Semiparametric Estimation of a Simultaneous Game with Incomplete Information

Andres Aradillas-Lopez*

(This version: October 18, 2004)

Abstract

We analyze a 2×2 simultaneous game. We start by showing that a likelihood function defined over the set of four observable outcomes and all possible variations of the game exists only if players have incomplete information. We assume a general incomplete information structure, where players' beliefs are conditioned on a vector of signals $\boldsymbol{Z}$ observable by the researcher but whose exact distribution is known only to the players. The resulting Bayesian-Nash equilibrium (BNE) is characterized as a vector of conditional moment restrictions. We show how to exploit the information contained in these equilibrium conditions efficiently. The proposal takes the form of a two-step estimator. The first step estimates the unknown equilibrium beliefs using semiparametric restrictions analog to the population BNE conditions. The second step maximizes a trimmed log-likelihood function using the estimates from the first step as plug-ins for the unknown equilibrium beliefs. The trimming set is an interior subset of the support of $\boldsymbol{Z}$ where the BNE conditions have a unique solution. The resulting estimator of the vector of structural parameters '$\boldsymbol{\theta}$' is $\sqrt{N}-$consistent and exploits all information in the model efficiently. We allow $\boldsymbol{Z}$ to include continuous and/or discrete random variables. Tests for uniqueness of equilibrium either for a given value of $\boldsymbol{Z}$ or for its entire support are also presented. As an empirical example we estimate a simple game of investment under uncertainty in industries with only two publicly traded firms. Results are consistent with a model in which the smaller firm has a comparatively greater incentive to predict the actions of the larger one, which bases its decisions mainly on private information and indicators of industry uncertainty, giving relatively less weight to the expected actions of the smaller firm.

---

*I would like to thank Professors James L. Powell and Guido W. Imbens for their help and advice. I also thank Professors Paul Ruud, Daniel McFadden, Thomas Rothenberg and Michael Jansson for their valuable comments on this and earlier versions of this paper. All remaining errors are responsibility of the author.
email: aradilla@econ.berkeley.edu. Correspondence: Andres Aradillas-Lopez, Department of Economics, 549 Evans Hall #3880, Berkeley, CA 94720-3880

1 Introduction

The econometric analysis of game-theoretic models has been an increasingly active area of research over the last decade. In these types of models, agents' actions are interdependent because each agent's utility function depends directly on others' choices and/or characteristics. These models have been used to study a wide variety of socioeconomic phenomena ranging from industry entry decisions to the role of neighborhood influences on socioeconomic outcomes such as education or marriage. The formulation and analysis of a game-theoretical model must be accompanied by an appropriately defined equilibrium solution, which is typically some variation of the notion of Nash Equilibrium[1]. Econometric analyses of these models generically assume that agents' observed actions constitute an equilibrium of the underlying game. As a consequence, given a set of stochastic assumptions of the model, the resulting equilibrium properties play a critical role in the econometric study of game-theoretic models. Specifically, given the primitives of the game, a well-defined likelihood function over the entire set of observable outcomes will not exist if, with strictly positive probability, the game has either multiple or no equilibria . Hence, econometric analysis of these models depends fundamentally on the equilibrium features of the underlying game.

In general, a researcher has two choices when it comes to estimating a game with multiple equilibria. The first option is to use some theory of equilibrium selection. An appropriately chosen equilibrium selection mechanism assures the existence of a well-defined likelihood function for the entire space of observable outcomes. Examples of papers which have assumed equilibrium selection rules in the estimation of games include those by Bjorn and Vuong

[1] In this paper we will assume that players maximize expected utility and their resulting optimal strategy profile constitutes a Nash Equilibrium. Alternatives to Nash equilibrium abound. For example, modern non-Nash solution concepts with learning and/or evolution foundations are detailed in Weibull (1997) and Fudenberg and Levine (1998). An elegant refutation to expected utility maximization can be found in Rabin (2000).

(1984, 1985) and Kooreman (1994) in games of complete information, and Sweeting (2004) in a game with incomplete information. The disadvantage of this approach is that while the Nash Equilibrium concept has been used extensively in many and diverse contexts, there is no generally accepted procedure for determining which equilibrium will be played when equilibrium is multiple[2]. Consistency of the estimation depends critically on the validity of the assumed selection rule. The second option is to redefine the game in a way that makes it estimable without the need for an equilibrium selection rule. One alternative is to redefine the space of outcomes of the game and transform it into one that exhibits uniqueness of equilibrium (Bresnahan and Reiss (1990, 1991)). More recently, Tamer (2003) used probability bounds for each outcome instead of their exact (not well-behaved) likelihood function. These alternatives are robust in the sense that they depend only on the concept of Nash Equilibrium, without developing a theory of equilibrium selection. The disadvantage of this type of approach is that the transformations/redefinitions result in some loss of resolution in the model. This in turn translates into efficiency losses. It also limits the the ability of the researcher to predict over the entire set of observable outcomes.

Conditions for uniqueness of equilibrium depend on the primitive elements that characterize the underlying game. Following Fudenberg and Tirole (1991), in non-cooperative games these elements consist of: (i) the set of players, (ii) the order of moves -i.e, who moves when, (iii) the players' payoffs as a function of their moves, (iv) the set of available choices at each move, (v) what each player knows when he makes his choices and (vi) the probability distributions over all exogenous events. This paper concentrates on the econometric

[2]One of the most thorough attempts to present a general equilibrium selection theory based on the same principles of rational behavior can be found in Harsanyi and Selten (1988). These authors propose a theory of equilibrium selection that selects a unique Nash equilibrium for any non-cooperative N-person game. The heart of their theory is given by a "tracing" procedure, a mathematical construction that adjusts arbitrary prior beliefs into equilibrium beliefs. A learning/evolutionary theory of equilibrium selection is presented in Samuelson (1998).

implications of (v) for a simultaneous game. We assume an incomplete information environment more general and flexible than those that have been previously employed in existing econometric work. First, we show that a well-behaved likelihood function for the entire space of observable outcomes exists under generically weaker conditions if players have incomplete information vis-à-vis perfect information. The game's resulting Bayesian-Nash equilibrium (BNE) conditions can be expressed as a vector of conditional moment restrictions. Then, we show how to exploit the information in the BNE conditions efficiently by imposing semiparametric restrictions analog to the BNE. In the end, the presence of incomplete information allows the econometrician to estimate the structural parameters of the model without losing resolution in the model. As we mentioned above, such losses are unavoidable in the perfect information version of the game unless some equilibrium selection rule is imposed.

Specifically, this paper focuses on a 2×2 simultaneous game proposed first by Bresnahan and Reiss in the context of industry entry models and later studied by Tamer. These authors analyzed the game assuming that players possess perfect information and that they only choose pure strategies. Under these assumptions, players' optimal strategies are described by a simultaneous discrete response system. Heckman (1978) studied the properties of such nonlinear systems[3]. Using his results, the aforementioned authors conclude that a well-defined likelihood function exists for the four observable outcomes only if the so-called "coherency" condition is satisfied. Imposing this condition eliminates the strategic interaction from the game. This negative result is a consequence of the presence of multiple equilibria. Bresnahan and Reiss, as well as Tamer propose different estimation techniques that avoid both the coherency condition and the use of equilibrium selection rules. These

[3]Other pioneering papers on systems of nonlinear simultaneous equations include those by Jorgenson and Laffont (1974), Amemiya (1974) and Schmidt (1981). Surveys of methods for estimation of nonlinear multivariate regressions and systems of nonlinear simultaneous equations can be found in Amemiya (1983).

methods result in some loss of resolution in the model, which translates into efficiency losses and reduces the ability to make predictions for all observable outcomes of the game.

Using the results of a companion paper (Aradillas-Lopez (2004)) we first show that if players have complete information and if mixed strategies are allowed, then a well-behaved likelihood function for the four observable outcomes exists under weaker assumptions than the coherency condition. However, we show that if players have complete information, nonexistence of a likelihood function prevails for an entire family of variations of this game, which we call "symmetric". We then concentrate on an incomplete information version of the game. In this setting players must use all relevant available information to construct beliefs about their opponent's expected behavior. Assuming expected utility maximization, in a Bayesian-Nash equilibrium (BNE) each player selects a best response against the expected action of his opponent. Equilibrium beliefs correspond to actual average behavior. Existing econometric literature on simultaneous games with incomplete information is relatively scarce. Existing papers include those of Seim (2002) in the context of an entry model and Sweeting (2004) in the context of a coordination game. Both authors assume that the only source of incomplete information among players is an idiosyncratic component which is unobservable to the econometrician. The BNE conditions in both cases can be expressed as (unconditional) moment restrictions.

This paper shows how to estimate efficiently a simultaneous game assuming a general form of incomplete information. First, instead of confining the source of incomplete information exclusively to an idiosyncratic component unobserved by the econometrician, we allow the possibility that some of the privately observed variables become available to the researcher after the game has been played. Second, we also allow the existence of a vector of publicly observed "signals" $\boldsymbol{Z}$ used by both players to construct their beliefs. These signals are assumed to be statistically related to some of the privately observed variables. They are also assumed to be available to the econometrician. Except for a set of smoothness assumptions, the exact distribution of $\boldsymbol{Z}$ is left unspecified. The game's resulting BNE can be expressed as a vector of conditional moment restrictions. We detail sufficient conditions for uniqueness of BNE and assume that these conditions hold at least inside a subset in the interior of the support of $\boldsymbol{Z}$[4]. Using this result, we show that conditions for existence of a well-defined likelihood function are generically weaker than in the complete information case. In particular, a well-defined likelihood function for the four observable outcomes of the game exists for a subset of symmetric variations of the game only if players have incomplete information. Equilibrium beliefs in our model are in fact conditional probabilities. Lack of knowledge about the distribution of $\boldsymbol{Z}$ implies that these equilibrium probabilities (beliefs) must be estimated using nonparametric methods. Replacing unknown conditional probabilities with nonparametric estimates in discrete choice models with uncertainty -but no strategic interaction- was suggested by Manski (1991, 1993) and thoroughly analyzed by Ahn and Manski (1993).

The estimation procedure takes the form of a trimmed quasi Maximum Likelihood maximization, where uniqueness of equilibrium prevails everywhere in the trimming set. Unknown equilibrium probabilities (beliefs) are replaced with semiparametric plug-ins. In an attempt to increase efficiency, we exploit the information about the structural parameter vector '$\boldsymbol{\theta}$' contained in the BNE conditions. Employing the usual (e.g kernel-based) nonparametric conditional probability estimators as plug-ins would be consistent, but would imply losing this information. Instead, we propose alternative plug-ins based on a semiparametric analog version of the BNE condition. We also show how to adapt this estimation procedure to the case in which uniqueness of equilibrium prevails everywhere in the support of the signals $\boldsymbol{Z}$. In this case, the proposed methodology allows us to use the entire support of $\boldsymbol{Z}$. We then characterize the asymptotic properties of the resulting estimator for $\boldsymbol{\theta}$ which is $\sqrt{N}-$consistent and exploits all available information. The methodology also

[4]We also provide sufficient conditions for uniqueness of BNE to hold everywhere in the support of $\boldsymbol{Z}$.

allows us to test the hypothesis of uniqueness of equilibrium, either for a given realization of $\boldsymbol{Z}$ or for its entire support. Even though the paper focuses on a particular game, the procedure can be adapted to game-theoretic models with more players and/or available actions. An immediate example would be the kind of Local Interaction Models surveyed by Brock and Durlauf (2001).

The paper proceeds as follows: section 2 describes the normal form representation of the game that will be analyzed here. Section 3 details the equilibrium properties of the game under complete and incomplete information. Section 4 focuses on the incomplete information case and presents two semiparametric quasi maximum likelihood estimators that exploit the information contained in the equilibrium conditions along with a detailed characterization of their asymptotic properties. Section 5 presents an empirical application of the game for an investment game in industries with two publicly traded firms. Section 6 includes some concluding remarks.

The proofs to all results can be found in the accompanying Mathematical Appendix.

2 Description of the game

We focus on a 2×2 simultaneous game with the following normal-form representation. As usual in game-theory, each entry in the matrix represents the Neumann-Morgenstern utility of each player for each one of the four outcomes

		PLAYER 2	
		$\boldsymbol{Y_2 = 1}$	$\boldsymbol{Y_2 = 0}$
PLAYER 1	$\boldsymbol{Y_1 = 1}$	$\boldsymbol{X}_1'\boldsymbol{\beta}_1 - \varepsilon_1 + \alpha_1$, $\boldsymbol{X}_2'\boldsymbol{\beta}_2 - \varepsilon_2 + \alpha_2$	$\boldsymbol{X}_1'\boldsymbol{\beta}_1 - \varepsilon_1$, 0
	$\boldsymbol{Y_1 = 0}$	0 , $\boldsymbol{X}_2'\boldsymbol{\beta}_2 - \varepsilon_2$	0 , 0

This payoff structure was first formally studied -in the context of empirical industry entry models- by Bresnahan and Reiss (1991), it was also the focus of Tamer(2003). Following the aforementioned authors, we will assume throughout that the econometrician observes the realization of the random variables $\boldsymbol{X}_1 \in \mathbb{R}^{k_1}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{k_2}$ but doesn't observe those of $\varepsilon_1 \in \mathbb{R}$ nor $\varepsilon_2 \in \mathbb{R}$. The focus of this paper will be to analyze the properties of the game according to the information available to each player. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2) \in \mathbb{R}^k$, with $k \equiv k_1 + k_2$ and denote $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$. Also denote the vector of parameters $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \alpha_1) \in \mathbb{R}^{k_1+1}$, $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \alpha_2) \in \mathbb{R}^{k_2+1}$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^{k+2}$, all of which are assumed as constants, unknown to the econometrician. According to the signs of α_1 and α_2 we say that the game is "symmetric" if $\alpha_1 \times \alpha_2 > 0$, "asymmetric" if $\alpha_1 \times \alpha_2 < 0$ and "not jointly strategic" if $\alpha_1 \times \alpha_2 = 0$.

3 Properties of the game under incomplete information

Assuming perfect knowledge of payoffs is a good approximation in some economic situations. When players do not have exact knowledge about the payoffs of their opponents the game is said to have "incomplete information". In this section we will assume that each player has complete information about his own payoff but has incomplete information about his opponent's payoff. Specifically, we will assume that the information structure satisfies the following properties:

3.1 Information assumptions

(I): 1.− The realizations of $(\boldsymbol{X}_1, \varepsilon_1)$ and $(\boldsymbol{X}_2, \varepsilon_2)$ are perfectly observed by players 1 and 2 respectively, who also know the value of $\boldsymbol{\theta}$.

2.– ε_1 and ε_2 are purely idiosyncratic shocks, privately observed by players 1 and 2 respectively.

3.– We allow some elements of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ to be publicly observed by both players, but we also allow the possibility that at least one element of $\boldsymbol{X}_1$ and one element of $\boldsymbol{X}_2$ are privately observed by players 1 and 2 respectively. We will assume the privately observed components of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ to be statistically independent of each other.

4.– There exist publicly observable variables $\boldsymbol{Z}_1 \in \mathbb{R}^{L_1}$ and $\boldsymbol{Z}_2 \in \mathbb{R}^{L_2}$ that are statistically related to the privately observed components of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ respectively. All publicly observable elements of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are included in $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$.

5.– Both players have perfect knowledge of the stochastic properties (probability distributions) of $\boldsymbol{\varepsilon}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ described below.

6.– Players' actions constitute a Bayesian Nash Equilibrium (BNE).

We will let $\boldsymbol{Y} \equiv (Y_1, Y_2)'$ and $\boldsymbol{Z} \equiv \boldsymbol{Z}_1 \cup \boldsymbol{Z}_2$. Denote the dimension of $\boldsymbol{Z}$ as L, so $\boldsymbol{Z} \in \mathbb{R}^L$, with $L \leq L_1 + L_2$. Assumptions (I.1)-(I.3) describe players' knowledge about their mutual payoffs. Instead of confining the source of incomplete information to the idiosyncratic components, these assumptions allow some of the variables available to the researcher to be privately observed at the time the game is played. Independence between the privately observed components of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is assumed merely to simplify the characterization of the equilibrium conditions. It permits both players to construct their equilibrium beliefs conditional on the same set of variables (namely, $\boldsymbol{Z}$). This assumption can be easily dropped from the model but will be maintained throughout.

Assumption (I.4) borrows from the Principal-Agent literature. The possibility of using publicly observable variables to learn more about privately observed individual characteristics has been extensively used in the field of contract theory[5]. Extensions of the basic principal-agent problem assume the existence of a verifiable signal available to the principal (i.e, a publicly observed variable) which is informative about the agent's privately observed characteristics[6]. Assumptions (I5) and (I6) assure that the equilibrium expected probabilities (beliefs) are equal to the actual probabilities. As we will see below, econometric estimation of $\boldsymbol{\theta}$ will rely on this result to "recover" (estimate) these unobservable beliefs using a well-defined sample analog of the population BNE conditions.

We next describe the stochastic assumptions to be used in this section. We will use these assumptions to study the BNE properties of the game. They will be strengthened in Section 4.2, which deals with the estimation of the model.

3.2 Stochastic assumptions

Throughout this paper we will use $\mathbb{S}(v)$ to denote the support of a random variable v. We will use the following stochastic assumptions in this section (they will be strengthened in Section 4.2).

Stochastic properties of $\boldsymbol{\varepsilon_1}$, $\boldsymbol{\varepsilon_2}$

$(\widetilde{S1})$: 1.− ε_1 and ε_2 are continuously distributed random variables, independent of each other, independent of $(\boldsymbol{X}, \boldsymbol{Z})$ and independent of any other publicly observable variable.

2.− We denote the cdf's of ε_1 and ε_2 as $G_1(\epsilon_1)$ and $G_2(\epsilon_2)$ respectively. We will denote their corresponding density functions by $g_1(\varepsilon_1)$ and $g_2(\varepsilon_2)$, which are assumed to

[5]If both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ were publicly observed, then we would have $\boldsymbol{Z}_1 = \boldsymbol{X}_1$ and $\boldsymbol{Z}_2 = \boldsymbol{X}_2$: players' only use of informational signals $\boldsymbol{Z}$ is to learn about the privately observed components of $\boldsymbol{X}$.

[6]Following the pioneering work by Spence (1973), Holmstrom (1979) showed that the principal should incorporate available signals in his optimal decision (contract design for the agent) as long as the signal is statistically related to the unobserved characteristics of the agent.

be bounded and strictly positive everywhere in $\mathbb{R}$ (i.e, $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$). Neither $G_1(\cdot)$ nor $G_2(\cdot)$ depend on $\boldsymbol{\theta}$.

Stochastic properties of $\boldsymbol{X}$, $\boldsymbol{Z}$

$(\widetilde{S2})$: 1.– Denote the conditional pdf's of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ given $\boldsymbol{Z}$ as $f_{\boldsymbol{X}_1|\boldsymbol{Z}}(\cdot)$ and $f_{\boldsymbol{X}_1|\boldsymbol{Z}}(\cdot)$ respectively. We will assume that both conditional pdf's are independent of $\boldsymbol{\theta}$.

Assumption $(\widetilde{S1}.1)$ is crucial for the model to be ultimately estimable: it assures that players' optimal beliefs are constructed conditional on variables observed by the econometrician[7]. Continuity of $G_1(\cdot)$ and $G_2(\cdot)$ (in assumption $(\widetilde{S1}.2)$) is necessary to show existence of equilibrium. The condition $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$ is not crucial. As we shall see, the results presented in this section hold even if these supports are bounded as long as a weaker condition is satisfied (see for example Lemma 3.1 and footnote 12 below). Assumption $(\widetilde{S2})$ simplifies the characterization of the BNE conditions. We will also use it to provide sufficient conditions for uniqueness of equilibrium.

Throughout the paper we will assume that after the game has been played, the econometrician observes $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$, but doesn't observe $\boldsymbol{\varepsilon}$. We will make precise assumptions concerning the econometrician's knowledge of the distribution functions in Section 4.2. The next section describes the characteristics of the BNE given our set of assumptions.

3.3 Equilibrium

In simultaneous (as opposed to sequential) games of incomplete information, players have no possibility to update their prior beliefs about their opponent's privately observed payoff-

[7]Manski (1991) 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.

relevant characteristics which determine players' actual choices. [8] Each player must construct beliefs about their opponent's expected action using all relevant, observable information. Given our assumptions, this implies that players' beliefs are constructed conditional on $\boldsymbol{Z}$. Specifically, let $\overline{\pi}_1^{(2)}(\boldsymbol{Z})$ = Player 2's expected probability that $Y_1 = 1$ given $\boldsymbol{Z}$ and $\overline{\pi}_2^{(1)}(\boldsymbol{Z})$ = Player 1's expected probability that $Y_2 = 1$ given $\boldsymbol{Z}$.

In a Bayesian Nash equilibrium (BNE) players maximize their expected utility conditional on their beliefs, which yields[9]

$$Y_1 = \mathbb{1}\Big\{\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\overline{\pi}_2^{(1)}(\boldsymbol{Z}) - \varepsilon_1 \geq 0\Big\} \quad \text{and} \quad Y_2 = \mathbb{1}\Big\{\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\overline{\pi}_1^{(2)}(\boldsymbol{Z}) - \varepsilon_2 \geq 0\Big\}$$

In a BNE, players' beliefs are equal to the actual probabilities. We will denote these equilibrium probabilities simply as $\pi_1^*(\boldsymbol{Z})$ and $\pi_2^*(\boldsymbol{Z})$. Take $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$. Take $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and define

$$\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) \equiv E\big[G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big] \quad \text{and} \quad \varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) \equiv E\big[G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big]$$

Then, equilibrium probabilities $\pi_1^*(\boldsymbol{Z})$ and $\pi_2^*(\boldsymbol{Z})$ solve (for π_1 and π_2) the equilibrium equations

$$\begin{aligned} \pi_1 - \varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) &= 0 \\ \pi_2 - \varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) &= 0. \end{aligned} \tag{1}$$

Clearly, equilibrium probabilities also depend on $\boldsymbol{\theta}$. From now on we will denote them as $\pi_1^*(\boldsymbol{Z}, \boldsymbol{\theta})$ and $\pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})$. Therefore in a BNE, players' optimal actions are described by the pair of threshold-crossing equations:

$$Y_1 = \mathbb{1}\Big\{\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta}) - \varepsilon_1 \geq 0\Big\} \quad \text{and} \quad Y_2 = \mathbb{1}\Big\{\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1^*(\boldsymbol{Z}, \boldsymbol{\theta}) - \varepsilon_2 \geq 0\Big\}. \tag{2}$$

[8]These privately observed payoff-relevant characteristics are usually called "types".

[9]The presence of incomplete information makes it impossible for players to randomize their actions to make their opponent exactly indifferent between $Y = 1$ and $Y = 0$. This is why optimal choice rules are described by these threshold equations. This contrasts with the complete information version of the game, where mixed-strategy Nash equilibria do exist.

The following section analyzes conditions for existence of a well-behaved likelihood function for the four observable outcomes of the game. As we shall see, these conditions are directly related to the existence and uniqueness properties of the solution to (1).

3.4 Conditions for existence of a likelihood function

In this section we examine conditions for existence of a well-defined conditional likelihood for the four observable outcomes of the game assuming that players choose equilibrium strategies. These conditions depend directly on the equilibrium properties (existence and uniqueness) of the game. We will also compare the results for the complete and the incomplete information versions of the game. As we shall see, conditions for existence of a well defined likelihood function are generically more stringent when players have perfect knowledge of their opponent's payoff realization. We begin by examining the complete information case.

3.4.1 Existence of likelihood function when players have complete information

Suppose $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$ are publicly observed by both players before choosing their actions. This corresponds to the complete information version of the game, which was analyzed previously by Bresnahan and Reiss (1990, 1991) and Tamer (2003). These authors outlined conditions for existence of a well-defined likelihood function $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ assuming the observed actions correspond to a pure-strategy Nash Equilibrium[10], ruling out mixed-strategies. If this is the case (only pure strategies are allowed) then the players' optimal actions can be expressed as a simultaneous discrete response system described by the pair of equations[11]

$$Y_1 = \mathbb{1}\Big\{\boldsymbol{X}'_1\boldsymbol{\beta}_1 + \alpha_1 Y_2 \geq 0\Big\} \quad \text{and} \quad Y_2 = \mathbb{1}\Big\{\boldsymbol{X}'_2\boldsymbol{\beta}_2 + \alpha_2 Y_1 \geq 0\Big\}.$$

[10] If players have perfect knowledge about their opponent's payoffs, there is no use for signals $\boldsymbol{Z}$ and the relevant conditional likelihood is simply $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$.

[11] These behavior equations replace (2), which describe players' optimal actions with incomplete information.

Heckman (1978) provided conditions for existence of a well-defined likelihood function of this model which he referred to as "principal conditions". Bresnahan and Reiss referred to them as conditions for existence of a "well-defined reduced form". Tamer later referred to these as "coherency conditions" . Aradillas-Lopez (2004) extended the results of Bresnahan and Reiss as well as Tamer to the case in which mixed-strategy Nash Equilibria are allowed. In this case, optimal strategies are no longer exactly described by a simultaneous discrete response system. The next Lemma summarizes the results in Aradillas-Lopez.

Lemma 3.1 *Suppose $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$ are publicly observed by both players and $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$. Let $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ denote the conditional likelihood of $\boldsymbol{Y}$ given $\boldsymbol{X}$. If the game is in equilibrium then*

(A) If mixed-strategies are allowed, a well defined $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ exists for the four outcomes of the game if and only if $\alpha_1\alpha_2 \leq 0$.

(B) If only pure-strategies are allowed, a well defined $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ exists for the four possible outcomes of the game if and only if $\alpha_1\alpha_2 = 0$.

See Aradillas-Lopez for details of the proof, which relies entirely on the Nash Equilibrium properties of the game. Tamer called $\alpha_1 \times \alpha_2 = 0$ the "coherency condition", which is necessary and sufficient for existence of a well-defined likelihood function for the four outcomes if we assume the game is in equilibrium and only pure-strategies are allowed. Once mixed-strategies are allowed, this condition can be relaxed to $\alpha_1 \times \alpha_2 \leq 0$. Using our early terminology we can summarize the result as "if players can choose mixed-strategies and the game is in equilibrium, a well defined likelihood function exists for the four possible outcomes if and only if the game is either asymmetric or not jointly strategic". The reason behind this result is simple: if $\alpha_1 \times \alpha_2 \leq 0$ then uniqueness of equilibrium is a generic property of the game.

If the game is symmetric (i.e, if $\alpha_1 \times \alpha_2 > 0$) and the support of $\boldsymbol{\varepsilon}$ is rich enough[12], then a well-defined $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ for the four outcomes does not exist even if we allow for mixed-strategies. The reason behind this result is once again a simple one: if $\alpha_1 \times \alpha_2 > 0$ then multiple equilibria is a generic property of the game. We should point out however, that if $\alpha_1 \times \alpha_2 > 0$ and mixed-strategies are ruled out, then $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ exists for a subset of the four outcomes of the game. This was first noted by Bresnahan and Reiss (1990,1991) and enabled them to treat multiple outcomes as one event, effectively transforming the model into one that predicts the joint equilibria. For example, if $\alpha_1 > 0, \alpha_2 > 0$ then a well-defined $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ exists for $\boldsymbol{Y} = (1,0)$ and $\boldsymbol{Y} = (0,1)$ whereas if $\alpha_1 < 0, \alpha_2 < 0$ then a well-defined $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\theta})$ exists for $\boldsymbol{Y} = (0,0)$ and $\boldsymbol{Y} = (1,1)$. Instead of using joint outcomes, Tamer proposed a semiparametric estimator based on the probability bounds for the multiple-equilibria outcomes implied by the model. Both alternatives avoid using an equilibrium selection theory at the cost of reducing the resolution of the game. Neither methodology is capable of making predictions (i.e, expected conditional probabilities) for the four observable outcomes of the game.

We now examine the incomplete information version of the game. We will show that a well-defined likelihood function exists under conditions generically weaker than in the complete information case.

3.4.2 Existence of likelihood function under incomplete information

As we mentioned above, after the game has been played the econometrician is assumed to observe $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$, but doesn't observe $\boldsymbol{\varepsilon}$. Denote the conditional likelihood of $\boldsymbol{Y}$ given $(\boldsymbol{X}, \boldsymbol{Z})$ as $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$. Existence of this likelihood function will depend on the

[12]Let $\mathcal{M}(\boldsymbol{X}, \boldsymbol{\theta}) = \big\{(\varepsilon_1, \varepsilon_2) : \ \mathrm{Min}\{\boldsymbol{X}_1'\boldsymbol{\beta}_1, \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\} \leq \varepsilon_1 \leq \mathrm{Max}\{\boldsymbol{X}_1'\boldsymbol{\beta}_1, \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\}$ and

$$\mathrm{Min}\{\boldsymbol{X}_2'\boldsymbol{\beta}_2, \boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\} \leq \varepsilon_2 \leq \mathrm{Max}\{\boldsymbol{X}_2'\boldsymbol{\beta}_2, \boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\}\big\}$$

Then the results of Lemma 3.1 hold if $\Pr\big\{(\varepsilon_1, \varepsilon_2) \in \mathcal{M}(\boldsymbol{X}, \boldsymbol{\theta})\big\} > 0$, which may be true even if $\mathbb{S}(\varepsilon_1) \neq \mathbb{R}$ or $\mathbb{S}(\varepsilon_1) \neq \mathbb{R}$. See Aradillas-Lopez (2004).

equilibrium properties of the game: Take $\boldsymbol{z} \in \mathbb{S}(\boldsymbol{Z})$. Then $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{z}, \boldsymbol{\theta})$ will exist if and only if the solution to (1) when $\boldsymbol{Z} = \boldsymbol{z}$ exists and is unique. We next examine the equilibrium properties (existence and uniqueness) of the game and the resulting conditions for existence of a well-behaved likelihood function $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$.

If assumptions $(\widetilde{S1})$ and $(\widetilde{S2})$ are satisfied, then $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ are monotonic, continuous and strictly bounded in $(0,1)$ for all π_1 and $\pi_2 \in \mathbb{R}$. They also satisfy:

$$\frac{d\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)}{d\pi_2} = \alpha_1 E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big] \quad \text{and} \quad \frac{d\varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)}{d\pi_1} = \alpha_2 E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big].$$

Figures 1 and 2 illustrate examples of $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy these properties for symmetric and asymmetric games respectively. As we can infer from Figures

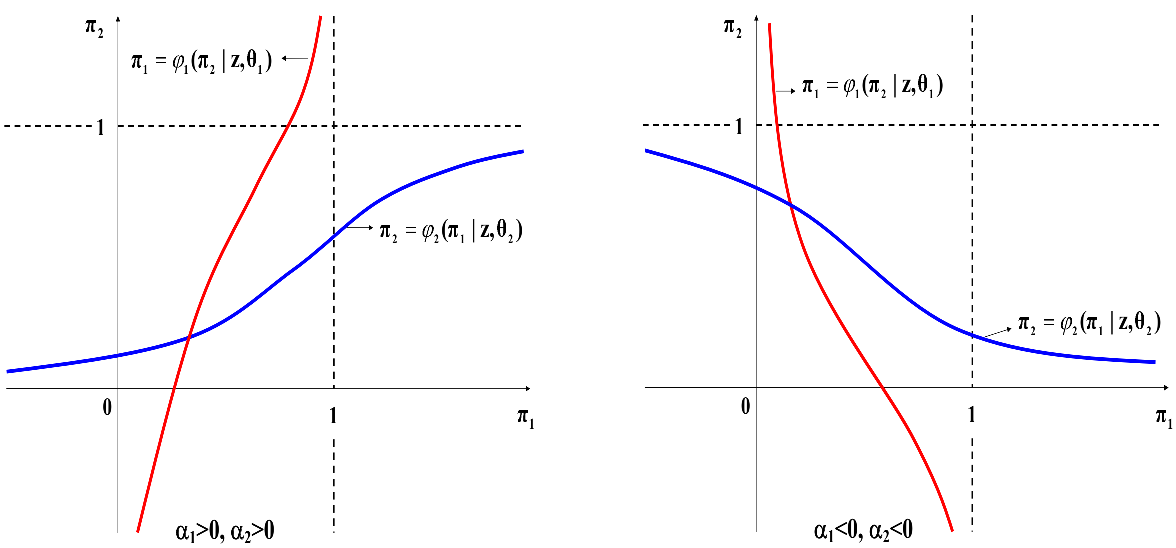

Figure 1: Examples of $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy assumptions $(\widetilde{S1})$ and $(\widetilde{S2})$ for symmetric games.

1 and 2, continuity and boundedness of $G_1(\cdot)$ and $G_2(\cdot)$ are enough to guarantee existence of equilibrium. Because $\mathbb{S}(\varepsilon_1)$ and $\mathbb{S}(\varepsilon_2)$ are assumed to be unbounded, the equilibrium is always strictly inside the unit-square $[0,1]^2$. Lemma 3.2 formalizes this existence result.

The proof uses a fixed-point argument and can be found in the accompanying Mathematical Appendix.

Lemma 3.2 (Existence of equilibrum) *Suppose assumptions* $(\widetilde{S1})$ *and* $(\widetilde{S2})$ *are satisfied. Then a solution to (1) exists for each* $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ *and each* $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$.

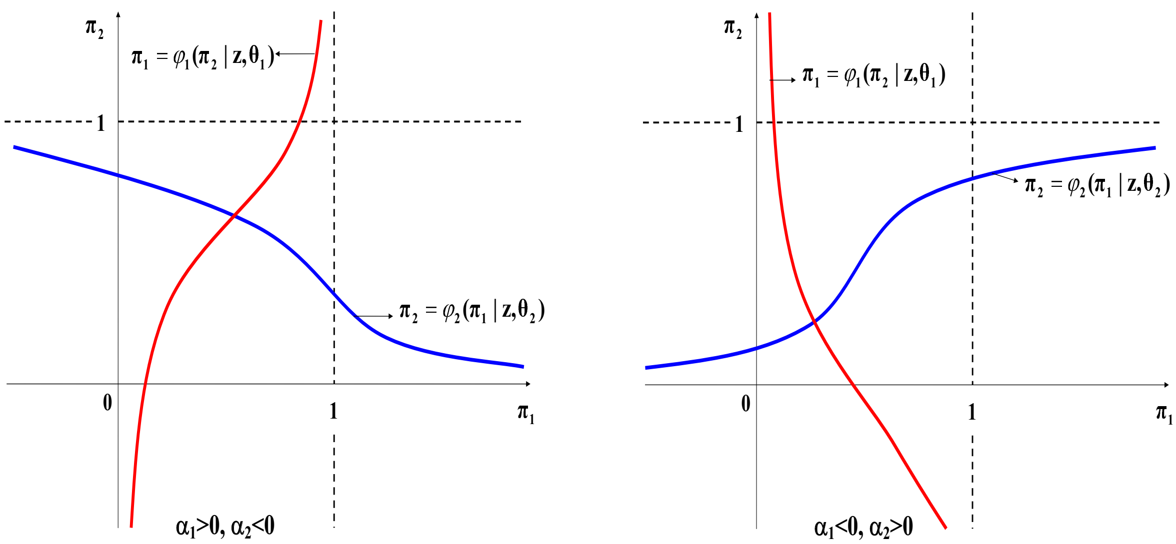

Figure 2: Examples of $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy assumptions $(\widetilde{S1})$-$(\widetilde{S2})$ for asymmetric games.

Having established existence of $\left(\pi_1^*(\boldsymbol{Z}, \boldsymbol{\theta}), \pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})\right)$ everywhere in $\mathbb{S}(\boldsymbol{Z})$, we now turn to the topic of uniqueness.

Looking at figures 3 and 4, we can see that if the game is symmetric we can easily find examples of $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy our assumptions and also yield multiple solutions to (1). Figure 4 presents cases with an infinite number of equilibria, all of which occur at points of tangency between the two curves. Borrowing from general equilibrium literature, we will refer to equilibria that occur at points of tangency as "critical". All other equilibria will be called "regular". From Figure 2 we can infer that examples of functions $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy our assumptions and also yield multiple equilibria can't be found if the game is asymmetric or not jointly strategic, which would imply that each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ has a unique equilibrium if $\alpha_1 \times \alpha_2 \leq 0$.

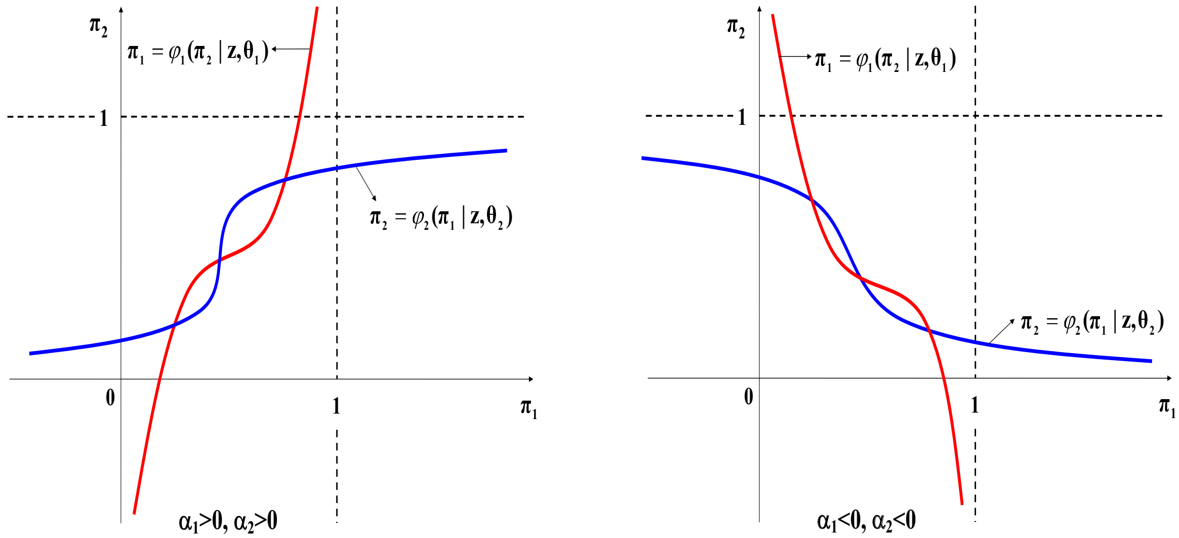

Figure 3: Examples of $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy assumptions $(\widetilde{S1})$-$(\widetilde{S2})$ but yield multiple equilibria if the game is symmetric.

The following lemma formalizes these arguments by describing a sufficient condition for uniqueness of equilibrium; such condition is satisfied by all asymmetric and not-jointly strategic games but it is also satisfied by a subset of symmetric games. The proof can be found in the accompanying Mathematical Appendix.

Lemma 3.3 (Uniqueness of equilibrium) *Take $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and suppose assumptions $(\widetilde{S1})$ and $(\widetilde{S2})$ are satisfied. In addition, suppose $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$ is such that*

$$\alpha_1\alpha_2 E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big] E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big] < 1 \;\forall(\pi_1, \pi_2) \in [0,1]^2,$$

then the equilibrium $\big(\pi_1^*(\boldsymbol{Z}, \boldsymbol{\theta}), \pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})\big)$ is unique and a well-defined conditional likelihood $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ exists.

We can provide some intuition behind the condition in Lemma 3.3. Take $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$, then it is easy to show that if $|\alpha_1| \cdot E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big] < 1$ and $|\alpha_2| \cdot E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big] < 1$

then $\left(\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1), \varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)\right)$ is a contraction mapping and consequently it has a unique fixed point. This last condition however is more restrictive than what we need. For example, $\alpha_1 \times \alpha_2 \leq 0$ then the fixed point is unique regardless of whether or not the right hand side of (1) is a contraction. There is also a geometric interpretation. If the condition of Lemma 3.3 is satisfied, then the slopes of the curves $\varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ and $\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ are different from each other for all $\pi_1 \in [0, 1]$. This puts a limit on the variability of the curves in figures 3 and 4 and restricts the "wiggliness" that gives rise to multiple crossing points (equilibria) and constitutes a sufficient condition for the two curves $\pi_1 = \varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\pi_2 = \varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ to cross only once.

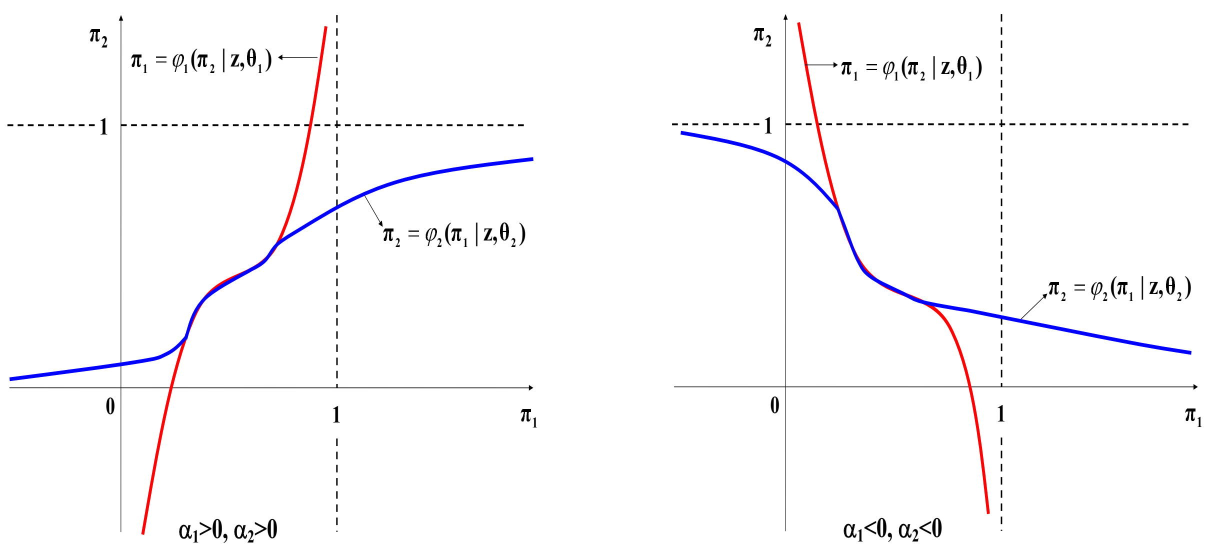

Figure 4: Examples of $\varphi_1(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ that satisfy assumptions $(\widetilde{S}1)$-$(\widetilde{S}2)$ but yield an infinite number of critical equilibria if the game is symmetric.

Lemma 3.3 provides sufficient conditions for uniqueness of equilibrium for a given realization of $\boldsymbol{Z}$. The next corollary builds upon the lemma and provides simple, sufficient conditions involving only (α_1, α_2) such that the conclusions of Lemma 3.3 hold everywhere in $\mathbb{S}(\boldsymbol{Z})$.

Corollary 1 (Uniqueness of equilibrium in $\mathbb{S}(\boldsymbol{Z})$) Suppose assumptions $(\widetilde{S1})$ and $(\widetilde{S2})$ are satisfied. Then the following holds*:*

1.- If the game is asymmetric or not jointly strategic, then there is a unique equilibrium $\big(\pi_1^*(\boldsymbol{Z},\boldsymbol{\theta}),\pi_2^*(\boldsymbol{Z},\boldsymbol{\theta})\big)$ for each $\boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z})$ and $\mathcal{F}(y_1,y_2\mid\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ exists for all $\boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z})$ and all $\boldsymbol{X}$.

2.- More generally, let $\overline{g}_{\varepsilon_1}=\underset{\varepsilon_1\in\mathbb{R}}{Max}\ g_1(\varepsilon_1)$ *and* $\overline{g}_2=\underset{\varepsilon_2\in\mathbb{R}}{Max}\ g_{\varepsilon_2}(\varepsilon_2)$ and suppose that $\boldsymbol{\theta}$ is such that $\alpha_1\times\alpha_2<1/(\overline{g}_1\overline{g}_2)$. Then there is a unique equilibrium $\big(\pi_1^*(\boldsymbol{Z},\boldsymbol{\theta}),\pi_2^*(\boldsymbol{Z},\boldsymbol{\theta})\big)$ for each $\boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z})$. Consequently, $\mathcal{F}(y_1,y_2\mid\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ exists for all $\boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z})$ and all $\boldsymbol{X}$.

If assumption $(\widetilde{S1})$ is satisfied, then $E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1+\alpha_1\pi_2)\mid\boldsymbol{Z}\big]\in[0,\overline{g}_1]$ and $E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2+\alpha_2\pi_1)\mid\boldsymbol{Z}\big]\in[0,\overline{g}_2]$ for all $(\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\pi})\in\mathbb{S}(\boldsymbol{Z})\times\mathbb{R}^{k+2}\times\mathbb{R}^2$. Consequently, $\alpha_1\times\alpha_2<1/(\overline{g}_1\overline{g}_2)$ is a sufficient (but not necessary) condition for the assumption of Lemma 3.3 to hold everywhere in $\mathbb{S}(\boldsymbol{Z})$. Thus, from Corollary 1 and Lemma 3.1 we conclude that if a well-defined likelihood function exists in both the complete and incomplete information cases if $\alpha_1\times\alpha_2\le 0$. However, if the game is symmetric then the likelihood function exists only if players have incomplete information.

The conditions in Lemma 3.3 and Corollary 1 are sufficient, but not necessary for uniqueness of the BNE in symmetric games. In general, the discussion in the preceding paragraphs shows that if the game is symmetric, the BNE will be unique if the strategic-interaction parameters α_1 and α_2 are small relative to the conditional supports $\mathbb{S}\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1\mid\boldsymbol{Z}\big)$ and $\mathbb{S}\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2\mid\boldsymbol{Z}\big)$ respectively. More precisely, we need them to be small enough so as to avoid the variability (wiggliness) of $\varphi_1(\pi_2\mid\boldsymbol{Z},\boldsymbol{\theta}_1)$ and $\varphi_2(\pi_1\mid\boldsymbol{Z},\boldsymbol{\theta}_2)$ in the interval $(\pi_1,\pi_2)\in[0,1]^2$ that is needed for multiple equilibria to prevail -see Figures 3 and 4-. The next part of the paper deals with the problem of estimating the structural parameter $\boldsymbol{\theta}$ when the game is played under incomplete information.

4 Estimation of the game with incomplete information

In this section we will present a methodology for estimating the structural parameter $\boldsymbol{\theta}$ under the assumption that players have incomplete information. First, we will see how to estimate the unobserved equilibrium probabilities (beliefs) using the BNE conditions. Then, we will show how to use these estimated equilibrium probabilities to estimate the structural parameter $\boldsymbol{\theta}$. The methodology exploits all information available to the econometrician. Due to the equilibrium characteristics of the game with incomplete information, we will be able to carry out the estimation without losing resolution in the model. The presence of incomplete information will enable us to make predictions for the four observable outcomes of the game.

Before proceeding, let us introduce some new notation. We will use '$-p$' to denote player p 's opponent. Trivially, we have: "$-p = 2$ if $p = 1$" and "$-p = 1$ if $p = 2$". As before, we will denote $\boldsymbol{Y} \equiv (Y_1, Y_2)' \in \mathbb{R}^2$, $\boldsymbol{X} \equiv (\boldsymbol{X}_1', \boldsymbol{X}_2')' \in \mathbb{R}^k$ and $\boldsymbol{Z} \equiv \boldsymbol{Z}_1 \cup \boldsymbol{Z}_2$, with $\boldsymbol{Z} \in \mathbb{R}^L$. We will use $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}$ to denote the true parameter value and the parameter space respectively. Except when noted otherwise, we will follow the existing convention and use upper and lower cases to distinguish between random variables and their realizations. Finally, we will define $M \equiv L+1$, where L is the number of signals $\boldsymbol{Z}$ used by the players to construct their beliefs. We next describe the set of assumptions that will be used through the rest of the paper.

4.1 Information assumptions

We will maintain assumption (I) exactly as described in Section 3.1.

Next, we strengthen the stochastic assumptions used in Section 3.1. Basically, we will impose smoothness assumptions as well as additional conditions that guarantee the existence of a well-behaved likelihood function. Some of the smoothness conditions we employ are similar or equivalent to those used by Ahn and Manski.

4.2 Stochastic assumptions

Stochastic properties of $\boldsymbol{\varepsilon_1}$, $\boldsymbol{\varepsilon_2}$

We will strengthen assumption $(\widetilde{S1})$ from Section by imposing additional "smoothness" conditions for $G_1(\cdot)$ and $G_2(\cdot)$. We will assume that:

(S1) 1.− ε_1 and ε_2 are continuously distributed random variables, independent of each other, independent of $(\boldsymbol{X}, \boldsymbol{Z})$ and independent of any other publicly observable variable.

2.− We denote the cdf's of ε_1 and ε_2 as $G_1(\epsilon_1)$ and $G_2(\epsilon_2)$ respectively. We will denote their corresponding density functions by $g_1(\varepsilon_1)$ and $g_2(\varepsilon_2)$ respectively, which are strictly positive everywhere in $\mathbb{R}$ (i.e, $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$). Neither $G_1(\cdot)$ nor $G_2(\cdot)$ depend on $\boldsymbol{\theta}$.

3.− $G_1(\epsilon_1)$ and $G_2(\epsilon_2)$ are $M+2$ times differentiable functions, with bounded $M+2$ derivatives everywhere in $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$. Both distribution functions are assumed to be known up to a finite dimensional parameter.

The only difference with respect to $(\widetilde{S1})$ has to do with the smoothness assumptions about $G_1(\cdot)$ and $G_2(\cdot)$. These conditions facilitate the approximations used to find the asymptotic distribution of our proposed estimator. Next, we describe the refinements to $(\widetilde{S2})$. We will now assume that $\boldsymbol{Z}$ is a continuously distributed random vector and impose smoothness assumptions for $f_{\boldsymbol{X}_1,\boldsymbol{Z}}(\boldsymbol{x}_1, \boldsymbol{z})$ and $f_{\boldsymbol{X}_2,\boldsymbol{Z}}(\boldsymbol{x}_2, \boldsymbol{z})$. We will also assume that $\mathbb{S}(\boldsymbol{X})$ is compact.

Stochastic properties of $\boldsymbol{X}$, $\boldsymbol{Z}$

Assumption $(\widetilde{S2})$ will also be strengthened by assuming that the vector of signals $\boldsymbol{Z}$ is continuously distributed and by introducing smoothness assumptions for $f_{\boldsymbol{X}_1,\boldsymbol{Z}}(\boldsymbol{x}_1,\boldsymbol{z})$ and $f_{\boldsymbol{X}_2,\boldsymbol{Z}}(\boldsymbol{x}_2,\boldsymbol{z})$. A compactness condition for $\mathbb{S}(\boldsymbol{X})$ will also be introduced. We will now assume that:

(S2) 1.– $\boldsymbol{Z}$ is a continuously distributed vector with density function denoted by $f_{\boldsymbol{Z}}(\boldsymbol{z})$. We will allow $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ to include continuous and/or discrete random variables and denote the joint pdfs with $\boldsymbol{Z}$ as $f_{\boldsymbol{X}_1,\boldsymbol{Z}}(\boldsymbol{x}_1,\boldsymbol{z})$ and $f_{\boldsymbol{X}_2,\boldsymbol{Z}}(\boldsymbol{x}_2,\boldsymbol{z})$ respectively. None of these functions depends on $\boldsymbol{\theta}$. All these density functions are unknown to the econometrician.

2.– $f_{\boldsymbol{X}_1,\boldsymbol{Z}}(\cdot\ ,\cdot)$, $f_{\boldsymbol{X}_2,\boldsymbol{Z}}(\cdot\ ,\cdot)$ and $f_{\boldsymbol{Z}}(\cdot)$ are bounded, M times differentiable functions of $\boldsymbol{Z}$, with bounded M derivatives everywhere in $\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}\times\mathbb{R}^{L}$.

3.– The supports $\mathbb{S}(\boldsymbol{X}_1)\subset\mathbb{R}^{k_1}$ and $\mathbb{S}(\boldsymbol{X}_2)\subset\mathbb{R}^{k_2}$ are compact sets.

Smoothness conditions (S2.2) are common in semi or non-parametric estimation problems. These conditions facilitate the approximations used to find the asymptotic distribution of our proposed estimator. Compactness of $\mathbb{S}(\boldsymbol{X})$ only needs to hold for the components that are privately observed. This boundedness condition is necessary to prove the uniform convergence results in Lemmas 4.2 and 4.3 which use Lemma 3 in Collomb and Hardle (1986). Indications are that compactness of $\mathbb{S}(\boldsymbol{X})$ can be relaxed in this setting[13]. However, we will maintain this assumption throughout the remaining sections.

According to our assumptions, after the game has been played the researcher observes the realizations of $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ but does not observe the realization of $\boldsymbol{\varepsilon}$. He also knows $G_1(\cdot)$ and $G_2(\cdot)$ -possibly up to a finite dimensional vector- but does not know $f_{\boldsymbol{X}_1,\boldsymbol{Z}}(\boldsymbol{x}_1,\boldsymbol{z})$, $f_{\boldsymbol{X}_2,\boldsymbol{Z}}(\boldsymbol{x}_2,\boldsymbol{z})$ nor $f_{\boldsymbol{Z}}(\boldsymbol{z})$, except for the smoothness assumptions outlined in (S2).

[13] See the proof of Corollary 4 in the accompanying Mathematical Appendix

Take $\boldsymbol{z} \in \mathbb{S}(\boldsymbol{Z})$, $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$ and $(\pi_1, \pi_2) \in \mathbb{R}^2$. We will follow the notation used in Section 3.3 and denote

$$\varphi_1(\pi_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1) = E\big[G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z} = \boldsymbol{z}\big]; \qquad \varphi_2(\pi_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2) = E\big[G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z} = \boldsymbol{z}\big]$$

In addition, we will define

$$\delta_1(\pi_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1) = E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z} = \boldsymbol{z}\big]; \qquad \delta_2(\pi_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2) = E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z} = \boldsymbol{z}\big].$$

The following assumption involves the parameter space. The first part assumes that $\boldsymbol{\Theta}$ is compact. The second part assumes that the necessary condition for uniqueness of equilibrium stated in Lemma 3.3 holds at least inside a compact set in the interior of $\mathbb{S}(\boldsymbol{Z})$:

(S3) 1.− The parameter space $\boldsymbol{\Theta}$ is compact.

2.− There exists a compact set $\boldsymbol{\mathcal{Z}}$ in the interior of $\mathbb{S}(\boldsymbol{Z})$ with $\inf_{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > 0$ such that

$$\alpha_1\alpha_2\delta_1(\pi_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1)\delta_2(\pi_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2) < 1 \quad \forall\ \boldsymbol{z} \in \boldsymbol{\mathcal{Z}},\ \forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \text{ and } \forall\ (\pi_1, \pi_2) \in [0,1]^2.$$

where the functions δ_1 and δ_2 are as defined above.

Assumption (S3.1) is common in econometric estimation models. (S3.2) follows from Lemma 3.3 and -combined with (I), (S1) and (S2)- assures uniqueness of equilibrium and existence of a well-defined likelihood function everywhere inside the compact set $\boldsymbol{\mathcal{Z}}$ [14]. The results of Corollary 1 apply here: If $\alpha_1\alpha_2 < 1/(\overline{g}_1\overline{g}_2)$ then the BNE is unique for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and (S3.2) holds with $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$.

From here on, we will denote $\boldsymbol{\pi} \equiv (\pi_1, \pi_2) \in \mathbb{R}^2$ and let:

$$\underset{2\times 1}{\boldsymbol{\varphi}(\boldsymbol{\pi}} \mid \boldsymbol{z}, \boldsymbol{\theta}) = \big(\varphi_1(\boldsymbol{\pi}_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1), \varphi_2(\boldsymbol{\pi}_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2)\big)'$$

$$\underset{2\times 2}{J\big(\boldsymbol{\pi}} \mid \boldsymbol{z}, \boldsymbol{\theta}\big) = \nabla_{\boldsymbol{\theta}}\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)$$

[14] From (S3.2) we have $\Pr\{\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}\} > 0$. Consequently, $\text{boundary}\big(\boldsymbol{\mathcal{Z}}\big) = \boldsymbol{\mathcal{Z}} \cap \text{cl}\big(\boldsymbol{\mathcal{Z}}^c\big)$ has Lebesgue measure zero in $\mathbb{R}^L$. Since $\boldsymbol{Z}$ is continuously distributed ($\boldsymbol{Z}$ is absolutely continuous with respect to Lebesgue measure), we have $\Pr\{\boldsymbol{Z} \in \text{boundary}\big(\boldsymbol{\mathcal{Z}}\big)\} = 0$.

The following lemma uses assumptions (S1), (S2.1-2) and (S3.2) to generalize the result of Lemma 3.3 in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

Lemma 4.1 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1), (S2) and (S3) are satisfied. For $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ let $\left(\pi_1^*(\boldsymbol{z}, \boldsymbol{\theta}), \pi_2^*(\boldsymbol{z}, \boldsymbol{\theta})\right)' \equiv \boldsymbol{\pi}^*(\boldsymbol{z}, \boldsymbol{\theta})$ denote the solution (for π_1 and π_2) to the system*

$$\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) = \mathbf{0}.$$

Then:

(A) Each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ has a unique solution $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \in (0,1)^2$.

(B) $\boldsymbol{\pi}^*$ is an M times differentiable function $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$ with bounded M derivatives everywhere in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. It also satisfies $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z}) \in (0,1)^2$ -strictly inside the unit square- for all $(\boldsymbol{\theta}, \boldsymbol{Z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

Part (A) of this lemma is a direct consequence of Lemma 3.3, while part (B) is a consequence of the smoothness assumptions in $(S1) - (S2)$ and the Implicit Function Theorem (IFT), which holds everywhere in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ since the Jacobian $\nabla_{\boldsymbol{\pi}}\left(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\right)$ is invertible for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and all $\boldsymbol{\pi} \in [0,1]^2$ by (S3.2). Another important property of $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$ stated in part (B) of the lemma is that it is strictly inside $(0,1)^2$ for all $(\boldsymbol{\theta}, \boldsymbol{Z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. This is a consequence of the compactness of $\boldsymbol{\Theta} \times \mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}}$ and the fact that $\mathbb{S}(\varepsilon_1) = \mathbb{S}(\varepsilon_2) = \mathbb{R}$, which implies that $G_1(v)$ and $G_1(v)$ are strictly inside $(0,1)$ for all $v \in \mathbb{R}$. Lastly, note that for all $\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}$

$$\begin{aligned} E\left[\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{\theta}\right] &= \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z}) \\ E\left[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}\right] &= \Big(G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1 \pi_2^*(\boldsymbol{\theta}, \boldsymbol{Z}))\ ,\ G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2 \pi_1^*(\boldsymbol{\theta}, \boldsymbol{Z}))\Big)' \end{aligned} \tag{3}$$

and therefore the conditional likelihood $\mathcal{F}(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ exists and is well defined for all $\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}$, all $\boldsymbol{X} \in \mathbb{S}(\boldsymbol{X})$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The next section deals with the estimation of the unobserved equilibrium probabilities $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$. We propose two alternative estimators, both of which exploit the information contained in the BNE conditions. The first one forces the data to satisfy a semiparametric condition analog to the BNE. The second one is a two-step estimator, based on a semiparametric linearization of the BNE.

4.3 Proposed estimators for equilibrium probabilities

We are interested in studying the properties of estimators that exploit the information about $\boldsymbol{\theta}_0$ contained in the equilibrium conditions (1). These conditions can be compactly expressed as

$$\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) - \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{\theta}_0, \boldsymbol{z}\big) = \mathbf{0}$$

Before proceeding, we present an alternative interpretation of $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$ as an extremum estimator.

4.3.1 Alternative interpretation of equilibrium conditions

Let $\boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv -\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)'\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big) \in \mathbb{R}_{-}$ and note that by definition, $\boldsymbol{Q}(\boldsymbol{\pi}^*(\boldsymbol{z}, \boldsymbol{\theta}) \mid \boldsymbol{z}, \boldsymbol{\theta}) = 0$ for all $(\boldsymbol{z}, \boldsymbol{\theta}) \in \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$. Naturally, for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ we have $\boldsymbol{\pi}^* \in \underset{\boldsymbol{\pi} \in \mathbb{R}^2}{\text{Argmax}}\, \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ if $\boldsymbol{\pi}^* - \boldsymbol{\varphi}(\boldsymbol{\pi}^* \mid \boldsymbol{z}, \boldsymbol{\theta}) = \mathbf{0}$. As we mentioned above, assumption (S3.2) implies that the Jacobian $\nabla_{\boldsymbol{\pi}}\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)$ is invertible for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and all $\boldsymbol{\pi} \in [0,1]^2$. From Lemma 4.1, we have $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \in (0,1)^2$. Therefore, for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ we also have: $\boldsymbol{\pi}^* \in \underset{\boldsymbol{\pi} \in [0,1]^2}{\text{Argmax}}\, \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ only if $\boldsymbol{\pi}^* - \boldsymbol{\varphi}(\boldsymbol{\pi}^* \mid \boldsymbol{z}, \boldsymbol{\theta}) = \mathbf{0}$. Combining both results, we can reinterpret the equilibrium conditions (1) as

"For all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$: $\boldsymbol{\pi}^* - \boldsymbol{\varphi}(\boldsymbol{\pi}^* \mid \boldsymbol{z}, \boldsymbol{\theta}) = \mathbf{0}$ if and only if $\boldsymbol{\pi}^* = \underset{\boldsymbol{\pi} \in [0,1]^2}{\text{Argmax}}\, \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$."

Invertibility of the Jacobian of the conditional moment restrictions (1) allows us to approach the estimation of the equilibrium probabilities as a (semiparametric) extremum estimation problem. We now present our first proposal to estimate $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$.

4.3.2 Semiparametric analog estimator

The first proposed estimator is one that solves a kernel-based sample analog of the BNE (1). Suppose we have a sample $\{\boldsymbol{Y}_n, \boldsymbol{X}_n, \boldsymbol{Z}_n\}_{n=1}^N$ of size N. Let h_N be a bandwidth sequence that depends on $N \in \mathbb{N}$ and let $K(\cdot) : \mathbb{R}^L \to \mathbb{R}$ be a Kernel function. Denote $K_{h_N}(\boldsymbol{\psi}) \equiv K\big(\boldsymbol{\psi}/h_N\big)$. We will assume that h_N and $K(\cdot)$ satisfy the following conditions:

(S4) 1.– $K(\cdot) : \mathbb{R}^L \to \mathbb{R}$ is everywhere continuous, bounded, symmetric around zero and satisfies

- (i) Lipschitz condition: $\exists \gamma > 0, c_k < \infty : \ | K(\boldsymbol{u}) - K(\boldsymbol{v}) | \leq c_k \big\|\boldsymbol{u} - \boldsymbol{v}\big\|^{\gamma} \ \forall \ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^L$
- (ii) $\big\|\boldsymbol{\Psi}\big\| \cdot |K(\boldsymbol{\Psi})| \to 0$ as $\big\|\boldsymbol{\Psi}\big\| \to \infty$.
- (iii) $\int K\big(\boldsymbol{\Psi}\big) d\boldsymbol{\Psi} = 1$.
- (iv) $\int \big\|\boldsymbol{\Psi}\big\|^M K\big(\boldsymbol{\Psi}\big) d\boldsymbol{\Psi} < \infty$. ($M^{th}$ moment of $K(\cdot)$ is bounded).
- (v) For all $0 < q_1 + \cdots + q_L < M$: $\int \Psi_1^{q_1} \cdots \Psi_L^{q_L} K(\boldsymbol{\Psi}) d\boldsymbol{\Psi} = 0$. (First $M-1$ moments of $K(\cdot)$ are zero).

2.– Let $M \equiv L + 1$. Then, as the sample size $N \to \infty$, the bandwidth h_N satisfies

- (i) $h_N \to 0$ and there exists $\varepsilon > 0$ such that $N^{1-2\varepsilon} h_N^{2L} \to \infty$.
- (ii) $N h_N^{2M} \to 0$.

3.– We have an iid sample $\{\boldsymbol{Y}_n, \boldsymbol{X}_n, \boldsymbol{Z}_n\}_{n=1}^N$.

Assumption (S4.1) is common in kernel-based semiparametric estimation problems. They are also sufficient to satisfy the corresponding conditions in Collomb and Hardle. (S4.2) was also employed by Ahn and Manski and it facilitates the uniform convergence results in Lemmas 4.2 and 4.3 (below). All these assumptions could be potentially relaxed, but they will be maintained throughout the remainder of the paper. For $p \in \{1,2\}$ define

$$\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

$$\widehat{\varphi}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{G_p\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p \pi_{-p}\big) K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})},$$

and denote

$$\widehat{\boldsymbol{\varphi}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv \big(\widehat{\varphi}_{1_N}(\boldsymbol{\pi}_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1), \widehat{\varphi}_{2_N}(\boldsymbol{\pi}_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2)\big)' \in \mathbb{R}^2$$

$$\widehat{\boldsymbol{Q}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv -\big(\boldsymbol{\pi} - \widehat{\boldsymbol{\varphi}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)'\big(\boldsymbol{\pi} - \widehat{\boldsymbol{\varphi}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big) \in \mathbb{R}$$

These are kernel-smoothed sample analogs for $\varphi(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ and $\boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ respectively. As we showed above, assumption (S3.2) implies that

$$\forall\ (\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}} :\ \boldsymbol{\pi}^* - \boldsymbol{\varphi}(\boldsymbol{\pi}^* \mid \boldsymbol{z}, \boldsymbol{\theta}) = \boldsymbol{0} \quad \text{if and only if} \quad \boldsymbol{\pi}^* = \underset{\boldsymbol{\pi} \in [0,1]^2}{\operatorname{argmax}}\, \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$$

Take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and let $\underset{2\times 1}{\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})}$ be defined as

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

We refer to $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$ as the semiparametric analog estimator of $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$. We want to trim $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$ in the set $[0,1]^2$ because assumption (S3.2) −which yields not only uniqueness of equilibrium and existence of a well-defined likelihood function in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ but also uniform boundedness of $\left\|J(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})^{-1}\right\|$ in $[0,1]^2 \times \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$− holds precisely in that set. From the results of Lemma 4.1, we get that $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \in (0,1)^2$ (is strictly inside the unit square) with probability approaching one uniformly in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. The details of this result are included in the accompanying Mathematical Appendix. The next lemma summarizes the asymptotic properties of $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$, $\underset{2\times(k+2)}{\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})}$ and $\underset{2(k+2)\times(k+2)}{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})}$. We focus on these three objects since -as we shall see below- the asymptotic properties of our proposed estimators for $\boldsymbol{\theta}$ depend on them to a first order of approximation.

Lemma 4.2 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and let $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) = \underset{\boldsymbol{\pi} \in [0,1]^2}{\operatorname{argmax}} \widehat{\boldsymbol{Q}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ .Then*

$$\text{(A)} \quad \sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \left\| \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \right\| = o_p(N^{-1/4}),$$

$$\text{(B)} \quad \sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \right\| = o_p(N^{-1/4}),$$

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \right\| = o_p(N^{-1/4}),$$

where for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ is the solution (for $\boldsymbol{\pi}$) to $\boldsymbol{\pi} - \varphi(\boldsymbol{\pi} \mid \boldsymbol{\theta}, \boldsymbol{z}) = \mathbf{0}$, which by (S3.2) is also the unique solution (for $\boldsymbol{\pi}$) to $\underset{\boldsymbol{\pi} \in [0,1]^2}{\operatorname{Max}} \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$.

The proof can be found in the Mathematical Appendix. Assumption (S3.2) and the result of Lemma 4.1 are equally important for the proof in the particular context of our model, since they assure that the norm of the inverse Jacobian matrix $\left\| J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z}) \mid \boldsymbol{Z}, \boldsymbol{\theta}\big)^{-1} \right\|$ is uniformly bounded in $\boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$. The smoothness conditions in (S1.3), (S4) and (S2.2) as well as the compactness of $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$ also play an important role. These results together allow us to use Lemma 3 of Collomb and Hardle, which establishes uniform rates of convergence of kernel estimators over compact sets. The details of the proof are a bit lengthy, as they require us to establish the uniform rate of convergence of a variety of kernel-smoothed objects. The results of Collomb and Hardle have been used previously to determine uniform rates of convergence over compact sets by Stoker (1991) and Ahn and Manski.

In the next section we present an alternative estimator that also uses the information contained in the equilibrium conditions (1). Instead of forcing the sample to satisfy the analog BNE conditions, it satisfies them asymptotically.

4.3.3 Linearized, two-step semiparametric estimator

As we did before, let $J(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ denote the Jacobian $\nabla_{\boldsymbol{\pi}}\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)$. Therefore, we have:

$$J(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) = \begin{pmatrix} 1 & -\alpha_1\delta_1(\pi_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1) \\ -\alpha_2\delta_2(\pi_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2) & 1 \end{pmatrix}.$$

From assumption (S3.2), $J(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ is invertible for all $(\boldsymbol{z}, \boldsymbol{\theta}) \in \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$ and all $\boldsymbol{\pi} \in [0,1]^2$. From (S3.2) and (S1.3) we get that $\left\| J\big(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \right\|$ is uniformly bounded in $(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{z}) \in [0,1]^2 \times \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Therefore, because $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \in [0,1]^2$ for all $(\boldsymbol{z}, \boldsymbol{\theta}) \in \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$, we have that $J(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta})$ is invertible and $\left\| J(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta})^{-1} \right\|$ is uniformly bounded everywhere in $\boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$. Now let $\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ and $J(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ denote the Jacobian $\nabla_{\boldsymbol{\pi}}\big(\boldsymbol{\pi} - \widehat{\varphi}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)$. Then $\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ is given by:

$$\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) = \begin{pmatrix} 1 & -\alpha_1\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1) \\ -\alpha_2\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2) & 1 \end{pmatrix}$$

where

$$\widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{g_p\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \quad \text{for } p \in \{1,2\}$$

which is in turn a kernel-smoothed sample analog of $\delta_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ for $p \in \{1,2\}$. Now let

$$\widetilde{\pi}_{p_N}(\boldsymbol{z}) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{Y_{p_n}K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \quad \text{for } p \in \{1,2\}$$

and note that $\widetilde{\pi}_{p_N}(\boldsymbol{z})$ is the usual nonparametric kernel estimator for $E[Y_p \mid \boldsymbol{Z} = \boldsymbol{z}]$ for $p \in \{1,2\}$. This estimator does not incorporate the information about $\boldsymbol{\theta}_0$ contained in the equilibrium conditions. However, we show in the Mathematical Appendix that it is uniformly consistent in $\boldsymbol{\mathcal{Z}}$. This suggests that we can use it as a first-step estimator in a linearized version of the analog estimator presented above. This linearized estimator would be computationally attractive relative to $\widehat{\boldsymbol{\pi}}_N(\boldsymbol{\theta}, \boldsymbol{z})$. Before proceeding, we define $\overline{\pi}_{p_N}(\boldsymbol{z}) = \text{Max}\left\{0, \text{Min}\left\{\widetilde{\pi}_{p_N}(\boldsymbol{z}), 1\right\}\right\}$ for $p \in \{1,2\}$ and let $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \equiv \big(\overline{\pi}_{1_N}(\boldsymbol{z}), \overline{\pi}_{2_N}(\boldsymbol{z})\big)'$. Take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, the proposed linearized estimator $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z})$ is given by

$$\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) = \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) + \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\big].$$

We trim $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z})$ in the set $[0,1]^2$ for the same reasons outlined for $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ in the paragraph previous to Lemma 4.2. Before proceeding, let

$$\rho(\boldsymbol{\theta},\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) + J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\big],$$

and note that by the equilibrium conditions $\rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. The next lemma summarizes the asymptotic properties of $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$, $\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$.

Lemma 4.3 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ and let $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ and $\rho(\boldsymbol{\theta},\boldsymbol{z})$ be as described above. Then*

$$\text{(A)}\quad \mathop{sup}_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\right\| = o_p(N^{-1/4}),$$

$$\text{(B)}\quad \mathop{sup}_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}}\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\right\| = o_p(N^{-1/4}),$$

$$\mathop{sup}_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\right\| = o_p(N^{-1/4}).$$

In particular

$$\text{(C)}\ \mathop{sup}_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\|\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\right\| = o_p(N^{-1/4})\ ,\ \mathop{sup}_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\|\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\right\| = o_p(N^{-1/4}).$$

Where for each $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$, $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$ are the equilibrium probabilities which solve (for $\boldsymbol{\pi}$) the system $\boldsymbol{\pi} - \varphi(\boldsymbol{\pi} \mid \boldsymbol{\theta}_0,\boldsymbol{z}) = \mathbf{0}$. By (S3.2), they are also the unique solution (for $\boldsymbol{\pi}$) to the problem $\mathop{\mathrm{Max}}_{\boldsymbol{\pi}\in[0,1]^2} \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}_0)$.

The proof is included in the accompanying Mathematical Appendix. It relies on the same technical conditions as those of the proof of Lemma 4.2. It is built upon some of the results of the proof of Lemma 4.3 and the uniform rate of convergence of $\overline{\boldsymbol{\pi}}(\boldsymbol{z})$ in $\boldsymbol{\mathcal{Z}}$. Once again, the result in Collomb and Hardle is crucial. By the result of Lemma 4.1 and assumption (S3.2), we have that $\|\rho(\boldsymbol{\theta},\boldsymbol{z})\|$, $\|\nabla_{\boldsymbol{\theta}}\rho(\boldsymbol{\theta},\boldsymbol{z})\|$ and $\|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\rho(\boldsymbol{\theta},\boldsymbol{z})\|$ are uniformly bounded in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Regarding part (C) of the lemma, we should point out that $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ does not converge to

$\nabla_{\boldsymbol{\theta\theta'}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$. This is a consequence of the fact that $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ is based on a linear (as opposed to second-order) approximation of the equilibrium conditions. As we will see below, this will not affect the asymptotic properties of the proposed estimator of $\boldsymbol{\theta}$.

4.4 Estimation of $\boldsymbol{\theta}$

In this section we present a proposal for estimating $\boldsymbol{\theta}$ based on a trimmed quasi maximum likelihood estimation, where the semiparametric estimators for $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})$ described previously are plugged in for the unknown $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})$. The trimming set is $\boldsymbol{\mathcal{Z}}$, where the likelihood function is well-behaved. Let us start by discussing some issues regarding identification.

4.4.1 Identification

Players' optimal actions are described by the system of threshold-crossing equations (2). Generically, identification in these types of models requires some normalization condition concerning the variance of $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ -see for example McFadden (1981)-. Given this normalization, the following condition will prove to be sufficient for identification of $\boldsymbol{\theta}$ everywhere in $\boldsymbol{\Theta}$:

(S5) Conditional on $\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}$, if $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$ with $\boldsymbol{\theta},\boldsymbol{\theta}_0\in\boldsymbol{\Theta}$ then

$$\Pr\Big\{\boldsymbol{\beta}_1'\boldsymbol{X}_1+\alpha_1\pi_2^*(\boldsymbol{\theta},\boldsymbol{Z})\neq\boldsymbol{\beta}_{1_0}'\boldsymbol{X}_1+\alpha_{1_0}\pi_2^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\Big\}>0$$
$$\Pr\Big\{\boldsymbol{\beta}_2'\boldsymbol{X}_2+\alpha_2\pi_1^*(\boldsymbol{\theta},\boldsymbol{Z})\neq\boldsymbol{\beta}_{2_0}'\boldsymbol{X}_2+\alpha_{2_0}\pi_1^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\Big\}>0.$$

As we will show below, if the previous assumptions are satisfied, then (S5) is sufficient for identification of $\boldsymbol{\theta}$. Define $\boldsymbol{W}\equiv(\boldsymbol{Y}',\boldsymbol{X}',\boldsymbol{Z}')'$. We will make a slight change in notation. Instead of using $\mathcal{F}(\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ as we did previously, we will now let $\mathcal{F}(\boldsymbol{W},\boldsymbol{\theta})$ denote the conditional probability function of $\boldsymbol{Y}$ given $(\boldsymbol{X},\boldsymbol{Z})$.

Using the results from Lemma 4.1, we know that $\mathcal{F}(\boldsymbol{W},\boldsymbol{\theta})$ exists and is well-defined for the four observable outcomes $\boldsymbol{Y}$ everywhere in $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$ and is given by

$$\mathcal{F}(\boldsymbol{W},\boldsymbol{\theta}) = G_1\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{\theta},\boldsymbol{Z})\big)^{Y_1}\big[1 - G_1\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]^{1-Y_1}$$
$$\times\, G_2\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1^*(\boldsymbol{\theta},\boldsymbol{Z})\big)^{Y_2}\big[1 - G_2\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]^{1-Y_2}.$$

By assumption (S1.3), we have that (S5) implies $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \;\Rightarrow \mathcal{F}(\boldsymbol{W},\boldsymbol{\theta}) \neq \mathcal{F}(\boldsymbol{W},\boldsymbol{\theta}_0)$ and by (S2.1), the structure of the model evaluated at $\boldsymbol{\theta}_0$ is not observationally equivalent to that evaluated at $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Consequently, $\boldsymbol{\theta}$ is globally identified in $\boldsymbol{\Theta}$ [15]. We can reinterpret assumption (S5) in terms of full-column rank condition of the matrices $\big(\boldsymbol{X}_1, \pi_1^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ and $\big(\boldsymbol{X}_2, \pi_2^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$. From assumption (I.3) we allow some elements of $\boldsymbol{X}_1$ or $\boldsymbol{X}_2$ to be included in $\boldsymbol{Z}$. In this case, assumption (S5) seems to rely on the nonlinearity of $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})$. We next examine a linear version of the model and show that even in the "worst case" scenario where $\boldsymbol{\pi}(\boldsymbol{\theta},\boldsymbol{Z})$ is a linear function of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, the parameter vector $\boldsymbol{\theta}$ can still be identified (condition (S5) is satisfied) by imposing a simple exclusion restriction. Lack of identification in a linear interactions-based model is known as the "reflection problem" and was first studied in Manski (1993). As we shall see next, a linear version of our game does not suffer from the reflection problem and therefore condition (S5) does not rely on the nonlinear nature of the equilibrium beliefs $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})$.

Identification and nonlinearity of $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})$

Suppose now that we momentarily drop assumptions (S1.2-3) and assume instead that $\varepsilon_1 \sim U[-1,1]$ and $\varepsilon_2 \sim U[-1,1]$. We also modify assumption (S2.3) and assume now that[16]

$$\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2 \in (-1,1) \quad \forall\, \boldsymbol{\theta}_1 \in \boldsymbol{\Theta},\, \forall\, \pi_2 \in [0,1],\, \forall\, \boldsymbol{X}_1 \in \mathbb{S}(\boldsymbol{X}_1)$$
$$\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1 \in (-1,1) \quad \forall\, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta},\, \forall\, \pi_1 \in [0,1],\, \forall\, \boldsymbol{X}_2 \in \mathbb{S}(\boldsymbol{X}_2).$$

Assumption (S3.2) now becomes simply $1 - (\alpha_1\alpha_2)/4 > 0\ \forall\, \boldsymbol{\theta} \in \boldsymbol{\Theta}$ which can be trivially

[15] See Definition 2.1 in Hsiao (1983).

[16] We will go back to our set of stochastic assumptions (S1)-(S3) immediately after this brief discussion.

re-expressed as $4-\alpha_1\alpha_2>0\ \forall\,\boldsymbol{\theta}\in\boldsymbol{\Theta}$. Take $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and $\boldsymbol{z}\in\mathbb{S}(\boldsymbol{Z})$, then the equilibrium probabilities $\pi^*(\boldsymbol{\theta},\boldsymbol{z})$ are the solution (for π_1 and π_2) to the pair of equations

$$\pi_1=\frac{E[\boldsymbol{X}_1\mid\boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_1+\alpha_1\pi_2+1}{2}\quad\text{and}\quad\pi_2=\frac{E[\boldsymbol{X}_2\mid\boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_2+\alpha_2\pi_1+1}{2},$$

which yields

$$\pi_1^*(\boldsymbol{\theta},\boldsymbol{z})=\frac{2\big[E[\boldsymbol{X}_1\mid\boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_1+1\big]+\alpha_1\big[E[\boldsymbol{X}_2\mid\boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_2+1\big]}{4-\alpha_1\alpha_2}$$
$$\pi_2^*(\boldsymbol{\theta},\boldsymbol{z})=\frac{2\big[E[\boldsymbol{X}_2\mid\boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_2+1\big]+\alpha_2\big[E[\boldsymbol{X}_1\mid\boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_1+1\big]}{4-\alpha_1\alpha_2}.$$

Therefore, we have

$$\boldsymbol{X}_1'\boldsymbol{\beta}_1+\alpha_1\pi_2^*(\boldsymbol{\theta},\boldsymbol{Z})=\delta_1+\boldsymbol{X}_1'\boldsymbol{\beta}_1+E[\boldsymbol{X}_1\mid\boldsymbol{Z}]'\boldsymbol{\gamma}_{1,1}+E[\boldsymbol{X}_2\mid\boldsymbol{Z}]'\boldsymbol{\gamma}_{1,2}$$
$$\boldsymbol{X}_2'\boldsymbol{\beta}_2+\alpha_2\pi_1^*(\boldsymbol{\theta},\boldsymbol{Z})=\delta_2+\boldsymbol{X}_2'\boldsymbol{\beta}_2+E[\boldsymbol{X}_1\mid\boldsymbol{Z}]'\boldsymbol{\gamma}_{2,1}+E[\boldsymbol{X}_2\mid\boldsymbol{Z}]'\boldsymbol{\gamma}_{2,2},$$

where δ_p is a function $\delta_p(\alpha_1,\alpha_2)$, $\boldsymbol{\gamma}_{p,1}$ is a function $\boldsymbol{\gamma}_{p,1}(\boldsymbol{\beta}_1,\alpha_1,\alpha_2)$ and $\boldsymbol{\gamma}_{p,2}$ is a function $\boldsymbol{\gamma}_{p,2}(\boldsymbol{\beta}_2,\alpha_1,\alpha_2)$ for $p\in\{1,2\}$. Note that the reduced forms given above are expressed in terms of $2(k+2)$ variables but we only have $k+2$ unknown parameters. We show in the Mathematical Appendix that a necessary and sufficient condition for identification of $\boldsymbol{\theta}$ is the existence of a pair of elements $X_{1,\ell_1}\in\boldsymbol{X}_1$ and $X_{2,\ell_2}\in\boldsymbol{X}_2$ such that $X_{1,\ell_1}\neq X_{2,\ell_2}$ and $\beta_{1,\ell_1}\neq 0$, $\beta_{2,\ell_2}\neq 0$. This simple exclusion restriction yields identification of all parameters -including constant terms in $\boldsymbol{X}_1$ and/or $\boldsymbol{X}_2$- even if $E[\boldsymbol{X}_2\mid\boldsymbol{Z}]=\boldsymbol{X}_2$ and $E[\boldsymbol{X}_1\mid\boldsymbol{Z}]=\boldsymbol{X}_1$. This shows that even in the "worst-case scenario" for identification in which equilibrium probabilities are linear functions of $\boldsymbol{X}$, we can still identify the parameter vector using a simple exclusion restriction. The nonlinear nature of the equilibrium probabilities that results from assumptions (S1) is not the source of identification in our model.

We now go back to our set of assumptions (S1)-(S4). Next, we describe the trimmed quasi maximum likelihood procedure to estimate the structural parameter $\boldsymbol{\theta}$.

4.4.2 Trimmed quasi maximum likelihood estimation

We estimate $\boldsymbol{\theta}$ in two steps. First, we estimate the unknown equilibrium probabilities (beliefs) $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$ incorporating the information about $\boldsymbol{\theta}$ contained in the equilibrium conditions (1). We then plug-in these estimators into a trimmed log-likelihood function and maximize it with respect to $\boldsymbol{\theta}$. Specifically, we study the properties of the estimators that result from plugging in either $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$ or $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$, both of which exploit all the information available about $\boldsymbol{\theta}$ from the equilibrium conditions (1). The trimmed set is $\boldsymbol{\mathcal{Z}}$, which -from assumption (S3.2)- yields uniqueness of equilibrium and also limits the influence of points in the boundary of $\mathbb{S}(\boldsymbol{Z})$. In a Section 4.6 we show how to modify the trimming if there is a unique equilibrium for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ -i.e, if $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$-.

This methodology is similar to that of Ahn and Manski, who studied a discrete choice model with uncertainty but without any element of strategic interaction. In their model there was no relationship to exploit between the unknown expectations and the parameter vector $\boldsymbol{\theta}$. Expectations were not derived from any equilibrium conditions. In our case, we plug-in semiparametric estimators that use the information contained in the BNE conditions of the game. As we did in Section 4.4.1, let $\mathcal{F}(\boldsymbol{W}, \boldsymbol{\theta})$ denote the conditional probability function of $\boldsymbol{Y}$ given $(\boldsymbol{X}, \boldsymbol{Z})$ and a particular value of $\boldsymbol{\theta}$. Define the trimmed conditional probability (likelihood) function $\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta}) = \mathcal{F}(\boldsymbol{W}, \boldsymbol{\theta})^{\mathbb{1}\{\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}\}}$. The next result shows that if (S5) holds -in addition to our previous assumptions-, then $\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta})$ satisfies the following information inequality result.

Lemma 4.4 *Suppose assumptions (I), (S1.1-2), (S2.1-2), (S3.2) and (S5) are satisfied, then* $E\left[\log \mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta})\right] < E\left[\log \mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta}_0)\right] \quad \forall\, \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,\ \boldsymbol{\theta} \in \boldsymbol{\Theta}.$

The proof can be found in the Mathematical Appendix. This result will prove to be useful to show consistency of our proposed estimator. Sharing a generic property of MLE problems, identification conditions will lead to consistency.

Let

$$\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}\big) = \mathbb{1}\{\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}\}\Big[Y_1 \log\, G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1+\alpha_1\pi_2)+(1-Y_1)\log\{1-G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1+\alpha_1\pi_2)\} \\ + Y_2 \log\, G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2+\alpha_2\pi_1)+(1-Y_2)\log\{1-G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2+\alpha_2\pi_1)\}\Big].$$

Note that $\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big) = \log\,\mathcal{F}_{\mathcal{Z}}(\boldsymbol{W},\boldsymbol{\theta})$ (the trimmed log-likelihood). The trimming index $\mathbb{1}\{\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}\}$ doesn't depend on $\boldsymbol{\theta}$. This was used to prove Lemma 4.4, and is also used (along with assumption (S2.1)) to show that the information identity applies to $\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ and we have

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

Details are shown in the appendix. Before proceeding, we will add the following assumption, which is standard in M-estimation problems:

(S6) 1.– The true parameter value $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$.

2.– The trimmed information matrix at $\boldsymbol{\theta}_0$,

$$\Im_{\mathcal{Z}} = -E\left[\frac{\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \quad \text{is invertible.}$$

We are ready to present the first proposed estimator. It uses the analog semiparametric estimator $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ as a plug-in. The corresponding estimator $\widehat{\boldsymbol{\theta}}$ is the solution to

$$\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{Max}}\ \frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}_n)\big).$$

Before outlining the asymptotic properties of $\widehat{\boldsymbol{\theta}}$, let $\nabla_{\boldsymbol{\theta}}\ell_{\mathcal{Z}}(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi})$ be the partial derivative of $\ell_{\mathcal{Z}}$ with respect to $\boldsymbol{\theta}$, with $\boldsymbol{\pi}$ constant. Let $\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi})$ be the partial derivative of $\ell_{\mathcal{Z}}$ with respect to $\boldsymbol{\pi}$, with $\boldsymbol{\theta}$ constant. Then, the score of our trimmed-log likelihood is given by

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

Now, let $\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}\big)/\partial\boldsymbol{\theta}\partial\boldsymbol{\pi}'$ denote the partial derivative of the score with respect to $\boldsymbol{\pi}$. Let $\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})$ be the expectation, conditional on $\boldsymbol{Z}$ of this cross-partial derivative evaluated at $\boldsymbol{\theta}_0$.

The exact expression for $\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})$ can be found in the appendix. As we have done throughout, let $J(\boldsymbol{\pi} \mid \boldsymbol{Z},\boldsymbol{\theta}) = \nabla_{\boldsymbol{\pi}}\big(\boldsymbol{\pi} - \varphi(\boldsymbol{\pi} \mid \boldsymbol{Z},\boldsymbol{\theta})\big)$ denote the Jacobian of the equilibrium conditions. We will define $J_0(\boldsymbol{Z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z}) \mid \boldsymbol{Z},\boldsymbol{\theta}_0\big)$ and $B_{\mathcal{Z}}(\boldsymbol{Z}) = \overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0(\boldsymbol{Z})^{-1}$. The next theorem provides the asymptotic properties of $\widehat{\boldsymbol{\theta}}$.

Theorem 1 *Suppose assumptions (I), (S1)-(S5) are satisfied and let* $\widehat{\boldsymbol{\theta}}$ *solve*

$$\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{Max}\ \frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}_n)\big),$$

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

(A) $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

(B) *If assumption (S6) is also satisfied, then:* $\sqrt{N}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \Im_{\mathcal{Z}}^{-1} + \Im_{\mathcal{Z}}^{-1}\Omega\Im_{\mathcal{Z}}^{-1}\big)$, *where*

$$\begin{aligned}\Omega &= E\Big[B_{\mathcal{Z}}(\boldsymbol{Z})E\Big[\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{Z}\big]\Big)\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{Z}\big]\Big)'\Big|\boldsymbol{Z}\Big]B_{\mathcal{Z}}(\boldsymbol{Z})'\Big] \\ &= E\Big[B_{\mathcal{Z}}(\boldsymbol{Z})\,Var\Big[E[\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z}]\Big|\boldsymbol{Z}\Big]B_{\mathcal{Z}}(\boldsymbol{Z})'\Big].\end{aligned}$$

The use of nonparametric methods to estimate the unknown equilibrium probabilities $\boldsymbol{\pi}^*(\cdot)$ increases the asymptotic variance by the term $\Im_{\mathcal{Z}}^{-1}\Omega\Im_{\mathcal{Z}}^{-1}$. If we knew exactly $f_{\boldsymbol{X},\boldsymbol{Z}}(\cdot)$, $f_{\boldsymbol{Z}}(\cdot)$ then we could solve (numerically) the equilibrium conditions (1), obtain the exact expression for $\boldsymbol{\pi}^*(\cdot)$ and the asymptotic variance would simply be $\Im_{\mathcal{Z}}$. The term $\overline{D}(\mathcal{Z})$ is a measure of interdependency between the problems of estimating the structural parameters $\boldsymbol{\theta}$ and the equilibrium probabilities (beliefs) $\boldsymbol{\pi}^*(\cdot)$. The assumption that the game is in equilibrium automatically relates both problems through the equilibrium conditions unless $\alpha_1 = \alpha_2 = 0$ in which case there is no strategic interaction between the players and $\overline{D}_{\mathcal{Z}}(\boldsymbol{Z}) = 0$ w.p.1. Consequently, if $\alpha_1 = \alpha_2 = 0$ then $B_{\mathcal{Z}}(\boldsymbol{Z}) = 0$, the asymptotic variance is simply $\Im_{\mathcal{Z}}$ and the estimation of $\boldsymbol{\theta}$ is adaptive (see Pagan and Ullah (1999), section 5.4 or Bickel (1982)). The term $J_0(\boldsymbol{Z})^{-1}\big(E[\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z}] - E[\boldsymbol{Y} \mid \boldsymbol{Z}]\big)$ is a linearization of the equilibrium conditions and is present because our semiparametric equilibrium probabilities estimators have an asymptotically linear representation.

The proof uses the results from Lemma 4.2. We go further by showing that if our assumptions are satisfied, then the objects described in such lemma have a uniform linear representation up to a term of order $o_p(\mathrm{N}^{-1/2})$. We combine this result with the first order conditions satisfied by $\widehat{\boldsymbol{\theta}}$ and rely on the properties of the Central Limit Theorem for U-Statistics (see Powell, Stock and Stoker (1989) or Pagan and Ullah, Appendix A.2). Details are a bit lengthy but are detailed in the accompanying Mathematical Appendix.

Efficiency:

The asymptotic variance of $\widehat{\boldsymbol{\theta}}$ satisfies the efficiency bound for the vector of moment conditions[17]

$$E\left[\frac{\partial \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})\big)}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}$$

$$E\Big[\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z}) - E[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}]\Big| \; \boldsymbol{Z}\Big] = \mathbf{0},$$

which is a combination of unconditional and conditional moment restrictions. These moment conditions summarize all relevant information about $\boldsymbol{\theta}$ contained in the model. Following the approach of Newey (1990), efficiency bounds for models with conditional moment restrictions can be found in Ai and Chen (2003). We apply their formulas in the Mathematical Appendix to find the efficiency bound for our model. This efficiency result should not come as a surprise, as the methodology is asymptotically equivalent to a constrained trimmed maximum likelihood estimation, where the constraint comes in the form of a conditional moment restriction. It is very important to note that the efficiency of $\widehat{\boldsymbol{\theta}}$ depends on the trimming set $\boldsymbol{\mathcal{Z}}$. In section 4.6 we will show how to make the asymptotic variance of $\widehat{\boldsymbol{\theta}}$ independent of any trimming set if the BNE is unique for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$.

[17] Recall that by definition, $\boldsymbol{\varphi}\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z}) \mid \boldsymbol{Z}, \boldsymbol{\theta}\big) = E\Big[E\big[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}\big] \;\Big|\; \boldsymbol{Z}\Big]$. See Equation 3.

Testing for uniqueness of equilibrium:

Our estimation procedure allows us to test sufficient conditions for uniqueness of equilibrium. First we show how to test if the BNE is unique for a given realization $\boldsymbol{Z} = \boldsymbol{z}$. Using the results from Lemma 4.2 and Theorem 1, it is not hard to show that if $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$ then

$$(Nh_N^L)^{1/2}\Big(\widehat{\delta}_{1_N}\big(\widehat{\boldsymbol{\pi}^*_{2_N}}(\widehat{\boldsymbol{\theta}},\boldsymbol{z}) \mid \boldsymbol{z},\widehat{\boldsymbol{\theta}}_1\big)\widehat{\delta}_{2_N}\big(\widehat{\boldsymbol{\pi}^*_{1_N}}(\widehat{\boldsymbol{\theta}},\boldsymbol{z}) \mid \boldsymbol{z},\widehat{\boldsymbol{\theta}}_2\big) - \delta_1\big(\boldsymbol{\pi}^*_2(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_{1_0}\big)\delta_2\big(\boldsymbol{\pi}^*_1(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_{2_0}\big)\Big)$$
$$\xrightarrow{d} \mathcal{N}\big(\mathbf{0},\mathrm{V}(\boldsymbol{z})\big),$$

where $\mathrm{V}(\boldsymbol{z})$ is a variance that depends on $\boldsymbol{z}$. Using this result we can construct a pivotal statistic to test the hypothesis $H_0 : \delta_1\big(\boldsymbol{\pi}^*_2(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_{1_0}\big)\delta_2\big(\boldsymbol{\pi}^*_1(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_{2_0}\big) = \kappa$ against the one-sided alternative $H_1 : \delta_1\big(\boldsymbol{\pi}^*_2(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_{1_0}\big)\delta_2\big(\boldsymbol{\pi}^*_1(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_{2_0}\big) > \kappa$. Failing to reject H_0 for some $\kappa < 1$ would be tantamount to failing to reject the hypothesis that equilibrium is unique when $\boldsymbol{Z} = \boldsymbol{z}$, or that $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Note that our pivotal statistic suffers from the so-called curse of dimensionality.

Using the results from Corollary 1, we can test for uniqueness of equilibrium everywhere in $\mathbb{S}(\boldsymbol{Z})$ by testing the hypothesis $H_0 : \alpha_1\alpha_2 = 1/(\overline{g}_1\overline{g}_2)$ against the one-sided alternative $H_1 : \alpha_1\alpha_2 < 1/(\overline{g}_1\overline{g}_2)$. In this case, rejecting the null hypothesis would be evidence that the game has a unique equilibrium for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$. However, failure to reject H_0 is not automatically indicative that the game has multiple equilibria for some realization of $\boldsymbol{Z}$ since the condition of Corollary 1 is sufficient, but not necessary for uniqueness to hold everywhere in $\mathbb{S}(\boldsymbol{Z})$. Due to the results from Theorem 1, the pivotal statistic used to test this hypothesis does not suffer from the curse of dimensionality since $\sqrt{N}\big(\widehat{\alpha}_1\widehat{\alpha}_2 - \alpha_1\alpha_2\big)$ is asymptotically normal with mean zero.

Next, we examine the properties of the trimmed quasi maximum likelihood estimator that uses the two-step linearized estimator $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{Z})$ as the plug-in. First, define

$$\widetilde{\mathcal{F}}(\boldsymbol{W},\boldsymbol{\theta}) = G_1\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\rho_2(\boldsymbol{\theta},\boldsymbol{Z})\big)^{Y_1}\big[1 - G_1\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\rho_2(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]^{1-Y_1}$$
$$\times G_2\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\rho_1(\boldsymbol{\theta},\boldsymbol{Z})\big)^{Y_2}\big[1 - G_2\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\rho_1(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]^{1-Y_2}.$$

Note that since $\rho(\boldsymbol{\theta}_0, \boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})$, we have $\widetilde{\mathcal{F}}(\boldsymbol{W}, \boldsymbol{\theta}_0) = \mathcal{F}(\boldsymbol{W}, \boldsymbol{\theta}_0)$ (the true conditional likelihood function). We will let $\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta}) = \widetilde{\mathcal{F}}(\boldsymbol{W}, \boldsymbol{\theta})^{\mathbb{1}\{\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}\}}$. If (S1.1-2) and (S2.1-2) are satisfied, then assumption (S3.2) precludes the situation $\boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Therefore, if (S5) is also satisfied we have that conditional on $\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}$, if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ with $(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \in \boldsymbol{\Theta}$ then $\Pr\left\{\boldsymbol{\beta}_1' \boldsymbol{X}_1 + \alpha_1 \rho_2(\boldsymbol{\theta}, \boldsymbol{Z}) \neq \boldsymbol{\beta}_{1_0}' \boldsymbol{X}_1 + \alpha_{1_0} \pi_2^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\right\} > 0$ and $\Pr\left\{\boldsymbol{\beta}_2' \boldsymbol{X}_2 + \alpha_2 \rho_1(\boldsymbol{\theta}, \boldsymbol{Z}) \neq \boldsymbol{\beta}_{2_0}' \boldsymbol{X}_2 + \alpha_{2_0} \pi_1^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\right\} > 0$. The next result is parallel to Lemma 4.4 and shows that $\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta})$ also satisfies an information-inequality result.

Lemma 4.5 *Suppose assumptions (I), (S1.1-2), (S2.1-2), (S3.2) and (S5) are satisfied, then*
$E\left[\log \widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta})\right] < E\left[\log \widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}, \boldsymbol{\theta}_0)\right] \quad \forall\, \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,\ \boldsymbol{\theta} \in \boldsymbol{\Theta}.$

The proof is included in the Mathematical Appendix. It relies on the nonzero probabilities described above and the fact that $\widetilde{\mathcal{F}}(\boldsymbol{W}, \boldsymbol{\theta}_0) = \mathcal{F}(\boldsymbol{W}, \boldsymbol{\theta}_0)$ everywhere in $\boldsymbol{\mathcal{Z}}$. We now study the properties of the estimator that uses the linearized semiparametric estimator $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$ as a plug-in. We denote this estimator $\widetilde{\boldsymbol{\theta}}$, which is the solution to

$$\operatorname*{Max}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{N} \sum_{n=1}^{N} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}, \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}_n)\big).$$

The next theorem presents the main result for $\widetilde{\boldsymbol{\theta}}$.

Theorem 2 *Suppose assumptions (I), (S1)-(S5) are satisfied. Let $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$ be as defined in Section 4.3.3, let $\widehat{\boldsymbol{\theta}}$ be as defined in Theorem 1 and let $\widetilde{\boldsymbol{\theta}}$ solve*

$$\operatorname*{\mathit{Max}}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{N} \sum_{n=1}^{N} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}, \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}_n)\big).$$

Then,

(A) $\widetilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

(B) *If assumption (S6) is also satisfied, then $\sqrt{N}\big(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}\big) \xrightarrow{p} \mathbf{0}$ and consequently $\sqrt{N}\big(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)$ has the asymptotic distribution given in Theorem 1 (B).*

This result shows that under the same set of assumptions, using the $\widehat{\boldsymbol{\pi}}_N^*(\cdot)$ or $\widetilde{\boldsymbol{\pi}}_N^*(\cdot)$ as plug-ins yields an estimator for $\boldsymbol{\theta}$ with the same asymptotic properties (to a first order of approximation). This extends to the discussions about efficiency and tests for uniqueness of equilibrium. Our analog semiparametric estimator $\widehat{\boldsymbol{\pi}}(\boldsymbol{\theta}, \boldsymbol{z})$ replicates asymptotically all the characteristics of $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ as a function of $\boldsymbol{\theta}$. As the proofs of Theorems 1 and 2 show, if our assumptions are satisfied then it is sufficient to achieve those properties asymptotically only to a first order of approximation. The computational advantage of $\widetilde{\boldsymbol{\pi}}_N^*(\cdot)$ makes it more attractive than its analog semiparametric counterpart $\widehat{\boldsymbol{\pi}}_N^*(\cdot)$. The proof relies on the results of Lemma 4.3 and follows linearization steps parallel to those used for the proof of Theorem 1. All details can be found in the Mathematical Appendix.

Until now, we have assumed that beliefs are constructed conditional on a vector of continuously distributed signals $\boldsymbol{Z}$. In a number of economic situations, $\boldsymbol{Z}$ may include variables with finite support (e.g, categorical variables). The next section states conditions under which the results from Theorems 1 and 2 can be extended to the case in which $\boldsymbol{Z}$ includes discrete random variables. These conditions simply require that assumptions (S2) and (S4) be appropriately modified.

4.5 Discrete conditioning signals $\boldsymbol{Z}$

Suppose we drop assumption (S2.1-2) and assume now that $\boldsymbol{Z}$ has finite support. Then, the following results hold:

Corollary 2 Suppose $\boldsymbol{Z}$ has a finite support and we drop assumptions (S2) and (S4) and modify (S3.2) correspondingly to assume now that $\boldsymbol{\mathcal{Z}}$ is a subset of elements in $\mathbb{S}(\boldsymbol{Z})$ all of which have strictly positive probability. Then the conclusions of Theorems 1 and 2 hold if for all $\boldsymbol{z}$ we replace $K_h(\boldsymbol{z}_n - \boldsymbol{z})$ with the indicator function $\mathbb{1}\{\boldsymbol{z}_n = \boldsymbol{z}\}$.

Corollary 3 Suppose $\boldsymbol{Z}$ can now be partitioned as $\boldsymbol{Z} = \left(\boldsymbol{Z}^{d'}, \boldsymbol{Z}^{c'}\right)'$, where $\boldsymbol{Z}^d \in \mathbb{R}^{L^d}$ has finite support and $\boldsymbol{Z}^c \in \mathbb{R}^{L^c}$ is continuously distributed. Suppose we replace $\boldsymbol{L}$ with $\boldsymbol{L^c}$ in all our assumptions and modify (S3.2) to assume now that $\boldsymbol{\mathcal{Z}}$ is a subset of $\mathbb{S}(\boldsymbol{Z})$ such that $f_{\boldsymbol{Z}^c}(\boldsymbol{z}^c) > 0$ and $\Pr\left(\boldsymbol{Z}^d = \boldsymbol{z}^d \mid \boldsymbol{Z}^c = \boldsymbol{z}^c\right) > 0$ for all $\boldsymbol{z} = \left(\boldsymbol{z}^{d'}, \boldsymbol{z}^{c'}\right)' \in \boldsymbol{\mathcal{Z}}$. Then the conclusions of Theorems 1 and 2 hold if for all $\boldsymbol{z}$ we replace $K_h\left(\boldsymbol{z}_n - \boldsymbol{z}\right)$ with $K_h\left(\boldsymbol{z}_n^c - \boldsymbol{z}^c\right)\mathbb{1}\left\{\boldsymbol{z}_n^d = \boldsymbol{z}^d\right\}$.

The proofs can be found in the Mathematical Appendix. Both of them rely on straightforward variations of the arguments used to prove Theorems 1 and 2. These results show that -if appropriately adapted- the methodology presented here is flexible enough to handle situations in which $\boldsymbol{Z}$ includes a mixture between continuous and discrete random variables. Note that we preserve the trimming index $\mathbb{1}\left\{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}\right\}$ even if $\boldsymbol{Z}$ includes only discrete random variables because it is the set in which the likelihood function is well defined. If equilibrium were unique everywhere in $\mathbb{S}(\boldsymbol{Z})$ (e.g, if $\alpha_1 \times \alpha_2 < 1/(\overline{g}_1\overline{g}_2)$), then trimming would not be necessary if $\boldsymbol{Z}$ included only discrete random variables. In this case, the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ would not depend on any trimming set. The case in which $\boldsymbol{Z}$ includes continuous and discrete random variables and equilibrium is unique everywhere in $\mathbb{S}(\boldsymbol{Z})$ is covered in the discussion of Section 4.6 (below).

4.6 Trimming when equilibrium is unique for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$

The expression for the asymptotic variance in Theorems 1 and 2 depends on the trimming set $\boldsymbol{\mathcal{Z}}$. As a consequence of the positive definiteness of $\Im_{\boldsymbol{\mathcal{Z}}}$ and the positive semi-definiteness of $B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z})\mathrm{Var}\left[E[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}]\middle|\boldsymbol{Z}\right]B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z})'$, we get that $\Im_{\boldsymbol{\mathcal{Z}}}^{-1}$ decreases -in the positive definite sense- and Ω increases as the set $\boldsymbol{\mathcal{Z}}$ increases. The overall effect on the asymptotic variance of our estimator(s) for $\boldsymbol{\theta}$ cannot be readily characterized. As we mentioned above, based on assumption (S3.2), we use the trimming set $\boldsymbol{\mathcal{Z}}$ to achieve two things: First, it allows us to remain in the subset of $\mathbb{S}(\boldsymbol{Z})$ where equilibrium is unique, the conditional likelihood is well defined and the results from Lemma 4.1 hold. Second, it also helps us limit the influence of points $\boldsymbol{z}$ in the boundary of $\mathbb{S}(\boldsymbol{Z})$. If equilibrium is unique for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$, we would like to modify the trimming in such a way that it still limits the influence of points in the boundary of $\mathbb{S}(\boldsymbol{Z})$, but $\boldsymbol{\mathcal{Z}} \to \mathbb{S}(\boldsymbol{Z})$, so that the distribution of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ do not depend on any trimming set. In other words, we would like to use all the information on $\mathbb{S}(\boldsymbol{Z})$, while avoiding the influence of points in its boundary. The proposal is to use the trimming index $\mathbb{1}\{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n) > b_N\}$ for an appropriately chosen sequence b_N. The following corollary extends the results of Theorems 1 and 2 to the case $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$ and makes the asymptotic distributions of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ independent from a trimming set.

Corollary 4 Suppose assumption (S3.2) holds everywhere in $\mathbb{S}(\boldsymbol{Z})$ (i.e, $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$) and the other assumptions hold as stated. Let ε be as defined in assumption (S4.2) and let b_N be a sequence that satisfies $b_N^2\left(N^{1-2\varepsilon}h_N^{2L}\right)^{1/4} \to \infty$. Take the set $\boldsymbol{\mathcal{Z}}_{b_N} = \left\{\boldsymbol{z} \in \mathbb{R}^L : f_{\boldsymbol{Z}}(\boldsymbol{z}) \geq b_N\right\}$ and define $\boldsymbol{z}^*_{b_N} = \sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}} \|\boldsymbol{z}\|$. Suppose that $\log\left(\boldsymbol{z}^*_{b_N}\right) = o_p(\mathrm{N}^{\varepsilon})$. Suppose the trimmed quasi maximum likelihood optimization is performed using the trimming index $\mathbb{1}\{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n) > b_N\}$. Then, the results of Theorems 1 and 2 hold with $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$.

This result includes the case in which b is a fixed, but also shows that if $b \to 0$ slowly enough, the asymptotic variance of our quasi-maximum likelihood estimators does not depend on any trimming set $\boldsymbol{\mathcal{Z}}$. If $b \to 0$ at a sufficiently slow rate, the limiting trimming set is $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$. This is a very convenient result, since -as we outlined in Corollary 1-, very simple, straightforward conditions for α_1 and α_2 guarantee uniqueness of equilibrium in $\mathbb{S}(\boldsymbol{Z})$ (e.g $\alpha_1 \times \alpha_2 < 1/(\overline{g}_1\overline{g}_2)$). The assumption $\log\left(\boldsymbol{z}^*_{b_N}\right) = o_p(\mathrm{N}^{\varepsilon})$ is related to the behavior of tails of $f_{\boldsymbol{Z}}(\boldsymbol{z})$. Given the condition $b_N^2\left(N^{1-2\varepsilon}h_N^{2L}\right)^{1/4} \to \infty$, a sufficient condition for $\log\left(\boldsymbol{z}^*_{b_N}\right) = o_p(\mathrm{N}^{\varepsilon})$ to hold is that the tails of $f_{\boldsymbol{Z}}(\cdot)$ go to zero geometrically or even polynomially with $\|\boldsymbol{z}\|$. It will also be satisfied if $E\left(\|\boldsymbol{Z}\|^{\lceil\varepsilon^{-1}\rceil}\right)$ is finite, where $\lceil\cdot\rceil$ denotes the ceiling function. However, this existence-of-moments requirement is not necessary[18]. The proof first shows that the steps that lead to the proofs of Theorems 1 and 2 can be extended to the set $\boldsymbol{\mathcal{Z}}_N \times \boldsymbol{\Theta}$, then we use a result recently shown by Ichimura (2004) to show that $\Pr\Big(\sup_n \Big|\mathbb{1}\{\widehat{f}_N(\boldsymbol{z}_n > b_N)\} - \mathbb{1}\{f(\boldsymbol{z}_n > b_N)\}\Big| \neq 0\Big) \to 0$. Lastly, note that if (S3.2) is not satisfied everywhere in $\mathbb{S}(\boldsymbol{Z})$ then letting $b \to 0$ would eventually lead to regions of $\mathbb{S}(\boldsymbol{Z})$ with multiple equilibria, where the conditional likelihood function is not well defined.

We have completed the discussion about the estimation of the game. We next present an example that applies this methodology to a simple investment game under uncertainty.

5 Example: A simple investment game under uncertainty

5.1 Short-term investment decisions as a simultaneous game

The arguments presented here are based on the Real Options approach to capital investment. This approach is a refinement to the traditional Net Present Value criteria of investment decisions. It was first proposed by Dixit and Pindyck (1994). We now summarize the basic premises of this theory. At any point in time there exist a finite number of investment opportunities for firms in a given industry. Having an investment opportunity is much like holding a financial "call" option: it gives the firm the right but not the obligation to "buy an asset", such asset in this case is the entitlement to the stream of profits from the investment opportunity in question. Exercising an option (capturing an investment opportunity) is an irreversible action. Thus, analyzing firms' investment decisions is equivalent to studying how firms exercise their options optimally. The driving force behind firms' investment decisions

[18]Take for example the univariate Cauchy density function $f(\boldsymbol{z}) = \{\pi(1+\boldsymbol{z})^2\}^{-1}$ which has no finite moments. Then $f(\boldsymbol{z}) = b_N$ if $\boldsymbol{z} = -1 \pm \sqrt{(b_N\pi)^{-1}}$. We have $N^{-\varepsilon}\log\left(\left|-1 \pm \sqrt{(b_N\pi)^{-1}}\right|\right) \to 0$ since $b_N N^{\varepsilon} \to \infty$. This in turn implies that $\log(\boldsymbol{z}^*_{b_N}) = o_p(\mathrm{N}^{\varepsilon})$.

is uncertainty about their future environment. A firm's expectation about the future market and/or technological conditions in its industry influence its decision about whether to delay or exercise the available investment option (capture the investment opportunity). Equally important is the firm's expectation about its rivals' actions: when a firm triggers an investment, future opportunities may be reduced or even lost for competing firms. This reduces the ability of firms to delay investment.

Strategic considerations can make it imperative for a firm to try and capture an investment opportunity and preempt investment by existing or potential competitors. At any point in time, this generates an incentive for firms to conceal their investment decisions from their competitors in an effort to prevent them from knowing which opportunities they are trying to exploit. In addition, each firm must act before having perfect knowledge about its opponents' investment plans, to prevent the investment opportunity from being taken away by a rival firm. In the long run of course, the investment opportunities captured by each firm become publicly observed. However, at any point in time short term investment decisions can be analyzed as a simultaneous game in which firms decide in advance which opportunities they will try to capture, before their rivals' investment projects become publicly observed. Thus, the principles of the Real Options approach to capital investment are consistent with the notion that short-term investment decisions can be modelled as a simultaneous game.

5.2 Timing of financial disclosure as a source of incomplete information

Aside from the simultaneous-game feature, an application to the methodology presented here requires the presence of an element of incomplete information. Publicly traded firms are required by law to disclose their financial statements periodically. However, such information is always made public long after the fact. This information lag has significant real-world implications, which were addressed by the Chairman of the U.S Securities and Exchange

Commission, Mr. Harvey Pitt in an "Op-Ed" for the Wall Street Journal on December 11, 2001:

> "Our current reporting and financial disclosure system has needed improvement and modernization for quite some time. Disclosures to investors are now required only quarterly or annually, and even then are issued long after the quarter or year has ended. This creates the potential for a financial 'perfect storm'. Information investors receive can be stale on arrival and mandated financial statements are often arcane and impenetrable."

A result of the existing regulations is that publicly traded firms compete against each other in an incomplete information environment. At any point in time a firm knows more about itself than the public does. By the same token, firms must anticipate their competitors' actions using incomplete -lagged- information about them. We are now ready to present the details of our proposed example.

5.3 Description of proposed example: Population, actions, timing and variables involved

In this section we provide a description of the game including the population who plays it, the actions to choose, the payoff variables $\boldsymbol{X}$, the conditioning signals $\boldsymbol{Z}$ as well as which elements of $\boldsymbol{X}$ are privately observed.

5.3.1 Population of players

The game is played in manufacturing industries that have only two publicly traded firms. All results should be interpreted as being conditional on that population of industries. Player 1 will be the firm with the largest market share and Player 2 will be the other firm. For a given industry, we could think of Player 1 as "leader" and Player 2 as "follower". In concordance with the game described in the previous sections, we assume that all leaders have the same payoff functions, as do all the followers. As we will see below, we will include a vector of industry-specific variables to control for structural differences among industries that influence firms' investment decisions.

5.3.2 Actions, timing, variables and information

At the end of a given year t, players (firms) are assumed to plan for year $t+1$. Specifically, we will assume that at the end of year t firms pre-commit to being "aggressive" or "passive" in their investment decision for year $t+1$. For firm $p \in \{1,2\}$ we will let $K_p(t)$ denote its total real stock of capital (physical and human) at the end of year t. At the end of Section 5.3.3 we explain how $K_p(t)$ was constructed. Let $\Delta\%K_p(t) = \big(K_p(t) - K_p(t-1)\big)/K_p(t-1)$. We will say that firm $p \in \{1,2\}$ decides to be aggressive in $t+1$ if both $\Delta\%K_p(t+1) > 0$ and $\Delta\%K_p(t+1) > \Delta\%K_p(t)$. We will let $Y_p(t+1) = 1$ if firm $p \in \{1,2\}$ is aggressive in year $t+1$ and let $Y_p(t+1) = 0$ otherwise. We will assume the game can be parameterized as the one analyzed in the previous sections of this paper. Next, we describe the payoff variables $\boldsymbol{X}$.

Payoff variables $\boldsymbol{X}$ They contain two kinds of variables: firm-specific and industry-specific characteristics. Firm-specific characteristics are publicly disclosed well after the end of the fourth quarter of year t, but are assumed to be privately known when firms make their choices: firms take advantage of financial disclosure regulations to devise their strategies before issuing its financial statements for the fourth quarter of the year. Let $Q_p(t)$ be firm p's Tobin's Q at the end of tear t and denote $\Delta Q_p(t) = Q_p(t) - Q_p(t-1)$. Tobin's Q compares the capitalized value of the marginal investment to its replacement cost. As such, it is effectively a measure of the firm's cost of capital. Investment theory asserts that increases in Q are accompanied by positive changes in investment. Tobin's Q' model of investment is formally equivalent to a theory of investment with marginal adjustment costs. We will include $\Delta Q_p(t)$ among the firm-specific payoff variables. At the end of Section 5.3.3 we explain how $Q_p(t)$ was constructed.

As we mentioned above, firms' strategies also have a long-term component since large-scale investment projects may take a long time to complete. To capture long-term strategies, we will include two firm-specific variables. We will include $Y_p(t)$ to indicate whether or not firm p was aggressive during year t. Adding in this variable also allows for the possibility of "state-dependent" strategies. We will also include $\Delta\%K_p(t)$ to capture the effect of the relative magnitude of changes in real capital on the firm's propensity to act aggressively. Including this variable also allows us to control for long-term investment strategies that involve large scale changes in capital. Let t_3 denote the end of the third quarter of year t. All the firm-specific variables presented here can be computed with financial data that is made public quarterly. We will assume that due to financial disclosure regulations, at the time choices are made $\Delta Q_p(t)$, $Y_p(t)$ and $\Delta\%K_p(t)$ are privately observed by firm p, but $\Delta Q_p(t_3)$, $Y_p(t_3)$ and $\Delta\%K_p(t_3)$ are publicly observable, where $\Delta Q_p(t_3) = Q_p\big(t_3\big) - Q_p\big((t-1)_3\big)$ and so on. Details of the construction of all firm-specific variables can be found in the appendix.

We will let I denote the industry to which firms 1 and 2 belong. In fact, I indexes a particular pair of players (firms). Following real-options investment theory, we want to include variables that reflect industry uncertainty [19]. We will include two measures of uncertainty:. First an indicator of market (demand) uncertainty known to both firms, which we will denote by $\mathrm{MKT}_\mathrm{I}(t)$. We use total real industry sales to construct a simple categorical variable that reflects two possible states: $\mathrm{MKT}_\mathrm{I}(t) = 1$ if the proportional annual growth of the industry's total real sales in the last year is greater than the average of the five previous years. $\mathrm{MKT}_\mathrm{I}(t) = 0$ otherwise. We also include a simple indicator of technological uncertainty which we denote by $\mathrm{TECH}_\mathrm{I}(t)$. We use data on patents and construct once

[19] Oriani, O. and M. Sobrero (2003) include measures of industry uncertainty in a model of firms' technological knowledge.

again a categorical variable for two possible states: $\mathrm{TECH_I}(t) = 1$ if the change in the number of patents (per employee) where industry I is considered either "industry of use" or "industry of manufacture" in the last year is greater than the average of the five previous years. $\mathrm{TECH_I}(t) = 0$ otherwise. Both measures of industry uncertainty are known to firms 1 and 2. Details of the construction of all firm-specific variables can be found in the appendix.

Conditioning signals $\boldsymbol{Z}$ As we have mentioned throughout, delay in financial information disclosure is the source for incomplete information in this model. All the firm-specific variables included here are affected by this information lag. On the other hand, the industry-specific uncertainty measures are assumed to be observed equally well by both firms. Therefore, the vector of conditioning signals $\boldsymbol{Z}$ is given by

$$\boldsymbol{Z} = \Big(\Delta Q_1(t_3), \Delta Q_2(t_3), \Delta\%K_1(t_3), \Delta\%K_2(t_3), Y_1(t_3), Y_2(t_3), \mathrm{MKT_I}(t), \mathrm{TECH_I}(t)\Big).$$

Consequently, the game can be summarized as

$$Y_p(t+1) = \mathbb{1}\Big\{\boldsymbol{X}(t)'\boldsymbol{\beta}_p - \alpha_p \pi^*_{-p}(\boldsymbol{\theta}, \boldsymbol{Z}) - \varepsilon_p(t) > 0\Big\} \text{ for } p \in \{1, 2\}$$

where $\boldsymbol{X}(t)'\boldsymbol{\beta}_p = \beta_{p,1}\Delta Q_p(t) + \beta_{p,2}\Delta\%K_p(t) + \beta_{p,3}Y_p(t) + \beta_{p,4}\mathrm{MKT_I}(t) + \beta_{p,5}\mathrm{TECH_I}(t)$ and $\boldsymbol{Z}$ is as described above.

5.3.3 Data source

Firm data

The source of all firm-specific variables is Standard & Poor's North America COMPUSTAT. These data files include information on publicly traded firms. These data covered approximately ninety percent of the employment in the manufacturing sector in 1995 but only about 1 percent of the total number of firms. To preserve the iid data assumption (S4.3), we pool together samples that are five years apart. The universe we use correspond to the years 1990 and 1995. For each year, COMPUSTAT included information for manufacturing firms in 458 industries at NAICS 6-digit level. Out of this universe, a total of 134 industries in 1990 and 1995 had only two publicly traded firms (roughly 14% of all industries in both years). This is the sample of firms we use. COMPUSTAT includes correspondences between NAICS 6-digit and 1987 4 digit SIC classifications for each firm.

Industry data

We used the following sources of industry data:

- NBER-CES Manufacturing Industry Database. This data set contains annual industry-level data on output, employment, payroll and other input costs, investment, capital stocks, TFP, and various industry-specific price indexes. The database covers all 4-digit manufacturing industries from 1958-1996. Detailed documentation is presented in Bartelsman and Gray (1996). This database was used to construct a price index for physical and R&D capital. The latter also used information from the National Science Foundation, as described below.

- National Science Foundation Industrial Research and Development Information System (IRIS). The National Science Foundation's (NSF) Industrial Research and Development Information System (IRIS) links an online interface to a historical database with more than 2,500 statistical tables containing all industrial research and development (R&D) data published by NSF since 1953. These tables are drawn from the results of NSF's annual Survey of Industrial Research and Development, the primary source for national-level data on U.S. industrial R&D. We used historical information on the composition of R&D costs, classified as "Wages", "Materials and Supplies" and "Other Costs". This information, combined with the price indexes in the NBER-CES Manufacturing Industry Database was used to construct a price index for R&D expenditures. The information exists at 2 and 4-digit SIC.

- US Historical Patent Set. This data set was put together by Daniel K.N. Johnson, it includes historical information on patents, classified by industry of use and industry of manufacture. It is based on the Wellesley Technology Concordance which uses information from over 1,500,000 patents granted in the US between 1975 and 1995 to build a concordance between the US Patent Classification system (USPC) and the International Patent Classification system IPC , with the probability that any given patent in a particular USPC will fall into a particular IPC. This data set contains information at 2-digit level SIC.

Next we present a brief description of how Tobin's Q and total capital stock were constructed for each firm.

Construction of Tobin's Q. We define market value as a firm's total market capitalization (the total market value of firm's outstanding securities). Following (Blundell et al., 1992, 1999) previous analysis on UK data , we calculate the market value of the firm adding the value of outstanding debt to market capitalization, therefore market value is the sum of a firm's common equity, preferred stock and outstanding debt. We follow Hall (1987) by defining the value of a firm's physical assets (a firm's book value) as the sum of net capital stock (net value of a firm's plant, property and equipment), inventories and other assets.

Construction of total capital stock. We construct a series of total real capital (R&D + Physical) by the end of year t for each firm. We deflated the net value of "property, plant and equipment" (COMPUSTAT) to measure the stock of physical capital of each firm. The stock of R&D capital -which we also call "human capital"- was computed as a perpetual inventory of the past real R&D expenditures (COMPUSTAT) with a constant depreciation rate, as described in detail by Mairesse and Griliches (1984) and Hall (1990). We used the R&D depreciation rates of Nadiri and Prucha (1993). R&D as well as physical capital price indices were constructed for each industry, at a 2-digit SIC level for the former and 4-digit SIC level for the latter. The composition of R&D costs in each industry was obtained from IRIS, while the composition for physical costs was obtained from NBER-CES.

5.3.4 Estimation results

The conditioning signals assumed to have a continuous distribution are: $\Delta Q_1(t_3)$, $\Delta Q_2(t_3)$,$\Delta\%K_1(t_3)$ and $\Delta\%K_2(t_3)$, so $L = 4$. The remaining signals $Y_1(t_3)$, $Y_2(t_3)$,$\mathrm{MKT_I}(t)$ and $\mathrm{TECH_I}(t)$ are categorical variables and thus clearly discrete. We estimate the model using the quasi maximum likelihood algorithm described above, with the modification outlined in Corollary 3 and using the linearized two-step semiparametric estimator as plug-in. In addition to the payoff variables $\boldsymbol{X}$ described above, a time dummy variable was included for $t = 1995$. Finally, we assumed that ε_1 and ε_2 were both normally distributed. We used a fifth-order multiplicative kernel of the form $K(\Psi) = k(\psi_1) \times \ldots \times k(\psi_L)$, where

$$k(\psi) = \frac{35}{32}\Big(\frac{5}{2} - 15\psi^2 + \frac{33}{2}\psi^4\Big)\phi(\psi) \quad \text{with} \quad \phi(\psi) = \frac{3}{4}(1-\psi^2)\mathbb{1}\Big\{\mid \psi \mid \leq 1\Big\}.$$

In addition to satisfying assumption (S4.2), following an idea by Fukunaga, we choose the bandwidth h_N in a way to keep the ratio $(\boldsymbol{z}_n - \boldsymbol{z}_m)/h_N$ "scale-free", we selected $h_N = \left|S^2(\boldsymbol{Z})\right|^{1/2} N^{\sigma}$, where $\left|S^2(\boldsymbol{Z})\right|$ is the determinant of the sample variance-covariance matrix $S^2(\boldsymbol{Z})$. We chose $\sigma = -1/12$, which satisfies assumption (S4.2) for any $\varepsilon \in (0, 1/6)$ with $L = 4$, $M \equiv L + 1 = 5$. We assumed a priori that condition (S3.2) is satisfied everywhere in $\mathbb{S}(\boldsymbol{Z})$. In concordance with this assumption and following the discussion in Section 4.6, the trimming index used was $\mathbb{1}\{\widehat{f}_{\boldsymbol{Z}}(\boldsymbol{z}_n) \geq 1/\log(N)\}$ where $N = 134$ is the sample size, see Section 5.3.3 above. We next describe how standard errors were obtained.

Estimation of standard errors

The correction matrix Ω was estimated as follows. First let

$$\widehat{\mathrm{Var}}\Big(E\big[Y \mid \boldsymbol{X}, \boldsymbol{Z}\big]\Big|\boldsymbol{Z} = \boldsymbol{z}_n\Big) = \frac{1}{N}\sum_{m=1}^{N}\bigg[\Big(\widehat{E}\big[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n\big] - \widehat{E}\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)\Big(\widehat{E}\big[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n\big] - \widehat{E}\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)'\bigg].$$

We used $\widehat{\Omega} = \frac{1}{N}\sum_{n=1}^{N} \widehat{B}_{\mathcal{Z}}(\boldsymbol{z}_n)\widehat{\mathrm{Var}}\Big(E\big[Y \mid \boldsymbol{X}, \boldsymbol{Z}\big]\Big|\boldsymbol{Z} = \boldsymbol{z}_n\Big)\widehat{B}_{\mathcal{Z}}(\boldsymbol{z}_n)'$, where our estimate for $\boldsymbol{\theta}$ was used to compute $\widehat{B}(\boldsymbol{z}_n)$. The information matrix $\Im_{\mathcal{Z}}$ was estimated as the numerical Hessian of $\frac{1}{N}\sum_{n=1}^{N} \ell_{\mathcal{Z}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)$. Table 1 summarizes the estimation results.

Table 1.

Estimation Results

	Player 1	**Player 2**
ΔQ	0.8124^*	0.7456^*
	(0.1007)	(0.1220)
$\triangle\%K$	-1.3721^*	-1.4106^*
	(0.5781)	(0.2823)
Y	0.1526	0.1411
	(0.1277)	(0.2044)
MKT	1.6147^*	0.9725^*
	(0.2292)	(0.3260)
$TECH$	0.9022^*	0.4121
	(0.2158)	(0.3575)
α	-1.1446	-2.1598^*
	(0.7996)	(0.5330)

Standard errors are shown in parenthesis. A superscript $(*)$ denotes statistical significance at a 95% level. The estimate for the time dummy variable for 1995 was -0.3427, with a standard error of 0.5781 and was not significant at a 95% level. We have $\overline{g}_1 = \overline{g}_2 = 1/\sqrt{2\pi}$. We tested the hypothesis $H_0 : \alpha_1\alpha_2 = 2\pi$ against the one-sided alternative $H_1 : \alpha_1\alpha_2 < 2\pi$. We were able to reject H_0 at a 99% significance level. This is statistically significant evidence that the game has a unique equilibrium everywhere in $\mathbb{S}(\boldsymbol{Z})$. We now discuss how the standard errors were obtained.

Overall, the estimation results seem to provide evidence that the large firm's investment decisions are driven mainly by its privately known characteristics ΔQ and $\Delta\%K$, as well as the industry-uncertainty measures MKT and $TECH$. Neither firm's decision to be aggressive seems to be significantly influenced by their aggressive/passive behavior the previous year. As for the strategic component of the game, results are consistent with the idea that the game is symmetric and is one in which both firms are affected when they are mutually aggressive. Most importantly, these results suggest that strategic considerations play a major role in the small firm's actions. The null hypothesis $H_0 : \alpha_1 = \alpha_2$ is rejected at a 95% confidence level against the alternative $H_1 : \alpha_2 > \alpha_1$. In fact, results show that $\widehat{\alpha}_1$ is not significant at a 95% level.

To summarize, these results are consistent with an investment game in which the large firm bases its actions on his privately observed characteristics as well as the state of industry uncertainty. Meanwhile, the small firm has a greater incentive to anticipate the actions of the large one, acting therefore as as a "follower". We must conclude by stressing that this example is meant to be an approximation of firms' actual behavior. We chose this particular approximation because it fits arguably well the description of a simultaneous game, where actions constitute precommitments. It would be interesting to see if the general results obtained here hold true if we estimate a game in which both firms choose their investment levels simultaneously. However, it would be problematic to represent it as a simultaneous game, as investment levels do not necessarily constitute irreversible pre-commitments.

The next section presents some concluding remarks for the paper.

6 Concluding remarks

This paper analyzed a 2×2 simultaneous game with incomplete information. We showed that a well-defined likelihood function exists for all possible variations of the game only if players have incomplete information. We assumed a general incomplete information structure and characterized the resulting Bayesian Nash equilibrium (BNE) conditions of the game. Players construct their equilibrium beliefs conditional on a vector of signals $\boldsymbol{Z}$, whose exact distribution is unknown to the researcher. BNE conditions take the form of conditional moment restrictions. We focused on estimation methods that exploit the information contained in the BNE. We presented two alternative semiparametric estimation procedures that achieved this goal. Both methods took the form of a two-step estimation procedure. The first step in both cases was to estimate the unknown equilibrium beliefs. The first method forced the sample to satisfy a semiparametric analog to the BNE conditions. The second one relied on a linearization of them.

The second step in both procedures used these semiparametric estimates as plug-ins for the unknown equilibrium beliefs in a trimmed maximum likelihood estimation. The trimming set is an interior subset of the support of $\boldsymbol{Z}$ where the BNE conditions have a unique solution. We found that the asymptotic distribution of the resulting estimators for the structural parameter vector $\boldsymbol{\theta}$ are the same for the two alternative procedures. We also found that they both exploit efficiently all available information to the researcher. The estimation is equivalent to a constrained MLE, where the constraint takes the form of a conditional moment restriction. We showed how to adapt the methodology for the cases in which $\boldsymbol{Z}$ includes continuous and/or discrete random variables. We also examined the case in which the BNE conditions have a unique solution everywhere in the support of $\boldsymbol{Z}$ and showed how to modify the methodology accordingly. Tests for uniqueness of equilibrium either for a given value of $\boldsymbol{Z}$ or for its entire support are also presented. A game of investment under uncertainty is estimated as an example. Results are consistent with a model in which the smaller firm has a comparatively greater incentive to predict the the actions of the larger one, which bases its decisions mainly on private information and indicators of industry uncertainty, giving relatively less weight to the expected actions of the smaller firm.

The presence of incomplete information in this game allows the researcher to estimate $\boldsymbol{\theta}$ efficiently and make predictions for the four observable outcomes in all possible variations of the game. If players have complete information, then a loss of resolution in the set of outcomes that can be predicted is inevitable in some variations of the game unless some equilibrium selection rule is imposed. The main lesson for applied work in general interactions-based models would be to carry out a careful inspection of the information conditions that prevailed when agents made their choices. The estimation procedure should be based -at least partially- on the information structure in an effort to exploit all available information in the game and achieve the greatest possible resolution for predictions. This paper can be extended in several ways. The first line of research would be to adapt the methodology to simultaneous games with more actions and/or players, or to sequential games. In all cases, we would first require to characterize the corresponding equilibrium and find conditions for uniqueness. These conditions must be testable and they should be as weak as possible, ideally weaker than contraction-mapping conditions. The estimation procedure would then exploit the information contained in the equilibrium conditions of the game. A second line of research is the design of tests aimed at selecting between alternative information structures of the game. This includes testing whether or not a particular game was played with complete or incomplete information, but should also include the design of tests that help us choose between alternative incomplete information structures.

References

[1] 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.

[2] Ai, C. and X Chen (2003) "Efficient estimation of models with conditional moment restrictions containing unknown functions" Econometrica 71: pp. 1795 – 1843.

[3] Amemiya, T. (1976) "Estimation in nonlinear simultaneous equation models" (Institut National de la Statisque et des Etudes Economiques, Paris).

[4] Amemiya, T. (1983) "Non-linear regression models" in: Z. Griliches and M.D Intrilligator, eds., Handbook of Econometrics, Vol. 1 (North Holland, Amsterdam) pp. 334 – 389.

[5] Amemiya, T. (1985) Advanced Econometrics (Harvard University Press, Cambridge).

[6] Andrews, D. and S.T Berry (2003) "On Placing Bounds on Parameters of Entry Games in the Presence of Multiple Equilibria", mimeo, Yale University.

[7] Aradillas-Lopez, A. (2004) "Existence of a Likelihood Function in a Simultaneous Game with Mixed Strategies", Unpublished manuscript.

[8] Bartelsman, E.J. and W. Gray (1996) "The NBER manufacturing productivity database" National Bureau of Economic Research Technical Working Paper 205.

[9] Berry, S. (1992), "Estimation of a Model of Entry in the Airline Industry", Econometrica 60 (4): pp. 889 – 917.

[10] Bjorn, P. and Q. Vuong (1984) "Simultaneous models for dummy endogenous variavles: a game theoretic formulation with an application to household labor force participation" Working paper, California Institute of Technology.

[11] Bjorn, P. and Q. Vuong (1985) "Econometric modelling of a Stackelberg game with an application to household labor force participation" Working paper, California Institute of Technology.

[12] Blundell, R. and J.R. Smith (1993) "Simultaneous Microeconometric Models with Censored or Qualitative Dependent Variables" in: G. S. Maddala, C. R. Rao and H. D. Vinod, eds. Handbook of Statistics, Vol. 11 (New York: North-Holland) pp. 117143.

[13] Blundell R, Bond S, Devereux M, Schiantarelli F (1992) "Investment and Tobin's Q" Journal of Econometrics 51: pp. $233-257$.

[14] Blundell R, Griffith R, Van Reenen J (1999) "Market share, market value and innovation in a panel of British manufacturing firms" Review of Economic Studies 66: pp. $529-554$.

[15] Blundell, R. and J.R. Smith (1994) "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables", Journal of Econometrics 64: pp. 355373.

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

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

[18] 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. $3297-3380$.

[19] 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. $77-89$.

[20] Dixit, A.K and R.S. Pindyck (1994) Investment Under Uncertainty (Princeton University Press, Princeton).

[21] Fudenberg, D. and K. Levine (1998) The Theory of Learning in Games (MIT Press, Cambridge, MA)

[22] Fudenberg, D. and J. Tirole (1991) Game Theory (MIT Press, Cambridge, MA)

[23] Griliches Z, J. Mairesse (1984) "Productivity and R&D at the firm level" in Griliches Z, ed. R&D, Patents, and Productivity. (The University of Chicago Press, Chicago).

[24] Hall, B.H. (1987) "The relationship between firm size and firm growth in the U.S manufacturing sector" The Journal of Industrial Economics 35: pp. $583-606$.

[25] Hall, B.H. (1990) "The manufacturing sector masterfile: 1959-1987" National Bureau of Economic Research Working Paper 3366.

[26] Hall, B.H and K. Vopel (1987) "Market value, market share, and innovation", mimeo, University of California, Berkeley.

[27] Harsanyi, J. and R. Selten (1988) A General Theory of Equilibrium Selection in Games. (MIT Press, Cambridge, MA).

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

[29] Holmstrom, B. (1979) "Moral hazard and observability" Bell Journal of Economics 10: pp. $324-340$.

[30] Hsiao, C. (1983) "Identification", in: Z. Griliches and M.D Intrilligator, eds., Handbook of Econometrics, Vol. 1 (North Holland, Amsterdam) pp. $223-283$.

[31] Ichimura, H. "Computation of asymptotic distribution for semiparametric GMM estimators" Unpublished manuscript.

[32] Jorgenson, D. W. and J. Laffont (1974) "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances", Annals of Economic and Social Measurement 3: pp. $615-640$.

[33] Klein, R.W. and R.H Spady (1993) "An efficient estimator of binary response models" Econometrica 61: pp. 387 – 421.

[34] Kooreman, P. "Estimation of econometric models of some discrete games" Journal of Applied Econometrics 9: pp. 255 – 268.

[35] 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).

[36] Manski, C.F (1993) "Identification of endogenous social effects: the reflection problem" Review of Economic Studies 60: pp. 531 – 542.

[37] Manski, C.F (1994) "Analog estimation of econometric models", in: R.Engle and D.McFadden, eds., Handbook of Econometrics, Vol. 4 (North Holland, Amsterdam) pp. 2559 – 2582.

[38] Manski, C.F (1997) "Identification of anonymous endogenous interactions" in W.B Arthur, S. Durlauf and D. Lane, eds. The Economy as a Complex Evolving System II (Addison-Wesley, Redwood City).

[39] Mas-Collel, A., Whinston, M. and J.R Green (1995) Microeconomic Theory (Oxford: Oxford University Press).

[40] 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).

[41] Nadiri, M.I and I.R. Prucha (1993) "Estimation of the depreciation rate of physical and R&D capital in the U.S. total manufacturing sector" National Bureau of Economic Research Working Paper 4591.

[42] Newey, W.K (1990)"Semiparametric efficiency bounds" Journal of Applied Econometrics 5: pp. 99 – 135.

[43] Newey, W.K (1994) “The asymptotic variance of semiparametric estimators” Econometrica 62: pp. $1349 - 1382$.

[44] 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. $2113 - 2245$.

[45] Oriani, O. and M. Sobrero (2003) “Assessing the market valuation of firms technological knowledge using a real options perspective.”, mimeo, University of Bologna.

[46] Pagan, A. and A. Ullah (1999) Nonparametric Econometrics (Cambridge University Press, Cambridge).

[47] Powell, J.L. (1994) “Estimation of semiparametric models” in: R.Engle and D.McFadden, eds., Handbook of Econometrics, Vol. 4 (North Holland, Amsterdam) pp. $2443 - 2521$.

[48] Powell, J.L, J.H Stock and T.M Stoker (1989) “Semiparametric estimation of index coefficients” Econometrica 57: pp. $1403 - 1430$.

[49] Rabin, M. (2000) “Risk Aversion and Expected-Utility Theory: A Calibration Theorem” Econometrica 68: pp. $1281 - 1292$.

[50] Samuelson, L. Evolutionary Games and Equilibrium Selection: Economic Learning and Social Evolution (MIT Press, Cambridge, MA).

[51] Schmidt, P. (1981) “Constraints on the Parameters in Simultaneous Tobit and Probit Models” in: C. Manski and D. McFadden , eds., Structural Analysis of Discrete Data with Econometric Applications (MIT Press, Cambridge, MA) pp. $422 - 434$.

[52] Seim, K. (2002) “An Empirical Model of Firm Entry with Endogenous Product-Type Choices”, mimeo, (Graduate School of Business, Stanford University).

[53] Spence, A.M. (1973) “Job Market Signaling” The Quarterly Journal of Economics 87 (3): pp. $355 - 74$.

[54] Sweeting, A. (2003) “Coordination Games, Multiple Equilibria and the Timing of Radio Commercials”, Unpublished Ph.D Job Market Paper (Economics Department, MIT).

[55] Tamer, E.T. (2002), “Empirical Strategies for Estimating Discrete Games with Multiple Equilibria”, mimeo, Princeton University.

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

[57] Wand, M.P and M.C Jones (1994) Kernel Smoothing (CRC Group, New York).

[58] Weibull, J (1997) Evolutionary Game Theory (MIT Press, Cambridge, MA).