{\mathcal{Z}} = 0$. The asymptotic variance of $\widehat{\theta}$ and $\widetilde{\theta}$ would simply be $\Im_{\mathcal{Z}}$. We now present a compact expression for the standard errors of our estimators.

[7]We would still use trimming to remain in regions of $\mathbb{S}(Z)$ where each BNE is unique.

3.2.3 Standard Errors

Define $\delta_p = E\big[g_p\big(X_p'\beta_{p0} + \alpha_{p0}\pi^*_{-p}(\theta_0, Z)\big)\big|\ Z\big]$; $\zeta_p = E\big[X_p g_p\big(X_p'\beta_{p0} + \alpha_{p0}\pi^*_{-p}(\theta_0, Z)\big)\big|\ Z\big]$. Let g_p and G_p denote $g_p\big(X_p'\beta_{p0} + \alpha_{p0}\pi^*_{-p}(\theta_0, Z)\big)$ and $G_p\big(X_p'\beta_{p0} + \alpha_{p0}\pi^*_{-p}(\theta_0, Z)\big)$ respectively. Define $\Delta_p = g_p/\big[G_p\big(1 - G_p\big)\big]$, $\Delta = (\Delta_1, \Delta_2)$ and $d = 1 - \delta_1\delta_2$. Using this notation, the inverse Jacobian of the BNE system evaluated at θ_0 is given by:

$$J_0^{-1} = d^{-1}\begin{pmatrix} 1 & \alpha_{1_0}\delta_1 \\ \alpha_{2_0}\delta_2 & 1 \end{pmatrix},$$

and $Y - E[Y|X, Z]$ is simply $\big(Y_1 - G_1, Y_2 - G_2\big)'$. To further simplify notation, we will use π^*_p and $\mathbb{1}_{\mathcal{Z}}$ to denote $\pi^*_p(\theta_0, Z)$ and $\mathbb{1}\big\{Z \in \mathcal{Z}\big\}$ respectively. With this in mind, let

$$\Lambda_{1\mathcal{Z}}\big(X, Z\big) = \begin{pmatrix} \Delta_1 X_1 & 0 \\ \Delta_1\pi_2^* & 0 \\ 0 & \Delta_2 X_2 \\ 0 & \Delta_2\pi_1^* \end{pmatrix} \times \mathbb{1}_{\mathcal{Z}};\ \Lambda_{2\mathcal{Z}}\big(X, Z\big) = d^{-1}\begin{pmatrix} \alpha_{1_0}\alpha_{2_0}\delta_2\Delta_1\zeta_1 & \alpha_{2_0}\Delta_2\zeta_1 \\ \alpha_{1_0}\alpha_{2_0}\delta_1\delta_2\Delta_1\pi_2^* & \alpha_{2_0}\delta_1\Delta_2\pi_2^* \\ \alpha_{1_0}\Delta_1\zeta_2 & \alpha_{1_0}\alpha_{2_0}\delta_1\Delta_2\zeta_2 \\ \alpha_{1_0}\delta_2\Delta_1\pi_1^* & \alpha_{1_0}\alpha_{2_0}\delta_1\delta_2\Delta_2\pi_1^* \end{pmatrix} \times \mathbb{1}_{\mathcal{Z}};$$

$$\Xi_{1\mathcal{Z}}\big(Z\big) = \begin{pmatrix} 0 & \alpha_{1_0}E\big[G_1(1-G_1)\Delta_1^2 X_1\big|\ Z\big] \\ 0 & \alpha_{1_0}E\big[G_1(1-G_1)\Delta_1^2\big|\ Z\big]\pi_2^* \\ \alpha_{2_0}E\big[G_2(1-G_2)\Delta_2^2 X_2\big|\ Z\big] & 0 \\ \alpha_{2_0}E\big[G_2(1-G_2)\Delta_2^2\big|\ Z\big]\pi_1^* & 0 \end{pmatrix} \times \mathbb{1}_{\mathcal{Z}}$$

$$\Xi_{2\mathcal{Z}}\big(Z\big) = d^{-1}\begin{pmatrix} \alpha_{2_0}^2 E\big[g_2\Delta_2\big|\ Z\big]\zeta_1 & \alpha_{1_0}^2\alpha_{2_0}E\big[g_1\Delta_1\big|\ Z\big]\delta_2\zeta_1 \\ \alpha_{2_0}^2 E\big[g_2\Delta_2\big|\ Z\big]\delta_1\pi_2^* & \alpha_{1_0}^2\alpha_{2_0}E\big[g_1\Delta_1\big|\ Z\big]\delta_1\delta_2\pi_2^* \\ \alpha_{1_0}\alpha_{2_0}^2 E\big[g_2\Delta_2\big|Z\big]\delta_1\zeta_2 & \alpha_{1_0}^2 E\big[g_1\Delta_1\big|\ Z\big]\zeta_2 \\ \alpha_{1_0}\alpha_{2_0}^2 E\big[g_2\Delta_2\big|\ Z\big]\delta_1\delta_2\pi_1^* & \alpha_{1_0}^2 E\big[g_1\Delta_1\big|\ Z\big]\delta_2\pi_1^* \end{pmatrix} \times \mathbb{1}_{\mathcal{Z}};$$

and define $\Lambda_{\mathcal{Z}}(X, Z) = \Lambda_{1\mathcal{Z}}\big(X, Z\big) + \Lambda_{2\mathcal{Z}}\big(X, Z\big)$ and $\Xi_{\mathcal{Z}}\big(Z\big) = \Xi_{1\mathcal{Z}}\big(Z\big) + \Xi_{2\mathcal{Z}}\big(Z\big)$. Then, we have $\Im_{\mathcal{Z}} = E\Big[\Lambda_{\mathcal{Z}}\big(X, Z\big)\text{Var}\big[Y\big|X, Z\big]\Lambda_{\mathcal{Z}}\big(X, Z\big)'\Big]$ and $B_{\mathcal{Z}}(Z) = \Xi_{\mathcal{Z}}\big(Z\big)J_0^{-1}$. Standard errors are obtained from these expressions, replacing θ_0 with its estimate ($\widehat{\theta}$ or $\widetilde{\theta}$), and using kernel-weighed estimates for each one of the conditional moments involved. To estimate $\Im_{\mathcal{Z}}$ and $\Omega_{\mathcal{Z}}$ respectively, we would use $\widehat{\Im}_{\mathcal{Z}} = \frac{1}{N}\sum_{n=1}^{N}\Big[\widehat{\Lambda}_{\mathcal{Z}}(X_n, Z_n)\widehat{\text{Var}}\big[Y\big|X_n, Z_n\big]\widehat{\Lambda}_{\mathcal{Z}}(X_n, Z_n)'\Big]$ and $\widehat{\Omega}_{\mathcal{Z}} = \frac{1}{N}\sum_{n=1}^{N}\widehat{B}_{\mathcal{Z}}(Z_n)\widehat{\text{Var}}\Big[E[Y|X, Z]\Big|\ Z_n\Big]\widehat{B}_{\mathcal{Z}}(Z_n)'$. Given our assumptions, the resulting standard errors are consistent.

3.2.4 A simple specification test

An interesting and useful byproduct of our estimation methodology is the estimated equilibrium probabilities $\widehat{\pi_N^*}(\widehat{\theta}, z)$. If the model is correctly specified, their difference with respect to the completely nonparametric choice probabilities $\overline{\pi}_N(z)$ should converge to zero in probability. We explore this idea formally in the next lemma.

Lemma 7 *Let $\overline{\pi}_N(z)$ be the nonparametric, kernel-weighed estimator of $E[Y|Z = z]$ described in Subsection 3.1.2. If the game is played according to our assumptions, then*

$$T_N(\widehat{\theta}) \equiv \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{Z_n \in \mathcal{Z}\}\big(\widehat{\pi_N^*}(\widehat{\theta}, Z_n) - \overline{\pi}_N(Z_n)\big) = \frac{1}{N}\sum_{n=1}^{N} \lambda_n + o_p(N^{-1/2})$$

where $\lambda_n = \mathbb{1}\{Z_n \in \mathcal{Z}\}\Big[J_0(Z_n)^{-1}\Big(E\big[Y\big|X_n, Z_n\big] - E\big[Y\big|\, Z_n\big]\Big) - \Big(Y_n - E\big[Y\big|Z_n\big]\Big)\Big] + E\Big[\mathbb{1}\{Z \in \mathcal{Z}\}\nabla_\theta E\big[Y\big|Z\big]\Big] \times \psi_n$*, where ψ_n is the influence function described in Theorem 1. Therefore, if the assumptions of the model are correct,* $\sqrt{N}T_N(\widehat{\theta}) \xrightarrow{d} \mathcal{N}\big(0, E[\lambda_n\lambda_n']\big)$.

A consistent estimator of $E[\lambda_n\lambda_n']$ can be constructed by noting that if the model is correct, then $\nabla_\theta E\big[Y\big|z\big]$ and $J_0(z)$ can be estimated by $\nabla_\theta\widehat{\pi_N^*}(\widehat{\theta}, z)$ and $\widehat{J}_N\big(\widehat{\pi_N^*}(\widehat{\theta}, z)\big|z, \widehat{\theta}\big)$ respectively. Assuming $E[\lambda_n\lambda_n']$ is invertible, then $NT_N(\widehat{\theta})'\widehat{E}[\lambda_n\lambda_n']^{-1}T_N(\widehat{\theta}) \xrightarrow{d} \chi^2_2$. This statistic could be the basis of a specification test. The same result holds if we use the two-step equilibrium belief estimator $\widetilde{\pi_N^*}(\widetilde{\theta}, z)$ instead of $\widehat{\pi_N^*}(\widehat{\theta}, z)$. We can construct alternative specification tests based on the estimated beliefs $\widetilde{\pi_N^*}(\widetilde{\theta}, z)$. For example, we could compare $N^{-1}\sum_{n=1}^{N} \mathbb{1}\big\{Z_n \in \mathcal{Z}\big\}\big(Y_{p_n} - \widehat{\pi^*}_{p_N}(\widehat{\theta}, Z_n)\big)^2$ and $N^{-1}\sum_{n=1}^{N} \mathbb{1}\big\{Z_n \in \mathcal{Z}\big\}\big(Y_{p_n} - \overline{\pi}_{p_N}(Z_n)\big)^2$. The properties of this and other simple tests could be studied by using Theorem 1 and Lemma 12 and Theorem 2 in the appendix, which present uniform asymptotic linear representation for $\widehat{\pi_N^*}(z, \theta) - \pi^*(z, \theta)$ in $\Theta \times \mathcal{Z}$. [8] The next subsection presents some additional comments on trimming.

3.2.5 On choosing the trimming set $\mathcal{Z}$

Assumption S2(3) states that any $z \in \mathcal{Z}$ satisfies equation (3) in Lemma 2. As we discussed previously, very simple conditions on the value of $\alpha_1\alpha_2$ guarantee that we can use any

[8] Those results could also help us study the properties of tests of the form $\sup_{z\in\mathcal{Z}}\big|\widehat{\pi_N^*}(\widehat{\theta}, z) - \overline{\pi}_N(z)\big|$.

$\mathcal{Z} \subset \mathbb{S}(Z)$. If we are completely agnostic about $\alpha_1\alpha_2$, we can still device a way to infer if a candidate set $\overline{\mathcal{Z}} \subset \mathbb{S}(Z)$ can be used as a trimming set. Let $\pi_p(Z) = E[Y_p|Z]$, $\Gamma_p(X, \pi_{\text{-}p}(Z)) = E\big[Y_p\big|X, \pi_{\text{-}p}(Z)\big]$ and $\delta_p(Z) = E\Big[\frac{\partial \Gamma_p(X,\pi_{\text{-}p}(Z))}{\partial \pi_{\text{-}p}(Z)}\Big|Z\Big]$. Take $z \in \mathbb{S}(Z)$, then condition (3) in Lemma 2 is satisfied iff $\mathcal{D}(z) < 1$. Let $\pi^p_{\text{-}p}$ be player p's subjective expectation that $Y_{\text{-}p} = 1$, then $\mathcal{D}(z) < 1$ yields BNE uniqueness in all games in which: (a) beliefs are conditioned on Z, (b) $0 < \Pr(Y_p = 1|X, \pi^p_{\text{-}p}) < 1$ for all $X \in \mathbb{S}(X)$ and all $\pi^p_{\text{-}p} \in [0, 1]$, (c) $\Pr(Y_p = 1|X, \pi^p_{\text{-}p})$ is a monotonic, $\mathcal{C}^1$ function of $\pi^p_{\text{-}p}$ for all $\pi^p_{\text{-}p} \in [0, 1]$ and (d) $E[Y_p|X, Z] = \Gamma_p(X, \pi^p_{\text{-}p}(Z))$. The game studied in this paper satisfies these conditions for all possible values of θ. Given assumption S2(3), any trimming set $\mathcal{Z}$ must satisfy the condition $\Pr\big(\mathcal{D}(Z) < 1\big|Z \in \mathcal{Z}\big) = 1$. Based on this condition, we can devise a way to evaluate a candidate trimming set $\mathcal{Z}$ without relying on $\widehat{\theta}$.

Suppose we maintain the assumption that $\{Y_n, X_n, Z_n\}$ are iid observations from the same data generating process –which may be completely unrelated to the game we studied here. If we assume that $\pi_p(Z)$, $\Gamma_p(X, \pi_{\text{-}p}(Z))$ and $\delta_p(Z)$ are well-defined and sufficiently smooth everywhere in $\text{int}\big(\mathbb{S}(Z)\big)$, we can estimate $\Pr\big(\mathcal{D}(Z) < 1\big|Z \in \mathcal{Z}\big)$ nonparametrically. Suppose X includes L_x continuously distributed elements and take a candidate trimming set $\widetilde{\mathcal{Z}}$. Consider for example the following kernel-smoothed estimators

$$\widehat{\pi}_p(z) = \frac{1}{N h_1^{L_z} \widehat{f}_z(z)} \sum_{j=1}^{N} Y_{p_j} K_{h_1}(Z_j - z)$$

$$\widehat{\Gamma}_p(x, \pi_{\text{-}p}) = \frac{1}{N h_2^{L_x+1} \widehat{f}_{X,\pi_{\text{-}p}}(x, \pi_{\text{-}p})} \sum_{\ell=1}^{N} Y_{p_\ell} K^X_{h_2}(X_\ell - x) K^\pi_{h_2}\big(\widehat{\pi}_{\text{-}p}(Z_\ell) - \pi_{\text{-}p}\big)$$

$$\widehat{\delta}_p(z) = \frac{1}{N h_3^{L_z} \widehat{f}_Z(z)} \sum_{m=1}^{N} \frac{\partial \widehat{\Gamma}_p(X_m, \widehat{\pi}_{\text{-}p}(Z_m))}{\partial \pi_{\text{-p}}} K_{h_3}(Z_m - z); \quad \widehat{D}(z) = \widehat{\delta}_1(z)\widehat{\delta}_2(z)$$

$$\widehat{P}_{\widetilde{\mathcal{Z}}} = \frac{1}{N_{\widetilde{\mathcal{Z}}} h_4} \int_{-\infty}^{1} \sum_{n=1}^{N} K_{h_4}\big(\widehat{D}(Z_n) - u\big) \mathbb{1}\big\{Z_n \in \widetilde{\mathcal{Z}}\big\} du; \quad \text{where } N_{\widetilde{\mathcal{Z}}} = \sum_{n=1}^{N} \mathbb{1}\big\{Z_n \in \widetilde{\mathcal{Z}}\big\}.$$

Suppose $\mathbb{S}(X)$ is compact[9] and that all the unknown densities are bounded away from zero everywhere in $\mathbb{S}(X), \mathcal{Z}$ (we could assume that $\mathbb{S}(Z)$ is also compact). Suppose in addition

[9]Compactness of $\mathbb{S}(X)$ is a strengthening of Assumption S1(D), we would need it here to avoid the influence of points x near the boundary of $\mathbb{S}(X)$. Without it, we would have to do additional trimming on X. Under some assumptions about the tails of $f_X(\cdot)$ we can make this trimming disappear asymptotically, see for example Lemma 8 below.

that all the unknown functions involved are continuous and bounded. By choosing the appropriate kernel functions and bandwidths we can show using standard nonparametric techniques that $\widehat{P}_{\widetilde{\mathcal{Z}}} \xrightarrow{p} \Pr\big(\mathcal{D}(Z) < 1 \big| Z \in \widetilde{\mathcal{Z}}\big) \equiv P_{\widetilde{\mathcal{Z}}}$. In addition, suppose the unknown functions involved are R-times differentiable with bounded derivatives everywhere in $\mathbb{S}(X, Z)$. Using bias-reducing kernels of appropriate order and choosing carefully the relative rates of convergence of the bandwidths involved, we can show that $\phi(N, h_1, \ldots, h_4)\big(\widehat{P}_{\widetilde{\mathcal{Z}}} - P_{\widetilde{\mathcal{Z}}}\big)$ is asymptotically normal for some function $\phi(\cdot)$. These results can be used to evaluate a candidate trimming set $\widetilde{\mathcal{Z}}$.

3.2.6 Trimming when each $z \in \mathbb{S}(Z)$ has a unique BNE.

Suppose we know that pair $(z, \theta) \in \mathbb{S}(Z) \times \Theta$ has a unique BNE. The model would be self-contained w.p.1. If some elements in Z have continuous support (S1(B)), even in this case we still need to use trimming to control the bias of $\widehat{\pi^*_N}(\cdot)$ and $\widetilde{\pi}_{N}(\cdot)$ near the boundary of $\mathbb{S}(Z)$.[10] The next result describes conditions under which we can make trimming disappear asymptotically while retaining the results of Theorem 1.

Lemma 8 *Suppose each pair $(z, \theta) \in \mathbb{S}(Z) \times \Theta$ has a unique BNE. Suppose $b_{N} \to 0$ is a positive sequence such that $N^{1-\sigma} h_{N}^{2L} b_{N}^2 \to \infty$, where σ is as defined in S3(B). Let $\mathcal{Z}_{N} = \big\{z \in \mathbb{R}^L : f_{Z}(z) > b_{N}\big\}$ and $\overline{z}_{N} = \sup_{\mathcal{Z}_N} \|z\|$. Suppose the trimming function $\mathbb{1}\{z_n \in \mathcal{Z}\}$ is replaced with $\mathbb{1}\{\widehat{f}_{Z_N}(z_n) > b_{N}\}$. Then, if there is no positive probability that $f_{Z}(Z) = b$ for any $b > 0$, if $\log \overline{z}_{N} = o_p(\log N)$ and if Assumptions I1-I3 and S1-S5 are satisfied, the asymptotic distribution results of Theorem 1 hold, with '1' replacing $\mathbb{1}\{Z \in \mathcal{Z}\}$.*

If $b_{N} = o_p(\log N)$, then polynomial tails for Z would suffice for the conditions in the previous lemma to hold. Exponential or sub-gaussian tails would allow b_{N} to go to zero at a faster rate. Consider the following generalization of our trimming function: $\mathbb{1}\{\widehat{f}_{Z_N}(z) > b_{N}\} \cdot \phi(z)$. We can show that $\sqrt{N}$-consistency still holds if $\phi(z) \in (0, 1)\ \forall\ z$ is M-times differentiable with bounded derivatives (the same is true for Theorem 1 if we replace $\mathbb{1}\{z_n \in \mathcal{Z}\}$ with $\mathbb{1}\{z_n \in \mathcal{Z}\}\phi(z)$). The asymptotic distribution would depend on the choice of $\phi(\cdot)$ –the

[10]Trimming could be dropped altogether in this case if we assume $\mathbb{S}(Z)$ to be compact with strictly positive density everywhere.

information identity would fail to hold in general if $\phi(\cdot)$ is not constant, see footnote 14 in the appendix–. A formal analysis between efficiency and the choice of $\phi(\cdot)$ is left for future research. We briefly discuss extensions to more complicated games next.

4 Extensions

The methodology examined here can be applied to simultaneous games with incomplete information with more players and/or actions as long as beliefs are conditioned on observables –to the econometrician–. For illustrative purposes, consider a more general version of our game in which both players decide simultaneously to continue or exit and if they continue, they choose $Y_p \in \{0, 1\}$, the extensive form of this game is depicted in Figure 1. Let $C_p = \mathbb{1}\{\text{Player 'p' Continues}\}$. Then, player p can choose the following actions: $\{C_p = 0\}$, $\{C_p = 1, Y_p = 0\}$ or $\{C_p = 1, Y_p = 1\}$. Normalizing the payoffs of Exit at zero would allow us to identify a richer payoff structure when our 2×2 game is actually played (i.e, when both players decide to continue). We could assume for example that $q_p = \beta'_{q_p} X_p - \varepsilon_{q_p}$, $r_p = \beta'_{r_p} X_p - \varepsilon_{r_p}$, $s_p = \beta'_{s_p} X_p - \varepsilon_{s_p}$ and $t_p = \beta'_{t_p} X_p - \varepsilon_{t_p}$, where q_p, r_p, s_p and t_p are the payoffs described in Figure 1 for $p = 1, 2$.

If we maintain our incomplete-information structure assuming that the ε's are privately observed idiosyncratic shocks, that some elements in X are private information and that beliefs are conditioned on informative signals Z, then BNE beliefs are once again the solution to a system of conditional moment restrictions (4×4 in this case), this BNE system is described in the appendix. The distinction between symmetric and asymmetric versions of this game becomes a bit more convoluted, but BNE uniqueness conditions can be stated in terms of conditions on the signs of the principal minors of the BNE Jacobian at all equilibrium solutions –a generalization of assumption S2(B)–. The methods discussed in Subsection 3.2.5 could be extended accordingly –they would be multiple conditions now, as the number of principal minors of the BNE Jacobian would increase in games with more actions and/or players–. In any case, if the support of the ε's is unbounded, then continuity of the distributions involved will *always* ensure BNE uniqueness in a region that includes $(\alpha_p, \delta_p) = (0, 0)$ as an interior point. Given a set of normalization assumptions concerning

Var(ε_p), the proposal would be once again to estimate equilibrium probabilities based on a semiparametric analog to the BNE conditions –see the BNE system for this game in the appendix– and maximize a trimmed quasi log-likelihood replacing the unobserved beliefs with these estimators.

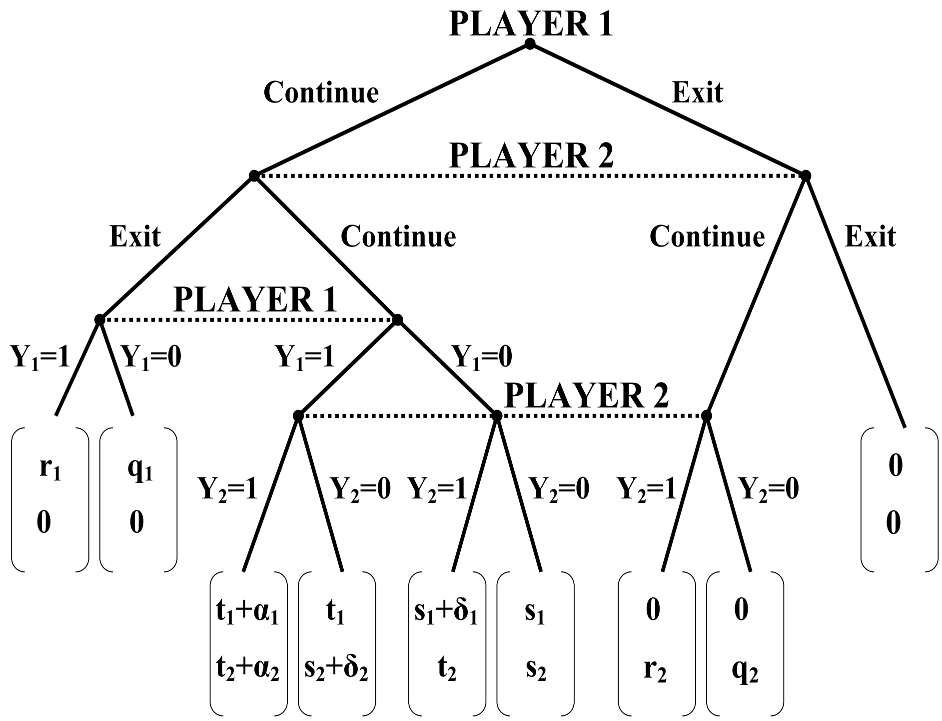

Figure 1: Extensive form of a simultaneous game with two-dimensional strategies.

5 Concluding remarks

The set of features that can be econometrically identified in a game-theoretic model can be determined by the knowledge players are assumed to possess about the primitives of the model. If players are assumed to observe only a noisy signal of the underlying state (the game's payoff realization in the case of this paper) then, if the amount of noise satisfies some conditions, it is possible to identify more features than in a complete information environment. In particular, if we rely exclusively on the assumption of Nash equilibrium and introduce incomplete information, the model produces a well-defined likelihood function for all outcomes of the game in cases in which this would be impossible with complete information and the model becomes amenable to maximum likelihood estimation. If players condition their beliefs on observable signals whose exact distribution is unknown to the econometrician, then the resulting likelihood function is only partially known. We showed here that this problem can be approached by estimating beliefs based on sample analog BNE conditions using semiparametric methods for conditional moment estimation. The cost paid in terms of efficiency for the lack of knowledge of the distributions involved depends on the degree to which the signals used by the players account for the variation of the privately observed payoff-relevant variables. Estimated equilibrium beliefs can also be used to construct specification tests for the model by comparing them against nonparametric choice probabilities. A thorough study of the relationship between agents' information, rationality and the set of identified features of an interactions-based model must become an integral part of the incomplete and partially-identified models literature.

A Mathematical appendix

A.1 Proof of Lemma 2

Fix $(z,\theta) \in \mathbb{S}(Z) \times \mathbb{R}^{k+2}$ and let $\Psi(\pi_1 \big| \, z,\theta) = \varphi_1\big(\varphi_2(\pi_1|z,\theta_2)\big| \, z,\theta_1\big)$. Then the pair $(\pi_1^*(z,\theta), \varphi_2(\pi_1^*(z,\theta)|z,\theta_2))$ is a BNE $\Leftrightarrow \Psi(\pi_1^*(z,\theta)\big| \, z,\theta) = \pi_1^*(z,\theta)$, which makes the BNE problem one-dimensional; all BNE occur when $\Psi(\pi_1\big| \, z,\theta)$ crosses the 45-degree line.

$$\frac{d\Psi(\pi_1|z,\theta)}{d\pi_1} = \alpha_1\alpha_2 \cdot \delta_1\big(\varphi_2(\pi_1|z,\theta_2)\big| \, z,\theta_1\big)\delta_2(\pi_1|z,\theta_2). \Rightarrow \text{sign}\bigg(\frac{d\Psi(\pi_1|z,\theta)}{d\pi_1}\bigg) = \text{sign}\big(\alpha_1\alpha_2\big)$$

If $\mathbb{S}(\varepsilon_p) = \mathbb{R}$ for $p = 1,2$, any fixed point π_1^* must be in $(0,1)$. If $\alpha_1\alpha_2 \leq 0$, then $\Psi(\pi_1\big| \, z,\theta)$ crosses the 45-degree line only once and the BNE must be unique. This case trivially satisfies $\alpha_1\alpha_2 \cdot \delta_1(\pi_2\big| \, z,\theta_1)\delta_2(\pi_1|z,\theta_2) < 1$ for all $\pi \in [0,1]^2$. Assume now that $\alpha_1\alpha_2 > 0$ and allow for multiple BNE. Let π_1^* and π_1^{**} be the smallest and largest BNE. $\mathbb{S}(\varepsilon_p) = \mathbb{R}$ yields $\Psi(0\big| \, z,\theta) > 0$ and $\Psi(1\big| \, z,\theta) < 1$. Then $\Psi(\pi_1\big| \, z,\theta) > \pi_1$ for all $\pi_1 \in [0,\pi_1^*)$ and $\Psi(\pi_1\big| \, z,\theta) < \pi_1$ for all $\pi_1 \in (\pi_1^{**},1]$. Since $\Psi(\pi_1^*\big| \, z,\theta) = \pi_1^*$ and $\Psi(\pi_1^{**}\big| \, z,\theta) = \pi_1^{**}$, there $\exists \;\epsilon > 0$ such that

$$\alpha_1\alpha_2 \cdot \delta_1\big(\varphi_2(\pi_1|z,\theta_2)\big| \, z,\theta_1\big)\delta_2(\pi_1|z,\theta_2) < 1 \quad \forall\, \pi_1 \in \big(\pi_1^* - \epsilon, \pi_1^*\big),\, \pi_1 \in \big(\pi_1^{**}, \pi_1^{**} + \epsilon\big). \quad \text{(A-1)}$$

BNE is multiple $\Leftrightarrow \pi_1^* \neq \pi_1^{**}$. From (A-1), there exists $\pi_1^{***} \in (\pi_1^*, \pi_1^{**})$ such that $\Psi(\pi_1^{***}\big| \, z,\theta) = \pi_1^{***}$ and $\alpha_1\alpha_2 \cdot \delta_1\big(\varphi_2(\pi_1^{***}|z,\theta_2)\big| \, z,\theta_1\big)\delta_2(\pi_1^{***}|z,\theta_2) \geq 1$. Thus, BNE is unique if $\alpha_1\alpha_2 \cdot \delta_1\big(\pi_2^*\big| \, z,\theta_1\big)\delta_2(\pi_1^*|z,\theta_2) < 1$ for all solutions (π_1^*, π_2^*). Now suppose that BNE is unique (π_1^*, π_2^*). From (A-1), $\exists \;\epsilon > 0$ such that $\alpha_1\alpha_2 \cdot \delta_1\big(\varphi_2(\pi_1|z,\theta_2)\big| \, z,\theta_1\big)\delta_2(\pi_1|z,\theta_2) < 1$ for all $\pi_1 \in \big(\pi_1^* - \epsilon, \pi_1^* + \epsilon\big) : \pi_1 \neq \pi_1^*$. Continuity of $\delta_1(\cdot)$, $\varphi_2(\cdot)$ and $\delta_2(\cdot)$ yields $\alpha_1\alpha_2 \cdot \delta_1\big(\varphi_2(\pi_1^*|z,\theta_2)\big| \, z,\theta_1\big)\delta_2(\pi_1^*|z,\theta_2) < 1$. Since $\pi_2^* = \varphi_2(\pi_1^*|z,\theta_2)$, the BNE is unique only if $\alpha_1\alpha_2 \cdot \delta_1\big(\pi_2^*\big| \, z,\theta_1\big)\delta_2(\pi_1^*|z,\theta_2) < 1$. $\square$

A.2 Proof of Lemma 3

The statement of the Lemma that involves the incomplete-information version of the game follows directly from Equation (1) and Lemma 2 (see also Corollary 1). If players have complete information and $\alpha_1\alpha_2 > 0$, the the game has two pure-strategy and one mixed-strategy Nash equilibria whenever $\varepsilon \in \mathcal{M}(X,\theta)$ (see for example Figure 1 in Tamer (2003)).

If $\mathbb{S}(\varepsilon_p) = \mathbb{R}$. If $\alpha_1\alpha_2 < 0$, there is a unique Nash equilibrium whenever $\varepsilon \in \mathcal{M}(X,\theta)$ given by $\Pr^*\big(Y_1 = 1|X,\varepsilon;\theta\big) = \big(\varepsilon_2 - X_2'\beta_2\big)/\alpha_2$ and $\Pr^*\big(Y_2 = 1|X,\varepsilon;\theta\big) = \big(\varepsilon_1 - X_1'\beta_1\big)/\alpha_1$. Equilibria are unique and involve only pure strategies whenever $\varepsilon \notin \mathcal{M}(X,\theta)$. The result of Lemma 3 for the complete information follows from these equilibrium-uniqueness features because $\Pr\big(Y_1, Y_2|X;\theta\big) = \int_{\mathbb{R}^2} \Pr\big(Y_1, Y_2|X,\varepsilon;\theta\big)g(\varepsilon)d\varepsilon$ and players' randomization is always mutually independent in Nash equilibria. □

A.3 Proof of Lemmas 4 and 5

We begin by proving a result that will be used to establish Lemmas 4 and 5.

A.3.1 A uniform linear representation result

Suppose $(X,Z) \in \mathbb{R}^P \times \mathbb{R}^L$ is a random vector with joint density $f_{X,Z}(x,z)$ and let $M \equiv L+1$. Assume throughout we have an iid sample $\{X_n, Z_n\}_{n=1}^N$. Fix two vectors $\gamma \in \mathbb{R}^D$ and $z \in \mathbb{R}^L$, consider a function $\eta : \mathbb{R}^P \times \mathbb{R}^L \times \mathbb{R}^D \to \mathbb{R}$, a kernel $K : \mathbb{R}^L \to \mathbb{R}$ and a bandwidth $h_N \to 0$. Let $K_{h_N}(\psi) = K\big(\psi/h_N\big)$ and define $R_N(z,\gamma) = \big(Nh_N^L\big)^{-1}\sum_{n=1}^N \eta(X_n,z,\gamma)K_{h_N}(Z_n - z)$, $\widehat{f}_{Z_N}(z) = \big(Nh_N^L\big)^{-1}\sum_{n=1}^N K_{h_N}(Z_n - z)$ and $\mu_N(z,\gamma) = R_N(z,\gamma)/\widehat{f}_{Z_N}(z)$. For any $z \in \mathbb{S}(Z)$ let $\mu(z,\gamma) = E\big[\eta(X,z,\gamma)\big|\, Z = z\big]$. Consider the following assumptions:

Assumptions

Assumption A1. (A) Z is absolutely continuous with respect to Lebesgue measure. (B) $f_{X,Z}(x,z)$ and $f_Z(z)$ are bounded, M times differentiable with respect to z with bounded derivatives everywhere in $\mathbb{S}(X,Z)$.

Assumption A2. There exist $\mathcal{Z} \subset \text{int}\big(\mathbb{S}(Z)\big)$ with $\inf_{z\in\mathcal{Z}} f_Z(z) > 0$, and $\Gamma \subset \mathbb{R}^D$ where the following conditions hold: (A) $\mu(z,\gamma)$ is M times differentiable with respect to z and γ with bounded derivatives for every $z \in \mathbb{S}(Z)$ and $\gamma \in \Gamma$. (B) There exists a function $\overline{\eta} : \mathbb{R}^P \to \mathbb{R}_+$ such that $\big|\eta(X,z,\gamma)\big| \leq \overline{\eta}(X)$ w.p.1 for all $X \in \mathbb{S}(X)$, $z \in \mathcal{Z}$, $\gamma \in \Gamma$; $E\big[\overline{\eta}(X)^2\big|\, Z = z\big]$ is a continuous function of z for all $z \in \mathbb{S}(Z)$, and $E\big[\overline{\eta}(X)^4\big] < \infty$. (C) There exists a function $\overline{\eta}_1 : \mathbb{R}^P \to \mathbb{R}_+$, and $\varphi_1 > 0$ such that $\big|\eta(X,z,\gamma) - \eta(X,z',\gamma)\big| \leq \overline{\eta}_1(X)\big\|z - z'\big\|^{\varphi_1}$ w.p.1 for all $X \in \mathbb{S}(X)$, $z, z' \in \mathcal{Z}$, $\gamma \in \Gamma$, and $E\big[\overline{\eta}_1(X)\big] < \infty$. (D) There exists a function $\overline{\eta}_2 : \mathbb{R}^P \to \mathbb{R}_+$, and $\varphi_2 > 0$ such that $\big|\eta(X,z,\gamma) - \eta(X,z,\gamma')\big| \leq \overline{\eta}_2(X)\big\|\gamma - \gamma'\big\|^{\varphi_2}$ w.p.1 for all $X \in \mathbb{S}(X)$, $z \in \mathcal{Z}$, $\gamma, \gamma' \in \Gamma$, and $E\big[\overline{\eta}_2(X)\big] < \infty$.

Assumption A3. (A) Define $\mathbb{S}(K(\cdot)) \equiv \big\{\psi \in \mathbb{R}^L : K(\psi) \neq 0\big\}$, then $\mathbb{S}(K(\cdot))$ is compact. $K(\cdot)$ is bounded and symmetric about zero, with $\int K(\psi)d\psi = 1$. Denote $\psi = (\psi_1, \ldots, \psi_L)'$, then $\int \big\|\psi\big\|^M \big|K(\psi)\big|d\psi < \infty$ and $\int \big(\psi_1^{q_1} \cdots \psi_L^{q_L}\big)K(\psi)d\psi_1 \ldots d\psi_L = 0$ for all $0 < q_1 + \cdots + q_L < M$. There exists $c_{_K}$ such that $\big\|K(\psi) - K(\psi')\big\| \leq c_{_K}\big\|\psi - \psi'\big\|$ $\forall\, \psi, \psi' \in \mathbb{R}^L$. (B) $h_{_N} \to 0$ satisfies: $Nh_{_N}^{L+2} \to \infty$; $Nh_{_N}^{2L}\big/\log(N) \to \infty$ and $Nh_{_N}^{2M} \to 0$. [11]

Theorem 2 *If assumptions A1-A3 are satisfied, then for any $z \in \mathcal{Z}$, $\gamma \in \Gamma$ we have*

$$\mu_{_N}(z,\gamma) - \mu(z,\gamma) = \frac{1}{f_{_Z}(z)}\frac{1}{Nh_{_N}^L}\sum_{n=1}^{N}\big[\eta(X_n,z,\gamma) - \mu(z,\gamma)\big]K_{h_{_N}}(Z_n - z) + \xi_N(z,\gamma)$$

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

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

Proof of Theorem 2: Let $\varphi = \text{Min}\,\big\{1, \varphi_{_1}, \varphi_{_2}\big\}$. Without loss of generality, suppose $\mathcal{Z} = [a_1, b_1] \times \cdots \times [a_L, b_L]$ and $\Gamma = [e_1, h_1] \times \cdots \times [e_D, h_D]$ where $a_\ell < b_\ell$ and $e_d < h_d$.[12] For $\ell = 1, \ldots, L$ and $d = 1, \ldots, D$, let $z_0^{(\ell)} = a_\ell$, $\gamma_0^{(d)} = e_d$, $z_i^{(\ell)} = \text{Min}\big\{z_0^{(\ell)} + i/N^{1/\varphi}\ , b_\ell\big\}$ and $\gamma_j^{(d)} = \text{Min}\big\{\gamma_0^{(d)} + j/N^{1/\varphi}\ , h_\ell\big\}$ where $i, j \in \mathbb{N}$. Define the sets $\mathcal{A}_{1_N} \subset \mathcal{Z}$ and $\mathcal{A}_{2_N} \subset \Gamma$ as $\mathcal{A}_{1_N} = \big\{z_0^{(1)}, \ldots, z_{Q_1}^{(1)}\big\} \times \cdots \times \big\{z_0^{(L)}, \ldots, z_{Q_L}^{(L)}\big\}$ and $\mathcal{A}_{2_N} = \big\{\gamma_0^{(1)}, \ldots, \gamma_{T_1}^{(1)}\big\} \times \cdots \times \big\{\gamma_0^{(D)}, \ldots, \gamma_{T_D}^{(D)}\big\}$. Let $z^* = \max_{z\in\mathcal{Z}}\|z\|$ and $\gamma^* = \max_{\gamma\in\Gamma}\|\gamma\|$. It follows that $Q_{_\ell} \leq \lceil 2z^*N^{1/\varphi}\rceil\ \forall\ \ell$, $T_{_d} \leq \lceil 2\gamma^*N^{1/\varphi}\rceil\ \forall\ d$; $\#\mathcal{A}_{1_N} < (2(z^*+1))^L N^{L/\varphi}$ and $\#\mathcal{A}_{2_N} < (2(\gamma^*+1))^D N^{D/\varphi}$ for all N. For any $(z,\gamma) \in \mathcal{Z}\times\Gamma$ we will denote from now on: $z_\kappa = \operatorname{argmin}_{u\in\mathcal{A}_{1_N}} \|u - z\|$ and $\gamma_\kappa = \operatorname{argmin}_{v\in\mathcal{A}_{2_N}} \|v - \gamma\|$. Note that $\sup_{z\in\mathcal{Z}} \|z - z_\kappa\| \leq \sqrt{L}\big/N^{1/\varphi}$ and $\sup_{\gamma\in\Gamma} \|\gamma - \gamma_\kappa\| \leq \sqrt{D}\big/N^{1/\varphi}$ by construction.

[11] If $L \geq 2$, $Nh_{_N}^{2L}\big/\log(N) \to \infty$ implies $Nh_{_N}^{L+2} \to \infty$.

[12] Every pair compact sets in $\mathbb{R}^L$ and $\mathbb{R}^D$ with Lebesgue measure greater than zero contains a set of the form $[a_1, b_1] \times \cdots \times [a_L, b_L]$ and $[e_1, h_1] \times \cdots \times [e_D, h_D]$ respectively, where $a_\ell < b_\ell$ and $e_d < h_d$.

Step 1 Take any pair of random variables $\mathcal{S}_{N}$, $\mathcal{R}_{N}$ such that: $\mathcal{S}_{N} \leq \mathcal{R}_{N}$ and $\mathcal{S}_{N} \in [0,1]$ w.p.1 $\forall\ N$. Suppose there exist $\varepsilon_1 \in (0,1)$, $\varepsilon_2 \in (0,1)$ and $\overline{N}$ such that $Pr(\mathcal{R}_{N} > \varepsilon_1) \leq \varepsilon_2\ \forall\ N \geq \overline{N}$. Then, $E[\mathcal{S}_N] \leq \varepsilon_1 + \varepsilon_2\ \ \forall\ N \geq \overline{N}$.

Proof: $E[S_N] \leq \varepsilon_1 \cdot \Pr(S_N \leq \varepsilon_1) + 1 \cdot \Pr(S_N > \varepsilon_1) \leq \varepsilon_1 \cdot 1 + 1 \cdot \Pr(R_{N} > \varepsilon_1) \leq \varepsilon_1 + \varepsilon_2\ \forall\ N \geq \overline{N}$.

Step 2 Define the following: $V_{1_N}(z) = \left(Nh_{N}^{L}\right)^{-1} \sum_{n=1}^{N} \overline{\eta}(X_n)^2 K_{h_N}(Z_n - z)^2$ and $V_{2N}(z) = N^{-1} \sum_{n=1}^{N} \overline{\eta}(X_n) \big| K_{h_N}(Z_n - z) \big|$. If the assumptions are satisfied, then $\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ V_{1_N}(z) = O_p(1)$ and $\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ V_{2N}(z) = O_p(1)$.

Proof: Due to the continuity of $E\big[\overline{\eta}(X) \big| Z\big]$ in $\mathbb{S}(Z)$ and boundedness of $K(\cdot)$, there exist $\overline{K}$ and $\overline{V}_1$ such that $\max_{\psi \in \mathbb{R}^L} \big| K(\psi) \big| < \overline{K}$ and $\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ EV_{1_N}(z)$. Define $W_{1_N} = \overline{K}^2 \overline{\eta}(X_n)^2 + h_{N}^{L} \overline{V}_1$ and $\overline{W_{1_N}^2} = N^{-1} \sum_{n=1}^{N} W_{1_N}^2$. Existence of $E\big[\overline{\eta}(X)^4)\big]$ implies that $\overline{W_{1_N}^2} = O_p(1)$. Take any $\overline{M} > 0$. Using Hoeffding's inequality and the fact that $\#\mathcal{A}_{1_N} < (2(z^* + 1))^L N^{L/\varphi}$, A1-A3 yield $\Pr\left(\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ \big| V_{1_N}(z) - EV_{1_N}(z) \big| > M \right) \leq \sum_{z \in \mathcal{A}_{1_N}} \Pr\Big(\big| V_{1_N}(z) - EV_{1_N}(z) \big| > M \Big) < 2(2(z^* + 1))^L N^{L/\varphi} \exp\Big\{ -\frac{1}{2} N h_{N}^{2L} M^2 \Big/ \overline{W_{1_N}^2} \Big\}$. Let $a_{1_N} = \log(2) + L \cdot \log(2(z^* + 1)) + (L/\varphi) \log(N)$. Take any $\varepsilon \in (0,1)$. Since $\overline{W_{1_N}^2} = O_p(1)$, there exists $\overline{N}_\varepsilon$ and $\Delta_\varepsilon > 0$ such that $\Pr\Big(\overline{W_{1_N}^2} > \Delta_\varepsilon \Big) < \varepsilon/2$ for all $N > \overline{N}_\varepsilon$. Define $M_\varepsilon = \sqrt{2\Delta_\varepsilon \big(a_{1_{\overline{N}_\varepsilon}} - \log(\varepsilon/2) \big) \big/ \overline{N}_\varepsilon h_{\overline{N}_\varepsilon}^{2L}}$. Since $N h_{N}^{2L} / \log(N) \to \infty$, we have $a_{1_N} - \frac{1}{2} N h_{N}^{2L} M_\varepsilon^2 \Big/ \Delta_\varepsilon < \log(\varepsilon/2)\quad \forall\ N > \overline{N}_\varepsilon$. Therefore $\forall\ \varepsilon \in (0,1)$, $\exists\ \ M_\varepsilon, \overline{N}_\varepsilon$ s.th $\Pr\left(2(2(z^* + 1))^L N^{L/\varphi} \exp\Big\{ -\frac{1}{2} N h_{N}^{2L} M_\varepsilon^2 \Big/ \overline{W_{1_N}^2} \Big\} > \varepsilon/2 \right) < \varepsilon/2$. Then $\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ V_{1_N}(z) = O_p(1)$ follows from Step 1 with $\mathcal{S}_{N} = \Pr\left(\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ \big| V_{1_N}(z) - EV_{1_N}(z) \big| > M_\varepsilon \right)$ and $\mathcal{R}_{N} = 2(2(z^* + 1))^L N^{L/\varphi} \exp\Big\{ -\frac{1}{2} N h_{N}^{2L} M_\varepsilon^2 \Big/ \overline{W_{1_N}^2} \Big\}$. The result $\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ V_{2N}(z) = O_p(1)$ follows more simply by noting that $\underset{z \in \mathcal{A}_{1_N}}{\text{Max}}\ V_{2_N}(z) \leq \overline{K} N^{-1} \sum_{n=1}^{N} \overline{\eta}(X_n) = O_p(1)$. □

Step 3 If Assumptions A1-A3 are satisfied, then there exists N' and $\overline{R}$ such that for all $N > N'$: $\underset{\substack{z \in \mathcal{Z} \\ \gamma \in \Gamma}}{sup} \Big| ER_{N}(z, \gamma) - f_{Z}(z) \mu(z, \gamma) \Big| \leq h_{N}^{M} \overline{R}$.

Proof: Fix $i \in \mathbb{N}$ and let $Q_i = \big\{ (q_1, \ldots, q_L) \in \mathbb{N}^L : q_1 + \cdots + q_L = i \big\}$. Define

$$\Delta_i(u, v) = \sum_{Q_i} \frac{\partial^i f_{x,z}(u, v)}{\partial z_1^{q_1} \cdots \partial z_L^{q_L}}.$$

Given our assumptions, the following Taylor approximation is valid for any $(z,\gamma) \in \mathcal{Z} \times \Gamma$:

$$ER_N(z,\gamma) = \int \eta(u,z,\gamma)\Big[f(u,z)\int K(\psi)d\psi + \sum_{i=1}^{M-1}(-1)^i \frac{h_{_N}^i}{i!}\Delta_i(u,z)\sum_{Q_i}\int \psi_1^{q_1}\cdots\psi_L^{q_L}K(\psi)d\psi$$
$$+ (-1)^M\frac{h_{_N}^M}{M!}\int\sum_{Q_M}\psi_1^{q_1}\cdots\psi_L^{q_L}\Delta_M(u,z+h_{_N}^*\psi)K(\psi)d\psi\Big]du,$$

with $h_{_N}^* \in (0,h_{_N})$. Since $\mathbb{S}(K(\cdot))$ and $\mathcal{Z}$ are compact and the latter is in the interior of $\mathbb{S}(Z)$, there exists N' such that $z + h_{_N}\psi \in \mathbb{S}(Z)$ for all $N > N'$ and $\psi \in \mathbb{S}(K(\cdot))$. Existence and uniform-boundedness of the m^{th} derivative of $\mu(z,\gamma)$ for $m = 1,\ldots M$ and the bias-reducing properties of $K(\cdot)$ imply that there exists $C > 0$ and N' such that for all $N > N'$,

$$\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\Big|ER_{_N}(z,\gamma) - f_{_Z}(z)\mu(z,\gamma)\Big| \le C\frac{h_{_N}^M}{M!}\Big|\int\sum_{Q_M}\psi_1^{q_1}\cdots\psi_L^{q_L}K(\psi)d\psi\Big|.$$

The result follows from the fact that $\int\|\psi\|^M|K(\psi)|d\psi < \infty$. $\Box$

Step 4 If A1-A3 are satisfied, then $\underset{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}{sup}\ \big(N^{1-\delta}h_{_N}^L\big)^{1/2}\Big|R_{_N}(z,\gamma) - ER_{_N}(z,\gamma)\Big| = O_p(1)$ for any $\delta > 0$.

Proof: Let z_κ and γ_κ be as defined prior to Step 1. Using the triangle inequality we have

$$\begin{aligned}\Big|R_{_N}(z,\gamma) - ER_{_N}(z,\gamma)\Big| \ &\le\ \Big|R_{_N}(z_\kappa,\gamma_\kappa) - ER_{_N}(z_\kappa,\gamma_\kappa)\Big| + \Big|R_{_N}(z,\gamma) - R_{_N}(z_\kappa,\gamma_\kappa)\Big| \\ &\quad + \Big|ER_{_N}(z,\gamma) - ER_{_N}(z_\kappa,\gamma_\kappa)\Big|\end{aligned} \tag{A-2}$$

By A1-A3: $\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\big(N^{1-\delta}h_{_N}^L\big)^{1/2}\ \Big|R_{_N}(z,\gamma) - R_{_N}(z_\kappa,\gamma_\kappa)\Big| \le c_k\big(N^{1+\delta}h_{_N}^{L+2}\big)^{-1/2}\sum_{n=1}^N\overline{\eta}(X_n)/N + \overline{K}/\big(N^{1+\delta}h_{_N}^L\big)^{-1/2}\Big[L^{\varphi_1/2}\cdot\sum_{n=1}^N\overline{\eta}_1(X_n)/N + L^{\varphi_2/2}\cdot\sum_{n=1}^N\overline{\eta}_1(X_n)/N\Big] = o_p(1)$. Step 3 yields $\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\big(N^{1-\delta}h_{_N}^L\big)^{1/2}\Big|ER_{_N}(z,\gamma) - ER_{_N}(z_\kappa,\gamma_\kappa)\Big| \le 2\big(N^{1-\delta}h_{_N}^{L+2M}\big)^{1/2}\overline{R} + \big(h_{_N}^L/N^{1+\delta}\big)^{1/2}\cdot\big[\overline{f}c_1 + c_2\big] = o(1)$. Equation A-2 becomes $\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\big(N^{1-\delta}h_{_N}^L\big)^{1/2}\Big|R_{_N}(z,\gamma) - ER_{_N}(z,\gamma)\Big| \le \max_{\substack{z\in\mathcal{A}_{1_N}\\ \gamma\in\mathcal{A}_{2_N}}}\big(N^{1-\delta}h_{_N}^L\big)^{1/2}\Big|R_{_N}(z,\gamma) - ER_{_N}(z,\gamma)\Big| + o_p(1)$.

Take any $M > 0$, then $\Pr\Big(\max_{\mathcal{A}_{1_N},\mathcal{A}_{2_N}}\big(N^{1-\delta}h_{_N}^L\big)^{1/2}\Big|R_{_N}(z,\gamma) - ER_{_N}(z,\gamma)\Big| > M\Big) \le \sum_{\gamma\in\mathcal{A}_{2_N}}\sum_{z\in\mathcal{A}_{1_N}}\Pr\Big(\big(N^{1-\delta}h_{_N}^L\big)^{1/2}\Big|R_{_N}(z,\gamma) - ER_{_N}(z,\gamma)\Big| > M\Big)$. Let $V_N(z) = V_{1_N}(z) + 2\big(h_{_N}^M\overline{R} +$

$\overline{f}\overline{\mu})V_{2_N}(z)+h_N^L\left(h_N^M\overline{R}+\overline{f}\overline{\mu}\right)^2$ and $V_N=\max\limits_{z\in\mathcal{A}_{1_N}} V_N(z)$, where $V_{1_N}(z)$ and $V_{2_N}(z)$ are as in Step 2, $\overline{\mu}=\sup\limits_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\left|\mu(z,\gamma)\right|$ and $\overline{f}$, $\overline{R}$ are as defined above. Using Steps 1, 2 and Hoeffding's inequality: $\Pr\Big(\left(N^{1-\delta}h_N^L\right)^{1/2}\Big|R_N(z,\gamma)-ER_N(z,\gamma)\Big|>M\Big)\leq\exp\Big\{-\frac{1}{2}NM^2\left(N^{1-\delta}h_N^L\right)^{-1}\Big/\frac{V_N(z)}{h_N^L}\Big\}=\exp\Big\{-\frac{1}{2}N^\delta M^2\Big/V_N(z)\Big\}\quad\forall\ z\in\mathcal{Z},\gamma\in\Gamma\leq\exp\Big\{-\frac{1}{2}N^\delta M^2\Big/V_N\Big\}\ \forall\ z\in\mathcal{A}_{1_N},\gamma\in\Gamma$. Since $\mathcal{A}_{2_N}\subset\Gamma$, this implies that

$$\Pr\Bigg(\max_{\substack{z\in\mathcal{A}_{1_N}\\ \gamma\in\mathcal{A}_{2_N}}}\left(N^{1-\delta}h_N^L\right)^{1/2}\Big|R_N(z,\gamma)-ER_N(z,\gamma)\Big|>M\Bigg)\leq\sum_{\gamma\in\mathcal{A}_{2_N}}\sum_{z\in\mathcal{A}_{1_N}}\exp\Big\{-\frac{1}{2}N^\delta M^2\Big/V_N\Big\}$$
$$<(2(z^*+1))^L(2(\gamma^*+1))^D N^{(L+D)/\varphi}\exp\Big\{-\frac{1}{2}N^\delta M^2\Big/V_N\Big\},\tag{A-3}$$

where z^* and γ^* were defined above. Using the results from Step 2, we have $V_N=O_p(1)$. Having noted this, we complete the proof by invoking the result of Step 1 using the same steps as those of Step 2, defining a_N and M_ε in the same fashion and letting $\mathcal{S}_N=\Pr\Bigg(\max\limits_{\substack{z\in\mathcal{A}_{1_N}\\ \gamma\in\mathcal{A}_{2_N}}}\left(N^{1-\delta}h_N^L\right)^{1/2}\Big|R_N(z,\gamma)-ER_N(z,\gamma)\Big|>M_\varepsilon\Bigg)$ and $\mathcal{R}_N=(2(z^*+1))^L(2(\gamma^*+1))^D N^{(L+D)/\varphi}\exp\Big\{-\frac{1}{2}N^\delta M^2\Big/V_N\Big\}$. □

Step 5 If A1-A3 are satisfied, then $\mathop{sup}\limits_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\left(N^{1-\delta}h_N^L\right)^{1/2}\Big|R_N(z,\gamma)-f_Z(z)\mu(z,\gamma)\Big|=O_p(1)$ for any $\delta>0$.

Proof: Follows immediately from Steps 3, 4 and the bandwidth condition $Nh_N^{2M}\to 0$. □

Step 6 (final step) Note that $\widehat{f}_{Z_N}(z)$ satisfies Assumptions A1-A3 with $\eta(X,z,\gamma)=1$ $\forall\ X,z,\gamma$. Step 4 yields $\sup\limits_{z\in\mathcal{Z}}\left(N^{1-\delta}h_N^L\right)^{1/2}\left|\widehat{f}_{Z_N}(z)-f_Z(z)\right|=O_p(1)$ for any $\delta>0$. Take any $z\in\mathcal{Z}$, $\gamma\in\Gamma$. Consider the second-order approximation

$$\mu_N(z,\gamma)-\mu(z,\gamma)=\frac{1}{f_Z(z)}\left[R_N(z,\gamma)-f_Z(z)\mu(z,\gamma)\right]-\frac{\mu(z,\gamma)}{f_Z(z)}\left[\widehat{f}_{Z_N}(z)-f_Z(z)\right]$$
$$+\frac{1}{2}\left[R_N(z,\gamma)-f_Z(z)\mu(z,\gamma)\ ,\ \widehat{f}_{Z_N}(z)-f_Z(z)\right]\underbrace{\begin{bmatrix}0 & \frac{-1}{\widetilde{f}_{Z_N}(z)^2}\\ \frac{-1}{\widetilde{f}_{Z_N}(z)^2} & \frac{2\widetilde{R}_N(z,\gamma)}{\widetilde{f}_{Z_N}(z)^3}\end{bmatrix}}_{\equiv\widetilde{H}_N(z,\gamma)}\begin{bmatrix}R_N(z,\gamma)-f_Z(z)\mu(z,\gamma)\\ \widehat{f}_{Z_N}(z)-f_Z(z)\end{bmatrix},$$

with $\widetilde{f}_{Z_N}(z)$ between $f_N(z)$ and $f_Z(z)$, and $\widetilde{R}_N(z,\gamma)$ between $R_N(z,\gamma)$ and $f_Z(z)\mu(z,\gamma)$. Using Step 5 and the characteristics of $\mathcal{Z}$ we get $\sup_{\substack{z\in\mathcal{Z}\\ \gamma\in\Gamma}}\left\|\widetilde{H}_N(z,\gamma)\right\| = O_p(1)$. Given this, the result of Theorem 2 follows immediately from Step 5. □

Take any $z \in \mathbb{R}^L$ and let $\widehat{f}_{Z_N}(z) = \left(Nh_N^L\right)^{-1}\sum_{n=1}^N K_h(Z_n - z)$. For any function $\phi(W)$ we will denote $\widehat{\mathbb{P}}_K(z)\phi(W) = \left(Nh_N^L\widehat{f}_{Z_N}(z)\right)^{-1}\sum_{n=1}^N \phi(W_n)K_h(Z_n - z)$. If $z \in \mathbb{S}(Z)$, let $\mathbb{P}(z)\phi(W) = E\left[\phi(W)\middle|Z = z\right]$. Fix any $\theta_p \in \mathbb{R}^{k_p+1}$ and $\pi \in \mathbb{R}$. Then $\widehat{\mathbb{P}}_K(z)\phi(W;\theta,\pi) = \left(Nh_N^L\widehat{f}_{Z_N}(z)\right)^{-1}\sum_{n=1}^N \phi(W_n;\theta,\pi)K_h(Z_n - z)$ and $\mathbb{P}(z)\phi(W;\theta,\pi) = E\left[\phi(W;\theta,\pi)\middle|Z = z\right]$. Define $\widehat{\varphi}_{pN}(\pi \mid z,\theta_p) = \widehat{\mathbb{P}}_K(z)G_p\left(X_p'\beta_p + \alpha_p\pi\right)$, $\widehat{\delta}^{(m)}_{pN}(\pi \mid z,\theta_p) = \widehat{\mathbb{P}}_K(z)g_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi\right)$, $\widehat{\zeta}^{(m)}_{pN}(\pi \mid z,\theta_p) = \widehat{\mathbb{P}}_K(z)X_pg_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi\right)$ and $\widehat{\xi}^{(m)}_{pN}(\pi \mid z,\theta_p) = \widehat{\mathbb{P}}_K(z)X_pg_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi\right)$ where (m) denotes the m^{th} derivative with $m \geq 0$. Let $\varphi(\pi \mid z,\theta_p) = \mathbb{P}(z)G_p\left(X_p'\beta_p + \alpha_p\pi\right)$, $\delta^{(m)}(\pi \mid z,\theta_p) = \mathbb{P}(z)g_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi\right)$, $\zeta^{(m)}(\pi \mid z,\theta_p) = \mathbb{P}(z)X_pg_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi\right)$ and $\xi^{(m)}(\pi \mid z,\theta_p) = \mathbb{P}(z)X_pX_p'G_p\left(X_p'\beta_p + \alpha_p\pi\right)$. We have the following result.

Lemma 9 *Take any three compact sets $A \in \mathbb{R}$, $B \in \mathbb{R}^{k_p+1}$ and $C \in \mathbb{R}^L$ with $C \in int\,\mathbb{S}(Z)$ and $inf_{z\in C}\, f_Z(z) > 0$. Suppose assumptions S1 and S3 are satisfied. Then for any $\delta > 0$, $m = 0, 1\ldots, M$ and $p = 1, 2$: (A)* $\sup_{z\in C}\left(N^{1-\delta}h_N^L\right)^{1/2}\left|\widehat{f}_{Z_N}(z) - f_Z(z)\right| = O_p(1)$

$$(B)\ \sup_{\substack{z\in C\\ \theta_p\in B\\ \pi\in A}}\left(N^{1-\delta}h_N^L\right)^{1/2}\left|\widehat{\varphi}_{pN}(\pi \mid z,\theta_p) - \varphi_p(\pi \mid z,\theta_p)\right| = O_p(1)$$

$$(C)\ \sup_{\substack{z\in C\\ \theta_p\in B\\ \pi\in A}}\left(N^{1-\delta}h_N^L\right)^{1/2}\left|\widehat{\delta}^{(m)}_{pN}(\pi \mid z,\theta_p) - \delta_p^{(m)}(\pi \mid z,\theta_p)\right| = O_p(1)$$

$$(D)\ \sup_{\substack{z\in C\\ \theta_p\in B\\ \pi\in A}}\left(N^{1-\delta}h_N^L\right)^{1/2}\left\|\widehat{\zeta}^{(m)}_{pN}(\pi \mid z,\theta_p) - \zeta_p^{(m)}(\pi \mid z,\theta_p)\right\| = O_p(1)$$

$$(E)\ \sup_{\substack{z\in C\\ \theta_p\in B\\ \pi\in A}}\left(N^{1-\delta}h_N^L\right)^{1/2}\left\|\widehat{\xi}^{(m)}_{pN}(\pi \mid z,\theta_p) - \xi_p^{(m)}(\pi \mid z,\theta_p)\right\| = O_p(1)$$

Proof: If S1 and S3 are satisfied, then each one of the objects described above satisfies the conditions of Theorem 2 with $\gamma \equiv (\theta,\pi)$: Using S1(A), there exists $\overline{D} < \infty$ such that w.p.1: $\left|G_p\left(X_p'\beta_p + \alpha_p\pi_{\text{-}p}\right)\right| \leq \overline{D}$; $\left|g_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi_{\text{-}p}\right)\right| \leq \overline{D}$ and $\left\|\left(X_pX_p'\right)g_p^{(m)}\left(X_p'\beta_p + \alpha_p\pi_{\text{-}p}\right)\right\| \leq \overline{D}\cdot\left\|X_pX_p'\right\|$. The conditions of Theorem 2 then follow from S1(B)-(D) and S3. □.

Lemma 10 *Let $\widehat{\pi_N^*}(z,\theta)$ be as defined in Section 3.1.1 and let $\pi^*(z,\theta)$ be the solution to the BNE conditions (2). If S1-S3 are satisfied, then*
$\underset{\mathcal{Z}\times\Theta}{sup} \left\|\widehat{\pi_N^*}(z,\theta) - \pi^*(z,\theta)\right\| = o_p(1).$

Proof:

Take $\delta > 0$, for each $(\theta, z) \in \Theta \times \mathcal{Z}$ let $M_{\theta,z} = \{\pi : \|\pi - \pi^*(\theta,z)\| < \delta\}$ and let $\overline{M}_{\theta,z}$ be its complement in $\mathbb{R}^2$. Now define the set $\mathcal{N}_{\theta,z} = \overline{M}_{\theta,z} \cap [0,1]^2$. Then $\mathcal{N}_{\theta,z} \in [0,1]^2$ is compact for all $(\theta, z) \in \Theta \times \mathcal{Z}$, and by continuity we get that $\max_{\pi\in\mathcal{N}_{\theta,z}} Q(\pi \mid \theta, z)$ exists for all $(\theta, z) \in \Theta \times \mathcal{Z}$. Now define $\varepsilon = \inf_{\mathcal{Z}\times\Theta} \left[Q(\pi^*(\theta,z) \mid \theta, z) - \max_{\pi\in\mathcal{N}_{\theta,z}} Q(\pi \mid \theta, z)\right]$. Then $\varepsilon > 0$, since $\pi^*(\theta, z)$ is the unique solution to $\max_{\pi\in\mathbb{R}^2} Q(\pi \mid \theta, z)$ for each $(\theta, z) \in \Theta \times Z$. Now let A_N be the event $\sup_{\substack{\mathcal{Z}\times\Theta \\ \pi\in[0,1]^2}} \left|\widehat{Q}_N(\pi \mid \theta, z) - Q(\pi \mid \theta, z)\right| < \varepsilon/2$. We have $\pi^*(\theta, z) \in [0,1]^2$ and $\widehat{\pi_N^*}(z,\theta) \in [0,1]^2$ for all $(\theta, z) \in \Theta \times \mathcal{Z}$. This yields the following two implications:

$$A_N \Rightarrow Q\big(\widehat{\pi_N^*}(\theta,z) \mid \theta, z\big) > \widehat{Q}_N\big(\widehat{\pi_N^*}(\theta,z) \mid \theta, z\big) - \frac{\varepsilon}{2} \qquad \forall\ (\theta, z) \in \Theta \times \mathcal{Z}$$
$$A_N \Rightarrow \widehat{Q}_N\big(\pi^*(\theta,z) \mid \theta, z\big) > Q\big(\pi^*(\theta,z) \mid \theta, z\big) - \frac{\varepsilon}{2} \qquad \forall\ (\theta, z) \in \Theta \times \mathcal{Z}.$$

Combining these the two implications with $\widehat{Q}_N\big(\widehat{\pi_N^*}(\theta,z) \mid \theta, z\big) \geq \widehat{Q}_N\big(\pi^*(\theta,z) \mid \theta, z\big)$ $\forall\ (\theta, z) \in \Theta \times \mathcal{Z}$ we get: $A_N \Rightarrow Q\big(\widehat{\pi_N^*}(\theta,z) \mid \theta, z\big) > Q\big(\pi^*(\theta,z) \mid \theta, z\big) - \varepsilon \quad \forall\ (\theta, z) \in \Theta \times \mathcal{Z}$. By definition of ε, we can conclude that $A_N \Rightarrow \|\widehat{\pi_N^*}(\theta,z) - \pi^*(\theta,z)\| < \delta$ for all $(\theta, z) \in \Theta \times \mathcal{Z}$ or equivalently, $A_N \Rightarrow \sup_{\mathcal{Z}\times\Theta}\|\widehat{\pi_N^*}(\theta,z) - \pi^*(\theta,z)\| < \delta$. As a consequence, we then have that $\Pr\left(\sup_{\mathcal{Z}\times\Theta}\|\widehat{\pi_N^*}(\theta,z) - \pi^*(\theta,z)\| < \delta\right) \geq \Pr(A_N)$. By Lemma 9(B), we have $\Pr(A_N) \to 1$. Therefore $\Pr\left(\sup_{\mathcal{Z}\times\Theta}\|\widehat{\pi_N^*}(\theta,z) - \pi^*(\theta,z)\| < \delta\right) \to 1$. Since δ is an arbitrary positive number, this implies that $\sup_{\mathcal{Z}\times\Theta} \|\widehat{\pi_N^*}(\theta,z) - \pi^*(\theta,z)\| = o_P(1)$, as claimed. □

Lemma 11 *Let $\widehat{\pi_N^*}(z,\theta)$ be as defined in Section 3.1.1. Define*

$$\widehat{d}_N\big(\widehat{\pi_N^*}(z,\theta)\big|\ z,\theta\big) = 1 - \alpha_1\alpha_2 \cdot \widehat{\delta}_{1_N}\big(\widehat{\pi_{2}^*}_N(z,\theta)\big|\ z,\theta_1\big)\widehat{\delta}_{2_N}\big(\widehat{\pi_{1}^*}_N(z,\theta)\big|\ z,\theta_2\big).$$

Where $\widehat{\delta}_{p_N}(\cdot) \equiv \widehat{\delta}^{(0)}_{p_N}(\cdot)$, with the latter defined as in Lemma 9. If Assumptions S1-S3 are satisfied, then $\underset{\mathcal{Z}\times\Theta}{sup} \left|\widehat{d}_N\big(\widehat{\pi_N^*}(z,\theta)\big|\ z,\theta\big)^{-1}\right| = O_p(1).$

Proof: Let $\inf_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}} d\big(\pi^*(\theta,z)|z,\theta\big) \equiv \underline{d} > 0$. We have $\sup_{\mathcal{Z}\times\Theta}\Big(\underline{d} - \widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big) \leq \sup_{\mathcal{Z}\times\Theta}\Big|d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big) - \widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big| + \sup_{\mathcal{Z}\times\Theta}\Big|d\big(\pi^*(z,\theta)\big|\, z,\theta\big) - d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big|$ –we used the fact that $\sup_{\mathcal{Z}\times\Theta}\big(\underline{d} - d\big(\pi^*(z,\theta)\big|\, z,\theta\big)\big) = 0$–. Take any $\epsilon > 0$. Then

$$\Pr\left(\inf_{\mathcal{Z}\times\Theta}\widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big) < \underline{d}-\epsilon\right) \leq \Pr\left(\sup_{\mathcal{Z}\times\Theta}\Big|d\big(\pi^*(z,\theta)\big|\, z,\theta\big) - d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big| > \epsilon/2\right)$$
$$+\Pr\left(\sup_{\mathcal{Z}\times\Theta}\Big|d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big) - \widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big| > \epsilon/2\right) \tag{A-4}$$

$\widehat{\pi^*_N}(z,\theta) \in [0,1]^2\ \forall\ \mathcal{Z}\times\Theta$. Therefore $\Pr\left(\sup_{\mathcal{Z}\times\Theta}\Big|d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big) - \widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big| > \epsilon/2\right) \leq \Pr\left(\sup_{\substack{\mathcal{Z}\times\Theta\\ \pi\in[0,1]^2}}\Big|d\big(\pi\big|\, z,\theta\big) - \widehat{d}_N\big(\pi\big|\, z,\theta\big)\Big| > \epsilon/2\right) \longrightarrow 0$, the (using Lemma 9(C)). For $\pi \in [0,1]^2$ define $\nabla_\pi d\big(\pi\big|\, z,\theta\big) = \Big(-\alpha_1\alpha_2^2\delta_1\big(\pi_2\big|\, z,\theta_1\big)\delta_2^{(1)}\big(\pi_1\big|\, z,\theta_2\big)\ ,\ -\alpha_1^2\alpha_2\delta_1^{(1)}\big(\pi_2\big|\, z,\theta_1\big)\delta_2\big(\pi_1\big|\, z,\theta_2\big)\Big)'$, where $\delta_p(\cdot) \equiv \delta_p^{(0)}(\cdot)$ and $\delta_p^{(1)}(\cdot)$ are as defined in Lemma 9(C). Using an intermediate value approximation along with assumption S1(A), there exists $\overline{M} < \infty$ such that $\sup_{\mathcal{Z}\times\Theta}\Big|d\big(\pi\big|\, z,\theta\big) - d\big(\pi'\big|\, z,\theta\big)\Big| \leq \overline{M}\cdot\big\|\pi-\pi'\big\|$ for any pair $\pi \in [0,1]^2$ and $\pi' \in [0,1]^2$. Therefore $\sup_{\mathcal{Z}\times\Theta}\Big|d\big(\pi^*(z,\theta)\big|\, z,\theta\big) - d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big| \leq \overline{M}\cdot\sup_{\mathcal{Z}\times\Theta}\Big\|\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\Big\|$, and using Lemma 10, we have $\Pr\left(\sup_{\mathcal{Z}\times\Theta}\Big|d\big(\pi^*(z,\theta)\big|\, z,\theta\big) - d\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)\Big| > \epsilon/2\right) \longrightarrow 0$. Going back to Equation A-4, these results yield $\Pr\left(\inf_{\mathcal{Z}\times\Theta}\widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big) < \underline{d}-\epsilon\right) \longrightarrow 0$ for any $\epsilon > 0$. Since $\overline{d} > 0$, this is enough to show that $\sup_{\mathcal{Z}\times\Theta}\Big|\widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)^{-1}\Big| = O_p(1)$. □

An immediate consequence of Lemma 11 is that $\sup_{\mathcal{Z}\times\Theta}\Big|\widehat{d}_N\big(\pi'_N(z,\theta)\big|\, z,\theta\big)^{-1}\Big| = O_p(1)$ for any sequence $\pi'_N(z,\theta)$ such that $\Big\|\pi'_N(z,\theta) - \pi^*(z,\theta)\Big\| \leq \Big\|\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\Big\|$. This fact will be useful in proving the next result.

Proof of Lemma 4: $\widehat{d}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big)$ is the determinant of $\widehat{J}_N\big(\widehat{\pi^*_N}(z,\theta)\big|\, z,\theta\big) = \nabla_\pi\Big(\pi - \widehat{\varphi}_N(\pi \mid z,\theta)\Big)\Big|_{\pi=\widehat{\pi^*_N}(z,\theta)}$. Lemma 11 implies that with probability approaching one uniformly in $\mathcal{Z}\times\Theta$, this Jacobian is invertible and $\widehat{\pi^*_N}(z,\theta)$ solves the first order conditions

$$\widehat{\pi^*_N}(z,\theta) - \widehat{\varphi}_N\big(\widehat{\pi^*_N}(z,\theta)\big|z,\theta\big) = 0. \tag{A-5}$$

A first order approximation yields $\widehat{J}_N\left(\pi'_N(z,\theta)\big|\, z,\theta\right)\Big(\widehat{\pi^*_N}(z,\theta)-\pi^*(z,\theta)\Big) = \widehat{\varphi}_N\left(\pi^*(z,\theta)\big|z,\theta\right)-\pi^*(z,\theta)$, where $\pi'_N(z,\theta)$ satisfies $\left\|\pi'_N(z,\theta)-\pi^*(z,\theta)\right\| \le \left\|\widehat{\pi^*_N}(z,\theta)-\pi^*(z,\theta)\right\|$. The discussion immediately after Lemma 11 implies that with probability approaching one uniformly in $\mathcal{Z}\times\Theta$, we can express $\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta) = \widehat{J}_N\left(\pi'_N(z,\theta)\big|\, z,\theta\right)^{-1}\Big(\widehat{\varphi}_N\left(\pi^*(z,\theta)\big|z,\theta\right) - \pi^*(z,\theta)\Big)$. Lemmas 9(C) and 11 yield $\sup_{\mathcal{Z}\times\Theta}\left\|\widehat{J}_N\left(\pi'_N(z,\theta)\big|\, z,\theta\right)^{-1}\right\| = O_p(1)$. Since $\pi^*(z,\theta) = \varphi\big(\pi^*(z,\theta)\big|z,\theta\big)$, we have $\sup_{\mathcal{Z}\times\Theta}\left\|\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\right\| = O_p(1)\sup_{\mathcal{Z}\times\Theta}\left\|\widehat{\varphi}_N\left(\pi^*(z,\theta)\big|z,\theta\right) - \varphi\big(\pi^*(z,\theta)\big|z,\theta\big)\right\|$. Then by Lemma 9(B): $\sup_{\mathcal{Z}\times\Theta}\left(N^{1-\delta}h^L_N\right)^{1/2}\left\|\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\right\| = O_p(1)$ for any $\delta>0$. This proves part (A) of the lemma. To prove part (B), we start once again with Equation A-5. Using the Implicit Function Theorem –which holds for Equation A-5 with probability approaching one uniformly in $\mathcal{Z}\times\Theta$ due to the result from Lemma 11– we can express $\left\|\nabla_\theta\widehat{\pi^*_N}(\theta,z) - \nabla_\theta\pi^*(\theta,z)\right\|$ and $\left\|\nabla_{\theta\theta'}\widehat{\pi^*_N}(\theta,z) - \nabla_{\theta\theta'}\pi^*(\theta,z)\right\|$ solely in terms of the objects described in Lemma 9(B)-(E) (with $m\le 2$). The arguments presented there, along with that of Lemma 11 yield the final result. □

Proof of Lemma 5: Since $Y_n \in \{0,1\}\times\{0,1\}$ w.p.1, using Theorem 2 we get $\sup_{z\in\mathcal{Z}}\left(N^{1-\delta}h^L_N\right)^{1/2}\left\|\overline{\pi}_N(z) - E\big[Y\big|\, Z=z\big]\right\| = O_p(1)$ for any $\delta>0$. If the game is in a BNE, then $E\big[Y\big|\, Z=z\big] = \pi^*(\theta_0,z)$. Unbounded support of $(\varepsilon_1,\varepsilon_2)$ implies that $\pi^*(\theta_0,z)\in(0,1)^2$ w.p.1 (i.e, $\pi^*_p(\theta_0,z) = \text{Max}\big\{0,\text{Min}\{\pi^*_p(\theta_0,z),1\}\big\}$ w.p.1). By the definition of $\overline{\overline{\pi}}_N(z)$ this yields $\sup_{z\in\mathcal{Z}}\left(N^{1-\delta}h^L_N\right)^{1/2}\left\|\overline{\overline{\pi}}_N(z) - \pi^*(\theta_0,z)\right\| = O_p(1)$ for any $\delta>0$. Given this uniform bound result, the rest of the proof proceeds by replicating the steps used to prove Lemma 4. In particular, the results of Lemma 9 hold because –as it was the case with $\widehat{\pi^*_N}(z,\theta)$– by construction, $\overline{\overline{\pi}}_N(z)\in[0,1]^2$ (a compact set) for all z, w.p.1. □

A.4 Proof of Lemma 6

Let $\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0) = L_{\mathcal{Z}}\big(W,\theta,\pi^*(\theta,Z)\big)/L_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)$. Similarly, define $\overline{\mathbb{L}}(W,\theta,\theta_0) = L\big(W,\theta,\pi^*(\theta,Z)\big)/L\big(W,\theta_0,\pi^*(\theta_0,Z)\big)$. Then $\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0) = \overline{\mathbb{L}}(W,\theta,\theta_0)$ if $Z\in\mathcal{Z}$ and $\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0) = 1$ otherwise. By Assumption S4, if $\theta\neq\theta_0$ then $\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0)$ is not constant with probability one. By construction $\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0)$ is always positive for every θ, θ_0. Therefore, Jensen's inequality yields $E\big[\log\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0)\big] < \log E\big[\overline{\mathbb{L}}_{\mathcal{Z}}(W,\theta,\theta_0)\big]$. Let w denote a particular realization of $W = (Y', X', Z')'$. If Assumptions I1-I3, S1 and S2 are satisfied, then

$$E\big[\mathbb{L}_{\mathcal{Z}}(W,\theta,\theta_0)\big] = \int_{z\in\mathcal{Z}}\int_{x\in\mathbb{S}(X)}\Big\{\sum_y L\big(w,\theta,\pi^*(\theta,z)\big)\Big\} f_{X,Z}(x,z)dxdz + \big(1-\Pr(Z\in\mathcal{Z})\big)$$
$$= \Pr(Z\in\mathcal{Z}) + (1-\Pr(Z\in\mathcal{Z})) = 1.$$

The last equality follows from the fact that $\sum_y L\big(w,\theta,\pi^*(\theta,z)\big) = \sum_y \Pr\big(Y=y\big|x,z,\theta\big) = 1$ for all (θ,x,z). Jensen's inequality therefore yields $E\big[\log\mathbb{L}_{\mathcal{Z}}(W,\theta,\theta_0)\big] < 0$ whenever $\theta\neq\theta_0$. Equivalently, $E\big[\ell_{\mathcal{Z}}(W,\theta,\pi^*(\theta,Z))\big] - E\big[\ell_{\mathcal{Z}}(W,\theta_0,\pi^*(\theta_0,Z))\big] < 0$ for all $\theta\neq\theta_0$. □

A.5 Proof of Theorem 1

We begin by characterizing an asymptotic linear representation for $\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)$.

Lemma 12 *Take any $(z,\theta)\in\mathcal{Z}\times\Theta$ and denote*

$$G\big(X,z,\theta\big) = \Big(G_1\big(X_1'\beta_1 + \alpha_1\pi_2^*(z,\theta)\big)\ ,\ G_2\big(X_2'\beta_2 + \alpha_2\pi_1^*(z,\theta)\big)\Big)'.$$

If Assumptions S1-S3 are satisfied then,

$$\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)$$
$$= J\big(\pi^*(z,\theta)\big|\ z,\theta\big)^{-1}\frac{1}{f_Z(z)}\frac{1}{Nh_N^L}\sum_{n=1}^{N}\Big(G\big(X_n,z,\theta\big) - \pi^*\big(z,\theta\big)\Big)K_h(Z_n-z) + \nu_N(z,\theta),$$

where $\sup_{\mathcal{Z}\times\Theta}\big\|\nu_N(z,\theta)\big\| = O_p\big(N^{\delta-1}h_N^{-L}\big)$ *for any* $\delta>0$.

Proof: A second order approximation of Equation (A-4) yields: $\pi^*(z,\theta) - \widehat{\varphi}_N\big(\pi^*(z,\theta)\big|\ z,\theta\big) + \widehat{J}_N\big(\pi^*(z,\theta)\big|\ z,\theta\big)\big(\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\big) + R_N(z,\theta) = 0$, where $R_N(z,\theta) = \Big(\big(\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\big)\otimes I_2\Big)\times H_N(z,\theta)\times\big(\widehat{\pi^*_N}(z,\theta) - \pi^*(z,\theta)\big)$ and

$$H_N(z,\theta) = -\frac{1}{2}\begin{pmatrix} 0 & 0 \\ -\alpha_2^2\widehat{\delta}^{(1)}_{2N}\big(\pi'_{1N}(z,\theta)\big|\ z,\theta\big) & 0 \\ 0 & -\alpha_1^2\widehat{\delta}^{(1)}_{1N}\big(\pi'_{2N}(z,\theta)\big|\ z,\theta\big) \\ 0 & 0 \end{pmatrix}$$

with $\left\|\pi^*(z,\theta) - \pi'_{_N}(z,\theta)\right\| \leq \left\|\pi^*(z,\theta) - \widehat{\pi^*_{_N}}(z,\theta)\right\|$. From Lemmas 9(C) and 4(A), we have $\sup_{\mathcal{Z}\times\Theta} \left\|R_{_N}(z,\theta)\right\| = O_p\big(N^{\delta-1}h_{_N}^{-L}\big)$ for any $\delta > 0$. Using Lemma 9(C), we obtain $\sup_{\mathcal{Z}\times\Theta} \big(N^{1-\delta}h_{_N}^L\big)^{1/2}\left\|\widehat{J}_{_N}\big(\pi^*(z,\theta)\big|\ z,\theta\big) - J\big(\pi^*(z,\theta)\big|\ z,\theta\big)\right\| = O_p(1)$ for any $\delta > 0$. By the BNE conditions, we have $\pi^*(z,\theta) - \varphi\big(\pi^*(z,\theta)\big|\ z,\theta\big) = 0$ which yields

$$0 = \varphi\big(\pi^*(z,\theta)\big|\ z,\theta\big) - \widehat{\varphi}_{_N}\big(\pi^*(z,\theta)\big|\ z,\theta\big) + J\big(\pi^*(z,\theta)\big|\ z,\theta\big)\big(\widehat{\pi^*_{_N}}(z,\theta) - \pi^*(z,\theta)\big) + \widetilde{\nu}_{_N}(z,\theta),$$

where $\sup_{\mathcal{Z}\times\Theta} \left\|\widetilde{\nu}_{_N}(z,\theta)\right\| = O_p\big(N^{\delta-1}h_{_N}^{-L}\big)$ for any $\delta > 0$. BNE uniqueness implies that $\sup_{\mathcal{Z}\times\Theta} \left\|J\big(\pi^*(z,\theta)\big|\ z,\theta\big)^{-1}\right\| = O(1)$. Using Theorem 2 and the BNE conditions we get

$$\widehat{\varphi}_{_N}\big(\pi^*(z,\theta)\big|\ z,\theta\big) - \varphi\big(\pi^*(z,\theta)\big|\ z,\theta\big) = \frac{1}{Nh_{_N}^L}\sum_{n=1}^{N}\Big(G\big(X_n,z,\theta\big) - \pi^*\big(z,\theta\big)\Big)K_h(Z_n - z) + \overline{\nu}_{_N}(z,\theta)$$

where $\sup_{\mathcal{Z}\times\Theta} \left\|\overline{\nu}_{_N}(z,\theta)\right\| = O_p\big(N^{\delta-1}h_{_N}^{-L}\big)$ for any $\delta > 0$. The final result follows if we define $\nu_{_N}(z,\theta) = \overline{\nu}_{_N}(z,\theta) + \widetilde{\nu}_{_N}(z,\theta)$. $\square$

There exists a parallel result for the linearized estimator $\widetilde{\pi^*_{_N}}(z,\theta)$ uniformly in $\mathcal{Z}$:

Lemma 13 *Let $G\big(X,z,\theta\big)$ be as defined in Lemma 12. Take any $z \in \mathcal{Z}$. If Assumptions S1-S3 are satisfied then,*

$$\begin{aligned}&\widetilde{\pi^*_{_N}}(z,\theta_0) - \pi^*(z,\theta_0)\\ &\quad = J\big(\pi^*(z,\theta_0)\big|\ z,\theta_0\big)^{-1}\frac{1}{f_Z(z)}\frac{1}{Nh_{_N}^L}\sum_{n=1}^{N}\Big(G\big(X_n,z,\theta_0\big) - \pi^*\big(z,\theta_0\big)\Big)K_h(Z_n - z) + \eta_{_N}(z),\end{aligned}$$

where $\sup_{z\in\mathcal{Z}} \left\|\eta_{_N}(z)\right\| = O_p\big(N^{\delta-1}h_{_N}^{-L}\big)$ for any $\delta > 0$.

Proof: In the proof of Lemma 5 we established –via Theorem 2– that $\overline{\overline{\pi}}_{_N}(z)$ satisfies $\sup_{z\in\mathcal{Z}} \big(N^{1-\delta}h_N^L\big)^{1/2}\left\|\overline{\overline{\pi}}_{_N}(z) - \pi^*(\theta_0,z)\right\| = O_p(1)$ for any $\delta > 0$. The rest of the proof follows essentially by replication of the steps in the proof of Lemma 12. We repeatedly rely on Lemma 5 and Theorem 2. $\square$

Lemma 14 *Take any* $(z,\theta) \in \mathcal{Z} \times \Theta$. *If Assumptions S1-S3 are satisfied, there exists* $\underset{2\times(k+2)}{\Gamma}\big(X,z,\theta\big)$ *such that*

$$\begin{aligned}\nabla_\theta \widehat{\pi_N^*}(\theta,z) - \nabla_\theta \pi^*(\theta,z) \\ = \frac{1}{N h_N^L f_Z(z)} \sum_{n=1}^{N} \Big(\Gamma\big(X_n,z,\theta\big) - E\big[\Gamma\big(X,z,\theta\big)\big|\ Z=z\big]\Big) K_h(Z_n - z) + \upsilon_N(z,\theta)\end{aligned}$$

where $\sup_{\mathcal{Z}\times\Theta} \big\|\upsilon_N(z,\theta)\big\| = O_p\big(N^{\delta-1}h_N^{-L}\big)$ *for any* $\delta > 0$.

Proof: We have $\nabla_\theta \widehat{\pi_N^*}(\theta,z) = \widehat{J}_N\left(\widehat{\pi_N^*}(\theta,z)\big|\ z,\theta\right)^{-1} \nabla_\theta \widehat{\varphi}_N\left(\widehat{\pi_N^*}(\theta,z)\big|\ z,\theta\right)$ and $\nabla_\theta \pi^*(\theta,z) = J\big(\pi^*(\theta,z)\big|\ z,\theta\big)^{-1} \nabla_\theta \varphi\big(\pi^*(\theta,z)\big|\ z,\theta\big)$. Lemmas 11 and 4(A) can be used to show that

$$\sup_{\mathcal{Z}\times\Theta} \left\| \widehat{J}_N\left(\widehat{\pi_N^*}(\theta,z)\big|\ z,\theta\right)^{-1} - J\big(\pi^*(\theta,z)\big|\ z,\theta\big)^{-1} \right\| = O_p\big(N^{\delta-1}h_N^{-L}\big)^{1/2}$$

for any $\delta > 0$. Using Lemmas 4(A) and 9, we can show that

$$\begin{aligned}\nabla_\theta \widehat{\varphi}_N\left(\widehat{\pi_N^*}(\theta,z)\big|\ z,\theta\right) - \nabla_\theta \varphi\big(\pi^*(\theta,z)\big|\ z,\theta\big) = \nabla_\theta \widehat{\varphi}_N\left(\pi^*(\theta,z)\big|\ z,\theta\right) - \nabla_\theta \varphi\big(\pi^*(\theta,z)\big|\ z,\theta\big) \\ + \nabla_\pi \text{vec}\Big(\nabla_\theta \varphi\big(\pi^*(\theta,z)\big|\ z,\theta\big)\Big)'\Big(\big(\widehat{\pi_N^*}(\theta,z) - \pi^*(\theta,z)\big) \otimes I_{(k+2)}\Big) + \overline{\upsilon}_N(z,\theta)\end{aligned}$$

where $\sup_{\mathcal{Z}\times\Theta} \big\|\overline{\upsilon}_N(z,\theta)\big\| = O_p\big(N^{\delta-1}h_N^{-L}\big)$ for any $\delta > 0$. The rest of the proof follows from the linear representations of Lemmas 9(B)-(E) (Theorem 2) and 12. $\square$

Lemma 15 *Fix* $z \in \mathcal{Z}$. *If S1-S3 are satisfied, there exists* $\underset{2\times(k+2)}{\Phi}\big(X,z,\theta_0\big)$ *such that*

$$\begin{aligned}\nabla_\theta \widetilde{\pi_N^*}(\theta_0,z) - \nabla_\theta \pi^*(\theta_0,z) \\ = \frac{1}{N h_N^L f_Z(z)} \sum_{n=1}^{N} \Big(\Phi\big(X_n,z,\theta_0\big) - E\big[\Phi\big(X,z,\theta_0\big)\big|\ Z=z\big]\Big) K_h(Z_n - z) + \varsigma_N(z)\end{aligned}$$

where $\sup_{z\in\mathcal{Z}} \big\|\varsigma_N(z)\big\| = O_p\big(N^{\delta-1}h_N^{-L}\big)$ *for any* $\delta > 0$.

Proof: $\nabla_\theta \widetilde{\pi_N^*}(\theta_0,z) = \Big(I_2 \otimes \Big(\widehat{\varphi}_N\big(\overline{\overline{\pi}}_N(z)\big|\ z,\theta_0\big) - \overline{\overline{\pi}}_N(z)\Big)\Big)' \nabla_\theta \Big(\text{vec}\Big(\widehat{J}_N\big(\overline{\overline{\pi}}_N(z)\big|\ z,\theta_0\big)^{-1}\Big)\Big) + \widehat{J}_N\big(\overline{\overline{\pi}}_N(z)\big|\ z,\theta_0\big)^{-1} \nabla_\theta \widehat{\varphi}_N\big(\overline{\overline{\pi}}_N(z)\big|\ z,\theta_0\big)^{-1}$. Using Theorem 2 we have $\overline{\pi}_N(z) - \pi^*(\theta_0,z) = \big(N h_N^L f_Z(z)\big)^{-1} \sum_{n=1}^{N} \big(Y_n - \pi^*(\theta_0,z)\big) K_h(Z_n - z) + \overline{\varsigma}_N(z)$ where $\sup_{z\in\mathcal{Z}} \big\|\overline{\varsigma}_N(z)\big\| = O_p\big(N^{\delta-1}h_N^{-L}\big)$

for any $\delta > 0$. Unbounded support of $\varepsilon_1, \varepsilon_2$ yields $\pi_p^*(\theta_0, z) = \text{Max}\big\{0, \text{Min}\{\pi_p^*(\theta_0, z), 1\}\big\}$ w.p.1, and $\overline{\overline{\pi}}_N(z)$ inherits this linear representation. Given this, the proof proceeds following the steps of the proof of Lemma 14 by invoking the linear representation results of Lemmas 9(B)-(E) (Theorem 2) and the fact that $\overline{\overline{\pi}}_N(z) \in [0,1]^2$ for all z by construction. □

A.5.1 Consistency

Let $U = \sup_{\substack{\theta\in\Theta \\ \pi\in[0,1]^2}} \Big|\ell_{\mathcal{Z}}\big(W, \theta, \pi\big)\Big|$ and $V = \sup_{\substack{\theta\in\Theta \\ \pi\in[0,1]^2}} \Big\|\nabla_\pi \ell_{\mathcal{Z}}\big(W, \theta, \pi\big)\Big\|$. Assumption S5(B) implies that $E[V] < \infty$. Combined with Lemma 4(A), this yields

$$\sup_{\theta\in\Theta}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(W_n, \theta, \widehat{\pi_N^*}(\theta, Z_n)\big) - \frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(W_n, \theta, \pi^*(\theta, Z_n)\big)\right| \xrightarrow{p} 0$$

By S5(B) we also have $E[U] < \infty$. Lemma 2.4 in Newey and McFadden (1994) yields

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

From Lemma 6 we know that $E\Big[\ell_{\mathcal{Z}}\big(W, \theta, \pi^*(\theta, Z)\big)\Big]$ is uniquely maximized at $\theta = \theta_0$. Given the results above, consistency of $\widehat{\theta}$ follows from Theorem 2.1 in Newey and McFadden. The proof of Consistency of $\widetilde{\theta}$ follows the exact same steps using Lemma 5(A) and the following result, which is equivalent to that of Lemma 6: Let $\rho(\theta, z)$ be as defined in Subsection 3.1.2. If Assumptions I1-I3, S1, S2 and S4 are satisfied, then for all $\theta \neq \theta_0,\ \theta \in \Theta$,

$$E\big[\ell_{\mathcal{Z}}\big(W, \theta, \rho(\theta, Z)\big)\big] < E\big[\ell_{\mathcal{Z}}\big(W, \theta_0, \rho(\theta_0, Z)\big)\big] = E\big[\ell_{\mathcal{Z}}\big(W, \theta_0, \pi^*(\theta_0, Z)\big)\big].$$

recall that $\rho(\theta_0, Z) = \pi^*(\theta_0, Z)$. If S4 is satisfied, then conditional on $Z \in \mathcal{Z}$, if $\theta \neq \theta_0$ then $\Pr\Big\{\beta_p' X_p + \alpha_p \rho_{\text{-}p}(\theta, Z) \neq \beta_{p0}' X_p + \alpha_{p0}\pi_{\text{-}p}^*(\theta_0, Z)\Big\} > 0$ for $p = 1, 2$. The rest of the proof is identical to that of Lemma 6 since $L_{\mathcal{Z}}\big(W, \theta_0, \rho(\theta_0, Z)\big) = L_{\mathcal{Z}}\big(W, \theta_0, \pi^*(\theta_0, Z)\big)$. □

A.5.2 Asymptotic normality

First, we present simplified expressions for the objects involved in the asymptotic distribution of $\widehat{\theta}$, $\widetilde{\theta}$. Let $\nabla_\theta \ell_{\mathcal{Z}}$ and $\nabla_\pi \ell_{\mathcal{Z}}$ denote the partial derivatives of $\ell_{\mathcal{Z}}$ w.r.t θ and π respectively, with all other arguments fixed. Let $\nabla_{\theta\theta'}\ell_{\mathcal{Z}}$, $\nabla_{\pi\pi'}\ell_{\mathcal{Z}}$ and $\nabla_{\theta\pi'}\ell_{\mathcal{Z}}$ be the second-derivatives.

Then,[13]

$$E\left[\frac{\partial^2 \ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)}{\partial\theta\partial\theta'}\right] = E\Big[\nabla_{\theta\theta'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big) + \nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)\nabla_{\theta}\pi^*(\theta_0,Z)$$
$$+ \nabla_{\theta}\pi^*(\theta_0,Z)'\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)' + \nabla_{\theta}\pi^*(\theta_0,Z)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)\nabla_{\theta}\pi^*(\theta_0,Z)\Big].$$

Recall that $\frac{\partial \ell_{\mathcal{Z}}(w,\theta,\pi^*(\theta,z))}{\partial\theta} = \nabla_{\theta}\ell_{\mathcal{Z}}\big(w,\theta,\pi^*(\theta,z)\big) + \nabla_{\theta}\pi^*(\theta,z)'\nabla_{\pi}\ell_{\mathcal{Z}}\big(w,\theta,\pi^*(\theta,z)\big)$. It is easy to show that

$$E\Big[\nabla_{\Box\Diamond'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)\Big|X,Z\Big] = -E\Big[\nabla_{\Box}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)\nabla_{\Diamond}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)'\Big|X,Z\Big]$$

for $\Box,\Diamond \in \{\theta,\pi\}$.[14] From these facts it is easy to show that $\ell_{\mathcal{Z}}$ satisfies an information identity result:

$$E\left[\frac{\partial^2 \ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)}{\partial\theta\partial\theta'}\right] = -E\left[\frac{\partial \ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)}{\partial\theta} \times \frac{\partial \ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)}{\partial\theta}'\right],$$

Let $\overline{D}_{\mathcal{Z}}(Z)$ be as defined in Theorem 1. Using iterated expectations we have

$$\overline{D}_{\mathcal{Z}}(Z) = E\Big[\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big) + \nabla_{\theta}\pi^*(\theta_0,Z)'\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big)\ \Big|Z\Big].$$

From now on and until we explicitly say otherwise, we will use lowercase letters to denote random variables. Since $\widehat{\theta} \in \Theta$ and $\theta_0 \in \text{int}\,\Theta$, Lemma 11 implies that with probability approaching one, the estimator $\widehat{\theta}$ satisfies the first order conditions

$$\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\theta}\ell_{\mathcal{Z}}\big(w_n,\widehat{\theta},\widehat{\pi^*_N}(\widehat{\theta},z_n)\big) + \nabla_{\theta}\widehat{\pi^*_N}(\widehat{\theta},z_n)'\nabla_{\pi}\ell_{\mathcal{Z}}\big(w_n,\widehat{\theta},\widehat{\pi^*_N}(\widehat{\theta},z_n)\big)\Big] = 0 \qquad \text{(A-6)}$$

Once again, Lemma 11 and the fact that $\widehat{\theta} \in \Theta$ imply that with probability approaching one, the following approximation to (A-6) is valid:

$$\widehat{\theta} - \theta_0 = -\left(\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2 \ell_{\mathcal{Z}}\big(w_n,\overline{\theta},\widehat{\pi^*_N}(\overline{\theta},z_n)\big)}{\partial\theta\partial\theta'}\right)^{-1}$$
$$\times\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\theta}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\widehat{\pi^*_N}(\theta_0,z_n)\big) + \nabla_{\theta}\widehat{\pi^*_N}(\theta_0,z_n)'\nabla_{\pi}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\widehat{\pi^*_N}(\theta_0,z_n)\big)\Big] \qquad \text{(A-7)}$$

[13] This expression does not depend on $\nabla_{\theta\theta'}\pi^*(\theta_0,Z)$ because $E\big[\nabla_{\pi}\ell_{\mathcal{Z}}\big(W,\theta_0,\pi^*(\theta_0,Z)\big) \mid X,Z\big] = 0$.

[14] Notice that a key to this result is the fact that $\mathbb{1}\{Z\in\mathcal{Z}\}^2 = \mathbb{1}\{Z\in\mathcal{Z}\}$.

where $\|\overline{\theta}-\theta_0\| \leq \|\widehat{\theta}-\theta_0\|$ and

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2 \ell_{\mathcal{Z}}\big(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n)\big)}{\partial\theta\partial\theta'} = \frac{1}{N}\sum_{n=1}^{N}\Bigg[\nabla_{\theta\theta'}\ell_{\mathcal{Z}}\big(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n)\big)$$
$$+\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n)\big)\nabla_{\theta}\widehat{\pi^*_{N}}(\theta,z_n)+\nabla_{\theta\theta'}\widehat{\pi^*_{N}}(\theta,z_n)'\big(\nabla_{\pi}\ell_{\mathcal{Z}}\big(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n)\big)\otimes I_{(k+2)}\big)$$
$$+\nabla_{\theta}\widehat{\pi^*_{N}}(\theta,z_n)'\Big(\nabla_{\pi\theta'}\ell_{\mathcal{Z}}\big(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n)\big)+\nabla_{\pi\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n)\big)\nabla_{\theta}\widehat{\pi^*_{N}}(\theta,z_n)\Big)\Bigg]$$

By S5(B) and Lemma 4, $\sup_{\theta\in\Theta}\left\|\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\mathcal{Z}}(w_n,\theta,\widehat{\pi^*_{N}}(\theta,z_n))}{\partial\theta\partial\theta'}-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\mathcal{Z}}(w_n,\theta,\pi^*(\theta,z_n))}{\partial\theta\partial\theta'}\right\| \xrightarrow{p} 0.$ Assumptions S5(A),(C) and the consistency of $\widehat{\theta}$ yields (via a dominated convergence argument)

$$-\left(\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\mathcal{Z}}\big(w_n,\overline{\theta},\widehat{\pi^*_{N}}(\overline{\theta},z_n)\big)}{\partial\theta\partial\theta'}\right)^{-1}\xrightarrow{p}\Im_{\mathcal{Z}}^{-1}. \tag{A-8}$$

Using Lemma 12 the following approximation is valid

$$\frac{1}{N}\sum_{n=1}^{N}\nabla_{\theta}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\widehat{\pi^*_{N}}(\theta_0,z_n)\big)=$$
$$\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\theta}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)+\frac{1}{N}\sum_{n=1}^{N}\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)\nu_{N}(z_n,\theta)\Big]$$
$$+\frac{1}{N^2h_{N}^{L}}\sum_{n=1}^{N}\sum_{m=1}^{N}\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)J_0(z_n)^{-1}}{f_Z(z_n)}\Big(G\big(x_m,z_n,\theta_0\big)-\pi^*\big(z_n,\theta_0\big)\Big)K_h(z_m-z_n)$$

We have $\big(N^2h_{N}^{L}\big)^{-1}\sum_{n=1}^{N}\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)J_0(z_n)^{-1}}{f_Z(z_n)}\Big(G\big(x_n,z_n,\theta\big)-\pi^*\big(z_n,\theta\big)\Big)=O_p(N^{\delta-1}h_{N}^{L})$ and $N^{-1}\sum_{n=1}^{N}\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)\nu_{N}(z_n,\theta)=O_p(N^{\delta-1}h_{N}^{L})$ for any $\delta>0$, where the last line follows from Assumption S5(B) and Lemma 12. If a sequence ξ_{N} satisfies $\xi_{N}=O_p(N^{\delta-1}h_{N}^{L})$ for all $\delta>0$, then using Assumption S3(B) we must have $\xi_{N}=o_p\big(N^{-1/2}\big)$ by letting $\delta=\sigma/2$. Therefore, Assumption S3(B) yields

$$\frac{1}{N}\sum_{n=1}^{N}\nabla_{\theta}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\widehat{\pi^*_{N}}(\theta_0,z_n)\big)$$
$$=\frac{1}{N}\sum_{n=1}^{N}\nabla_{\theta}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)+\binom{N}{2}^{-1}\sum_{m<n}T_{N}\big(w_n,w_m\big)+o_p\big(N^{-1/2}\big)$$

where

$$T_{{}_N}\big(w_n, w_m\big) = \frac{N-1}{2N}\Bigg[\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)J_0(z_n)^{-1}}{h_{{}_N}^L f_{{}_Z}(z_n)}\Big(G\big(x_m, z_n, \theta_0\big) - \pi^*\big(z_n, \theta_0\big)\Big)$$

$$+ \frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_m, \theta_0, \pi^*(\theta_0, z_m)\big)J_0(z_m)^{-1}}{h_{{}_N}^L f_{{}_Z}(z_m)}\Big(G\big(x_n, z_m, \theta_0\big) - \pi^*\big(z_m, \theta_0\big)\Big)\Bigg] \times K_h(z_m - z_n).$$

Let $\overline{G}\big(z_n, \theta_0\big|\ z_m\big) = E\Big[G\big(x_m, z_n, \theta_0\big)\Big|\ z_n, z_m\Big]$. Then

$$E\Bigg[\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)J_0(z_n)^{-1}}{h_{{}_N}^L f_{{}_Z}(z_n)}\Big(G\big(x_m, z_n, \theta_0\big) - \pi^*\big(z_n, \theta_0\big)\Big)K_h(z_m - z_n)\Bigg|\ w_n\Bigg] =$$

$$\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)J_0(z_n)^{-1}}{f_{{}_Z}(z_n)}\int\Big(\overline{G}\big(z_n, \theta_0\big|z_n + h_{{}_N}\psi\big) - \pi^*(z_n, \theta_0)\Big)K(\psi)f_{{}_Z}(z_n + h_{{}_N}\psi)d\psi$$

By Assumptions S1 and S2(B), $\overline{G}\big(u, \theta_0\big|\ v\big)f_{{}_Z}(v)$ is M-times differentiable with respect to v with bounded M derivatives for all $u, v \in \mathbb{S}(Z)$. Assumption S3 and an Mth order Taylor expansion yields $\int \overline{G}\big(z_n, \theta_0\big|z_n + h_{{}_N}\psi\big)f_{{}_Z}(z_n + h_{{}_N}\psi)K(\psi)d\psi = \overline{G}\big(z_n, \theta_0\big|z_n\big)f_{{}_Z}(z_n) + \zeta'_n$ and $\int K(\psi)f_{{}_Z}(z_n + h_{{}_N}\psi)d\psi = f_{{}_Z}(z_n) + \zeta''_n$, where $\sup_n \|\zeta'_n\| = o_p\big(N^{-1/2}\big)$ and $\sup_n \|\zeta''_n\| = o_p\big(N^{-1/2}\big)$. BNE conditions imply $\overline{G}\big(v, \theta_0\big|v\big) = \pi^*(v, \theta_0)$ for all $v \in \mathcal{Z}$. These results combined yield $\int\Big(\overline{G}\big(z_n, \theta_0\big|z_n + h_{{}_N}\psi\big) - \pi^*(z_n, \theta_0)\Big)K(\psi)f_{{}_Z}(z_n + h_{{}_N}\psi)d\psi = \zeta_n$, where $\sup_n \|\zeta_n\| = o_p\big(N^{-1/2}\big)$. We move to the next term:

$$E\Bigg[\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_m, \theta_0, \pi^*(\theta_0, z_m)\big)J_0(z_m)^{-1}}{h_{{}_N}^L f_{{}_Z}(z_m)}\Big(G\big(x_n, z_m, \theta_0\big) - \pi^*\big(z_m, \theta_0\big)\Big)K_h(z_n - z_m)\Bigg|\ w_n\Bigg]$$

$$= \int B_{\mathcal{Z}}(z_n + h_{{}_N}\psi)\Big(G\big(x_n, z_n + h_{{}_N}\psi, \theta_0\big) - \pi^*(z_n + h_{{}_N}\psi, \theta_0)\Big)K(\psi)d\psi,$$

where $B_{\mathcal{Z}}(\cdot)$ is as defined in Theorem 1. By Assumptions S1(A)-(C) and S2(B) and the Implicit Function Theorem, $\pi^*(v, \theta_0)$ and $G\big(x_n, v, \theta_0\big)$ are M-times differentiable with bounded M derivatives for all $v \in \mathcal{Z}$. Since Z is absolutely continuous with respect to Lebesgue measure, this is also true w.p.1 for $B_{\mathcal{Z}}(v)$. Assumption S3 and an Mth order approximation yield $\int B_{\mathcal{Z}}(z_n + h_{{}_N}\psi)\Big(G\big(x_n, z_n + h_{{}_N}\psi, \theta_0\big) - \pi^*(z_n + h_{{}_N}\psi, \theta_0)\Big)K(\psi)d\psi = B_{\mathcal{Z}}(z_n)\Big(G\big(x_n, z_n, \theta_0\big) - \pi^*(z_n, \theta_0)\Big) + \upsilon_n$, where $\sup_n \|\upsilon_n\| = o_p\big(N^{-1/2}\big)$. Letting ζ_n be as described above, these results combined imply that

$$\frac{1}{N}\sum_{n=1}^{N} E\big[T_{_N}(w_n,w_m)\big|w_n\big] = \left(\frac{N-1}{2N}\right)\times\left[\frac{1}{N}\sum_{n=1}^{N}\left(\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)J_0(z_n)^{-1}}{f_{_Z}(z_n)}\right)\zeta_n + \frac{1}{N}\sum_{n=1}^{N}\Big[B_{\mathcal{Z}}(z_n)\Big(G\big(x_n,z_n,\theta_0\big)-\pi^*(z_n,\theta_0)\Big)+\upsilon_n\Big]\right].$$

By Assumption S2(B), there exists $\overline{C}$ such that $\sup_{v\in\mathcal{Z}}\left\|\Big(f_{_Z}(v)J_0(v)\Big)^{-1}\right\|\le\overline{C}$ and therefore

$$\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)J_0(z_n)^{-1}}{f_{_Z}(z_n)}\right\|\cdot\sup_n\big\|\zeta_n\big\|\le o_p\big(N^{-1/2}\big)\cdot O_p(1)=o_p\big(N^{-1/2}\big).$$

Similarly, $\frac{1}{N}\sum_{n=1}^{N}\|\upsilon_n\| = o_p\big(N^{-1/2}\big)$ and we obtain

$$\frac{1}{N}\sum_{n=1}^{N}E\big[T_{_N}(w_n,w_m)\big|w_n\big]=\frac{1}{2N}\sum_{n=1}^{N}B_{\mathcal{Z}}(z_n)\Big(G\big(x_n,z_n,\theta_0\big)-\pi^*(z_n,\theta_0)\Big)+o_p\big(N^{-1/2}\big)$$

Then, by Assumptions S2(B) and S3(A)-(B) and S5(B), there exist $\overline{C}$, $\overline{K}$ such that

$$E\left[\left\|\frac{\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)J_0(z_n)^{-1}}{h_{_N}^L f_{_Z}(z_n)}\Big(G\big(x_m,z_n,\theta_0\big)-\pi^*\big(z_n,\theta_0\big)\Big)K_h(z_m-z_n)\right\|^2\right]$$
$$\le\left(\frac{\overline{C}\cdot\overline{K}}{h_{_N}^L}\right)^2\times E\left[\left\|\nabla_{\theta\pi'}\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)\Big(G\big(x_m,z_n,\theta_0\big)-\pi^*\big(z_n,\theta_0\big)\Big)\right\|^2\right]=o(N).$$

Therefore, Lemma A.3 in Ahn and Powell (1993) yields[15]

$$\binom{N}{2}^{-1}\sum_{m<n}T_{_N}\big(w_n,w_m\big)=\frac{1}{N}\sum_{n=1}^{N}B_{\mathcal{Z}}(z_n)\Big(G\big(x_n,z_n,\theta_0\big)-\pi^*(z_n,\theta_0)\Big)+o_p\big(N^{-1/2}\big)$$

and consequently,

$$\frac{1}{N}\sum_{n=1}^{N}\nabla_\theta\ell_{\mathcal{Z}}\big(w_n,\theta_0,\widehat{\pi_{_N}^*}(\theta_0,z_n)\big)$$
$$=\frac{1}{N}\sum_{n=1}^{N}\nabla_\theta\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)+\frac{1}{N}\sum_{n=1}^{N}B_{\mathcal{Z}}(z_n)\Big(G\big(x_n,z_n,\theta_0\big)-\pi^*(z_n,\theta_0)\Big)+o_p\big(N^{-1/2}\big). \tag{A-9}$$

Using Lemmas 4(A)-(B) and 14 and Assumptions S3(B) and S5(B), we have

$$\frac{1}{N}\sum_{n=1}^{N}\nabla_\theta\widehat{\pi_{_N}^*}(\theta_0,z_n)'\nabla_\pi\ell_{\mathcal{Z}}\big(w_n,\theta_0,\widehat{\pi_{_N}^*}(\theta_0,z_n)\big)=\frac{1}{N}\sum_{n=1}^{N}\nabla_\theta\pi^*(\theta_0,z_n)'\nabla_\pi\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)$$
$$+\sum_{n=1}^{N}\sum_{m=1}^{N}\frac{\Big(\Gamma\big(x_m,z_n,\theta_0\big)-E\big[\Gamma\big(x,z,\theta_0\big)\big|\ z_n\big]\Big)'}{N^2h_{_N}^L\ f_{_Z}(z_n)}K_h(z_m-z_n)\nabla_\pi\ell_{\mathcal{Z}}\big(w_n,\theta_0,\pi^*(\theta_0,z_n)\big)+o_p(N^{-1/2})$$

[15] $T_{_N}\big(w_n,w_m\big)$ can also be shown to be Euclidean and satisfy Theorem 3 in Sherman (1994).

$f_Z(z)^{-1}$ is uniformly bounded in $\mathcal{Z}$, and $E\Big[\nabla_\pi \ell_{\mathcal{Z}}\big(w_m, \theta_0, \pi^*(\theta_0, z_m)\big)\Big|\, x_m, z_m\Big] = 0$. Then,

$$E\left[\frac{\Big(\Gamma\big(x_n, z_m, \theta_0\big) - E\big[\Gamma\big(x, z, \theta_0\big)\big|\, z_m\big]\Big)'}{h_N^L\ f_Z(z_m)} K_h(z_n - z_m)\nabla_\pi \ell_{\mathcal{Z}}\big(w_m, \theta_0, \pi^*(\theta_0, z_m)\big)\Bigg|\, w_n, x_m, z_m\right] = 0$$

Assumption S2(B) yields a uniform bound $\sup_{v\in\mathcal{Z}} \left\|\Big(f_Z(v)J_0(v)\Big)^{-1}\right\| \le \overline{C}$. Assumption S1(A)-(C) and an M^{th} order approximation yields

$$E\left[\frac{\big(\Gamma\big(x_m, z_n, \theta_0\big) - E\big[\Gamma\big(x, z, \theta_0\big)\big|\, z_n\big]\big)'}{h_N^L f_Z(z_n)} K_h(z_m - z_n)\Bigg|\, w_n\right] = o_p(N^{-1/2})$$

where the last equality follows Assumption S3. Using Assumptions S2(B), S3(A)-(B) and S5(B) we have

$$E\left[\left\|\frac{\Big(\Gamma\big(x_m, z_n, \theta_0\big) - E\big[\Gamma\big(x, z, \theta_0\big)\big|\, z_n\big]\Big)'}{h_N^L\ f_Z(z_n)} K_h(z_m - z_n)\nabla_\pi \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)\right\|^2\right] = o(N).$$

Using Lemma A.3 in Ahn and Powell, we can show that these results combined yield

$$\sum_{n=1}^{N}\sum_{m=1}^{N} \frac{\Big(\Gamma\big(x_m, z_n, \theta_0\big) - E\big[\Gamma\big(x, z, \theta_0\big)\big|\, z_n\big]\Big)'}{N^2 h_N^L\ f_Z(z_n)} K_h(z_m - z_n)\nabla_\pi \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big) = o_p(N^{-1/2})$$

and consequently,

$$\frac{1}{N}\sum_{n=1}^{N} \nabla_\theta \widehat{\pi_N^*}(\theta_0, z_n)' \nabla_\pi \ell_{\mathcal{Z}}\big(w_n, \theta_0, \widehat{\pi_N^*}(\theta_0, z_n)\big)$$
$$= \frac{1}{N}\sum_{n=1}^{N} \nabla_\theta \pi^*(\theta_0, z_n)' \nabla_\pi \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big) + o_p(N^{-1/2}). \tag{A-10}$$

Recall that $\frac{\partial \ell_{\mathcal{Z}}(w_n, \theta_0, \pi^*(\theta_0, z_n))}{\partial \theta} = \nabla_\theta \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big) + \nabla_\theta \pi^*(\theta_0, z_n)' \nabla_\pi \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)$. Combining this with (A-8), (A-9) and (A-10), Equation (A-7) becomes

$$\widehat{\theta} - \theta_0 =$$
$$\Im_{\mathcal{Z}}^{-1} \times \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\partial \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)}{\partial \theta} + B_{\mathcal{Z}}(z_n)\Big(G\big(x_n, z_n, \theta_0\big) - \pi^*(z_n, \theta_0)\Big)\right] + o_p(N^{-1/2})$$
$$= \Im_{\mathcal{Z}}^{-1} \times \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\partial \ell_{\mathcal{Z}}\big(w_n, \theta_0, \pi^*(\theta_0, z_n)\big)}{\partial \theta} + B_{\mathcal{Z}}(z_n)\Big(E\big[y_n\big|\, x_n, z_n\big] - E\big[y_n\big|\, z_n\big]\Big)\right] + o_p(N^{-1/2})$$

(A-7')

The first result in Theorem 1 follows from this and Assumption S5(A). To prove the second result in Theorem 1 we start with the equivalent to (A-7), which is satisfied with probability approaching one by $\widetilde{\theta}$. The equivalent of (A-8) follows from Lemma 5,[16] Assumption S5 and the consistency of $\widetilde{\theta}$. The approximations (A-9) and (A-10) follow from Lemmas 13 and 15, and Assumptions S2(B), S3(A)-(B) and S5(B) [17]. The linear representation $(A-7')$ is therefore satisfied by $\widetilde{\theta}-\theta_0$. This concludes the proof of Theorem 1. □

A.6 Proof of Lemma 7

Let $\mathbb{I}_{\mathcal{Z}}(z_n) = \mathbb{1}\{z_n \in \mathcal{Z}\}$ and define $T_{1_N}(w_n, w_m) = \mathbb{I}_{\mathcal{Z}}(z_n) f_Z(z_n)^{-1} J_0(z_n)^{-1}\big(E[y|x_m, z_n] - E[y|z_n]\big) K_h(z_m - z_n)$ and $T_N(w_n, w_m) = h_N^{-L}\big(T_{1_N}(w_n, w_m) + T_{1_N}(w_m, w_n)\big)$. Let $R(w_n, w_\ell) = \mathbb{I}_{\mathcal{Z}}(z_n)\nabla_\theta E[y|z_n]\psi_m + \mathbb{I}_{\mathcal{Z}}(z_m)\nabla_\theta E[y|z_m]\psi_n$, $A_{1_N}(w_n, w_m) = \mathbb{I}_{\mathcal{Z}}(z_n) f_Z(z_n)^{-1}\Big(\big(E[y|x_m, z_n] - E[y|z_n]\big)' K_h(z_m - z_n) \otimes I_2\Big)\frac{\partial}{\partial\theta}\big[J_0(z_n)^{-1}\big]$, $A_{2_N}(w_n, w_m) = \mathbb{I}_{\mathcal{Z}}(z_n) J_0(z_n)^{-1}\frac{\partial}{\partial\theta}\Big(E[y|x_m, z_n] - E[y|z_n]\Big) K_h(z_m - z_n)$ and $A_N(w_n, w_m) = A_{1_N}(w_m, w_n) + A_{2_N}(w_m, w_n)$. Take $i, j, k \in \mathbb{N}$ and let $C_2(i,j) = \big\{(i,j,j), (j,j,i), (j,i,j), (j,i,i), (i,i,j), (i,j,i)\big\}$ and $C_3(i,j,k) = \big\{(i,j,k), (i,k,j), (j,i,k), (j,k,i), (k,i,j), (k,j,i)\big\}$. Let ψ be the influence function of Theorem 1 and define $S_{1_N}(w_n, w_m) = h_N^{-L}\sum_{i,j,k} A_N(w_i, w_j)\psi_k \;:\; (i,j,k) \in C_2(n,m)$ and $S_{2_N}(w_n, w_m, w_\ell) = h_N^{-L}\sum_{i,j,k} A_N(w_i, w_j)\psi_k \;:\; (i,j,k) \in C_3(n,m,\ell)$. Using assumptions S1(A),(D), S2(B), S3(A),(B), Theorem 1, Lemma 12, and the fact that $G(x, z, \theta_0) = E[y|x,z]$ and $\pi^*(z, \theta_0) = E[y|z]$ we get $N^{-1}\sum_n \mathbb{I}_{\mathcal{Z}}(z_n)\widehat{\pi^*_N}(\widehat{\theta}, z_n) = N^{-1}\sum_n \mathbb{I}_{\mathcal{Z}}(z_n)\pi^*(\theta_0, z_n) + N^{-2}\sum_{n<m} T_N(w_n, w_m) + N^{-2}\sum_{n<m} R(w_n, w_m) + N^{-3}\sum_{n<m} S_{1_N}(w_n, w_m) + N^{-3}\sum_{n<m<\ell} S_{2_N}(w_n, w_m, w_\ell) + o_p(N^{-1/2})$. Assumption S3(A),(B), Theorem 1 and the fact that $E[y|x,z], E[y|z] \in [0,1]^2$ yields $N^{-1}\sum_n \mathbb{I}_{\mathcal{Z}}(z_n)\overline{\pi}_N(z_n) = N^{-1}\sum_n \mathbb{I}_{\mathcal{Z}}(z_n)\pi^*(\theta_0, z_n) + N^{-2}\sum_{n<m} U_N(w_n, w_m) + o_p(N^{-1/2})$, where $U_N(w_n, w_m) = h_N^{-L}\Big[\mathbb{I}_{\mathcal{Z}}(z_n) f_Z(z_n)^{-1}\big(y_m - E[y|z_n]\big) + \mathbb{I}_{\mathcal{Z}}(z_m) f_Z(z_m)^{-1}\big(y_n - E[y|z_m]\big)\Big] \times K_h(z_n - z_m)$.

By assumptions S1, S2 and S3, $N^{-3}\sum_{n<m} S_{1_N}(w_n, w_m) = o_p(N^{-1/2})$. Ignore the terms involving $K_h(\cdot)$ in T_N, S_{2_N} and U_N. By Assumptions S1(A)-(C), S2(B) and absolute

[16] See footnote 13 to see why the difference between $\nabla_{\theta\theta'}\pi^*(\theta_0, Z)$ and $\nabla_{\theta\theta'}\rho(\theta_0, Z)$ has no effect –to first order of approximation– on the asymptotic distributions of $\sqrt{N}\big(\widehat{\theta} - \theta_0\big)$ and $\sqrt{N}\big(\widetilde{\theta} - \theta_0\big)$.

[17] Recall that $\nabla_\theta\rho(\theta_0, Z) = \nabla_\theta\pi^*(\theta_0, Z)$ for all $Z \in \mathcal{Z}$.

continuity of Z w.r.t Lebesgue measure, the remaining terms in these objects are M-times differentiable with probability one w.r.t z with bounded M derivatives. Since $f_{X,Z}(\cdot)$ does not depend on θ, we have $E\Big[\frac{\partial}{\partial\theta}G(x,z,\theta)\Big|z\Big] = \frac{\partial}{\partial\theta}E\big[G(x,z,\theta)\big|z\big]\ \forall\ \theta$. Using all these facts, an Mth-order approximation and Assumption S3 yield $E\big[T_N(w_n,w_m)\big|w_n\big] = \mathbb{I}_{\mathcal{Z}}(z_n)f_Z(z_n)^{-1}J_0(z_n)^{-1}\big(E[y|x_n,z_n]-E[y|z_n]\big)+o_p(N^{-1/2})$, $E\big[S_{2_N}(w_n,w_m,w_\ell)\big|w_n\big] = o_p(N^{-1/2})$ and $E\big[U_N(w_n,w_m)\big|w_n\big] = \mathbb{I}_{\mathcal{Z}}(z_n)f_Z(z_n)^{-1}\big(y_n - E[y|z_n]\big) + o_p(N^{-1/2})$. It is easy to see that $E\big[R_N(w_n,w_m,w_\ell)\big|w_n\big] = E\big[\mathbb{I}_{\mathcal{Z}}(z)\nabla_\theta E[y|z]\big]\psi_n$. Assumptions S1, S2 and S3 yield $E\big[\big\|T_N(w_n,w_n,w_n)\big\|^2\big] = o(N)$, $E\big[\big\|R(w_n,w_n,w_n)\big\|^2\big] = o(N)$, $E\big[\big\|S_{2_N}(w_n,w_n,w_n)\big\|^2\big] = o(N)$ and $E\big[\big\|U_N(w_n,w_n,w_n)\big\|^2\big] = o(N)$. From Lemma A.3 in Ahn and Powell (1993) we get $N^{-2}\sum_{n<m}T_N(w_n,w_m) = N^{-1}\sum_{n=1}^N\mathbb{I}_{\mathcal{Z}}(z_n)f_Z(z_n)^{-1}J_0(z_n)^{-1}\big(E[y|x_n,z_n] - E[y|z_n]\big) + o_p(N^{-1/2})$, $N^{-2}\sum_{n<m}R_N(w_n,w_m) = N^{-1}\sum_{n=1}^N E\big[\mathbb{I}_{\mathcal{Z}}(z)\nabla_\theta E[y|z]\big]\psi_n + o_p(N^{-1/2})$ and we also have $N^{-3}\sum_{n<m<\ell}S_{2_N}(w_n,w_m,w_\ell) = o_p(N^{-1/2})$ and finally, $N^{-2}\sum_{n<m}U_N(w_n,w_m) = N^{-1}\sum_{n=1}^N\mathbb{I}_{\mathcal{Z}}(z_n)f_Z(z_n)^{-1}\big(y_n - E[y|z_n]\big) + o_p(N^{-1/2})$. This concludes the proof. □

A.7 Proof of Lemma 8

The first step of the proof is to show that the results of Theorem 2 still apply for $\mathcal{Z}_N$. If the assumptions of Lemma 8 are satisfied, we can show that the right-hand side of Equation A-3 converges in probability to zero when we replace z^* with $\overline{z}_N$. Once this is established, Step 2 in the proof of Theorem 2 holds for $\mathcal{Z}_N$. The remaining steps in the proof of such theorem follow from this and the fact that $f_Z(z_n) > b_N$ for all N and $n \leq N$ and $N^{1-2\sigma}h_N^{2L}b_N^2 \to \infty$. The next step of the proof of Lemma 8 relies on the fact that its assumptions are compatible with those of Lemma 26 in Ichimura (2004). Letting $\widehat{I}_N(z) = \mathbb{1}\{\widehat{f}_{Z_N}(z) > b_N\}$ and $I_N(z) = \mathbb{1}\{f_Z(z) > b_N\}$, this means that $\Pr\big(\widehat{I}_N(z_n) \neq I_N(z_n)$ for at least one $n\big) \to 0$. The proof of Theorem 1 proceeds with $\mathbb{1}\{z \in \mathcal{Z}\}$ replaced with $I_N(z)$. The final result of Lemma 8 follows from a dominating convergence argument. □

A.8 Equilibrium conditions of game in Section 4

Fix $\pi \in [0,1]$ and for $p = 1,2$ define $G_p^A(u \mid \pi) = \Pr\big(\pi\varepsilon_{t_p} + (1-\pi)\varepsilon_{r_p} < u\big)$, $G_p^B(u \mid \pi) = \Pr\big(\pi\varepsilon_{s_p} + (1-\pi)\varepsilon_{q_p} < u\big)$, $G_p^{A,B}(u,v \mid \pi) = \Pr\big(\pi\varepsilon_{t_p} + (1-\pi)\varepsilon_{r_p} < u\ ,\ \pi\varepsilon_{s_p} + (1-\pi)\varepsilon_{q_p} < v\big)$,

$G_p^{A,B}(u,v \mid \pi) = \Pr\big(\pi\varepsilon_{t_p} + (1-\pi)\varepsilon_{r_p} < u \ , \pi\varepsilon_{s_p} + (1-\pi)\varepsilon_{q_p} < v\big)$. Take $z \in \mathbb{S}(Z)$ and define the following equilibrium probabilities for $p = 1,2$: $\pi^*_{c_p}(z,\theta) = \Pr\big(C_p = 1 \mid Z = z;\theta\big)$ and $\pi^*_{y_p,c_p}(z,\theta) = \Pr\big(Y_p = 1, C_p = 1 \big| \ Z = z;\theta\big)$. The probability of the third available strategy is immediately determined. Now define $\beta_{h-j} = \beta_h - \beta_j$, $\gamma_p = \alpha_p - \delta_p$ and let

$$\begin{aligned}\mathcal{P}_p^1(x_p,\pi_a,\pi_b,\theta) = \ & G_p^A\Big(\beta'_{r_p} x_p + \beta'_{t_p\text{-}r_p} x_p\pi_a + \alpha_p\pi_a\pi_b \ \Big|\pi_a\Big) + \ G_p^B\Big(\beta'_{q_p} x_p + \beta'_{s_p\text{-}q_p} x_p\pi_a + \delta_p\pi_a\pi_b \ \Big|\pi_a\Big) \\ & - G_p^{A,B}\Big(\beta'_{r_p} x_p + \beta'_{t_p\text{-}r_p} x_p\pi_a + \ \alpha_p\pi_a\pi_b \ , \beta'_{q_p} x_p + \beta'_{s_p\text{-}q_p} x_p\pi_a + \delta_p\pi_a\pi_b \ \Big|\pi_a\Big)\end{aligned}$$

$$\mathcal{P}_p^{1,1}(x_p,\pi_a,\pi_b,\theta) =$$

$$G_p^{A-B,B}\Big(\beta'_{r_p\text{-}q_p} x_p + \big[\beta_{t_p\text{-}r_p} - \beta_{s_p\text{-}q_p}\big]' x_p\pi_a + \gamma_p\pi_a\pi_b, \ \ \beta'_{r_p} x_p + \beta'_{t_p\text{-}r_p} x_p\pi_a + \alpha_p\pi_a\pi_b \ \Big|\pi_a\Big).$$

The BNE beliefs solve the following 4×4 system for $p = 1,2$:

$$\begin{aligned}\pi^*_{c_p}(z,\theta) &= E\Big[\ \mathcal{P}_p^1\Big(X, \pi^*_{c\text{-}p}(z,\theta), \pi^*_{c\text{-}p,y\text{-}p}(z,\theta)/\pi^*_{c\text{-}p}(z,\theta), \theta\Big) \ \big|Z = z\Big] \\ \pi^*_{c_p,y_p}(z,\theta) &= E\Big[\ \mathcal{P}_p^{1,1}\Big(X, \pi^*_{c\text{-}p}(z,\theta), \pi^*_{c\text{-}p,y\text{-}p}(z,\theta)/\pi^*_{c\text{-}p}(z,\theta), \theta\Big) \ \Big|Z = z\Big].\end{aligned}$$

References

[1] Ahn, H. (1995) "Nonparametric two-stage estimation of conditional choice probabilities in a binary choice model under uncertainty" Journal of Econometrics 67: pp. 337 – 378.

[2] Ahn, H. and C.F. Manski (1993) "Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations" Journal of Econometrics 56: pp. 291 – 321.

[3] Ahn, H. and J.L. Powell (1993) "Semiparametric estimation of censored selection models with a nonparametric selection mechanism" Journal of Econometrics 58: pp. 3 – 29.

[4] Andrews, D.W.K., Berry, S. and Jia, P. (2004) "Confidence regions for parameters in disrete games with multiple equilibria with an application to discount chain store location", Unpublished manuscript (Department of Economics, Yale University).

[5] Bresnahan, T. F. and P.C. Reiss (1990) "Entry in Monopoly Markets", Review of Economic Studies, 57 (4):pp. 531 – 553.

[6] Bresnahan, T. F. and P.C. Reiss (1991) "Empirical Models of Discrete Games", Journal of Econometrics 48: pp. 57 – 81.

[7] Brock, W. and S. Durlauf (2001) "Interactions-based models", in: J. Heckman and E. Leamer, eds., Handbook of Econometrics, Vol. 5 (North Holland, Amsterdam) pp. 3297 – 3380.

[8] Carlsson, H. and E. van Damme (1993) "Global games and equilibrium selection", Econometrica 61: pp. 989 – 1018.

[9] Chernozhukov, V., Hong, H. and E. Tamer (2004) "Inference for identified parameter sets in econometric models", Unpublished manuscript (Department of Economics, M.I.T.).

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

[11] Gourieroux, J., Laffont and A. Monfort (1981) "Coherency conditions in simultaneous linear equation models with endogenous switching regimes", Econometrica, 48: pp. 675 – 696.

[12] Heckman, J. (1978) "Dummy Endogenous Variables in a Simultaneous Equation System", Econometrica, 46: pp. 931 – 959.

[13] Ichimura, H (2004) "Computation of asymptotic distribution for semiparametric GMM estimators" Unpublished manuscript (Economics Department, UCL).

[14] Jovanovic, B. (1989) "Observable implications of models with multiple equilibria", Econometrica 57: pp. 1431 − 1438.

[15] Lewbel, A. (2005) "Coherency of structural models containing a dummy endogenous variable" Unpublished manuscript (Economics Department, Boston College).

[16] Manski, C.F (1991) "Nonparametric estimation of expectations in the analysis of discrete choice under uncertainty", in: William Barnett, James Powell, and George Tauchen, eds., Nonparametric and Semiparametric Methods in Econometrics and Statistics (Cambridge University Press, Cambridge).

[17] Manski, C.F. and Tamer, E. (2002) "Inference on regressions with interval data on a regressor or outcome" Econometrica 70: pp. 519 − 546.

[18] McFadden, D. (1981) "Econometric models of probabilistic choice", in C. Manski and D. McFadden , eds., Structural Analysis of Discrete Data: With Econometric Applications (MIT Press, Cambridge, MA).

[19] Morris, S.E. and H. Shin. (2001) "Global games: Theory and applications" Cowles Foundation Discussion Paper No. 1275R.

[20] Newey, W.K and D. McFadden (1994) "Large sample estimators and hypothesis testing" in: R.Engle and D.McFadden, eds., Handbook of Econometrics, Vol. 4 (North Holland, Amsterdam) pp. 2113 − 2245.

[21] Pakes, A., Porter, J., Ho, K., and J. Ishii (2005) "Moment inequalities and their application", Unpublished manuscript (Economics Department, Harvard).

[22] Powell, J.L, J.H Stock and T.M Stoker (1989) "Semiparametric estimation of index coefficients", Econometrica 57: pp. 1403 − 1430.

[23] Seim, K. (2002) "An Empirical Model of Firm Entry with Endogenous Product-Type Choices", Unpublished manuscript, (Graduate School of Business, Stanford University).

[24] Sherman, R. (1994a) "U-processes in the analysis of a generalized semiparametric regression estimator", Econometric Theory 10: pp. 372 − 395.

[25] Sherman, R. (1994b) "Maximal inequalities for degenerate U-processes with applications to optimization estimators", Annals of Statistics 22(1): pp. 439 − 459.

[26] Stoker, T. (1991) "Equivalence of direct and indirect estimators of average derivatives", in: W. Barnett, J.L. Powell and G. Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics (Cambridge University Press, Cambridge).

[27] Sweeting, A. (2003) "Coordination Games, Multiple Equilibria and the Timing of Radio Commercials", Unpublished manuscript (Economics Department, MIT).

[28] Tamer, E.T. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", Review of Economic Studies 70(1): pp. 147 – 165.