Abstract

This paper studies an econometric modeling of a signaling game with two players where one player has one of two types. In particular, we develop an estimation strategy that identifies the payoffs structure and the distribution of types from data of observed actions. We can achieve uniqueness of equilibrium using a refinement, which enables us to identify the parameters of interest. In the game, we consider non-strategic public signals about the types. Because the mixing distribution of these signals is nonparametrically specified, we propose to estimate the model using a sieve conditional MLE. We achieve the consistency and the asymptotic normality of the structural parameters estimates. As an alternative, we allow for the possibility of multiple equilibria, without using an equilibrium selection rule. As a consequence, we adopt a set inference allowing for multiplicity of equilibria.

Keywords: Semiparametric Estimation, Signaling Game, Set Inference, Infinite Dimensional Parameters, Sieve Simultaneous Conditional MLE

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

1 Introduction

The econometric modeling of game theories has been of significant interest over the last decade. It has been one of most vivid research topics in the field of empirical industrial organization including studies on industry entry decisions (Bresnahan and Reiss (1990, 1991), Berry (1992), Toivanen and

*This paper is based on my doctoral dissertation. I am truly grateful to my advisor, Jinyong Hahn for his advice and encouragement throughout my doctoral study in UCLA. I would like to thank Joe Hotz, Moshe Buchinsky, Dan Ackerberg, John Riley, Kathleen McGarry, Raffaella Giacomini, Patrik Guggenberger, and William R. Zame for their useful comments and suggestions. I would also like to thank seminar participants at Brown, Rochester, SMU, UCL, University of Washington, WU in St. Louis, University of Wisconsin-Madison. All remaining errors are mine. Correspondence: Kyoo il Kim, School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903, Tel: +65 6828-0876, Fax: +65 6828-0833, E-mail: kyookim@smu.edu.sg.
Waterson (2000), Seim (2002), Ciliberto and Tamer (2003)) and others (Mazzeo (2002), Sweeting (2004), Davis (2005)). Most of econometric analysis of game theoretic models has been focused on simultaneous games with complete information (Bjorn and Vuong (1984, 1985), Bresnahan and Reiss (1990, 1991), Tamer (2003), Bajari, Hong and Ryan (2004)) or with incomplete information (Brock and Durlauf (2001, 2003), Seim (2002), Sweeting (2004), Aradillas-Lopez (2005)). More recently, there are also many studies on dynamic games (Aguirregabiria and Mira (2003), Bajari, Benkard, and Levin (2003), Berry, Ovstrovsky, and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003)).

However, although the information asymmetry problem has been noted widely both theoretically or empirically (Riley (2001) provides a valuable survey on this issue), a formal econometric analysis of a signaling game appears only in few studies\footnote{Ackerberg (2003) provides a dynamic learning model of consumer behavior allowing for the signaling effect of advertising on experience goods. Brown (2002) provides a structural estimation of prestige effects in IPO underwriting within a signaling game context where a firm’s owner has an incentive to signal the firm’s type by choosing prestigious underwriters.} for several reasons\footnote{We admit that there exist many studies that empirically test testible implications of a particular signaling game but there are very few studies that actually estimate the game’s primitives.}. What has hindered the econometric implementation of these games is the presence of multiple equilibria (or even absence of any equilibrium), which results in the nonexistence of a well-defined likelihood function over the entire set of observable outcomes. Even when these problems are overcome, it is generally hard to set up a flexible econometric model that is able to deal with a generic family of signaling games. This paper studies the identification and the estimation of a signaling game with two players where one player, informed player, has one of two types. In particular, we develop an estimation strategy that identifies the payoffs structure and the distribution of types from data of observed actions by the two players. Though multiple equilibria arise in the game we study, an equilibrium refinement such as intuitive criterion by Cho and Kreps (1987) enables us to obtain the uniqueness of equilibrium.

Providing more empirical relevance to the basic model, we introduce some non-strategic public signals about the types, which the informed player can not manipulate technically, or at least has no incentive to do so. These signals are observed by all players and econometricians. Under the separating equilibrium, the action of the informed player reveals the true type and thus the uninformed party has no incentive to use the additional information. However, under the pooling equilibrium, the uninformed party will update her belief on types using these signals following the Bayes’ rule. We specify the mixing distribution of non-strategic signals on types nonparametrically and thus estimate the model using a sieve conditional MLE (a conditional MLE version of Wong and Severini (1991) or Shen (1997)) where the infinite dimensional parameters are approximated by sieves. Noting that the conditional probability of choosing a certain combination of actions can be written in terms of several conditional moment restrictions, as an alternative, we can also estimate the model using the sieve minimum distance (MD) estimation developed by Ai and Chen.
(2003). In both methods, we can obtain the consistency and the asymptotic normality of structural parameters estimates. As noted earlier, in the signaling game, the existence of multiple equilibria naturally arises given some realization of payoffs. We have resolved the multiple equilibria issue using a refinement of equilibrium (which is an equilibrium selection rule) as common in the literature. Even though players may select one equilibrium out of multiple equilibria using a selection rule, it is often hard to model or justify such a selection mechanism. Examples of papers that handle the multiple equilibria problem in other games under some strong assumption on the equilibrium selection rules include Bjorn and Vuong (1984, 1985), Kooreman (1994), and Bajari, Hong and Ryan (2004) in games of complete information. Another example is Sweeting (2004) in a game with incomplete information. The disadvantage of this approach is that there is no generally accepted procedure to determine which equilibrium will be played among multiple equilibria even though an equilibrium selection of a signaling game is more acceptable than that of a simultaneous game, at least theoretically (See Banks and Sobel (1987) and Cho and Kreps (1987)). Other options dealing with multiple equilibria include the redefinition of a game in a way that makes it estimable without requiring an equilibrium selection rule. Some studies redefine the space of outcomes of the game so that it exhibits uniqueness of equilibrium (Bresnahan and Reiss (1990, 1991)). These approaches have the merit that we do not need additional assumptions to justify the equilibrium selection but it is noted that such redefinitions may result in some loss of information in the game.

Inspired by important work by Manski and co-authors (Manski (1990), Horowitz and Manski (1995), Manski and Tamer (2002)) on bound analysis, some researchers have started to develop set inferences rather than relying on point estimation, without attempting to resolve the equilibrium selection (Sutton (2000), Ciliberto and Tamer (2003), Andrews, Berry, Jia (2004)). We note that for signaling games, the self-fulfilling property is essential, which yields multiple equilibria depending on different beliefs on plays of the other party. Thus, as an attractive alternative to the previous approach, we allow for the possibility of multiple equilibria, without attempting to resolve the equilibrium selection problem. In particular, we consider the model where some asymptotic inequalities may define a region of parameters rather than a single point in the parameter space. By definition, when there are multiple equilibria, there exist regions of unobservables that are consistent with the necessary conditions for more than one equilibrium. Therefore, the probability implied by the necessary condition for a given event is greater than or equal to the true probability of the event. Some necessary conditions for a perfect Bayesian equilibrium (PBE) provide a set of inequality constraints on the parameters for the games we study. We illustrate that we can allow for the multiplicity of equilibria using Andrews, Berry, and Jia (2004)’s results.

The structure of this paper is as follows. Section 2 describes the game we study. In Section 3, we characterize equilibria of the game. In Section 4, we consider public information about the
distribution of types. In Section 5, we estimate the model using a sieve conditional MLE and a sieve MD. The consistency and the asymptotic normality of the structural parameters estimates are presented. Section 6 briefly considers the set inference of the model. We conclude in Section 7. Some technical details and mathematical proofs are presented in the Appendix.

2 Description of the Game

To introduce the game of our interest, we first define a Bayesian extensive game with observable actions. The game models a situation where each player observes the actions of all players including herself but has uncertainty about payoffs that are affected by types of other players where the type of each player is known only to that player and not to the other players. We define this game formally following Osborne and Rubinstein (1994). First, we say that chance or nature selects types for the players and refer to player $i$ after she receives the information $t_i$ as type $t_i$.

**Definition 2.1** A Bayesian extensive game with observable actions is a tuple $(\Gamma, (\Xi_i), (p_i), (u_i))$ where

- $\Gamma = (N, H, P)$ is an extensive game form with perfect information and simultaneous moves where $N$ denotes the set of players; $H$ is a history which collects actions taken by players. The set of terminal histories is denoted by $T$; $P$ is a player function that assigns a player who takes an action after the history $h \in H$.
- $\Xi_i$ is a finite set of possible types of player $i$; we write $\Xi = \times_{i \in N} \Xi_i$.
- $p_i$ is a probability measure on $\Xi_i$ for which $p_i(t_i) > 0$ for all $t_i \in \Xi_i$, and the measures $p_i$ are stochastically independent across $i$ ($p_i(t_i)$ is the probability that player $i$ is selected to be of type $t_i$).
- $u_i : \Xi \times T \rightarrow \mathbb{R}$ is a von Neumann-Morgenstern utility function ($u_i(t, h)$ is player $i$’s payoff when the profile of types is $t$ and the terminal history of $\Gamma$ is $h$).

Now we define a signaling game as

**Definition 2.2** A signaling game is a Bayesian extensive game with observable actions given by a tuple $(\Gamma, (\Xi_i), (p_i), (u_i))$ in which $\Gamma$ is a two-player game form where Player 1 takes an action then Player 2 takes an action, and the set of Player 2’s type $\Xi_2$ is a singleton.

What makes a signaling game interesting is that Player 2 (uninformed party or receiver) controls the action and Player 1 (informed party or sender) controls the information. The receiver has an incentive to try to deduce the sender’s type from the sender’s message (signal), and the sender may have an incentive to mislead the receiver. The solution concept we will use here is a perfect Bayesian equilibrium (PBE), instead of a sequential equilibrium (SE). The former is simpler than the latter.
We do so without loss of generality since the game of interest here is a finite Bayesian extensive game with observable actions. Every sequential equilibrium of the extensive game associated with a finite Bayesian extensive game with observable actions is equivalent to a perfect Bayesian equilibrium of the Bayesian extensive game, in the sense that they induce the same behavior and beliefs\(^4\). The formal definition of a PBE can be found in the Appendix, which is borrowed from Osborne and Rubinstein (p.233, 1994). In particular, we are interested in the following structure of the game \(G\) with two players where Player 1 has one of two types, \textit{strong} or \textit{weak}:

\[
\langle \Gamma, (\Xi_1 = \{"strong", "weak"\}), (p_1("strong") = p), (u_i) \rangle \quad \text{with} \quad \Gamma = \langle \{1, 2\}, H, P \rangle.
\]

Figure 1 illustrates the structure of the game. We design this game such that it is as simple as possible but still contains all the essence of the signaling game. This is an econometric modelling of the \textit{beer-quiche} game in Cho and Kreps (1987). In this game, we have two players. Player 1 has either of two types \{strong, weak\} with the probability of being the strong type equal to \(p\) and knows her type. After observing her type, Player 1 moves first sending one of two messages \{B, Q\} to Player 2. Then, Player 2 chooses an action \("F"\) or \("NF"\) after observing the signal sent by Player 1. After the play, a payoff is realized according to actions chosen by two players. The payoffs of this game have the following properties that are common in signaling games in general.

- The payoffs of Player 2 (\textit{uninformed party}) given her action are determined by the type of Player 1 (\textit{informed party}) not by Player 1’s action (signal)
- Given Player 2’s action, each type of Player 1 has bigger payoffs by choosing the signal corresponding to Player 1’s true type. \("B\) is the signal for \textit{strong} and \("Q\) is the signal for \textit{weak} by construction (\(\phi_{1s} > 0, \phi_{1w} > 0\))
- The \textit{strong} type of Player 1 has an incentive to signal its true type
- The \textit{weak} type of Player 1 has an incentive to mislead Player 2

The payoffs of the game are characterized by seven parameters:

- \(u_{1s}\): difference of payoffs between \("NF\) and \("F\) outcomes for the \textit{strong} type of Player 1
- \(u_{1w}\): difference of payoffs between \("NF\) and \("F\) outcomes for the \textit{weak} type of Player 1
- \(\phi_{1s}\): mimicking cost of the \textit{strong} type of Player 1 (cost of signalling \textit{falsely})
- \(\phi_{1w}\): mimicking cost of the \textit{weak} type of Player 1 (cost of signalling \textit{falsely})
- \(u_{2s}\): difference of payoffs choosing between \("NF\) and \("F\) for Player 2 when Player 1 is \textit{strong}

\(^4\)Note that the reverse statement is not true in general (see page 234-235, Osborne and Rubinstein (1994)). However, Fudenberg and Tirole (1991) note that for a finite Bayesian extensive game with two types or with two periods, every PBE is also equivalent to a sequential equilibrium. Thus, for the game of Figure 1&2, these two equilibrium concepts are equivalent each other.
• $u_{2w}$: difference of payoffs choosing between "NF" and "F" for Player 2 when Player 1 is weak
• $p$: distribution of types

It is of our interest to identify these seven parameters that characterize the game from observed outcomes of actions. It is obvious that we can only identify up to three parameters nonparametrically if there is no exclusion or parametric restrictions since we have only three independent conditional probabilities out of four possible outcomes \{B&NF, B&F, Q&NF, Q&F\}. To ensure the identification\(^5\), here we will adopt the following parametric specifications of the payoffs and we impose an exclusion restriction that $X_2$ contains at least one variable that does not enter in $X_1$, which has the support of at least three values.\(^6\)

- $u_{1s} = \mu_s + X_1^1 \beta_s - \varepsilon_1$
- $u_{1w} = \mu_w + X_1^1 \beta_w - \varepsilon_1$
- $u_{2s} = X_2^2 \beta_2 - \varepsilon_2 + \phi_2s$
- $u_{2w} = X_2^2 \beta_2 - \varepsilon_2 - \phi_2w$

---

\(^5\)There are other possible ways of obtaining the identificaition of seven parameters. We may even identify some parameters nonparametrically utilizing exclusion restrictions or functional form restrictions. This issue is one of our future research.

\(^6\)For example, suppose we have no $X_1$ and suppose $X_2$ is a scalar random variable that can have one of three values \{-1, 0, 1\}. For each value of $X_2$, we have three linearly independent conditional probabilities. Thus, we have three sets of three linearly independent conditional probabilities (nine moment conditions), from which we can identify seven parameters of interest.
To make discussion simple as possible, here we only consider symmetric payoffs as in Figure 2\textsuperscript{7} where we impose $\mu \equiv \mu_s = \mu_w$ and $\beta_1 \equiv \beta_s = \beta_w$.\textsuperscript{8} The game with asymmetric payoffs are considered in Appendix B. Here we note that

- $\phi_{1s} > 0$ measures the potential cost of the strong type for mimicking the weak type given a fixed response of Player 2.
- $\phi_{1w} > 0$ measures the potential cost of the weak type for mimicking the strong type given a fixed response of Player 2.
- $\phi_{2s}$ and $\phi_{2w}$ measure degrees of Player 2’s incentive to single out a particular type of Player 1.

Throughout the paper, we assume that $\phi_{2s}$ and $\phi_{2w}$ are not negative. Note that this assumption is innocuous in the sense that “strong” and “weak” types are just labels\textsuperscript{9} unless we give some structure to it. By imposing $\phi_{2s} \geq 0$, we mean that Player 2 is more likely to be better off by singling out the strong type for the action “$NF$” and to be better off by singling out the weak type for the action “$F$”.

The game tells that if Player 1 is the strong type and if Player 1 chooses “$Q$” and Player 2 chooses “$NF$”, Player 1 obtains $\mu + X_1^t \beta_1 - \varepsilon_1 - \phi_{1s}$ and Player 2 obtains $X_2^t \beta_2 - \varepsilon_2 + \phi_{2s}$, respectively. If Player 1 is the strong type and if Player 1 chooses “$Q$” and Player 2 chooses “$F$”, they earn $-\phi_{1s}$ and 0, respectively. Other payoffs can be read in the same way. It is noted that each player’s payoffs are not only determined by her own action but also by the other player’s action or type, which depicts interactions between players.

We let $X = (X_1^t, X_2^t) \in \mathbb{R}^k$ with $k = k_1 + k_2$ and let $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$. We denote the vectors of parameters as $(\mu, \beta_1) \in \mathbb{R}^{k_1+1}$, $\beta_2 \in \mathbb{R}^{k_2}$, $\theta_{sg} = (\phi_{1s}, \phi_{1w}, \phi_{2s}, \phi_{2w})' \in (0, \infty)^2 \times [0, \infty)^2$, and let $(\theta, p) = (\mu, \beta_1', \beta_2', \theta_{sg}', p)' \in \mathbb{R}^{k_1+1} \times (0, \infty)^2 \times [0, \infty)^2 \times (0, 1)$. These parameters of interest are perfectly known to players but unknown to econometricians.

Note that this stochastic payoffs game is different from the deterministic payoffs game in that by adding the unobserved heterogeneity $(\varepsilon_1, \varepsilon_2)$, we allow such cases that players with the same observed characteristics $(X_1, X_2)$ can show different outcomes of the game. We also note that the payoffs structure of the game $G$ is somewhat restrictive in the sense that we let the following two payoffs\textsuperscript{10} be the same. One is the payoff of the strong type Player 1 when she chooses “$B$” and

\begin{itemize}
    \item[7.] One might notice that the payoffs structure of Figure 2 is not nested by Figure 1 seeing the payoffs of Player 2 when she plays “$F$” and Player 1 is the weak type. However, note that what matters in the game is only the payoffs differential. We choose to use Figure 2 as an illustrational purpose to show that Player 2 is more likely to be better off by choosing “$F$” when Player 1 is the weak type.
    \item[8.] Here we assume that $\phi_{1s}, \phi_{1w}, \phi_{2s}$, and $\phi_{2w}$ are constant but we may extend the model such that these parameters also depend on characteristics of Player 1 and Player 2, respectively.
    \item[9.] Note that depending on realizations of $\varepsilon_1, \varepsilon_2$ given $X_1, X_2$, the strong type can also have an incentive to mimic the weak type.
    \item[10.] Actually, the payoffs differences between the outcomes of “$NF$” and “$F$” due to the normalization.
\end{itemize}
Player 2 plays “NF”. The other is the payoff of the weak type Player 1 when she chooses “Q” and Player 2 chooses “NF”. This structure may be justified in some cases but might be too restrictive in general. We relax this restriction in Appendix B. It turns out that the games with/without asymmetric payoffs are very similar in terms of equilibrium characterization.

2.1 The Information Structure (IS)

The game we study has incomplete information since players do not have exact knowledge about the payoffs of their opponents. It is also a signaling game since the true type of Player 1 is only known to Player 1 herself and Player 1 signals her type to Player 2 by choosing some action.

Assumption 2.1 (IS)

1 Player 1 knows her true type but Player 2 knows only the distribution of Player 1’s types ($p$ is known to Player 2).
2 The realizations of $(X_1, \varepsilon_1)$ and $(X_2, \varepsilon_2)$ are perfectly observed by both Players 1 and 2.
3 $\varepsilon_1$ and $\varepsilon_2$ are pure shocks commonly observed by Player 1 and Player 2. They are independent of each other and of any other variables in the game. $\varepsilon_1$ is also independent of the type of Player 1.
4 Players’ actions and beliefs constitute a Perfect Bayesian Equilibrium (Sequential Equilibrium). Whenever there exist multiple equilibria, only one equilibrium is chosen out of these according to some equilibrium refinements. Players are assumed to play actions and hold beliefs about this unique equilibrium.

Note that the generic uniqueness of the PBE is ensured by some refinements of the equilibrium concept.

2.2 Stochastic Assumptions (SA-1)

We impose the following distributional assumptions on the random variables of the game. We first consider the simplest structure and generalize it later. Hereafter we denote the support of a random variable by $S(\cdot)$.

Assumption 2.2 (SA-1)

1 $\varepsilon_1$ and $\varepsilon_2$ are continuously distributed, statistically independent of each other and of $X$.
2 The cdf’s of $\varepsilon_1$ and $\varepsilon_2$ are continuous and denoted by $G_1(\varepsilon_1)$ and $G_2(\varepsilon_2)$ with corresponding density functions $g_1(\varepsilon_1)$ and $g_2(\varepsilon_2)$, respectively. The density functions are assumed to be bounded and strictly positive on their supports $\mathbb{R}$. The density functions do not depend on the model parameter $(\theta, p)$ nor on the type of Player 1.
3 Both $X_1$ and $X_2$ can be continuous or discrete random variables. Both $X_1$ and $X_2$ are independent of the type of Player 1. We denote the density of $X_1$ and $X_2$ as $f_{X_1}(\cdot)$ and $f_{X_2}(\cdot)$, respectively. Neither $f_{X_1}(\cdot)$ or $f_{X_2}(\cdot)$ depends on the structural parameter $(\theta, \rho)$.

4 $X_2$ contains at least one variable that does not enter in $X_1$, which has the support of at least three values.

By imposing Assumption SA-1.1, we ensure that Player 2’s equilibrium beliefs are constructed conditional on variables observed by the econometrician. The exclusion restriction of Assumption SA-1.4 is easily satisfied for the game model we consider since there exist inherent exclusions between $X_1$ and $X_2$. In other words, $X_1$ and $X_2$ are different variables for a typical signaling game. We will strengthen these stochastic assumptions to ensure the validity of the econometric modelling in the later section.

3 Refinement and Uniqueness of Equilibrium

In this section we characterize equilibria of the game under PBE and then show we can achieve the uniqueness of equilibrium using a refinement of Cho and Kreps (1987). For the game we study, it turns out that we have multiple equilibria for some realizations of payoffs, which can be removed adopting a stronger equilibrium concept as an equilibrium selection. The disadvantage of this approach noted in the literature is that there is neither generally accepted or empirically testable procedure to determine which equilibrium will be played among multiple equilibria, especially for simultaneous move games. However, we note that an equilibrium selection of a signaling game is more acceptable than that of a simultaneous game, at least theoretically, which may be a justification of our approach but we relax this in Section 6.

We introduce some notation here. Let $u_i(t_1; A_1, A_2)$ denote the payoffs of player $i \in \{1, 2\}$ when the true type of Player 1 is $t_1 \in \{t_s \equiv \text{strong}, t_w \equiv \text{weak}\}$, Player 1 chooses action (signal) $A_1 \in \{B, Q\}$ and Player 2 chooses action $A_2 \in \{NF, F\}$. We also let $A_1(t_1)$ denote the action taken by a particular type of Player 1. With $(A_1, A'_1)$, we mean $(A_{1t_s}, A_{1t_w}) = (A_1, A'_1)$ where $A_1, A'_1 \in \{B, Q\}$. Note $p = \Pr(t_1 = t_s)$ is the prior belief of Player 2 on the type of Player 1. This is also the population distribution of types. We let $E_1[u_1(t_1; A_1, A_2)]$ be the expected payoff of Player 2 on the type of Player 1 after observing the action (signal) of $A_1$. We also let $Y_2|A_1$ be an indicator function that has the value 1 when Player 2 chooses the action “NF” after observing the signal $A_1$. Similarly, $A_2|A_1$ denotes the action of Player 2 after observing $A_1$. Throughout this paper, we will use this notation. In the Appendix D.2, we determine regions of $(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$, where each PBE exists given a realization of $X$. Each PBE is obtained using the four conditions of Definition D.1 in the Appendix: sequential rationality, correct initial belief, action-determined beliefs, and Bayesian updating.
Figure 3\textsuperscript{11} illustrates regions of \((\varepsilon_1, \varepsilon_2)\) where a particular equilibrium is supported\textsuperscript{12} in terms of observations, assuming \(\phi_{1s} > 0, \phi_{1w} > 0, \phi_{2s} \geq 0, \) and \(\phi_{2w} \geq 0\) (but \(\phi_{2s} \cdot \phi_{2w} \neq 0\)). It is noted that under \(\varepsilon_2 > X'_{2\beta_2} + \phi_{2s}\), Player 2 is better off by choosing \(F\) regardless of Player 1’s type or action and will choose \(F\). Thus, Player 1 will choose a signal corresponding to her type since given Player 2’s action, Player 1 is better off by choosing the signal that corresponds to her type. Similarly, under \(\varepsilon_2 < X'_{2\beta_2} - \phi_{2w}\), Player 2 is better off by choosing \(N\) no matter what and thus under this region, Player 1 is willing to reveal her true type. We also note that in many empirical studies, researchers tend to focus on the separating equilibrium where Player 1 reveals her type and Player 2 chooses different actions according to different types of Player 1 (such as region \(S_3 \equiv \{(\varepsilon_1, \varepsilon_2) | \mu + X'_{1\beta_1} - \phi_{1w} < \varepsilon_1 < \mu + X'_{1\beta_1} + \phi_{1s} \) and \(X'_{2\beta_2} - \phi_{2w} < \varepsilon_2 < X'_{2\beta_2} + \phi_{2s}\)} in Figure 3) by imposing some conditions or by simply asserting a separating equilibrium is more reasonable. However, Figure 3 illustrates that other kinds of equilibria can arise depending on the realizations of \((\varepsilon_1, \varepsilon_2)\). For example, in region \(\{(\varepsilon_1 < \mu + X'_{1\beta_1} - \phi_{1w} \) and \(X'_{2\beta_2} + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} < \varepsilon_2 < X'_{2\beta_2} + \phi_{2s}\},\) we have a semi-separating equilibrium where the strong type plays \(B\) and the weak type mixes between \(B\) and \(Q\). Similarly, in region \(\{\varepsilon_1 > \mu + X'_{1\beta_1} + \phi_{1s} \) and \(X'_{2\beta_2} - \phi_{2w} < \varepsilon_2 < X'_{2\beta_2} + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}\},\) we have another semi-separating equilibrium where the weak type plays \(Q\) and the strong type mixes between \(B\) and \(Q\).

The following two theorems are about the existence of PBE for all regions of \((\varepsilon_1, \varepsilon_2)\) given \(X\) and conditions to achieve uniqueness of equilibrium.

**Theorem 3.1 (Existence of Equilibrium)**

Suppose Assumptions IS and SA-1 hold. Suppose also that \(\phi_{1s} > 0, \phi_{1w} > 0, \phi_{2s} \geq 0, \) and \(\phi_{2w} \geq 0\) (but \(\phi_{2s} \cdot \phi_{2w} \neq 0\)). Then, there exist PBE for all regions of \((\varepsilon_1, \varepsilon_2)\) given \(X = x\).

See the Appendix D.2 for the proof. We note that there are regions where multiple equilibria arise. However, using the refinement of Cho and Kreps (1987), we can achieve uniqueness of equilibria for each region of \((\varepsilon_1, \varepsilon_2)\) given \(X\). Figure 3\textsuperscript{14} shows that there exist three regions where

\textsuperscript{11}For the game of Figure I in Cho and Kreps (1987), we have two pooling PBE. In one equilibrium, both types of Player 1 choose \(B\), and Player 2 does not fight if she observes \(B\) and she fights if she observes \(Q\) with out-of-equilibrium belief \(\mu_2(t_s|Q) \leq 0.5\). In the other equilibrium, both types of Player 1 choose \(Q\), Player 2 chooses not to fight if she observes \(Q\) and she fights if she observes \(B\) with out-of-equilibrium belief \(\mu_2(t_s|B) \leq 0.5\). We note that the example of Cho and Kreps (1987) corresponds to region \(S_3 \equiv \{(\varepsilon_1, \varepsilon_2) | \varepsilon_1 < \mu + X'_{1\beta_1} - \phi_{1w} \) and \(X'_{2\beta_2} - \phi_{2w} < \varepsilon_2 < X'_{2\beta_2} + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}\}\) in Figure 3 and 4.

\textsuperscript{12}To define a PBE, we also specify the out-of-equilibrium belief that supports a particular equilibrium. A specific out-of-equilibrium belief for each PBE of the game can be found in the Mathematical Appendix. Here we do not present such out-of-equilibrium belief to make discussion simple.

\textsuperscript{13}We note that this region gets larger as \(\phi_{1s}, \phi_{1w}, \phi_{2s},\) and \(\phi_{2w}\) become larger. This implies that the larger the cost of mimicking and the larger Player 2’s incentives of singling out a particular type of Player 1, the larger the region that supports this separating equilibrium.

\textsuperscript{14}Figure 3 depicts the case \(\phi_{1w} > \phi_{1s}\) but this is not necessary at all.
we have multiple equilibria. In region $A_a \equiv \{(\varepsilon_1, \varepsilon_2)|\varepsilon_1 > \mu + X'_1\beta_1 + \phi_{1w}$ and $X'_2\beta_2 + (\phi_{2s} + \phi_{2w})p - \phi_{2w} < \varepsilon_2 < X'_2\beta_2 + \phi_{2s}\}$, we have two equilibria. One is pooling $(B, B)$ with $A_{2B} = F$ and the other is pooling $(Q, Q)$ with $A_{2Q} = F$ or separating equilibrium (when $\phi_{1w} < \phi_{1s}$). In region $A_b \equiv \{\varepsilon_1 < \mu + X'_1\beta_1 - \phi_{1s}$ and $X'_2\beta_2 - \phi_{2w} < \varepsilon_2 < X'_2\beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}\}$, we also have two equilibria. One is pooling $(Q, Q)$ with $A_{2Q} = NF$ and the other is pooling $(B, B)$ with $A_{2B} = NF$ or separating equilibrium. In Appendix A, using the intuitive criterion of Cho and Kreps (1987)\textsuperscript{15}, we show that the pooling $(B, B)$ fails the refinement in region $A_a$, the pooling $(Q, Q)$ fails the refinement in $A_b$\textsuperscript{16}. Figure 4 illustrates the uniqueness of equilibrium based on this refinement result.

\textbf{FIGURE 3. Multiple Equilibria of the Game} \hfill \textbf{FIGURE 4. Uniqueness of Equilibrium}

We need to distinguish an equilibrium from its realized outcomes. For example, in region $S_1 \equiv \{(\varepsilon_1, \varepsilon_2)|\varepsilon_1 \in \mathbb{R}$ and $\varepsilon_2 > X'_2\beta_2 + \phi_{2s}\}$, we have one equilibrium where Player 1 plays the separating equilibrium with $(B, Q)$ and Player 2 chooses $F$. However, in terms of realized outcomes, we have two possible outcomes in this region. With probability $p$ (when Player 1 is the strong type), we will observe the $B&F$ combination but we can also observe the $Q&F$ combination with probability $1 - p$ (when Player 1 is the weak type). Likewise even though we have only one equilibrium (semi-separating where the strong type plays $B$ and the weak type mixes between $B$ and $Q$) in region $S_2 \equiv \{(\varepsilon_1, \varepsilon_2)|\varepsilon_1 < \mu + X'_1\beta_1 - \phi_{1w}$ and $X'_2\beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} < \varepsilon_2 < X'_2\beta_2 + \phi_{2s}\}$, we can observe all four possible outcomes with certain probabilities, respectively.

\textsuperscript{15}The intuitive criterion is based on the concept of “equilibrium dominance” which tells that a certain type should not be expected to use a certain strategy. For example, in region $S_4 \equiv \{(\varepsilon_1, \varepsilon_2)|\varepsilon_1 > \mu + X'_1\beta_1 + \phi_{1s}$ and $X'_2\beta_2 + p(\phi_{2s} + \phi_{2w}) + \phi_{2w} < \varepsilon_2 < X'_2\beta_2 + \phi_{2s}\}$, Player 1 plays a pooling equilibrium strategy with $(B, B)$ or $(Q, Q)$. However, for the pooling equilibrium with $(B, B)$ under $S_4$, it is not reasonable to believe that a deviation is played by the strong type because she is better off by choosing $B$ no matter what Player 2 chooses (see Appendix). Thus, Player 2 will think a deviation is played by the weak type for sure when she observes $Q$. Under this refined belief, now we can eliminate the pooling with $(B, B)$ in region $S_4$.

\textsuperscript{16}For the game of Figure I in Cho and Kreps (1987), the pooling PBE with $Q$ fails the intuitive criterion. This corresponds to region $S_5$ in Figure 3 and 4.
Theorem 3.2 (Uniqueness of Equilibrium)

Suppose Assumptions IS and SA-1 hold and that \( \phi_{1s} > 0, \phi_{1w} > 0, \phi_{2s} \geq 0, \phi_{2w} \geq 0 \) (but \( \phi_{2s} \cdot \phi_{2w} \neq 0 \)). Suppose each Player plays only one equilibrium that survives the refinement of Cho and Kreps (1987), when there exist multiple equilibria. Then, there exist unique equilibrium for each region of \((\varepsilon_1, \varepsilon_2)\) given \(X\).

Theorem 3.2 enables us to obtain a well-defined likelihood function. One may think that we cannot identify \(\mu\) separately from \(\phi_{1s}\) and \(\phi_{1w}\) under unique equilibrium as Figure 4 suggests\(^{17}\). However, we actually can identify \(\mu\) separately from \(\phi_{1s}\) and \(\phi_{1w}\) since \(\phi_{1s}\) and \(\phi_{1w}\) appear separately in the model conditional probabilities due to the semi-separating equilibria (see Appendix C). From the result of Theorem 3.2, we can define the conditional probabilities for four possible observed outcomes.

\[ P_{jl}(X, \theta, p) = \Pr(Y_1 = j, Y_2 = l|X) \text{ for } j, l \in \{0, 1\}. \]

The specific forms of those conditional probabilities are provided in Appendix C. Now using these conditional probabilities, one can estimate the parameter \(\theta\), using the conditional ML method such that

\[ (\hat{\theta}_{CML}, \hat{p}_{CML}) = \arg\max_{\theta \in \Theta, p \in (0, 1)} \frac{1}{n} \sum_{i=1}^{n} \log L (y_i | x_i, \theta, p) \]  \hspace{1cm} (1)

where

\[ \log L (y_i | x_i, \theta, p) = y_{1i} y_{2i} \log P_{11}(x_i, \theta, p) + y_{1i} (1 - y_{2i}) \log P_{10}(x_i, \theta, p) 
+ (1 - y_{1i}) y_{2i} \log P_{01}(x_i, \theta, p) + (1 - y_{1i}) (1 - y_{2i}) \log P_{00}(x_i, \theta, p). \]

The estimator \((\hat{\theta}_{CML}, \hat{p}_{CML})\) will be consistent and asymptotically normal under suitable conditions. Those conditions can be found in Newey and McFadden (1994), for example. Note that this conditional ML estimator is identical to the ML estimator since the density function of \(X\) does not depend on the parameter \((\theta, p)\) by Assumption SA-1.3.

4 Introducing Public Information about the Type

We consider public signals about the type of Player 1. We denote such signals by \(Z\), which are observable to all the players of the game and econometricians. Until now, we have assumed that \(X_1\) is independent of the type of Player 1. However, it is likely that at least some of observed
characteristics of Player 1 will reveal information regarding the type of Player 1. In the beer or quiche game story, characteristics of Player 1 such as muscle intensity, height, or age will tell how likely Player 1 is the strong type. Thus, \( Z \) can include all the variables in \( X_1 \) or a subset of \( X_1 \). In the job signaling game (Spence (1974)), parents’ education is a possible public signal for the ability of a job candidate.

For a public signal, we require that Player 1 cannot strategically choose the signal \( Z \) when a game is played\(^{18} \) or at least Player 1 does not have an incentive to do so. This means that in the game \( G \), only the action \( A_1 \) plays the role of the strategic signal. The public signal \( Z \in \mathbb{R}^{\text{dim}(Z)} \) has a mixing distribution \( F_Z(\cdot) \) with a mixing variable \( p \). We let the density of \( Z \) as

\[
 f_Z(z) = p f_{(st)}(z) + (1 - p) f_{(we)}(z)
\]

where \( p \) is known to the players of the game but not known to econometricians. Player 2 has an incentive to use these signals while playing the game. When players play a separating equilibrium, these additional signals have no additional information for Player 1’s type since Player 1’s type is perfectly inferred from her action. When the players play a pooling equilibrium, Player 2 will use these additional signals to update her belief on Player 1’s type using the Bayes’ rule. Therefore, under a separating equilibrium, we have

\[
 \mu_2(t_1 = t_s | A_1, Z) = \mu_2(t_1 = t_s | A_1)
\]

but under a pooling equilibrium, we have

\[
 \mu_2(t_1 = t_s | A_1, Z = z) = \frac{p f_{(st)}(z)}{p f_{(st)}(z) + (1 - p) f_{(we)}(z)}
\]

which is also the conditional probability of being strong type given \( Z = z \). We let \( p(z) \) denote this conditional probability, \( \Pr(t_1 = t_s | Z = z) = \frac{p f_{(st)}(z)}{p f_{(st)}(z) + (1 - p) f_{(we)}(z)} \). Thus, we have \( p(Z) = \mu_2(t_1 = t_s | A_1, Z) \) under a pooling equilibrium. Note that equation (3) and \( p(z) \) become the prior \( p \), when \( f_{(st)}(z) = f_{(we)}(z) \) (no mixture). We will maintain the following assumptions:

**Assumption 4.1 (IS-A)**

1. Assumption IS holds.
2. The public signal \( Z \) about the types of Player 1 is perfectly known to both Player 1 and Player 2.

**Assumption 4.2 (SA-1A)**

1. \( \varepsilon_1 \) and \( \varepsilon_2 \) are continuously distributed, statistically independent of each other and of \( W \equiv X \cup Z \).

\(^{18}\)It means that in the beer-quiche game, Player 1 cannot work out her muscle when the game is played.
2 The cdf’s of \( \varepsilon_1 \) and \( \varepsilon_2 \) are continuous and denoted by \( G_1(\varepsilon_1) \) and \( G_2(\varepsilon_2) \) with corresponding density functions \( g_1(\varepsilon_1) \) and \( g_2(\varepsilon_2) \), respectively. The density functions are assumed to be bounded and strictly positive on their supports \( \mathbb{R} \). The density functions do not depend on the model parameter \((\theta, p)\) or on the type of Player 1.

3 Both \( X_1 \) and \( X_2 \) can be continuous or discrete random variables. Both \( X_1 - Z \) and \( X_2 \) are independent of the type of Player 1. We denote the density of \( X_1 - Z \) and \( X_2 \) as \( f_{X_1 - Z}(\cdot) \) and \( f_{X_2}(\cdot) \), respectively. Neither \( f_{X_1 - Z}(\cdot) \) or \( f_{X_2}(\cdot) \) depends on the structural parameter \((\theta, p)\).

4 \( Z \) is a continuously distributed random vector with the mixing density \( f_Z(z) = pf_{(st)}(z) + (1 - p)f_{(we)}(z) \). Neither \( f_{(st)}(z) \) or \( f_{(we)}(z) \) depends on the structural parameter \((\theta, p)\).

5 \( X_2 \) contains at least one variable that does not enter in \( X_1 \), which has the support of at least three values.

In the game \( G \) under Assumptions IS-A and SA-1A, it is not difficult to see that we will obtain exactly the same equilibria under Assumptions IS and SA-1 except for replacing \( \mu_2(t_1 = t_s|A_1) \) with \( \mu_2(t_1 = t_s|A_1, Z = z) \). This means that whenever Player 2’s belief is involved, we replace \( p \) with \( p(z) \). Note that we should distinguish the population distribution of types \((\Pr(t_1 = t_s) = p)\) from the posterior belief of Player 2 after observing the signal \( Z = z \) under a pooling equilibrium, which should also be the conditional probability of being the strong type given \( Z = z \) in equilibrium. Regardless of the public signal \( Z \), the overall distribution of types \((p)\) is fixed by nature. We apply the refinement of Cho and Kreps (1987) to these games with public signal \( Z \) and obtain the same uniqueness of equilibrium with Theorem 3.2.

We will use a logistic specification for the posterior belief of Player 2 under a pooling equilibrium noting that equation (3) can be rewritten as

\[
p(z) = \frac{pf_{(st)}(z)}{pf_{(st)}(z) + (1 - p)f_{(we)}(z)} = \frac{\frac{p - f_{(st)}(z)}{1 - p f_{(we)}(z)}}{1 + \frac{p - f_{(st)}(z)}{1 - p f_{(we)}(z)}} \tag{4}
\]

and thus, the belief only depends on \( p \) and the ratio \( f_{(st)}(z)/f_{(we)}(z) \). The relationship in (4) means that for updating, we do not need to know \( f_{(st)} \) and \( f_{(we)} \) individually but only the ratio between these two is necessary.

Therefore, by letting \( h^0(z) = \log \left( \frac{f_{(st)}(z)}{f_{(we)}(z)} \right) \), we can rewrite (3) as a logistic specification with \( L(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)} \),

\[
p(Z = z) = \frac{\exp \left( \log \left( \frac{p}{1 - p} \right) + h^0(z) \right)}{1 + \exp \left( \log \left( \frac{p}{1 - p} \right) + h^0(z) \right)} = L(\log \left( \frac{p}{1 - p} \right) + h^0(z)) \tag{5}
\]

\(^{19}\)For the game with additional incomplete information presented in the Mathematical Appendix E, we also obtain the same result with Theorem E.3.
recalling that the posterior belief of Player 2 under a pooling equilibrium equals to the conditional probability of being strong type given $Z = z$.

This specification reflects the separability of the mixing variable and the component functions. If there is no mixture, we will have $h^o(z) = \log \left( f_{(st)}(z) / f_{(we)}(z) \right) = 0$. It suggests that by examining whether or not $h^o(z) = 0$, we may test the existence of a mixing distribution.

5 Estimation of the Signaling Game

In this section, we provide several estimation methods for the game $G$ with public signals. We note that the mixing distribution of public signals is nonparametrically specified. The information structure of IS-A is maintained but the stochastic assumptions are strengthened to facilitate the estimation. To preserve uniqueness of equilibrium, we also maintain the assumption that each player plays only one equilibrium that survives the refinement of Cho and Kreps (1987), when there exist multiple equilibria. We also let $d_l$ denote the dimension of a vector $l$. We let $Y_1$ denote the indicator function that has the value one when Player 1 chooses "B" such that $Y_1(A_1 = B)$. Similarly, we let $Y_2 \equiv 1(A_2 = NF)$. We also let $Y = (Y_1, Y_2)$. We let $C(\cdot), C_1(\cdot), C_2(\cdot)$, and so on denote generic positive constants or functions. For a positive number $k$, we let $\lfloor k \rfloor$ denote the largest integer smaller than $k$. Finally, we let the upper case stand for a random variable and the lower case stand for a realization of it. We use the subscript "o" to denote the true value of parameters. Throughout the paper, we assume that econometricians observe the realizations of the random variables $X_1 \cup Z, X_2, Y_1,$ and $Y_2$ but do not observe $\varepsilon_1$ or $\varepsilon_2$. We let $W \equiv X \cup Z$ and let $\theta = (\mu, \beta_1', \beta_2', \phi_1, \phi_2, \phi_{1w}, \phi_{2w})' \in \mathbb{R}^{1+k_1+k_2} \times (0, \infty)^2 \times [0, \infty)^2$, $\alpha^o = (\theta, p, h^o) \in \mathcal{A}^o \equiv \Theta \times (0,1) \times \mathcal{H}^o$. The $\alpha^o$ is the parameter of interest that econometricians want to estimate.

5.1 Regularity Conditions

Now we impose stronger conditions than SA-1A by adding some "smoothness" conditions for $G_1(\cdot)$ and $G_2(\cdot)$.

Assumption 5.1 (SA-2)

1. $\varepsilon_1$ and $\varepsilon_2$ are continuously distributed, statistically independent of each other and of $W$.
2. The cdf's of $\varepsilon_1$ and $\varepsilon_2$ are continuous and denoted by $G_1(\varepsilon_1)$ and $G_2(\varepsilon_2)$ with corresponding density functions $g_1(\varepsilon_1)$ and $g_2(\varepsilon_2)$, respectively. The density functions are assumed to be bounded and strictly positive on their supports $\mathbb{R} (S(\varepsilon_1) = S(\varepsilon_2) = \mathbb{R})$. The density functions do not depend on the model parameter $\alpha^o$ or on the type of Player 1.
3. $G_1(\varepsilon_1)$ and $G_2(\varepsilon_2)$ are $\nu + 3$ times continuously differentiable with bounded $\nu + 3$ derivatives everywhere in $S(\varepsilon_1) = S(\varepsilon_2) = \mathbb{R}$. The density of $g_1$ and $g_2$ are known to Player 1 but an
econometrician knows $g_1$ and $g_2$ up to a finite dimensional parameter\textsuperscript{20}, respectively.

4 Both $X_1$ and $X_2$ can be continuous or discrete random variables. Both $X_1 - Z$ and $X_2$ are independent of the type of Player 1. We denote the density of $X_1 - Z$ and $X_2$ as $f_{X_1-Z}(\cdot)$ and $f_{X_2}(\cdot)$, respectively. Neither $f_{X_1-Z}(\cdot)$ or $f_{X_2}(\cdot)$ depends on the model parameter $\alpha^o$.

5 In particular, we assume that $S(X_1) \subset \mathbb{R}^{k_1}$ and $S(X_2) \subset \mathbb{R}^{k_2}$ are compact.

6 $Z$ is a continuously distributed random vector with density $f_Z(z)$. $S(Z)$ is compact with nonempty interior. The density $f_Z(z) = pf_{(st)}(z) + (1 - p)f_{(we)}(z)$ is bounded and bounded away from zero. Neither $f_{(st)}(z)$ or $f_{(we)}(z)$ depends on the structural parameter $(\theta, p)$.

7 $X_2$ contains at least one variable that does not enter in $X_1$, which has the support of at least three values.

All the assumptions are standard in the literature. In particular, Assumption SA-2.7 is an order condition for the identification of model parameters. Again we note that the exclusion restriction of Assumption SA-2.7 is easily satisfied due to the inherent exclusions between $X_1$ and $X_2$. $X_1$ and $X_2$ are allowed to include continuous and/or discrete random variables. Noting that some or all of variables in $X_1$ can be included in $Z$. We let $W = X \cup Z$ and denote the joint pdf as $f_W(w)$.

The joint density $f_W(w)$ is unknown to the econometrician except for some smoothness conditions.

5.2 Identification and Estimation

Now we let $\theta = (\mu, \beta_1', \beta_2', \phi_1, \phi_{1w}, \phi_{2w}, \phi_2, \phi_{2w})' \in \mathbb{R}^{1+k_1+k_2} \times (0, \infty)^2 \times [0, \infty)^2$, $\alpha = (\theta, h) \in A \equiv \Theta \times \mathcal{H}$ where $h = \log \left( \frac{p}{1-p} \right) + h^o$. Then, we let $L_{y_1y_2}(W, \alpha)$ denote the conditional probabilities of observed outcomes such that

$$L_{y_1y_2}(W, \alpha) \equiv \Pr(Y_1 = y_1, Y_2 = y_2|W, \alpha) \text{ for } (y_1, y_2) = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$$

where $\Pr(Y_1 = y_1, Y_2 = y_2|W, \alpha)$ takes the rather complicated form and is defined in Appendix C. Recall that we need to distinguish the population probability $p$ from the posterior belief $p(Z)$ of Player 2, which is also the conditional probability of being strong type conditional on $Z$. We note that one may think we cannot identify $p$ without a restriction on $h^o$ since the conditional probabilities depend on $p(Z)$, not on $p$ separately. However, it turns out that we can still identify $p$ using the relationship

$$E[p(Z)] = \int p(z) f_Z(z)dz = \int \frac{p f_{(st)}(z)}{f_Z(z)} f_Z(z)dz = p \int f_{(st)}(z) dz = p. \quad (6)$$

We will discuss this issue later (see Section 5.3). Now we let $\alpha_0 \equiv (\theta_0, h_0)$ denote the true value of $\alpha$. Later we show that the following condition is sufficient for identification of $\alpha_0$.

\textsuperscript{20}Here we assume that those parameters are known, i.e. the functional forms of $g_1(\cdot)$ and $g_2(\cdot)$ are known to simplify our notation without loss of generality as Aradillas-Lopez (2005).
Assumption 5.2 \textbf{(SA-3)} Conditional on $W$, if $\alpha \neq \alpha_0$ for $\alpha, \alpha_0 \in A$, then

$$\Pr (\mathcal{L}_{y_1 y_2}(W, \alpha) \neq \mathcal{L}_{y_1 y_2}(W, \alpha_0)) > 0$$

for $(y_1, y_2) = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$.

Now we let

$$l(Y|W, \alpha) \equiv l(Y|W, \theta, h)$$

$$= \begin{bmatrix} Y_1 Y_2 \log \mathcal{L}_{11}(W, \theta, h) + Y_1 (1 - Y_2) \log \mathcal{L}_{10}(W, \theta, h) \\ + (1 - Y_1) Y_2 \log \mathcal{L}_{01}(W, \theta, h) + (1 - Y_1) (1 - Y_2) \log \mathcal{L}_{00}(W, \theta, h) \end{bmatrix}$$

denote the single observation conditional log likelihood function. We will estimate the parameters of interest using the conditional sieve ML based on this conditional likelihood function. In this estimation, we approximate the unknown function $h$ with some sieves. For this purpose, we need to restrict the space of functions $\mathcal{H}$ where $h_0$ belong. We consider the Hölder space $\Lambda^{\nu_1}(S(Z))$ with order $\nu_1 > 0$ as Ai and Chen (2003). The Hölder space is a space of functions $g : S(Z) \rightarrow \mathbb{R}$ such that the first $\nu_1$ derivatives are bounded, and the $\nu_1$-th derivatives are Hölder continuous with the exponent $\nu_1 - \nu_1 \in (0, 1]$, where $\nu_1$ is the largest integer smaller than $\nu_1$. The Hölder space becomes a Banach space when endowed with the Hölder norm:

$$\|g\|_{\Lambda^{\nu_1}} = \sup_z |g(z)| + \max_{a_1 + a_2 + \cdots + a_d = \nu_1} \sup_{z \neq z'} \frac{|\nabla^a g(z) - \nabla^a g(z')|}{(||z - z'||_E)^{\nu_1 - \nu_1}} < \infty,$$

where $\nabla^a g(z) \equiv \frac{\partial^{a_1 + a_2 + \cdots + a_d}}{\partial z_1^{a_1} \cdots \partial z_d^{a_d}} g(z)$. The Hölder ball (with radius $C_1$) $\Lambda^{\nu_1}_{C_1}(S(Z))$ is defined accordingly as

$$\Lambda^{\nu_1}_{C_1}(S(Z)) \equiv \{g \in \Lambda^{\nu_1}(S(Z)) : \|g\|_{\Lambda^{\nu_1}} \leq C_1 < \infty\}.$$

In the literature, it is known that functions in $\Lambda^{\nu_1}_{C_1}(S(Z))$ can be approximated well by various sieves such as power series, Fourier series, splines, and wavelets. We let $\mathcal{H} = \Lambda^{\nu_1}_{C_1}(S(Z))$. In particular, we approximate $h(\cdot)$ by power series. According to Theorem 8, p.90 in Lorentz (1986) (Also see Timan (1963)), if a function $f$ is $s$-times continuously differentiable, then there exist a $K$-vector $\gamma_K$ and a triangular array of polynomials $R^K(z)$ (Particularly $E[R^K(Z)R^K(Z)'] = I_K$) on the compact set $Z$ such that

$$\sup_{z \in Z} |f(z) - R^K(z)' \gamma_K| \leq C_1 K^{-\frac{s}{r}}.$$

We consider the tensor-product power series, $\mathcal{H}_n$, as a sieve space such that

$$\mathcal{H}_n = \{h(z)|h(z) = R^K(z)' \pi \text{ for all } \pi \text{ satisfying } \|h\|_{\Lambda^{\nu_1}} \leq C_1\}$$

where we let the first element of $R^K(z)$ equal to the constant 1. \footnote{For comprehensive discussion of tensor-product linear sieves including power series, see Chen (2005).}

\footnote{Noting $h = \log \left(\frac{1}{1-p}\right) + h''$ and thus, $h(0) = \log \left(\frac{1}{1-p}\right) + h''(0) \neq 0$ in general.}
Then, from the result of (7), we have a \( h_K(\cdot) \equiv R_K(\cdot)' \pi_K \) such that

\[
\sup_{z \in S(Z)} |h_0(z) - R_K(z)' \pi_K| = O \left( K^{-\frac{1}{2}} \right).
\] (9)

We obtain the estimator \( \tilde{\theta}_n \) and \( \tilde{h}_n \) as

\[
\tilde{\alpha}_n = \left( \tilde{\theta}_n, \tilde{h}_n \right) = \arg\max_{(\theta, h) \in \mathcal{A}_n \equiv \Theta \times \mathcal{H}_n} \hat{Q}(\theta, h) \equiv \frac{1}{n} \sum_{i=1}^{n} l(y_i | w_i, \theta, h(z_i)).
\] (10)

We call \( \tilde{\alpha}_n \) the exact sieve conditional ML estimator. Because the complexity (in the sense defined in the Appendix F.3) of the sieve space \( \mathcal{A}_n \) increases with \( n \) and because the maximizer of (10) is often obtained numerically, we do not require the maximization of \( \hat{Q}(\theta, h) \) over \( \mathcal{A}_n \) need to be exact. An approximated estimator, \( \hat{\alpha}_n \), is enough for the asymptotic results, such that

\[
\hat{Q}(\tilde{\alpha}_n) \geq \sup_{\alpha \in \mathcal{A}_n} \hat{Q}(\alpha) - O_p(\varepsilon_n), \text{ with } \varepsilon_n = o(1).
\] (11)

We call \( \tilde{\alpha}_n \) in (11) approximate sieve ML estimator. Note that we choose the order of \( \varepsilon_n \) such that it can justify a desirable asymptotics result. Now define

\[
Q(\alpha) \equiv Q(\theta, h) = E \left[ l(Y | W, \theta, h) \right].
\] (12)

The following lemma shows that if Assumption \( \text{SA-3} \) holds, then \( l(Y | W, \alpha) \) satisfies an information inequality result which is useful to prove the consistency of our proposed estimator \( \tilde{\alpha}_n \) in (11) together with the uniform convergence of \( \hat{Q}(\alpha) \) to \( Q(\alpha) \).

**Lemma 5.1 (Identification)**

Suppose Assumptions \( \text{IS-A} \) and \( \text{SA-2} \) hold. Further suppose Assumption \( \text{SA-3} \) holds. Then, \( Q(\alpha) < Q(\alpha_0) \) for all \( \alpha \neq \alpha_0 \in \mathcal{A} \).

The proof can be found in the Appendix.

It is worthwhile to note that the conditional ML estimator \( \tilde{\alpha}_n \) does not use all the information we have. In particular, we do not utilize the information about the mixing distribution of \( Z \). The conditional ML will be numerically identical to the ML if the density of \( W \) does not depend on the parameter \( \alpha \). However, we know that the density of \( Z \) depends on \( h^o \). If we assume that \( Z \) is independent with \((X_1 - Z) \cup X_2 \) (or we already eliminate such dependence), the full ML estimator will be given by

\[
\tilde{\alpha}_{\text{Full}}^o = \arg\max_{(\theta, p, h^o) \in \mathcal{A}_n \equiv \Theta \times (0,1) \times \mathcal{H}_n} \left[ \frac{1}{n} \sum_{i=1}^{n} \log(p \exp(h^o(z_i)) + (1 - p)) + \frac{1}{n} \sum_{i=1}^{n} l(y_i | w_i, \theta, p, h^o(z_i)) \right]
\] (13)

noting that neither \( f_{X_1-Z} \) or \( f_{X_2} \) depends on \( h^o \), that

\[
\log \left( pf_{(st)}(z) + (1 - p)f_{(we)}(z) \right) = \log f_{(we)}(z) + \log(p \exp(h^o(z_i)) + (1 - p)),
\]
and that \( f_{(we)}(z) \) does not depend on \( \alpha^0 \). Therefore, the conditional ML estimator \( \tilde{\alpha}_n \) is different from the corresponding ML and will be less efficient (see Gourieroux and Monfort (1995), Section 7.5.3) since it is obtained by dropping the first term in the log likelihood function. In the Appendix G, we discuss how to recover this lost information using a \textit{pseudo EM algorithm}. Note that in the ML, we may estimate all of the parameters at the same time since \( p \) appears separately from \( p(Z) \) in the likelihood function. Considering these advantages of ML, one may want to use ML instead of the conditional ML. However, we note that the kind of transformation in the likelihood function of (13) is not valid if \( f_{(st)}(\cdot) \) and \( f_{(we)}(\cdot) \) depend on some model parameters. If this is the case, the full ML requires to estimate \( f_{(st)}(\cdot) \) and \( f_{(we)}(\cdot) \) together with other model parameters, which we want to avoid\(^{23}\). For this reason, we will focus on the conditional ML. Though our proposed sieve conditional ML estimator do not achieve the efficiency, we note that the asymptotic results of the sieve ML estimator in the literature will be identically applied to the sieve conditional ML. The results obtained for the conditional ML in the following subsections can be applied to the ML estimator \( \tilde{\alpha}_n^{full} \) in a similar manner.

5.2.1 Consistency and Convergence Rates of the Sieve Conditional ML

The consistency of sieve MLE was derived in Wong and Severini (1991) and Geman and Hwang (1992) for i.i.d data. Some consistency results of sieve M-estimators can be found in Gallant (1987) and Gallant and Nychka (1987). Chen (2005) presented a consistency result of sieve extremum estimators allowing for non-compact infinite-dimensional \( A \), which is an extension of Theorem 2.1 in Newey and McFadden (1994) and of Lemma A1 in Newey and Powell (2003). Using Theorem 3.1 in Chen (2005), we establish the consistency under a pseudo metric \( \| \cdot \|_s \) defined by

\[
\| \alpha_1 - \alpha_2 \|_s = \| \theta_1 - \theta_2 \|_E + \| h_1 - h_2 \|_\infty
\]

where \( \| \cdot \|_E \) is the Euclidean norm and \( \| h \|_\infty = \sup_{z \in S(Z)} | h(z) | \). We make additional assumptions.

**Assumption 5.3 (SA-4)**

(i) \( \{ Y_{1i}, Y_{2i}, W_i \}_{i=1}^n \) are iid; (ii) \( \alpha_0 = (\theta_0, h_0) \in A = \Theta \times \mathcal{H} \); (iii) \( \Theta \) is compact with nonempty interior and \( \mathcal{H} = \Lambda_{C_1}^\nu (S(Z)) \); (iv) \( K_n \to \infty \) and \( K_n/n \to 0 \).

**Lemma 5.2 (Consistency)**

Suppose Assumptions \textbf{SA-2}, \textbf{SA-3}, and \textbf{SA-4} hold and suppose Condition 9 (Lipschitz) in the Appendix holds. Then, \( \| \hat{\alpha}_n - \alpha_0 \|_s = o_p(1) \).

The proof can be found in the Appendix. This consistency result is obtained combining the identification condition and the uniform convergence of the criterion function as in the case of the parametric extremum estimation.

\(^{23}\)A semiparametric estimation of a mixing distribution with nonparametric mixing components has not been established yet in the literature except some identification result in Kitamura (2004).
Now we consider the convergence rate of the sieve conditional ML estimator under a weaker metric. We present a convergence rate using Theorem 3.2 of Chen (2005) which is a version of Chen and Shen (1996) for i.i.d data. We redefine, \( \hat{\alpha}_n \), the approximate sieve ML that satisfies

\[
\hat{Q}(\alpha) - O_p(\varepsilon_n^2), \quad \text{with} \quad \varepsilon_n = o(1).
\]

(14)

Now suppose that \( A = \Theta \times \mathcal{H} \) is convex in \( \alpha_0 \) so that \( \alpha_0 + \tau(\alpha - \alpha_0) \in A \) for all small \( \tau \in [0, 1] \) and all fixed \( \alpha \in A \). Suppose that the pathwise derivative of \( l(\cdot) \) at the direction \( [\alpha - \alpha_0] \) is well-defined for almost all \( w \times y \) in the support of \( S(W) \times S(Y) \). We denote the pathwise first derivative at the direction \( [\alpha - \alpha_0] \) evaluated at \( \alpha_0 \) by

\[
\frac{d l(y_i|w_i, \alpha_0)}{d \alpha} \equiv \lim_{\tau \to 0} \frac{d l(y_i|w_i, (1 - \tau)\alpha_0 + \tau \alpha)}{d \tau} = \frac{d l(y_i|w_i, \alpha_0)}{d \theta} (\theta - \theta_0) + \frac{d l(y_i|w_i, \alpha_0)}{d h} [h - h_0].
\]

Now we define the \( L_2(P_0) \)-norm, \( \|\alpha - \alpha_0\|_2 \), based on the pathwise derivative of \( l(\cdot) \) evaluated at \( \alpha_0 \), i.e.

\[
\|\alpha - \alpha_0\|_2 = \sqrt{E \left[ \left( \frac{d l(Y_i|W_i, \alpha_0)}{d \alpha} \left[\alpha - \alpha_0\right] \right)^2 \right]}.
\]

(15)

This is the ML version of Ai and Chen (2003)'s \( L_2(P_0) \)-metric and the conditional ML version of the metric used by Wong and Severini (1991).

**Proposition 5.1** Let \( \hat{\alpha}_n \) be the approximate sieve ML defined in (14). Suppose Assumptions \( SA-2, SA-3, \) and \( SA-4 \) hold and suppose Conditions 10-12 (Lipschitz) in the Appendix hold. Then, we have

\[
\|\hat{\alpha}_n - \alpha_0\|_2 = O_p \left( \max \left\{ O \left( \frac{K_n}{n} \right), O \left( K_n^{-\nu_1/d_x} \right) \right\} \right).
\]

Moreover, with \( K_n = n^{1/(2\nu_1/d_x+1)} \) and \( \nu_1/d_x > 1/2 \), we have \( \|\hat{\alpha}_n - \alpha_0\|_2 = O_p \left( n^{-\frac{\nu_1}{2\nu_1/d_x+1}} \right) \).

The proof of this proposition can be found in the Appendix.

### 5.2.2 Asymptotic Normality

In this section, we derive the \( \sqrt{n} \)-asymptotic normality of the structural parameters \( \hat{\theta}_n \). The following discussion and notations are based on Theorem 4.3 of Chen (2005), which is a simplified

---


25For the ML context, this norm is very natural since it is the Fisher norm (see Wong and Severini (1991)).

26Wong and Severini (1991) established the \( \sqrt{n} \)-asymptotic normality and efficiency of smooth functionals of non-parametric MLE with parameter space \( \mathcal{A}_n \equiv \mathcal{A} = \Theta \times \mathcal{H} \). Shen (1997) extended their results to sieve MLE allowing for highly curved (nonlinear) least favorable directions while Chen and Shen (1998) extended Shen (1997) to general sieve M-estimation with stationary weakly dependent data.
version of Shen (1997) and Chen and Shen (1998).

Suppose the functional of interest, $f : A \to \mathbb{R}$, is smooth in the sense that

$$
\frac{df(\alpha_0)}{d\alpha}[\alpha - \alpha_0] \equiv \lim_{\tau \to 0} \frac{f(\alpha_0 + \tau(\alpha - \alpha_0)) - f(\alpha_0)}{\tau}
$$

is well defined and

$$
\left\| \frac{df(\alpha_0)}{d\alpha} \right\|_{\alpha - \alpha_0 > 0} \left\| \alpha - \alpha_0 \right\|_2 < \infty.
$$

Now let $\overline{V}$ denote the closure of the linear span of $A - \alpha_0$ under the metric $\left\| \alpha - \alpha_0 \right\|_2$. Then, $(\overline{V}, \left\| \cdot \right\|_2)$ is a Hilbert space with inner product

$$
\langle v_1, v_2 \rangle = E \left[ \left( \frac{dl(Y_i|W_i, \alpha_0)}{d\alpha} [v_1] \right) \left( \frac{dl(Y_i|W_i, \alpha_0)}{d\alpha} [v_2] \right) \right].
$$

Then, by the Riesz representation theorem, there exists $v^* \in \overline{V}$ such that, for any $\alpha \in A$,

$$
\frac{df(\alpha_0)}{d\alpha}[\alpha - \alpha_0] = \langle \alpha - \alpha_0, v^* \rangle \iff \left\| \frac{df(\alpha_0)}{d\alpha} \right\|_{\alpha - \alpha_0 > 0} < \infty. \tag{16}
$$

In particular, we let $f(\alpha) = \lambda' \theta$ for any fixed and nonzero $\lambda \in \mathbb{R}^{d_\theta}$. Then, $f(\alpha) \equiv \lambda' \theta$ is a linear functional on $\overline{V}$. To estimate $f(\alpha) \equiv \lambda' \theta$ at a $\sqrt{n}$ rate, $f(\alpha)$ has to be bounded ($\sup_{0 \neq \alpha - \alpha_0 \in \overline{V}} |f(\alpha) - f(\alpha_0)| / \left\| \alpha - \alpha_0 \right\|_2 < \infty$) according to Van der Vaart (1991) and Shen (1997).

Now let $\overline{\mathcal{H}} - h_0$ denote the closure of the linear span of $\mathcal{H} - h_0$. Define $b^*_j \in B \equiv \overline{\mathcal{H}} - h_0$ for each component $\theta_j$ of $\theta$ such that

$$
b^*_j = \arg\min_{b_j \in B} E \left[ \left( \frac{dl(Y|W, \alpha_0)}{d\theta_j} - \frac{dl(Y|W, \alpha_0)}{dh} [b_j] \right)^2 \right]. \tag{17}
$$

Now define $b^* = (b^*_1, \ldots, b^*_{d_\theta})$,

$$
\frac{dl(Y|W, \alpha_0)}{dh} [b^*] = \left( \frac{dl(Y|W, \alpha_0)}{dh} [b^*_1], \ldots, \frac{dl(Y|W, \alpha_0)}{dh} [b^*_d] \right),
$$

and $D_{b^*}(Y, W) \equiv \frac{dl(Y|W, \alpha_0)}{dh} [b^*]$. We impose

**Assumption 5.4 (SA-5)**

(i) $\theta_0 \in \text{int}(\Theta)$; (ii) $E [D_{b^*}(Y, W)^{\prime} D_{b^*}(Y, W)]$ is positive definite; (iii) each element $b^*_j(Z)$ belongs to the Hölder space $\Lambda_{C_j}^{m_j}(\mathcal{S}(Z))$ with $m_j > d_\zeta/2$.

Now note that $\frac{df(\alpha_0)}{d\alpha}[\alpha - \alpha_0] = (\theta - \theta_0)' \lambda$ which implies that

$$
f(\alpha) - f(\alpha_0) - \frac{df(\alpha_0)}{d\alpha}[\alpha - \alpha_0] = 0. \tag{18}
$$
In addition, we can show that for \( f(\alpha) \equiv X'\theta \) with \( \lambda \in \mathbb{R}^d_+, \lambda \neq 0, \)

\[
\sup_{0 \neq \alpha - \alpha_0 \in \mathcal{V}} \frac{|f(\alpha) - f(\alpha_0)|^2}{\|\alpha - \alpha_0\|^2} = \lambda' \left( E \left[ D_{b'}(Y, W)' D_{b'}(Y, W) \right] \right)^{-1} \lambda
\]

which implies \( f(\alpha) = X'\theta \) is bounded (in the sense of \( \sup_{0 \neq \alpha - \alpha_0 \in \mathcal{V}} |f(\alpha) - f(\alpha_0)| / \|\alpha - \alpha_0\| < \infty \) if and only if \( E \left[ D_{b'}(Y, W)' D_{b'}(Y, W) \right] \) is positive-definite (Assumption SA-5 (ii)). For this case, there exists \( v^* \in \mathcal{V} \) such that

\[
f(\alpha) - f(\alpha_0) \equiv \lambda' (\theta - \theta_0) = (v^*, \alpha - \alpha_0) \quad \text{for all } \alpha \in \mathcal{A}
\]

by (18) and by the Riesz representation theorem (see (16)). We find that \( v^* \equiv (v^*_b, v^*_h) \in \mathcal{V} \) satisfies (19) with \( v^*_b = (E \left[ D_{b'}(Y, W)' D_{b'}(Y, W) \right])^{-1} \lambda \) and \( v^*_h = -b^* \times v^*_b \).

The following theorem states that we can achieve the \( \sqrt{n} \)-asymptotic normality for the structural parameters.

**Theorem 5.1** Suppose Assumptions SA-2, SA-3, SA-4, and SA-5 hold and suppose Conditions 10-12 (Lipschitz) in the Appendix hold. Then, we have

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Omega_*)
\]

where \( \Omega_* = \{ E \left[ D_{b'}(Y, W)' D_{b'}(Y, W) \right]\}^{-1} \).

The proof of this theorem can be found in the Appendix.

### 5.2.3 A Consistent Covariance Estimator

To do a statistical inference of the structural parameters based on Theorem 5.1, we need a consistent estimator of \( \Omega_* \). First, we need a consistent estimator of \( b^* \). We estimate \( b^*_j \) by \( \hat{b}^*_j, j = 1, \ldots, d_\theta \) such that

\[
\hat{b}^*_j = \arg\min_{b_j \in \mathcal{N}_n} \sum_{i=1}^n \left( \frac{dl(y_i|w_i, \hat{\alpha}_n)}{d\theta_j} - \frac{dl(y_i|w_i, \hat{\alpha}_n)}{dh} [b_j] \right)^2.
\]

Note that for linear sieves, \( \hat{b}^*_j \) is easily obtained by regressing the derivatives of \( l(\cdot | \cdot, \hat{\alpha}_n) \) with respect to \( \theta_j \) on the derivatives of \( l(\cdot | \cdot, \hat{\alpha}_n) \) with respect to \( h \). Finally, we estimate \( \Omega_* \) by

\[
\hat{\Omega}_* = \left( \frac{1}{n} \sum_{i=1}^n D_{b^*}(y_i, w_i, \hat{\alpha}_n)' D_{b^*}(y_i, w_i, \hat{\alpha}_n) \right)^{-1}
\]

where \( \hat{b}^* = (\hat{b}^*_1, \ldots, \hat{b}^*_d_\theta) \) and \( D_{b^*}(Y, W, \alpha) = \frac{\partial \hat{H}(Y|W, \alpha)}{\partial \theta} - \frac{\partial \hat{H}(Y|W, \alpha)}{\partial h} [b^*] \). We note that \( \hat{\Omega}_* \) is consistent under suitable conditions.

**Proposition 5.2** Suppose Assumptions SA-2, SA-3, SA-4, and SA-5 hold and suppose Conditions 10-12 (Lipschitz) in the Appendix hold. Then, \( \hat{\Omega}_* = \Omega_* + o_p(1) \).
5.2.4 Alternative Sieve Minimum Distance Estimator (SMD)

Under uniqueness of equilibrium, the true conditional probabilities of observed outcomes should be identical with the conditional probabilities implied by a model evaluated at the true parameter. This implies that the game models can be represented by a set of conditional moment conditions. To be precise, we have

\[
0 = E[Y_1Y_2 - \Pr(Y_1 = 1, Y_2 = 1|W, \alpha_0)|W] \equiv E[\rho_1(Y_1, Y_2, W, \theta_0, h_0)|W] \\
0 = E[Y_1(Y_2 - 1) - \Pr(Y_1 = 1, Y_2 = 0|W, \alpha_0)|W] \equiv E[\rho_2(Y_1, Y_2, W, \theta_0, h_0)|W] \\
0 = E[(Y_1 - 1)Y_2 - \Pr(Y_1 = 0, Y_2 = 1|W, \alpha_0)|W] \equiv E[\rho_3(Y_1, Y_2, W, \theta_0, h_0)|W].
\]

Based on these conditional moment restrictions, we can estimate the parameters of interest using the sieve minimum distance estimation proposed by Ai and Chen (2003) as an alternative to the sieve conditional ML. Note that only three outcomes are independent out of the four possible outcomes of \((Y_1, Y_2)\) since the sum of probabilities of four outcomes is always one. Now let \(\rho(\cdot) = (\rho_1(\cdot), \rho_2(\cdot), \rho_3(\cdot))^t\) and then

\[
E[\rho(Y, W, \theta_0, h_0)|W] = 0.
\]

Ai and Chen (2003) provides a general framework dealing with an estimation for models of conditional moment restrictions such as (22) and obtain the consistency and \(\sqrt{n}\)-asymptotic normality of the estimator for \(\theta_0\).

Now let \(m(W, \alpha)\) be the conditional mean function of the moment equation \(E[\rho(Y_1, Y_2, W, \alpha)|W]\) and denote by \(\hat{m}(W, \alpha)\), a linear sieve estimator of \(m(\cdot, \cdot)\) as

\[
\hat{m}_l(W, \alpha) = \sum_{j=1}^{n} \rho_l(Y_j, W_j, \alpha)p^{k_\alpha}(W_j)^t(P'P)^{-1}p^{k_\alpha}(W), \quad l = 1, \ldots, \text{dim}(\rho) = 3,
\]

where \(p^{k_\alpha}(W)\) is a tensor-product sieve (such as power series, Fourier series, or B-splines) and \(P = (p^{k_\alpha}(W_1), \ldots, p^{k_\alpha}(W_n))^t\). Also denote by \(\hat{\Sigma}(W)\), a consistent estimator of a positive definite matrix \(\Sigma(W)\) for any given \(W\). Then, an SMD estimator based on the conditional moment restriction of (22)\textsuperscript{27} is obtained by

\[
\hat{\alpha}_{MD} = (\hat{\theta}_{MD}, \hat{h}_{MD}) = \arg\min_{\alpha=(\theta, h) \in \mathcal{A}_n \times \Theta \times H_n} \frac{1}{n} \sum_{i=1}^{n} \hat{m}(W_i, \alpha)^t[\hat{\Sigma}(W_i)]^{-1}\hat{m}(W_i, \alpha)
\]

where \(H_n\) is the sieve space defined in (8)\textsuperscript{28}.

\textsuperscript{27} Alternatively, the conditional moment conditions of (22) can be estimated using a penalized empirical likelihood estimation (PELE) proposed by Otsu (2003). The main difference between these two estimators is that in the PELE, we do not restrict the class of functions \(\mathcal{H}\) to be compact and do not approximate the function space \(\mathcal{H}\) with sieves \(\mathcal{H}_n\). Instead, the maximization is taken over the function space \(\mathcal{H}\) while controlling the physical plausibility (such as roughness) of functions using a penalty function. Otsu (2003) shows that the optimally weighted kernel smoothed version of MD (the original version of the SMD considered in Ai and Chen (2003)'s working paper version) and the PELE are asymptotically equivalent in the first order sense. From this result, we expect that the PELE is first order asymptotically equivalent to the optimally weighted SMD. However, we conjecture that the PELE is preferable to the SMD in terms of higher order asymptotics from the relationship between parametric GMM and EL estimators (see Newey and Smith (2004)).

\textsuperscript{28} The asymptotic property of this estimator is provided in the dissertation version of this paper.
5.3 Estimation of the Type Distribution

In Section 5.2, we argue that we can identify the parameter $p$ from $p(Z)$ by (6). From this result, we propose an estimator for $p$ as

$$
\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} L(\hat{h}_n(z_i)) = \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(\hat{h}_n(z_i))}{1 + \exp(\hat{h}_n(z_i))}
$$

(24)

where $\hat{h}_n(\cdot)$ is obtained from (10) or (23). Applying the mean value theorem with $\hat{h}_n$ that lies between $\hat{h}_n$ and $h_0$, we have

$$
\hat{p}_n - p_0 = \frac{1}{n} \sum_{i=1}^{n} \left( L(\hat{h}_n(Z_i)) - L(h_0(Z_i)) \right) + \frac{1}{n} \sum_{i=1}^{n} \left( L(h_0(Z_i)) - E[L(h_0(Z_i))] \right)
$$

(25)

$$
= \frac{1}{n} \sum_{i=1}^{n} L(\hat{h}_n(Z_i))(1 - L(\hat{h}_n(Z_i)))(\hat{h}_n(Z_i) - h_0(Z_i)) + \frac{1}{n} \sum_{i=1}^{n} (L(h_0(Z_i)) - E[L(h_0(Z_i))])
$$

$$
\leq \frac{1}{4} \left\| \hat{h}_n - h_0 \right\|_{\infty} + o_p(1)
$$

where the second equality is obtained by $L'(\cdot) = L(\cdot)(1 - L(\cdot))$ and the last result holds since $L(\cdot)(1 - L(\cdot)) \leq \frac{1}{4}$ uniformly and the second term in the RHS of (25) is $o_p(1)$ by LLN noting $|L(h_0(\cdot))| < 1$ uniformly and $\{Z_i\}_{i=1}^{n}$ are iid.

Thus, $\hat{p}_n$ is consistent as long as $\hat{h}_n$ is consistent. Now we can derive the asymptotic distribution of $\hat{p}_n$ similarly with Chen, Linton, and van Keilegom (2003) since (6) can be written as a moment condition

$$
m(p, h) = p - L(h), E[m(p_0, h_0)] = 0
$$

and since we have an initial estimator $\hat{h}_n$ such that $\left\| \hat{h}_n - h_0 \right\|_{\infty} = o_p(n^{-1/4})$ from the sieve conditional ML or the SMD\textsuperscript{29}. We obtain the following result. We let $M(h) = \int_{S(Z)} L(h)dF_Z$ and $M_n(h) = \frac{1}{n} \sum_{i=1}^{n} L(h(Z_i))$.

**Proposition 5.3** Suppose (i) $\left\| \hat{h}_n - h_0 \right\|_{\infty} = o_p(n^{-1/4})$ and (ii)

$$
\sqrt{n} \left( \int_{S(Z)} L(h_0)(1 - L(h_0)) \, dF_Z + M_n(h_0) - M(h_0) \right) \xrightarrow{d} N(0, V_p).
$$

(26)

Then, $\sqrt{n} \left( \hat{p}_n - p_0 \right) \xrightarrow{d} N(0, V_p)$.

Note that the first term in the LHS of (26) appears due to the fact that we use an initial estimator $\hat{h}_n$ and will disappear if $\hat{h}_n = h_0$. The proof of this Proposition can be found in the Appendix. In

\textsuperscript{29}Combining with the result of Lemma 2 in Chen and Shen (1998), which shows that $\left\| h - h_0 \right\|_{\infty} \leq \left\| h - h_0 \right\|_{2^{p+1}+c}$. 

24
the Appendix, we also show that the condition (ii) of Proposition 5.3 holds for the sieve conditional
ML estimator under Assumptions in Theorem 5.1. We note that even if an explicit form of \(V_p\)
can be derived, a feasible estimator of \(V_p\) may be difficult to calculate. Alternatively, we can use
the ordinary nonparametric bootstrap. The following proposition shows that we can consistently
estimate the distribution of \(p_n(b_p)\) using the bootstrap under some conditions. We use “*” to
denote the bootstrap counterpart of the original sample \(Z_i\). We let \(M_n(h) = \frac{1}{n} \sum_{i=1}^{n} L(h(Z_i^*))\).

**Proposition 5.4** Suppose (i) with \(P^*\)-probability tending to one, \(\hat{h}^*_n \in \mathcal{H},\) and \(\|\hat{h}^*_n - \hat{h}_n\|_\infty = o_P^*(n^{-1/4})\); (ii) \(\|\hat{h}_n - h_0\|_\infty = o(n^{-1/4})\) a.s.; (iii) for any positive sequence \(\delta_n = o(1),\)
\[
\sup_{\|h-h_0\|_\infty \leq \delta_n} |M_n(h) - M(h) - M_n(h_0) + M(h_0)| = o(n^{-1/2}) \text{ a.s.};
\]
\[
(iv) \sqrt{n} \left( \int_{S(Z)} L(\hat{h}_n)(1 - L(\hat{h}_n)) \left( \hat{h}^*_n - \hat{h}_n \right) dF_Z + M_n^*(\hat{h}_n) - M_n(\hat{h}_n) \right) = N(0, V_p) + o_P^*(1).
\]

Then, \(\sqrt{n} (\hat{p}_n - p_n)\) converges in distribution to \(N(0, V_p)\) in \(P^*\)-probability.

6 Discussion: Set Identification

In previous sections, to achieve uniqueness of equilibrium, we have used an equilibrium refinement
developed by Cho and Kreps (1987). For signaling games, however, the “self-fulfilling property”
is essential, which can yield multiple equilibria depending on different beliefs on plays of the other
party. In this section, we will consider the possibility of multiple equilibria under PBE without
attempting to resolve the equilibrium selection problem. As a consequence, we give up point
identification since we do not have a well-defined likelihood function and adopt a set identification
approach. Here we consider the model where some asymptotic inequalities may define a region
of parameters rather than a single point in the parameter space. The idea is as follows. When
multiple equilibria arise, there are regions of unobservables that are consistent with necessary
conditions for more than one equilibrium. Therefore, the probability of an event, implied by the
necessary condition, is greater than or equal to the true probability of the event. As a consequence,
these necessary conditions will provide a set of inequality constraints on the parameters rather
than a set of equality conditions. Interestingly, we note that as suggested in Figure 3 compared to
Figure 4, the multiplicity of equilibria may help the identification of parameters (though they are
set-identified) since there are more variations of equilibria.

We can adopt a simple estimation strategy that utilizes the sample analog of these population
necessary conditions as in Chernozhukov, Hong, and Tamer (2003) and Andrews, Berry, and Jia
(2004). The approach we take is from Andrews, Berry, and Jia (2004). Noting that Andrews, Berry,
and Jia (2004) do not allow for infinite dimensional parameters\(^{30}\), we consider a parameterization of \(h(\cdot)\) for the game with public signals. A separate study of Kim (2006) illustrates that we can actually allow for infinite dimensional parameters in this set estimation and inference for a fully symmetric version of the signaling game. Now let \(h(\cdot) = h(\cdot, \delta_0)\) for \(\delta_0 \in \mathcal{D}\) and let \(\vartheta_0 = (\theta'_0, \delta_0)'\). Recall \(Y_i = (Y_{1i}, Y_{2i}), y = (y_1, y_2), \) and \(\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})\). Denote \(\Omega(Y = y, W, \vartheta_0)\) to be the region of \(\varepsilon\) under which \(Y\) takes the value \(y\) given \(W\) and \(\vartheta_0\). To be precise,

\[
\Omega(y, w, \vartheta_0) \equiv \Omega(Y = y, W = w, \vartheta_0) = \{\varepsilon | Y_1 = y_1, Y_2 = y_2 \text{ given } W = w \text{ and } \vartheta_0\}.
\]

Then, the probability that the necessary conditions for \(Y = y\) holds will equal the probability that \(\varepsilon\) belongs to \(\Omega(y, w, \vartheta_0)\) given \(W = w\) and \(\vartheta_0\). The sieve ML or SMD approaches taken in the previous sections typically proceed by identifying a one-to-one mapping between the possible discrete outcomes and regions of the unobservables. Under the existence of multiple equilibria, probabilities of observed events are not necessarily equal to the probabilities of the associated regions of unobservables. Now for any \((y, w) \in \mathcal{S}(Y) \times \mathcal{S}(W)\), the probability is defined to be

\[
P(y|w, \vartheta) = \Pr(\varepsilon \in \Omega(y, w, \vartheta))
\]

When \(\vartheta = \vartheta_0\), this is a simple \(\varepsilon\)-orthant probability. Note that at the true parameter value \(\vartheta = \vartheta_0\), the probabilities of the necessary conditions must be at least as large as the true probabilities of the events \(y \in \mathcal{S}(Y)\) given \(W = w\), denoted by \(P_0(y|w)\):

\[
P(y|w, \vartheta_0) \geq P_0(y|w), \forall (y, w) \in \mathcal{S}(Y) \times \mathcal{S}(W).
\]

Notice that this inequality follows from the fact that the outcome \(y\) implies the necessary conditions for \(y\) but the necessary condition need not imply the outcome \(y\). Now based on the population inequality conditions of (28), we can follow Andrews, Berry, and Jia (2004)’s approach.

From the results of Section 3 (Figure 3), it is not difficult to construct the model probabilities for each discrete outcome of the game using (27). We note that though the usage of a mixed strategy by players is well justified theoretically, it is still a matter of question whether a mixed strategy is used in the real world. Some experimental\(^{31}\) and empirical\(^{32}\) studies are trying to answer this question but findings are mixed. As an attractive alternative, one may proceed the analysis assuming that any observed outcome can arise without any restriction in the regions where we have semi-separating equilibria in the game. This may still provide a tight bound for the estimate of the parameters. We may compare this result with that of allowing the semi-separating equilibria.

---

\(^{30}\)It is because ABJ utilizes finite numbers of cells to facilitate the estimation, which is not compatible with infinite dimensional parameters. Simply it violates the order condition for identification.


\(^{32}\)See Walker and Wooders (2001) and Chiappori, Levitt, and Groseclose (2002).
6.1 Set Estimator

The model probabilities \( \{P(y|W_i, \vartheta) : i = 1, \ldots, n\} \) induced by the games we have studied have analytic closed form expressions. This makes our problem simpler since we do not need to consider the simulation of the probabilities. However, we still need to consider the construction of analytic closed form expressions. This makes our problem simpler since we do not need to consider the model probabilities.

Now we briefly review the data-dependent construction of \( W \) cells following Andrews, Berry, and Jia (2004). Consider a set \( \{q_\gamma : \gamma \in \Gamma\} \) of real-valued weight functions on the support \( S(W) \) of \( W_i \), where \( \gamma \) is a subset of \( S(W) \) and \( \Gamma \) is a collection of subsets of \( S(W) \). In particular, for each \( y_j \), we consider \( M_j \) subsets of \( S(W) \) denoted by \( \gamma_{j,m} : \Gamma = \{\gamma_{j,m} \subset S(W) : (j, m) \in I_{J_Y,M}\} \), where

\[
I_{J_Y,M} = \{ (j, m) : m = 1, \ldots, M_j, j = 1, \ldots, J_Y \}. \tag{29}
\]

The functions \( \{q_\gamma : \gamma \in \Gamma\} \) aggregate and/or weight the necessary conditions for an equilibrium over different values of \( \vartheta \). Now let

\[
\tilde{\Gamma}_n = \{\tilde{\gamma}_{n,j,m} \subset S(W) : (j, m) \in I_{J_Y,M}\},
\]

where \( \tilde{\gamma}_{n,j,m} \) is a random subset of \( S(W) \). For the consistency of the set estimator, we require that \( \tilde{\Gamma}_n \to \Gamma_0 \) under certain metric where

\[
\Gamma_0 = \{\gamma_{0,j,m} \subset S(W) : (j, m) \in I_{J_Y,M}\}. \tag{30}
\]

The set estimator proposed by Andrews, Berry, and Jia (2004) is obtained as follows. Define

\[
c_0(j, \gamma, \vartheta) = \int (P(y_j|w, \vartheta) - P_0(y_j|w)) q_\gamma(w) dF_W(w) \quad \text{and} \quad \hat{c}_n(j, \gamma, \vartheta) = n^{-1} \sum_{i=1}^n (P(y_j|W_i, \vartheta) - 1[Y_i = y_j]) q_\gamma(W_i). \tag{31}
\]

Necessary conditions for \( \vartheta \) to be the true parameters are

\[
P(y|w, \vartheta) - P_0(y|w) \geq 0, \forall (y, w) \in S(Y) \times S(W) \tag{32}
\]

which implies that

\[
c_0(j, \gamma_{0,k,m}, \vartheta) \geq 0, \forall (j, m) \in I_{J_Y,M}. \tag{33}
\]

Define

\[
\Theta_0 = \{\vartheta \in \Theta \times D : (31) \text{ holds}\} \quad \text{and} \quad \Theta_+ = \{\vartheta \in \Theta \times D : (32) \text{ holds}\}.
\]

By definition, the set \( \Theta_0 \) is the smallest set of parameter values that necessarily includes the true value \( \vartheta_0 \). By construction, \( \Theta_+ \supset \Theta_0 \) since (31) implies (32). Now the set estimator is obtained as

\[
\hat{\Theta}_n = \{\vartheta \in \Theta \times D : \vartheta \text{ minimizes } Q_n(\vartheta) \text{ over } \Theta \times D\}
\]
where
\[ Q_n(\vartheta) = \sum_{(j,m) \in I_{j_\vartheta,m}} [\hat{\varphi}_n(j, \hat{\gamma}_{n,j,m}, \vartheta)] \cdot 1[\hat{\varphi}_n(j, \hat{\gamma}_{n,j,m}, \vartheta) \leq 0]. \quad (33) \]

Provided that Assumptions 1-6 of Andrews, Berry, and Jia (2004) are satisfied, we have
\[ d(\hat{\Theta}_n, \Theta_+) \to 0 \]
where \( d(\cdot, \cdot) \) is the Hausdorff metric that measures the distance between two sets. For the inference of this set estimator, Andrews, Berry, and Jia (2004) provide confidence intervals for individual parameters and confidence regions, whose critical values are obtained using the bootstrap methods.

6.2 Set Estimation of the Type Distribution

We have noted that in the conditional probabilities of observed outcomes, implied by the model, \( p \) does not appear separately from \( p(Z) \). However, from (6), we can still identify the type distribution parameter \( p \). Using the relationship in (6), we obtain a set estimator of \( p_0 \) such that
\[
\hat{p}_n = \left\{ p : p = \frac{1}{n} \sum_{i=1}^{n} L(h(z_i, \delta)) = \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(h(z_i, \delta))}{1 + \exp(h(z_i, \delta))} \right\}
\]
for each \( \delta \in \hat{\Theta}_n \).

We note that as long as \( d(\hat{\Theta}_n, \Theta_+) = o_p(1) \), \( \hat{p}_n \) converges to its population counterpart \( p_+ \) defined by
\[
P_+ = \left\{ p : p = E[L(h)] = E \left[ \frac{\exp(h(Z, \delta))}{1 + \exp(h(Z, \delta))} \right] \right\}
\]
for each \( \delta \in \Theta_+ \).

**Proposition 6.1** Suppose \( d(\hat{\Theta}_n, \Theta_+) = o_p(1) \) and \( \sup_{\delta \in D, z \in S(Z)} \left\| \frac{\partial h(z, \vartheta)}{\partial \vartheta} \right\|_E < \infty \).

Then, \( d(\hat{p}_n, p_+) = o_p(1) \).

7 Concluding Remarks

This paper develops an econometric modeling of a signaling game with two players where one player, *informed party*, has private information summarized by types. In particular, we provide an estimation strategy that identifies the payoffs structure and the distribution of types from data of observed actions. Though multiplicity of equilibria arises in the game, we show that uniqueness of equilibrium given a realization of payoffs can be achieved as long as players play a PBE and choose only one equilibrium out of multiple equilibria using the refinement of Cho and Kreps (1987). This uniqueness enables us to derive well-defined conditional probabilities that are useful for the estimation.

To provide some empirical relevance, we consider public signals about the type of a player. Technically, these signals cannot be manipulated by the *informed* player or at least the player has no incentive to manipulate the signals. Therefore, the *uninformed party* will use this information
to update her belief on types after observing an action of the *informed party* when a pooling equilibrium is played. Since the mixing distribution of these non-strategic signals on types is nonparametrically specified, we estimate the model using a sieve conditional MLE where the infinite dimensional parameters are approximated by sieves. Noting that the conditional probability of choosing a certain combination of actions can be written in terms of several conditional moment restrictions, as an alternative, we estimate the model using the sieve minimum distance (MD) estimation. In both methods, we obtain the consistency and the root $n$-asymptotic normality of structural parameters estimates.

We note that in the signaling game, multiple equilibria naturally arise given a realization of payoffs due to the *self-fulfilling* property. We resolve this problem by refining the equilibrium using an equilibrium selection rule that may be arbitrary and cannot be justified. As an attractive alternative to the previous approach, we allow for the possibility of multiple equilibria, without attempting to resolve the equilibrium selection problem. As a consequence, we give up point identification since we do not have a well-behaved likelihood function and adopt the set identification approach. In particular, we consider the model where some asymptotic inequalities may define a region of parameters rather than a single point in the parameter space. We adopt the important work of Andrews, Berry, and Jia (2004) for the set estimation.
Appendix I

A Equilibrium Refinement and Uniqueness of Equilibrium (Proof of Theorem 3.2)

• Under region $A_a \equiv \{(\varepsilon_1, \varepsilon_2) | \varepsilon_1 > \mu + \chi_1' \beta_1 + \phi_{1w} \text{ and } \chi_2' \beta_2 + (\phi_{2w} + \phi_{2w})p - \phi_{2w} < \varepsilon_2 < \chi_2' \beta_2 + \phi_{2w}\}$, we show that the pooling with $(B, B)$ cannot survive the Intuitive Criterion of Cho and Kreps (1987) while the pooling with $(Q, Q)$ survives it.

**Pooling with $(A_{1s}, A_{1w}) = (B, B)$ with $A_{2B} = F$**

Note $u_1(t_s; B, F) > u_1(t_s; Q, NF)$ since $\mu + \chi_1' \beta_1 - \varepsilon_1 - \phi_{1s} < 0$ under $A_a$ and note $u_1(t_w; B, F) > u_1(t_w; Q, F)$ since $\phi_{1w} > 0$. This means that the strong type has no incentive to deviate in any case. Thus, Player 2 will assign $\mu_2(t_1 = t_s|Q) = 0$ and hence Player 2 will choose $F$ after observing the deviation play $Q$ under region $A_a$. Note $u_1(t_w; B, F) = -\phi_{1w} < u_1(t_w; Q, F) = 0$.

Thus, the weak type will deviate for sure. Therefore, the equilibrium outcome fails the Intuitive Criterion.

**Pooling with $(A_{1s}, A_{1w}) = (Q, Q)$ with $A_{2Q} = F$**

Note $u_1(t_s; Q, F) > u_1(t_w; B, NF)$ since $\mu + \chi_1' \beta_1 - \varepsilon_1 - \phi_{1s} < 0$ under $A_a$ and note $u_1(t_w; Q, F) > u_1(t_w; B, F)$ since $\phi_{1w} > 0$. This means that the weak type has no incentive to deviate in any case. Thus, Player 2 will assign $\mu_2(t_1 = t_s|B) = 1$ and hence Player 2 will choose $NF$ after observing the deviation play $B$ under region $A_a$. Now we need to check whether the weak type is better off by deviation under this situation. Note $u_1(t_s; Q, F) = -\phi_{1s} > u_1(t_s; B, NF) = \mu + \chi_1' \beta_1 - \varepsilon_1$

under $S_1$. Thus, the strong type will not deviate. Therefore, the equilibrium outcome survives the Intuitive Criterion.

• Under region $A_b = \{\varepsilon_1 < \mu + \chi_1' \beta_1 - \phi_{1s} \text{ and } \chi_2' \beta_2 - \phi_{2w} < \varepsilon_1 < \chi_2' \beta_2 + (\phi_{2w} + \phi_{2w})p - \phi_{2w}\}$, we show that the pooling with $(Q, Q)$ cannot survive the Intuitive Criterion of Cho and Kreps (1987) while the pooling with $(B, B)$ survives it.

**Pooling with $(A_{1s}, A_{1w}) = (Q, Q)$ with $A_{2Q} = NF$**

Note $u_1(t_s; Q, NF) > u_1(t_w; B, F)$ since $\mu + \chi_1' \beta_1 - \varepsilon_1 + \phi_{1w} > 0$ under $A_b$ and note $u_1(t_w; Q, NF) > u_1(t_w; B, NF)$ since $\phi_{1w} > 0$. This means that the weak type has no incentive to deviate in any case. Thus, Player 2 will assign $\mu_2(t_1 = t_s|B) = 1$ and hence Player 2 will choose $NF$ after observing the deviation play $B$ under region $A_b$. Now we need to check whether the strong type is better off by deviation under this situation. Note $u_1(t_s; Q, NF) = \mu + \chi_1' \beta_1 - \varepsilon_1 - \phi_{1s} < u_1(t_s; B, NF) = \mu + \chi_1' \beta_1 - \varepsilon_1$

since $\phi_{1s} > 0$. Thus, the strong type will deviate for sure. Therefore, the equilibrium outcome fails the Intuitive Criterion.

**Pooling with $(A_{1s}, A_{1w}) = (B, B)$ with $A_{2B} = NF$**

Note $u_1(t_s; B, NF) > u_1(t_s; Q, NF)$ since $\phi_{1s} > 0$ and note $u_1(t_s; B, NF) > u_1(t_s; Q, F)$ under $A_b$. 30
This means that the strong type has no incentive to deviate in any case. Thus, Player 2 will assign \( \mu_2(t_1 = t_s | Q) = 0 \) and hence Player 2 will choose \( F \) after observing the deviation play \( Q \) under region \( A_b \). Now we need to check whether the weak type is better off by deviation under this situation. Note

\[
u_1(t_w; B, NF) = \mu + X_1^t \beta_1 - \varepsilon_n - \phi_1 w > u_1(t_w; Q, F) = 0
\]

under region \( S_5 \). Thus, the weak type will not deviate. Therefore, the equilibrium outcome survives the Intuitive Criterion.

- The separating equilibrium \((B, Q)\) with \((Y_1|B = 1, Y_1|Q = 0)\) under region \( S_3 \) cannot fail the Intuitive Criterion since none of Player 1 wants to deviate regardless of Player 2’s action.

### B Game with Asymmetric Payoffs

For the game presented in Figure 2, we note that the payoffs structure is restrictive. We relax this restriction as in the game of Figure A1 by allowing \( \mu \) and \( \beta_1 \) to be different for different types. It turns out that we can achieve uniqueness of equilibrium even under asymmetric payoffs by imposing Conditions 1-2 (see the Appendix D.3).

**Condition 1** \( \mu_w + X_1^t \beta_w - \phi_1 w < \mu_s + X_1^t \beta_s + \phi_1 s \) for all \( X_1 \in S(X_1) \).

**Condition 2** \( \mu_s + X_1^t \beta_s - \phi_1 s < \mu_w + X_1^t \beta_w + \phi_1 w \) for all \( X_1 \in S(X_1) \).

These conditions hold immediately in the symmetric payoffs case \((\mu_w = \mu_s, \beta_w = \beta_s)\) since \( \phi_1 s, \phi_1 w > 0 \).

We summarize the result as

**Corollary B.1** Suppose Conditions 1-2 hold. Further suppose Assumptions IS and SA-1 hold and that \( \phi_1 s > 0, \phi_1 w > 0, \phi_2 s > 0, \) and \( \phi_2 w \geq 0 \) (but \( \phi_2 s \cdot \phi_2 w \neq 0 \)). Suppose each Player plays only one equilibrium that survives the refinement of Cho and Kreps (1987), when there exist multiple equilibria. Then, there exist unique equilibrium for each region of \((\varepsilon_1, \varepsilon_2)\) given \( X \).

See the Appendix D.3 for the proof. Figure A2 illustrates uniqueness of equilibrium for the game with IS and asymmetric payoffs. We note that all the estimation strategies considered in this paper are still valid for the asymmetric payoffs case as long as Assumption SA-Asym holds. However, even if Assumption SA-Asym is violated for certain observations in the data, we can still estimate the game model using a trimming device that trims out those observations violating Assumption SA-Asym under certain conditions.

**Assumption B.1** (SA-Asym)

\[
\mu_w + X_1^t \beta_w - \phi_1 w < \mu_s + X_1^t \beta_s + \phi_1 s \ \text{and} \ \mu_s + X_1^t \beta_s - \phi_1 s < \mu_w + X_1^t \beta_w + \phi_1 w \ \text{hold for all} \ X_1 \in S(X_1) \ \text{and for all} \ (\mu_s, \mu_w, \beta_s, \beta_w, \phi_1 s, \phi_1 w) \ \text{in the parameter space}.
\]

We note that this assumption is embedded when we derive the appropriate conditional probabilities in the following section. Again note that Assumption SA-Asym holds immediately in the symmetric payoffs game \((\mu_w = \mu_s, \beta_w = \beta_s)\) since \( \phi_1 s, \phi_1 w > 0 \).
C Conditional Probabilities of Four Observed Outcomes

Summarizing the result derived in the Appendix D.2, here we provide the conditional probabilities of four observed outcomes in the game with IS-A and SA-2 under asymmetric payoffs of Figure A1. For the game with IS and SA-1, we obtain the same conditional probabilities in replace of \( p(Z) \) with \( p \) and \( W \) with \( X \), respectively. Also note that the corresponding conditional probabilities of the symmetric payoffs game (Figure 2) are easily obtained by replacing \( \beta_s \) & \( \beta_w \) with \( \beta_1 \) and \( \mu_s \) & \( \mu_w \) with \( \mu \), respectively.

1. \((Y_1 = 1, Y_2 = 1): (B, NF)\)

It happens under \( S_5 \) with probability one (pooling), under \( S_3 \cup S_7 \) with probability \( p(Z) \) (separating), and under \( S_2 \cup S_6 \) (semi-separating). From these, we have

\[
\Pr(Y_1 = 1, Y_2 = 1|W, \alpha) = \frac{G_1(\mu_w + X_1\beta_w - \phi_{1w}) (G_2(X_2\beta_2 + p(Z)(\phi_{2s} + \phi_{2w}) - \phi_{2w}) - G_2(X_2\beta_2 - \phi_{2w}))}{p(Z)G_2(X_2\beta_2 - \phi_{2w})} \\
+ \int_0^1 p(Z) \mu^*_B g_1 \left( X_2\beta_2 + \left[ \frac{p(Z)}{p(Z) + (1-p(Z))\mu^*_B} \right] \phi_{2s} + \phi_{2w} \right) \frac{p(Z)(1-p(Z)) (\phi_{2s} + \phi_{2w})}{(p(Z) + (1-p(Z))\mu_B^*)^2} d\mu_B^* \\
+ \int_0^1 \mu^*_B g_2 \left( X_2\beta_2 + \left[ \frac{(1-\mu^*_B)p(Z)}{(1-\mu^*_B)p(Z) + (1-p(Z))} \right] \phi_{2s} + \phi_{2w} \right) \frac{p(Z)(1-p(Z)) (\phi_{2s} + \phi_{2w})}{(1-\mu^*_B)^2} d\mu_B^* \\
\times \frac{\phi_{2w}}{1-\sigma_2^2} \frac{\phi_{2s}}{\sigma_2^2} d\sigma_2
\]

2. \((Y_1 = 1, Y_2 = 0): (B, F)\)
It happens under $S_1$ with probability $p(Z)$ (separating) and under $S_2 \cup S_6$ (semi-separating). From these, we obtain

\[
\Pr(Y_1 = 1, Y_2 = 0|W, \alpha) = p(Z) + G_2(X'_2\beta_2 + \phi_2w) + \int_0^1 (1 - p(Z)) (1 - \mu_B) g_2\left(X'_2\beta_2 + \frac{p(Z)}{p(Z)+(1-p(Z))} \left(\phi_2s + \phi_2w\right) - \phi_2w\right) \frac{p(Z)(1-p(Z)) \left(\phi_2s + \phi_2w\right)}{(1-\mu_B^2p(Z))^2} \, d\mu_B^n
\]

\[
\times \int_0^1 (1-\sigma_2) g_1\left(\mu_w + X'_1\beta_w - \frac{\phi_1w}{\sigma_2}\right) \frac{\phi_1w}{\sigma_2^2} d\sigma_2
\]

It happens under $S_7$ with probability $1 - p(Z)$ (separating), and under $S_2 \cup S_6$ (semi-separating). From these, we have

\[
\Pr(Y_1 = 0, Y_2 = 1|W, \alpha) = (1 - p(Z)) + G_2(X'_2\beta_2 - \phi_2w) + \int_0^1 (1 - p(Z)) (1 - \mu_B) g_2\left(X'_2\beta_2 + \frac{p(Z)}{p(Z)+(1-p(Z))} \left(\phi_2s + \phi_2w\right) - \phi_2w\right) \frac{p(Z)(1-p(Z)) \left(\phi_2s + \phi_2w\right)}{(1-\mu_B^2p(Z))^2} \, d\mu_B^n
\]

\[
\times \int_0^1 (1-\sigma_2) g_1\left(\mu_w + X'_1\beta_w - \frac{\phi_1w}{\sigma_2}\right) \frac{\phi_1w}{\sigma_2^2} d\sigma_2
\]

4 (Y_1 = 0, Y_2 = 0) : (Q, F)

It happens under $S_4$ with probability one (pooling), under $S_1 \cup S_3$ with probability $1 - p(Z)$ (separating), and under $S_2 \cup S_6$ (semi-separating). From these, we conclude

\[
\Pr(Y_1 = 1, Y_2 = 0|W, \alpha) = (1 - G_1(\mu_s + X'_1\beta_s + \phi_1s)) (G_2(X'_2\beta_2 + \phi_2s) - G_2(X'_2\beta_2 + p(Z) \left(\phi_2s + \phi_2w\right) - \phi_2w))
\]

\[
+(1 - p(Z)) (1 - G_2(X'_2\beta_2 + \phi_2s)) + \int_0^1 (1 - p(Z)) G_1(\mu_s + X'_1\beta_s + \phi_1s) - G_1(\mu_w + X'_1\beta_w - \phi_1w)) (G_2(X'_2\beta_2 + \phi_2s) - G_2(X'_2\beta_2 - \phi_2w))
\]

\[
\int_0^1 (1 - p(Z)) (1 - \mu_B) g_2\left(X'_2\beta_2 + \frac{p(Z)}{p(Z)+(1-p(Z))} \left(\phi_2s + \phi_2w\right) - \phi_2w\right) \frac{p(Z)(1-p(Z)) \left(\phi_2s + \phi_2w\right)}{(1-\mu_B^2p(Z))^2} \, d\mu_B^n
\]

\[
\times \int_0^1 (1-\sigma_2) g_1\left(\mu_w + X'_1\beta_w - \frac{\phi_1w}{\sigma_2}\right) \frac{\phi_1w}{\sigma_2^2} d\sigma_2
\]

\[
\int_0^1 (1 - \mu_B^2p(Z)) g_2\left(X'_2\beta_2 + \frac{1(1-\mu_B^2p(Z))}{1-\mu_B^2p(Z)} \left(\phi_2s + \phi_2w\right) - \phi_2w\right) \frac{p(Z)(1-p(Z)) \left(\phi_2s + \phi_2w\right)}{(1-\mu_B^2p(Z))^2} \, d\mu_B^n
\]

\[
\times \int_0^1 (1-\sigma_2) g_1\left(\mu_s + X'_1\beta_s + \frac{\phi_1s}{1-\sigma_2}\right) \frac{\phi_1s}{(1-\sigma_2)^2} d\sigma_2
\]
Appendix II

D  Perfect Bayesian Equilibrium

D.1  Definition of PBE

We let \( P(h) \) and \( A(h) \) be the player function and the action set at the history \( h \), respectively.

**Definition D.1 (Osborne and Rubinstein (p.233, 1994))**

Let \( (\Gamma, (\Xi_i), (p_i), (u_i)) \) be a Bayesian extensive game with observable actions, where \( \Gamma = (N, H, P) \). A pair \( ((\sigma_i), (\mu_i)) = \left( (\sigma_i(t_i))_{i \in N, t_i \in \Xi_i}, (\mu_i(h))_{i \in N, h \in H \setminus T} \right) \), where \( \sigma_i(t_i) \) is a behavioral strategy of player \( i \) in \( \Gamma \), \( \mu_i(h) \) is a probability measure on \( \Xi_i \), and \( T \) is the terminal history, is a perfect Bayesian equilibrium of the game if the following conditions are satisfied

- Sequential Rationality: For every nonterminal history \( h \in H \setminus T \) every player \( i \in P(h) \), every \( t_i \in \Xi_i \), \( u_i(\sigma_{-i}, \sigma_i(t_i), \mu_{-i}|h) \) is at least as good for type \( t_i \) as \( u_i(\sigma_{-i}, s_i, \mu_{-i}|h) \) for any strategy \( s_i \) of player \( i \) in \( \Gamma \).
- Correct initial beliefs: \( \mu_i(\emptyset) = p_i \) for each \( i \in N \)
- Action-determined beliefs: If \( i \notin P(h) \) and \( a \in A(h) \) then \( \mu_i(h, a) = \mu_i(h) \); if \( i \in P(h) \), \( a \in A(h) \), \( a' \in A(h) \), and \( a_i = a_i' \) then \( \mu_i(h, a) = \mu_i(h, a') \).
- Bayesian updating: If \( i \in P(h) \) and \( a_i \) is in the support of \( \sigma_i(t_i|h) \) for some \( t_i \) in the support of \( t_i(h) \) then for any \( t'_i \in \Xi_i \) we have

\[
\mu_i(h, a)(t'_i) = \frac{\sigma_i(t'_i)(h)(a_i) \cdot \mu_i(h)(t'_i)}{\sum_{t_i \in \Xi_i} \sigma_i(t_i)(h)(a_i) \cdot \mu_i(h)(t_i)}.
\]

D.2  PBE of the Game G with IS

We derive equilibria of the game with asymmetric payoffs as in Figure A1. Equilibria of the game with symmetric payoffs \( (\beta_s = \beta_w, \mu_s = \mu_w) \) are easily obtained from the results in this section. We impose \( \phi_{1s} > 0, \phi_{1w} > 0, \phi_{2s} \geq 0 \), and \( \phi_{2w} \geq 0 \) (but \( \phi_{2s} \cdot \phi_{2w} \neq 0 \)). We let \( E_i[u_i(t_1; A_1, A_2)] \) be the expected payoffs of Player \( i \) based on Player \( i \)'s information for \( i \in \{1, 2\} \).

D.2.1  Pooling Equilibria

P-1) Pooling Equilibrium (both \( t_s \) and \( t_w \) choose \( B \)):

Then, Player 2 does not update its belief and thus the posterior equals to the prior belief: \( \mu_2(t_1 = t_s|B) = p \) and \( \mu_2(t_1 = t_w|B) = 1 - p \). Thus, the expected payoffs of Player 2 from choosing each action on the equilibrium path will be

\[
E_2[u_2(t_1; B, NF)] = X_2^s \beta_2 + p \cdot \phi_{2s} - \varepsilon_2 \quad \text{and} \quad E_2[u_2(t_1; B, F)] = (1 - p) \cdot \phi_{2w}
\]
and hence \( Y_{2|B} = 1 \{ X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} - \varepsilon_2 \geq 0 \} \) on the equilibrium path. The resulting payoffs of each type of Player 1 on the equilibrium path will be

\[
\begin{align*}
Y_{2|B} &= 1 \\
Y_{2|B} &= 0
\end{align*}
\]

\[
\begin{align*}
u_1(t_s; B, \cdot) &\quad \mu_s + X'_1 \beta_s - \varepsilon_1 \\
u_1(t_w; B, \cdot) &\quad \mu_w + X'_1 \beta_w - \varepsilon_1 - \phi_{1w} - \phi_{1w}
\end{align*}
\]

1) To have \( Y_{2|B} = 1 \) as an equilibrium with pooling \((A_{1t_s}, A_{1t_w}) = (B, B)\):

**P1-11:** \( t_s \) should have no incentive to deviate. If \( u_1(t_s; B, NF) = \mu_s + X'_1 \beta_s - \varepsilon_1 \geq u_1(t_s, Q, F) = -\phi_{1s} \), then no matter what the value of \( t_1(t_s = t_s(Q), t_s \) has no incentive to deviate by construction. Now assume \( \mu_s + X'_1 \beta_s - \varepsilon_1 \leq -\phi_{1s} \). Player 1 expect that Player 2 with \( \overline{p}_2 | = \mu_s(t_1 = t_s(Q) \text{ as off-the-equilibrium belief will choose } Y_{2|Q} = 1 \{ X'_2 \beta_2 + \overline{p}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w} - \varepsilon_2 \geq 0 \} \). To make Player 1 not deviate, the belief should induce Player 2 chooses \( NF \) after observing \( Q \). It is required that \( Y_{2|Q} = 1 \) while \( Y_{2|B} = 1 \). This holds when \( \min (X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}, X'_2 \beta_2 + \overline{p}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w}) \geq \varepsilon_2 \).

**P1-12:** \( t_w \) should have no incentive to deviate. Since \( u_1(t_w; B, NF) < u_1(t_w; Q, NF) \), Player 1 of type \( t_w \) has an incentive to deviate unless \( u_1(t_w; B, NF) \geq u_1(t_w; Q, F) \) and Player 2 chooses \( F \) after seeing \( Q \) (while choosing \( NF \) if seeing \( B \)). Not-to-deviate conditions require that \( \mu_w + X'_1 \beta_w - \varepsilon_1 \geq \phi_{1w} \geq 0 \) and \( X'_2 \beta_2 + \overline{p}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w} \leq \varepsilon_2 \leq X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} \).

Combining P1-11 and P1-12, we conclude that \( Y_{2|B} = 1 \) and \( Y_{2|Q} = 0 \) with pooling \((A_{1t_s}, A_{1t_w}) = (B, B)\) can be supported as an equilibrium where Player 2 assigns a belief with \( \overline{p}_2 = \mu_s(t_1 = t_s(Q) \) under

\[
\begin{align*}
\varepsilon_1 &\leq \min (\mu_w + X'_1 \beta_w - \phi_{1w}, \mu_s + X'_1 \beta_s + \phi_{1s}) \\
X'_2 \beta_2 + \overline{p}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w} &\leq \varepsilon_2 \leq X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w},
\end{align*}
\]

which requires \( \overline{p}_2 < p \) for the existence of this equilibrium. Note that for any \( \overline{p}_2 < p \) satisfying (34), the equilibrium is supported\(^{\text{34}}\) and hence regions of \((\varepsilon_1, \varepsilon_2)\) that support the equilibrium are actually

\[
\begin{align*}
\varepsilon_1 &\leq \min (\mu_w + X'_1 \beta_w - \phi_{1w}, \mu_s + X'_1 \beta_s + \phi_{1s}) \\
X'_2 \beta_2 - \phi_{2w} &\leq \varepsilon_2 \leq X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}.
\end{align*}
\]

2) To have \( Y_{2|B} = 0 \) as an equilibrium with pooling \((A_{1t_s}, A_{1t_w}) = (B, B)\):

We first assume that Player 1 expects that Player 2 will have \( \overline{p}_2 = \mu_s(t_1 = t_s(Q) \) as off-the-equilibrium belief after observing \( Q \).

**P1-21:** \( t_s \) should have no incentive to deviate. Player 1 of type \( t_s \) will have no incentive to deviate from the equilibrium in either of two cases

i. \( u_1(t_s; Q, NF) \leq u_1(t_s; B, F) \). This holds when \( \varepsilon_1 \geq \mu_s + X'_1 \beta_s - \phi_{1s} \).

ii. \( u_1(t_s; Q, NF) \geq u_1(t_s; B, F) \) but \( Y_{2|Q} = 0 \). This holds when \( \varepsilon_1 \leq \mu_s + X'_1 \beta_s - \phi_{1s} \) and \( X'_2 \beta_2 + \overline{p}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w} \leq \varepsilon_2 \).

**P1-22:** \( t_w \) should have no incentive to deviate. Since \( u_1(t_w; B, F) < u_1(t_w; Q, F) \), Player 1 of type \( t_w \) has an incentive to deviate unless \( u_1(t_w; Q, NF) \leq u_1(t_w; B, F) \) and \( Y_{2|Q} = 1 \). This requires \( \varepsilon_1 \geq \mu_w + X'_1 \beta_w + \phi_{1w} \) and \( X'_2 \beta_2 + \overline{p}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w} \geq \varepsilon_2 \) for not-to-deviate play.

\(^{\text{33}}\)Note that \( \phi_{2s} + \phi_{2w} > 0 \).

\(^{\text{34}}\)This means that for any realization of \( X_2 \) and \( \varepsilon_2 \) satisfying \( \varepsilon_2 \leq X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} \), there exist values of \( \overline{p}_2 = \overline{p}_2(X'_2 \beta_2, p, \phi_{2s}, \phi_{2w}, \varepsilon_2) \), \( 0 \leq \overline{p}_2 < p \) such that (34) holds.
Combining P1-21 and P1-22, we conclude that under

\[ \varepsilon_1 \geq \max (\mu_s + X'_1 \beta_s - \phi_1s, \varepsilon_1, \mu_w + X'_1 \beta_w + \phi_1w) \] \( \text{and} \)

\[ X'_2 \beta_2 + p (\phi_2s + \phi_2w) - \phi_2w \leq \varepsilon_2 \leq X'_2 \beta_2 + \bar{\mu}_2 (\phi_2s + \phi_2w) - \phi_2w, \] \( (35) \)

the equilibrium of \( Y_{2|B'} = 0 \) and \( Y_{2|Q} = 1 \) with pooling \((A_{1t_s}, A_{1t_w}) = (Q, Q)\) can be supported as an equilibrium where Player 2 holds off-the-equilibrium belief \( \bar{\mu}_2 \equiv \mu_2(t_1 = t_s | Q) \). Note that for any \( \bar{\mu}_2 > p \) that satisfies \((35)\), the equilibrium is supported\(^{35}\) and hence regions of \((\varepsilon_1, \varepsilon_2)\) that support the equilibrium are actually

\[ \varepsilon_1 \geq \max (\mu_s + X'_1 \beta_s - \phi_1s, \mu_w + X'_1 \beta_w + \phi_1w) \] \( \text{and} \)

\[ X'_2 \beta_2 + p (\phi_2s + \phi_2w) - \phi_2w \leq \varepsilon_2 \leq X'_2 \beta_2 + \bar{\mu}_2 (\phi_2s + \phi_2w) - \phi_2w. \]

**P-2** Pooling Equilibrium (both \( t_s \) and \( t_w \) choose \( Q \)):

In this case, Player 2 does not update its belief and thus the posterior equals to the prior belief: \( \mu_2(t_1 = t_s | Q) = p \) and \( \mu_2(t_1 = t_w | Q) = 1 - p \). Thus, the expected payoffs of Player 2 from choosing each action on the equilibrium path will be

\[ E_2[u_2(t_1; Q, NF)] = X'_2 \beta_2 + p \cdot \phi_2s - \varepsilon_2 \text{ and } E_2[u_2(t_1; Q, F)] : (1 - p) \cdot \phi_2w \]

and hence \( Y_{2|Q} = 1 \{ X'_2 \beta_2 + p \{ \phi_2s + \{ \phi_2w \} - \phi_2w - \varepsilon_2 \geq 0 \} \) on the equilibrium path. The resulting payoffs of each type of Player 1 on the equilibrium path will be

\[
\begin{align*}
Y_{2|Q} &= 1 & Y_{2|Q} &= 0 \\
\mu_s + X'_1 \beta_s - \varepsilon_1 - \phi_1s & -\phi_1s & u_1(t_s; Q, \cdot) & \cdot \\
\mu_w + X'_1 \beta_w - \varepsilon_1 & 0 & u_1(t_w; Q, \cdot) & \cdot 
\end{align*}
\]

1) To have \( Y_{2|Q} = 1 \) as an equilibrium with pooling \((t_s, t_w) = (Q, Q)\):

**P2-11**: \( t_w \) should have no incentive to deviate. If \( u_1(t_w; Q, NF) = \mu_w + X'_1 \beta_w - \varepsilon_1 \geq u_1(t_w; B, F) = -\phi_1w \), then no matter what the value of \( \mu_2(t_1 = t_w | B) \), \( t_w \) has no incentive to deviate by construction. Now assume \( \mu_w + X'_1 \beta_w - \varepsilon_1 \leq -\phi_1w \). Player 2 with \( \bar{\mu}_2 \equiv \mu_2(t_1 = t_w | B) \) as off-the-equilibrium belief will choose \( Y_{2|B} = 1 \{ X'_2 \beta_2 + \bar{\mu}_2 (\phi_2s + \phi_2w) - \phi_2w - \varepsilon_2 \geq 0 \} \). Thus, Player 1 of type \( t_w \) will have no incentive to deviate from the equilibrium if Player 2 chooses \( NF \) after observing \( B \). This requires that \( Y_{2|B} = 1 \) together with \( Y_{2|Q} = 1 \):

\[ \varepsilon_2 \leq \min (X'_2 \beta_2 + p (\phi_2s + \phi_2w) - \phi_2w, X'_2 \beta_2 + \bar{\mu}_2 (\phi_2s + \phi_2w) - \phi_2w). \]

**P2-12**: \( t_s \) should have no incentive to deviate. Since \( u_1(t_s; Q, NF) < u_1(t_s; B, NF) \), Player 1 of type \( t_s \) has an incentive to deviate unless \( u_1(t_s; Q, NF) \geq u_1(t_s; B, F) \) and Player 2 chooses \( F \) after seeing \( B \) (while choosing \( NF \) if seeing \( Q \)). Not-to-deviate conditions require that

\[ \mu_s + X'_1 \beta_1 - \varepsilon_1 - \phi_1s \geq 0 \text{ and } \]

\[ X'_2 \beta_2 + \bar{\mu}_2 (\phi_2s + \phi_2w) - \phi_2w \leq \varepsilon_2 \leq X'_2 \beta_2 + p (\phi_2s + \phi_2w) - \phi_2w. \]

\(^{35}\)Similarly as before, this means that for any realization of \( X_2 \) and \( \varepsilon_2 \) satisfying \( \varepsilon_2 \geq X'_2 \beta_2 + p (\phi_2s + \phi_2w) - \phi_2w \), there exist values of \( \bar{\mu}_2 = \bar{\mu}_2 (X'_2 \beta_2, p, \phi_2w, \varepsilon_2) \), \( p < \bar{\mu}_2 \leq 1 \) such that \((35)\) holds.
Similarly to the previous case, note that for any $\phi_{2w} > 0$ supporting $A_{1s}$, $A_{1w}$, the equilibrium is supported similarly as before and hence regions of $(\varepsilon_1, \varepsilon_2)$ that support the equilibrium are actually

$$
\varepsilon_1 \leq \min (\mu_s + X'_1 \beta_s - \phi_{1w}, \mu_w + X'_1 \beta_w + \phi_{1w}) \quad \text{and} \quad X'_2 \beta_2 + \bar{\mu}_2 (\phi_{2s} + \phi_{2w}) - \phi_{2w} \leq \varepsilon_2 \leq X'_2 \beta_2 + p (\phi_{2s} + \phi_{2w}) - \phi_{2w},
$$

which requires $\bar{\mu}_2 < p$ for the existence of this equilibrium. Note that for any $\bar{\mu}_2 < p$ that satisfies (36), the equilibrium is supported similarly as before and hence regions of $(\varepsilon_1, \varepsilon_2)$ that support the equilibrium are actually

$$
\varepsilon_1 \leq \min (\mu_s + X'_1 \beta_s - \phi_{1w}, \mu_s + X'_1 \beta_w + \phi_{1w}) \quad \text{and} \quad X'_2 \beta_2 - \phi_{2w} \leq \varepsilon_2 \leq X'_2 \beta_2 + p (\phi_{2s} + \phi_{2w}) - \phi_{2w},
$$

2) To have $Y_{2|Q} = 0$ as an equilibrium with pooling $(A_{1s}, A_{1w}) = (Q, Q)$:

We first assume that Player 1 expect that Player 2 will have $\bar{\mu}_2 = \mu_2(t_1 = t_s|B)$ as off-the-equilibrium belief after observing $B$.

P2-21: $t_w$ should have no incentive to deviate. Player 1 of type $t_w$ will have no incentive to deviate from the equilibrium in either of two cases

i. $u_1(t_w; B, NF) \leq u_1(t_w; Q, F)$. This holds when $\varepsilon_1 \geq \mu_w + X'_1 \beta_w - \phi_{1w}$.

ii. $u_1(t_w; Q, NF) \geq u_1(t_w; Q, F)$ but $Y_{2|B} = 0$.

P2-22: $t_s$ should have no incentive to deviate. Since $u_1(t_s, Q, F) < u_1(t_s, B, F)$, Player 1 of type $t_s$ has an incentive to deviate unless $u_1(t_s, B, NF) \leq u_1(t_s, Q, F)$ and $Y_{2|B} = 1$.

Combining P2-21 and P2-22, we conclude that $Y_{2|Q} = 0$ and $Y_{2|B} = 1$ with pooling $(A_{1s}, A_{1w}) = (Q, Q)$ can be supported as an equilibrium under

$$
\varepsilon_1 \geq \max (\mu_w + X'_1 \beta_w - \phi_{1w}, \mu_s + X'_1 \beta_s + \phi_{1s}) \quad \text{and} \quad X'_2 \beta_2 + p (\phi_{2s} + \phi_{2w}) - \phi_{2w} \leq \varepsilon_2 \leq X'_2 \beta_2 + \bar{\mu}_2 (\phi_{2s} + \phi_{2w}) - \phi_{2w}.
$$

Similarly to the previous case, note that for any $\bar{\mu}_2 > p$ satisfying (37), the equilibrium is supported and hence regions of $(\varepsilon_1, \varepsilon_2)$ that support the equilibrium are actually

$$
\varepsilon_1 \geq \max (\mu_w + X'_1 \beta_w - \phi_{1w}, \mu_s + X'_1 \beta_s + \phi_{1s}) \quad \text{and} \quad X'_2 \beta_2 + p (\phi_{2s} + \phi_{2w}) - \phi_{2w} \leq \varepsilon_2 \leq X'_2 \beta_2 + \phi_{2s}.
$$

D.2.2 Separating Equilibria

S-1) Separating Equilibrium ($t_s$ chooses $B$ and $t_w$ choose $Q$):

In the separating equilibrium, Player 2 has complete information once it observes the signal. Thus, $\mu_2(t_1 = t_s|B) = 1$ and $\mu_2(t_1 = t_s|Q) = 0$. Upon seeing $B$ Player 2 will choose $NF$ if $u_2(t_s; B, NF) \geq u_2(t_s; B, F)$ and choose $F$ otherwise. This implies $Y_{2|B} \equiv 1 \{X'_2 \beta_2 + \phi_{2s} - \varepsilon_2 \geq 0\}$. Similarly upon seeing $Q$ Player 2 will have $Y_{2|Q} = 1 \{X'_2 \beta_2 - \varepsilon_2 \geq \phi_{2w}\}$.

1) To have $(Y_{2|B} = 1, Y_{2|Q} = 1)$ as an equilibrium with separating $(A_{1s}, A_{1w}) = (B, Q)$: (When $\varepsilon_2 \leq X'_2 \beta_2 - \phi_{2w}$)

S1-11: $t_s$ should have no incentive to deviate. It requires $u_1(t_s; B, NF) = \mu_s + X'_1 \beta_s - \varepsilon_1 \geq u_1(t_s; Q, NF) = \mu_s + X'_1 \beta_s - \varepsilon_1 - \phi_{1s}$, which holds for all $\varepsilon_1$. 37
S1-12: \(t_w\) should have no incentive to deviate. It requires \(u_1(t_w; Q, NF) = \mu_w + X'_1\beta_w - \varepsilon_1 \geq u_1(t_w; B, NF) = \mu_w + X'_1\beta_w - \varepsilon_1 - \phi_{1w}\), which also holds for all \(\varepsilon_1\). 

Combining S1-11 and S1-12, we conclude that \((Y_{2|B} = 1, Y_{2|Q} = 1)\) with separating \((A_{1t_s}, A_{1t_w}) = (B, Q)\) can be supported as an equilibrium as long as \(\varepsilon_2 \leq X'_2\beta_2 - \phi_{2w}\).

2) To have \((Y_{2|B} = 1, Y_{2|Q} = 0)\) as an equilibrium with separating \((A_{1t_s}, A_{1t_w}) = (B, Q)\): (When \(X'_2\beta_2 - \phi_{2w} \leq \varepsilon_2 \leq X'_2\beta_2 + \phi_{2s}\))

S1-21: \(t_s\) should have no incentive to deviate. It requires \(u_1(t_s; B, NF) = \mu_s + X'_1\beta_s - \varepsilon_1 \geq u_1(t_s; Q, F) = -\phi_{1s}\), which holds for \(\varepsilon_1 \leq \mu_s + X'_1\beta_s + \phi_{1s}\).

S1-22: \(t_w\) should have no incentive to deviate. It requires \(u_1(t_w; Q, F) = 0 \geq u_1(t_w; B, NF) = \mu_w + X'_1\beta_w - \varepsilon_1 - \phi_{1w}\), which holds for \(\varepsilon_1 \geq \mu_w + X'_1\beta_w - \phi_{1w}\).

Combining S1-21 and S1-22, we conclude that \((Y_{2|B} = 1, Y_{2|Q} = 0)\) with separating \((A_{1t_s}, A_{1t_w}) = (B, Q)\) can be supported as an equilibrium under\(^{36}\)

\[
X'_2\beta_2 - \phi_{2w} \leq \varepsilon_2 \leq X'_2\beta_2 + \phi_{2s} \quad \text{and} \quad \mu_w + X'_1\beta_w - \phi_{1w} \leq \varepsilon_1 \leq \mu_s + X'_1\beta_s + \phi_{1s}.
\]

3) To have \((Y_{2|B} = 0, Y_{2|Q} = 1)\) as an equilibrium with separating \((A_{1t_s}, A_{1t_w}) = (B, Q)\): There is no such \(\varepsilon_2\) that supports this equilibrium since \(\phi_{2s} \geq 0, \phi_{2w} \geq 0, \text{ and } \phi_{2s} \cdot \phi_{2w} \neq 0\).

4) To have \((Y_{2|B} = 0, Y_{2|Q} = 0)\) as an equilibrium with separating \((A_{1t_s}, A_{1t_w}) = (B, Q)\): (When \(\varepsilon_2 \geq X'_2\beta_2 + \phi_{2s}\))

S1-41: \(t_s\) should have no incentive to deviate. It requires \(u_1(t_s; B, F) = 0 \geq u_1(t_s; Q, F) = -\phi_{1s}\), which holds for all \(\varepsilon_1\).

S1-42: \(t_w\) should have no incentive to deviate. It requires \(u_1(t_w; Q, F) = 0 \geq u_1(t_w; B, F) = -\phi_{1w}\), which also holds for all \(\varepsilon_1\).

Combining S1-41 and S1-42, we conclude that \((Y_{2|B} = 0, Y_{2|Q} = 0)\) with separating \((A_{1t_s}, A_{1t_w}) = (B, Q)\) can be supported as an equilibrium as long as \(\varepsilon_2 \geq X'_2\beta_2 + \phi_{2s}\).

**S-2) Separating Equilibrium (\(t_s\) chooses \(Q\) and \(t_w\) choose \(B\)):**

In the separating equilibrium, Player 2 has complete information once it observes the signal. Thus, \(\mu_2(t_1 = t_s|B) = 0\) and \(\mu_2(t_1 = t_s|Q) = 1\). Upon seeing \(B\) Player 2 will choose \(NF\) if \(u_2(t_w, B, NF) \geq u_2(t_w, B, F)\) and choose \(F\) otherwise. This implies \(Y_{2|B} \equiv \mathbb{1}\{X'_2\beta_2 - \varepsilon_2 \geq \phi_{2w}|B\}\). Similarly upon seeing \(Q\) Player 2 will have \(Y_{2|Q} = \mathbb{1}\{X'_2\beta_2 - \varepsilon_2 + \phi_{2w} \geq 0\}\).

1) To have \((Y_{2|B} = 1, Y_{2|Q} = 1)\) as an equilibrium with separating \((A_{1t_s}, A_{1t_w}) = (Q, B)\): (When \(\varepsilon_2 \leq X'_2\beta_2 - \phi_{2w}\))

\(t_s\) should have no incentive to deviate. It requires \(u_1(t_s; Q, NF) = \mu_s + X'_1\beta_s - \varepsilon_1 - \phi_{1s} \geq u_1(t_s; B, NF) = \mu_s + X'_1\beta_s - \varepsilon_1\). Hence there is no such \(\varepsilon_1\) that satisfies this condition since \(\phi_{1s} > 0\). Thus, \(Y_{2|B} = 1, Y_{2|Q} = 1\) with separating \((A_{1t_s}, A_{1t_w}) = (Q, B)\) cannot be an equilibrium.

2) To have \((Y_{2|B} = 1, Y_{2|Q} = 0)\) as an equilibrium with separating \((A_{1t_s}, A_{1t_w}) = (Q, B)\): There is no such \(\varepsilon_2\) that supports this equilibrium since \(\phi_{2s} \geq 0, \phi_{2w} \geq 0, \text{ and } \phi_{2s} \cdot \phi_{2w} \neq 0\).

3) To have \((Y_{2|B} = 0, Y_{2|Q} = 1)\) as an equilibrium with separating \((A_{1t_s}, A_{1t_w}) = (Q, B)\): (When \(X'_2\beta_2 - \phi_{2w} \leq \varepsilon_2 \leq X'_2\beta_2 + \phi_{2s}\))

\(^{36}\)If \(\mu_w + X'_1\beta_w - \phi_{aw} > \mu_s + X'_1\beta_s + \phi_{as}\), this equilibrium does not exist.
S2-21: $t_s$ should have no incentive to deviate. It requires $u_1(t_s; Q, NF) = \mu_s + X'_1\beta_s - \varepsilon_1 - \phi_{1s} \geq u_1(t_s; B, F) = 0$, which holds for $\varepsilon_1 \leq \mu_s + X'_1\beta_s - \phi_{1s}$.

S2-22: $t_w$ should have no incentive to deviate. It requires $u_1(t_w; B, F) = -\phi_{1w} \geq u_1(t_w; Q, NF) = \mu_w + X'_1\beta_w - \varepsilon_1$, which holds for $\varepsilon_1 \geq \mu_w + X'_1\beta_w + \phi_{1w}$.

Combining S1-21 and S1-22, we conclude that there is no such $\varepsilon_1$ satisfying both S1-21 and S2-22 if $\mu_s + X'_1\beta_s - \phi_{1s} < \mu_w + X'_1\beta_w + \phi_{1w}$.

4) To have $(Y_{2|B} = 0, Y_{2|Q} = 0)$ as an equilibrium with separating $(A_{1t_s}, A_{1t_w}) = (Q, B)$, (When $\varepsilon_2 \geq X'_2\beta_2 + \phi_{2s}$) $t_s$ should have no incentive to deviate. It requires $u_1(t_s; Q, F) = -\phi_{1s} \geq u_1(t_s; B, F) = 0$. Hence there is no such $\varepsilon_1$ that satisfies this condition since $\phi_{1s} > 0$. Thus, $Y_{2|B} = 0, Y_{2|Q} = 0$ with separating $(A_{1t_s}, A_{1t_w}) = (Q, B)$ cannot be an equilibrium.

**D.2.3 Semi-Separating Equilibria**

We characterize the semi-separating equilibria in this section.

SS-1: **The weak type of Player 1 plays the separating equilibrium with $Q$ and the strong type randomizes with $\sigma_s(B) = \mu_s^*$.**

Then,

$$
\mu_2(t_1 = t_s|B) = \frac{\mu^*_2 \cdot p}{\mu^*_s \cdot p} = 1 \text{ and } \mu_2^{ss} \equiv \mu_2(t_1 = t_s|Q) = \frac{(1 - \mu^*_2) p}{(1 - \mu^*_s) p + (1 - p)} \in (0, p)
$$

which implies that the expected payoff of Player 2 for each action will be

$$
E_2[u_2(t_1; Q, NF)] = X'_2\beta_2 + \mu_2^{ss} \phi_{2s} - \varepsilon_2, \quad E_2[u_2(t_1; Q, F)] = (1 - \mu_2^{ss}) \phi_{2w},
$$

$$
E_2[u_2(t_1; B, NF)] = X'_2\beta_2 + \phi_{2s} - \varepsilon_2, \quad \text{and } E_2[u_2(t_1; B, F)] = 0.
$$

Player 2 will be indifferent between $NF$ and $F$ after observing $Q$ if $\varepsilon_2 = X'_2\beta_2 + \mu_2^{ss} \phi_{2s} + \phi_{2w} - \phi_{2w}$ and under this, Player 2 will choose $NF$ after seeing $B$ since $\varepsilon_2 = X'_2\beta_2 + \mu_2^{ss} \phi_{2s} + \phi_{2w} - \phi_{2w}$ and $\mu_2^{ss} \in (0, p)$ implies that $E_2[u_2(t_1; B, NF)] > E_2[u_2(t_1; B, F)]$. Therefore, for the strong type of Player 1 to be indifferent between choosing $B$ and $Q$, we require that

$$
\mu_s + X'_1\beta_s - \varepsilon_1 = u_1(t_s; B, NF) = E_1[u_1(t_s; Q, \cdot)] = \sigma_2(NF)(\mu_s + X'_1\beta_s - \varepsilon_1 - \phi_{1s}) + (1 - \sigma_2(NF))(\phi_{1s})
$$

where $\sigma_2(NF)$ is the probability that Player 2 chooses $NF$. It follows that we require $\varepsilon_1 = \mu_s + X'_1\beta_s + \frac{\phi_{1s}}{1 - \sigma_2(NF)}$ where $0 \leq \sigma_2(NF) \leq 1$. Noting $\varepsilon_1$ varies from $\mu_s + X'_1\beta_s + \phi_{1s}$ to infinity and $\varepsilon_2$ varies from $X'_2\beta_2 - \phi_{2w}$ to $X'_2\beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}$ since $\mu_2^{ss} \in (0, p)$, this implies that under region $S_6$, there exist the semi-separating of SS-1. Finally note that we have

$$
\varepsilon_1 = \mu_s + X'_1\beta_s + \frac{\phi_{1s}}{1 - \sigma_2(NF)} \quad \text{and } \varepsilon_2 = X'_2\beta_2 + \left(\frac{(1 - \mu_2^*) p}{(1 - \mu_2^*) p + (1 - p)}\right) (\phi_{2s} + \phi_{2w}) - \phi_{2w}. \quad (38)
$$
SS-2: The strong type of Player 1 plays the separating equilibrium with $B$ and the weak type randomizes with $\sigma_{sw}(B) = \mu_B^w$:

Then,

$$\tilde{\mu}_{1s}^w = \mu_2(t_1 = t_s|B) = \frac{p}{p + \mu_B^w \cdot (1-p)} \in (p, 1) \text{ and } \mu_2(t_1 = t_s|B) = \frac{0}{(1 - \mu_B^w \cdot (1-p))} = 0$$

which implies that the expected payoff of Player 2 for each action will be

$$E_2[u_2(t_1; Q, NF)] = X'_2 \beta_2 - \varepsilon_2, \ E_2[u_2(t_1; Q, F)] = \phi_{2w},$$

$$E_2[u_2(t_1; B, NF)] = X'_2 \beta_2 + \tilde{\mu}_{1s}^w \phi_{2s} - \varepsilon_2, \text{ and } E_2[u_2(t_1; B, F)] = (1 - \tilde{\mu}_{1s}^w) \phi_{2w}.$$  

Player 2 will be indifferent between $NF$ and $F$ after observing $B$ if $\varepsilon_2 = X'_2 \beta_2 + \tilde{\mu}_{1s}^w (\phi_{2s} + \phi_{2w}) - \phi_{2w}$ and under this, Player 2 will choose $F$ after seeing $Q$ since $\varepsilon_2 = X'_2 \beta_2 + \tilde{\mu}_{1s}^w (\phi_{2s} + \phi_{2w}) - \phi_{2w}$ and $\tilde{\mu}_{1s}^w \in (p, 1)$ implies that $E_2[u_2(t_1; Q, NF)] < E_2[u_2(t_1; Q, F)]$. Therefore, for the weak type of Player 1 to be indifferent between choosing $B$ and $Q$, we require that

$$0 = u_1(t_w; Q, F) = E_1[u_1(t_w; B, \cdot)] = \bar{\sigma}_2(NF)(\mu_w + X'_1 \beta_w - \varepsilon_1 - \phi_w) + (1 - \bar{\sigma}_2(NF))(-\phi_w)$$

where $\bar{\sigma}_2(NF)$ is the probability that Player 2 chooses $NF$. It follows that we require $\varepsilon_1 = \mu_w + X'_1 \beta_w - \frac{\phi_{1w}}{2(NF)}$ where $0 \leq \bar{\sigma}_2(NF) \leq 1$. This implies that under region $S_2$, there exist the semi-separating of SS-2 noting $\varepsilon_1$ varies from $\mu_w + X'_1 \beta_w - \phi_w$ to negative infinity and $\varepsilon_2$ varies from $X'_2 \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}$ to $X'_2 \beta_2 + \phi_{2s}$ since $\tilde{\mu}_{1s}^w \in (p, 1)$. Finally note

$$\varepsilon_1 = \mu_w + X'_1 \beta_w - \frac{\phi_{1w}}{\bar{\sigma}_2(NF)} \text{ and } \varepsilon_2 = X'_2 \beta_2 + \left(\frac{p}{p + \mu_B^w \cdot (1-p)}\right)(\phi_{2s} + \phi_{2w}) - \phi_{2w}.$$  \hspace{1cm} (39)

D.2.4 Existence of Well-defined Likelihood for Semi-Separating

We let $\sigma_2 \equiv \sigma_2(NF)$ and $\bar{\sigma}_2 \equiv \bar{\sigma}_2(NF)$. In the equilibrium of SS-1 under the region of $S_6$, four observed outcomes can arise with following probabilities, respectively, denoted by

$$\Pr(Y_1 = 1, Y_2 = 1; S_6, \mu_B^w, \sigma_2) = \int_{S_6} p \cdot \mu_B^w(\varepsilon_2) \cdot \sigma_2(\varepsilon_1) g_1(\varepsilon_1) g_2(\varepsilon_2) d\varepsilon_1 d\varepsilon_2$$

$$\Pr(Y_1 = 1, Y_2 = 0; S_6, \mu_B^w, \sigma_2) = \int_{S_6} p \cdot \mu_B^w(\varepsilon_2) \cdot (1 - \sigma_2(\varepsilon_1)) g_1(\varepsilon_1) g_2(\varepsilon_2) d\varepsilon_1 d\varepsilon_2$$

$$\Pr(Y_1 = 0, Y_2 = 1; S_6, \mu_B^w, \sigma_2) = \int_{S_6} (p - \mu_B^w(\varepsilon_2)) \cdot (1 - \sigma_2(\varepsilon_1)) g_1(\varepsilon_1) g_2(\varepsilon_2) d\varepsilon_1 d\varepsilon_2$$

$$\Pr(Y_1 = 0, Y_2 = 0; S_6, \mu_B^w, \sigma_2) = \int_{S_6} (p - \mu_B^w(\varepsilon_2)) \cdot (1 - \sigma_2(\varepsilon_1)) g_1(\varepsilon_1) g_2(\varepsilon_2) d\varepsilon_1 d\varepsilon_2.$$  

For example, $\Pr(Y_1 = 1, Y_2 = 1; S_6, \mu_B^w, \sigma_2)$ is obtained using the following facts. Conditional on $(\varepsilon_1, \varepsilon_2)$, $(B, NF)$ is observed when the nature draw the strong type with the probability $p$, the strong type plays $B$ with the probability $\mu_B^w(\varepsilon_2)$, and Player 2 plays $NF$ with probability $\sigma_2(\varepsilon_1)$ after observing $B$. Thus, the conditional probability of observing $(B, NF)$ conditional on $(\varepsilon_1, \varepsilon_2)$ is given by

$$p \cdot \mu_B^w(\varepsilon_2) \cdot \sigma_2(\varepsilon_1)$$  \hspace{1cm} (40)

and we obtain the unconditional probability by taking expectation of (40) with respect to $(\varepsilon_1, \varepsilon_2) \in S_6$. Similarly, $\Pr(Y_1 = 0, Y_2 = 1; S_6, \mu_B^w, \sigma_2)$ is obtained from the followings. Conditional on $(\varepsilon_1, \varepsilon_2)$, $(Q, NF)$ is observed in two cases: 1) the nature draw the strong type with the probability $p$, the strong type plays
Q with the probability $1 - \mu_B^*(\varepsilon_2)$, and Player 2 plays NF with probability $\sigma_2(\varepsilon_1)$ after observing Q and 2) the nature draw the weak type with the probability $1 - p$, the weak type plays Q with the probability one (separating), and Player 2 plays NF with probability $\sigma_2(\varepsilon_1)$ after observing Q. Thus, the conditional probability of observing $(Q, NF)$ conditional on $(\varepsilon_1, \varepsilon_2)$ is given by

$$(p(1 - \mu_B^*(\varepsilon_2)) + (1 - p)) \cdot \sigma_2(\varepsilon_1)$$

(41)

and we obtain the unconditional probability by taking expectation of (41) with respect to $(\varepsilon_1, \varepsilon_2) \in S_6$. Other probabilities can be interpreted similarly.

Some changes of variables using the relationship of $\mu_B^*$ and $\varepsilon_2$ and that of $\sigma_2$ and $\varepsilon_1$ as shown in (38) give us

$$\mathrm{Pr}(Y_1 = 1, Y_2 = 1; S_6, \mu_B^*, \sigma_2) = \int \mu_B^* \, g_2 \left( X_2^* \beta_2 + \frac{(1 - \mu_B^*)p}{(1 - \mu_B^*)p + (1 - p)} (\phi_2s + \phi_2w) - \phi_2w \right) \frac{p(1 - p)(\phi_2s + \phi_2w)}{(1 - \mu_B^*)^2} \, d\mu_B^*$$

$$\times \int (1 - \sigma_2) g_1 \left( \mu_s + X_1 \beta_s + \frac{\phi_1s}{(1 - \sigma_2)^3} \right) \frac{\phi_1s}{(1 - \sigma_2)^3} \, d\sigma_2$$

$$\mathrm{Pr}(Y_1 = 0, Y_2 = 1; S_6, \mu_B^*, \sigma_2) = \int \mu_B^* \, g_2 \left( X_2^* \beta_2 + \frac{(1 - \mu_B^*)p}{(1 - \mu_B^*)p + (1 - p)} (\phi_2s + \phi_2w) - \phi_2w \right) \frac{p(1 - p)(\phi_2s + \phi_2w)}{(1 - \mu_B^*)^2} \, d\mu_B^*$$

$$\times \int (1 - \sigma_2) g_1 \left( \mu_s + X_1 \beta_s + \frac{\phi_1s}{(1 - \sigma_2)^3} \right) \frac{\phi_1s}{(1 - \sigma_2)^3} \, d\sigma_2$$

$$\mathrm{Pr}(Y_1 = 1, Y_2 = 0; S_6, \mu_B^*, \sigma_2) = \int \mu_B^* \, g_2 \left( X_2^* \beta_2 + \frac{(1 - \mu_B^*)p}{(1 - \mu_B^*)p + (1 - p)} (\phi_2s + \phi_2w) - \phi_2w \right) \frac{p(1 - p)(\phi_2s + \phi_2w)}{(1 - \mu_B^*)^2} \, d\mu_B^*$$

$$\times \int (1 - \sigma_2) g_1 \left( \mu_s + X_1 \beta_s + \frac{\phi_1s}{(1 - \sigma_2)^3} \right) \frac{\phi_1s}{(1 - \sigma_2)^3} \, d\sigma_2$$

where we use the fact that $d\varepsilon_2 = \frac{p(1 - p)}{(1 - \mu_B^*)^2} (\phi_2s + \phi_2w) \, d\mu_B^*$, $d\varepsilon_1 = \frac{\phi_1s}{(1 - \sigma_2)^3} \, d\sigma_2$, and the independence of $\varepsilon_1$ and $\varepsilon_2$. Similarly, in the equilibrium of SS-2 under the region of $S_2$, four observed outcomes can arise with following probabilities, respectively, denoted by

$$\mathrm{Pr}(Y_1 = 1, Y_2 = 1; S_2, \mu_B^*, \sigma_2) = \int S_2 \left( p + (1 - p)\mu_B^*(\varepsilon_2) \right) \cdot \sigma_2(\varepsilon_1) g_1(\varepsilon_1) g_2(\varepsilon_2) \, d\varepsilon_1 \, d\varepsilon_2$$

$$\mathrm{Pr}(Y_1 = 1, Y_2 = 0; S_2, \mu_B^*, \sigma_2) = \int S_2 \left( p + (1 - p)\mu_B^*(\varepsilon_2) \right) \cdot (1 - \sigma_2(\varepsilon_1)) g_1(\varepsilon_1) g_2(\varepsilon_2) \, d\varepsilon_1 \, d\varepsilon_2$$

$$\mathrm{Pr}(Y_1 = 0, Y_2 = 1; S_2, \mu_B^*, \sigma_2) = \int S_2 \left( p \right) \cdot \sigma_2(\varepsilon_1) g_1(\varepsilon_1) g_2(\varepsilon_2) \, d\varepsilon_1 \, d\varepsilon_2$$

$$\mathrm{Pr}(Y_1 = 0, Y_2 = 0; S_2, \mu_B^*, \sigma_2) = \int S_2 \left( 1 - p \right) \cdot (1 - \sigma_2(\varepsilon_1)) g_1(\varepsilon_1) g_2(\varepsilon_2) \, d\varepsilon_1 \, d\varepsilon_2$$

Some changes of variables using the relationship of $\mu_B^*$ and $\varepsilon_2$ and that of $\sigma_2$ and $\varepsilon_1$ as shown in (39) give us

$$\mathrm{Pr}(Y_1 = 1, Y_2 = 1; S_2, \mu_B^*, \sigma_2) = \int \mu_B^* \, g_2 \left( X_2^* \beta_2 + \frac{p}{p + \mu_B^*(1 - p)} (\phi_2s + \phi_2w) - \phi_2w \right) \frac{p(1 - p)(\phi_2s + \phi_2w)}{(p + (1 - p)\mu_B^*)} \, d\mu_B^*$$

$$\times \int (1 - \sigma_2) g_1 \left( \mu_w + X_1 \beta_w - \frac{\phi_1w}{\sigma_2} \right) \frac{\phi_1w}{\sigma_2} \, d\sigma_2$$
\[ \text{Pr}(Y_1 = 1, Y_2 = 0; S_2, \mu_B^w, \sigma_2) = \]
\[ \int_0^1 \left( p + (1 - p)\mu_B^w \right) g_2 \left( X_2' \beta_2 + \left( \frac{p}{p + \mu_B^w (1 - p)} \right) \left( \phi_2 s + \phi_2 w \right) - \phi_2 w \right) \frac{p(1-p)(\phi_2 s + \phi_2 w)}{(p + (1 - p)\mu_B^w)^2} \, d\mu_B^w \]
\[ \times \int_0^1 (1 - \sigma_2) g_1 \left( \mu_w + X_1 \beta_2 - \frac{\phi_1 w}{\sigma_2} \right) \frac{\phi_1 w}{\sigma_2^2} \, d\sigma_2 \]
\[ \text{Pr}(Y_1 = 0, Y_2 = 1; S_2, \mu_B^w, \sigma_2) = \]
\[ \int_0^1 (1 - p) \left( 1 - \mu_B^w \right) g_2 \left( X_2' \beta_2 + \left( \frac{p}{p + \mu_B^w (1 - p)} \right) \left( \phi_2 s + \phi_2 w \right) - \phi_2 w \right) \frac{p(1-p)(\phi_2 s + \phi_2 w)}{(p + (1 - p)\mu_B^w)^2} \, d\mu_B^w \]
\[ \times \int_0^1 \sigma_2 g_1 \left( \mu_w + X_1 \beta_2 - \frac{\phi_1 w}{\sigma_2} \right) \frac{\phi_1 w}{\sigma_2^2} \, d\sigma_2 \]
\[ \text{Pr}(Y_1 = 0, Y_2 = 0; S_2, \mu_B^w, \sigma_2) = \]
\[ \int_0^1 (1 - p) \left( 1 - \mu_B^w \right) g_2 \left( X_2' \beta_2 + \left( \frac{p}{p + \mu_B^w (1 - p)} \right) \left( \phi_2 s + \phi_2 w \right) - \phi_2 w \right) \frac{p(1-p)(\phi_2 s + \phi_2 w)}{(p + (1 - p)\mu_B^w)^2} \, d\mu_B^w \]
\[ \times \int_0^1 (1 - \sigma_2) g_1 \left( \mu_w + X_1 \beta_2 - \frac{\phi_1 w}{\sigma_2} \right) \frac{\phi_1 w}{\sigma_2^2} \, d\sigma_2 \]
where we use the fact that \( d\varepsilon_2 = -\frac{p(1-p)}{(p + \mu_B^w (1 - p))^2} (\phi_2 s + \phi_2 w) d\mu_B^w \), \( d\varepsilon_1 = \frac{\phi_1 w}{\sigma_2^2} d\sigma_2 \), and the independence of \( \varepsilon_1 \) and \( \varepsilon_2 \).

### D.3 Equilibrium Refinement and Uniqueness of Equilibrium with Asymmetric Payoffs

We make the following assumption to obtain uniqueness of equilibrium.

**Condition 3** \( \mu_w + X_1' \beta_w - \phi_1 w < \mu_s + X_1' \beta_s + \phi_1 s \) for all \( X_1 \in S(X_1) \).

**Condition 4** \( \mu_s + X_1' \beta_s - \phi_1 s < \mu_w + X_1' \beta_w + \phi_1 w \) for all \( X_1 \in S(X_1) \).

These conditions are sufficient for uniqueness of equilibria with the refinement of Cho and Kreps (1987). Note that Condition 3 holds immediately when \( \beta_w = \beta_s \) and \( \mu_s = \mu_w \) (symmetric payoffs) since \( \phi_1 s, \phi_1 w > 0 \). Note that Condition 3 ensures the existence of the separating equilibrium \((B, Q)\) with \((Y_{2|B} = NF, Y_{2|Q} = F)\) (see S1-21 & S1-22 in Section D.2.2) and Condition 4 eliminates the separating equilibrium \((Q, B)\) with \((Y_{2|B} = NF, Y_{2|Q} = F)\) (see S2-21 & S2-22 in Section D.2.2).

- Under region \( A \equiv \{(\varepsilon_1, \varepsilon_2) | \mu_s + X_1' \beta_s - \phi_1 s < \varepsilon_1 \text{ and } X_2' \beta_2 + p(\phi_2 s + \phi_2 w) - \phi_2 w < \varepsilon_2 < X_2' \beta_2 + \phi_2 s \} \), we show that the pooling \((B, B)\) with \( A_{2|B} = F \) cannot survive the Intuitive Criterion of Cho and Kreps (1987).

**Pooling with \((B, B)\) with \( A_{2|B} = F \)**

Note \( u_1(t_s; B, F) > u_1(t_s; Q, NF) \) since \( \mu_s + X_1' \beta_1 - \varepsilon_1 - \phi_1 s < 0 \) under \( A \) and note \( u_1(t_s; B, F) > u_1(t_s; Q, F) \) since \( \phi_1 s > 0 \). This means that the strong type has no incentive to deviate in any case. Thus, Player 2 will assign \( \mu_2(t_1 = t_s|Q) = 0 \) and hence Player 2 will choose \( F \) after observing the deviation play \( Q \) under \( A \). Now we need to check whether the weak type is better off by deviation under this situation. Note

\[ u_1(t_w; B, F) = -\phi_1 w < u_1(t_w; Q, F) = 0. \]

Thus, the weak type will deviate for sure. Therefore, the equilibrium outcome fails the Intuitive Criterion.
Under region $\bar{A} \equiv \{(\varepsilon_1, \varepsilon_2) | \mu_s + X_1^t \beta_s + \phi_s < \varepsilon_1 \text{ and } X_2^t \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} < \varepsilon_2 < X_2^t \beta_2 + \phi_{2s}\}$, we show that the pooling $(Q, Q)$ with $A_{2|Q} = F$ survives the Intuitive Criterion of Cho and Kreps (1987).

**Pooling with $(Q, Q)$ with $A_{2|Q} = F$**

Note $u_1(t_w; Q, F) > u_1(t_w; B, NF)$ since $\mu_w + X_1^t \beta_w - \varepsilon_1 - \phi_{1w} < 0$ under $\bar{A}$ by Condition 3 and note $u_1(t_w; Q, F) > u_1(t_w; B, F)$ since $\phi_{1w} > 0$. This means that the weak type has no incentive to deviate in any case. Thus, Player 2 will assign $\mu_2(t_1 = t_s|B) = 1$ and hence Player 2 will choose NF after observing the deviation play $B$ under region $\bar{A}$. Now we need to check whether the strong type is better off by deviation under this situation. Note

$$u_1(t_s; Q, F) = -\phi_{1s} > u_1(t_s; B, NF) = \mu_s + X_1^t \beta_s - \varepsilon_1$$

under $\bar{A}$. Thus, the strong type will not deviate. Therefore, the equilibrium outcome survives the Intuitive Criterion.

Under region $\bar{B} \equiv \{(\varepsilon_1, \varepsilon_2) | \mu_w + X_1^t \beta_w + \phi_w > \varepsilon_1 \text{ and } X_2^t \beta_2 - \phi_{2w} < \varepsilon_2 < X_2^t \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}\}$, the pooling $(Q, Q)$ with $A_{2|Q} = NF$ cannot survive the Intuitive Criterion of Cho and Kreps (1987).

**Pooling with $(Q, Q)$ with $A_{2|Q} = NF$**

Note $u_1(t_w; Q, NF) > u_1(t_w; B, F)$ since $\mu_w + X_1^t \beta_w - \varepsilon_1 + \phi_{1w} > 0$ under $\bar{B}$ and note $u_1(t_w; Q, NF) > u_1(t_w; B, NF)$ since $\phi_{1w} > 0$. This means that the weak type has no incentive to deviate in any case. Thus, Player 2 will assign $\mu_2(t_1 = t_s|B) = 1$ and hence Player 2 will choose NF after observing the deviation play $B$ under region $\bar{B}$. Now we need to check whether the strong type is better off by deviation under this situation. Note

$$u_1(t_s; Q, NF) = \mu_s + X_1^t \beta_s - \varepsilon_1 - \phi_{1s} < u_1(t_s; B, NF) = \mu_s + X_1^t \beta_s - \varepsilon_1$$

since $\phi_{1s} > 0$. Thus, the strong type will deviate for sure. Therefore, the equilibrium outcome fails the Intuitive Criterion.

Under region $\tilde{B} \equiv \{(\varepsilon_1, \varepsilon_2) | \mu_w + X_1^t \beta_w + \phi_w > \varepsilon_1 \text{ and } X_2^t \beta_2 - \phi_{2w} < \varepsilon_2 < X_2^t \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w}\}$, the pooling $(B, B)$ with $A_{2|B} = NF$ survives the Intuitive Criterion of Cho and Kreps (1987).

**Pooling $(B, B)$ with $A_{2|B} = NF$**

Note $u_1(t_s; B, NF) > u_1(t_s; Q, NF)$ since $\phi_{1s} > 0$ and note $u_1(t_s; B, NF) > u_1(t_s; Q, F)$ under $\tilde{B}$ by Condition 3. This means that the strong type has no incentive to deviate in any case. Thus, Player 2 will assign $\mu_2(t_1 = t_s|Q) = 0$ and hence Player 2 will choose F after observing the deviation play $Q$ under region $\tilde{B}$. Now we need to check whether the weak type is better off by deviation under this situation. Note

$$u_1(t_w; B, NF) = \mu_w + X_1^t \beta_w - \varepsilon_1 - \phi_w > u_1(t_w; Q, F) = 0$$

under region $\tilde{B}$. Thus, the weak type will not deviate. Therefore, the equilibrium outcome survives the Intuitive Criterion.
E Alternative Information Structure (IS-2)

Here we introduce an additional incomplete information to the game. Namely, Player 2 has some private information on her payoff, not observed by Player 1.

Assumption E.1 (IS-2)

1. Player 1 knows its true type but Player 2 knows only the distribution of Player 1’s types (i.e., \( p \) is known to Player 2).
2. The realizations of \((X_1, \varepsilon_1)\) and \((X_2, \varepsilon_2)\) are perfectly observed by Player 2 but Player 1 only observes \((X_1, \varepsilon_1)\) and \(X_2\). \((X_1, X_2)\) are public information.
3. \( \varepsilon_1 \) and \( \varepsilon_2 \) are pure shocks observed by Player 2, but Player 1 only observe her own \( \varepsilon_1 \). They are independent of each other and of any other variables in the game. \( \varepsilon_1 \) is also independent of the type of Player 1.
4. Players’ actions and beliefs constitute a Perfect Bayesian Equilibrium (Sequential Equilibrium). Whenever there exist multiple equilibria, only one equilibrium is chosen out of these according to some equilibrium refinements. Players are assumed to play actions and hold beliefs of this unique equilibrium.

Under the game with IS, both players can perfectly observe \((X_1, \varepsilon_1)\) and \((X_2, \varepsilon_2)\), respectively but under this game with IS-2, Player 1 cannot observe \( \varepsilon_2 \) while Player 2 observes \((X_1, \varepsilon_1)\) and \((X_2, \varepsilon_2)\) perfectly. It makes Player 1 take expectation over \( \varepsilon_2 \) when she derive her expected payoffs depending on her choice of signals under the game with IS-2.

E.1 Equilibrium of the Game

In Section E.2, we characterize all possible PBE and obtain regions of \((\varepsilon_1, \varepsilon_2)\) given \(X\) where a particular PBE exist. We use a simplifying notation \( G_2^t(a) = G_2(X_2^t, \beta_2 - a(\phi_{2s} + \phi_{2w}) - \phi_{2w}) \). Figure A3 summarizes the result. Figure A3 depicts the case that \( \phi_{1w} > \phi_{1s} \) and \( \frac{\phi_{1s}}{\phi_{1w}} > \frac{G_2^t(p) - G_2^t(0)}{G_2^t(1) - G_2^t(0)} \). However, none of these conditions are necessary to obtain a PBE. We, again, impose a structure \( \phi_{1s} > 0, \phi_{1w} > 0, \phi_{2s} \geq 0, \) and \( \phi_{2w} \geq 0 \) (but \( \phi_{2s} \cdot \phi_{2w} \neq 0 \)), which are innocuous since a meaningful signaling game requires that (i) each signal corresponds to a particular type of Player 1 and (ii) Player 2 has an incentive to single out a particular type.

We conclude that

Theorem E.1 (Existence of Equilibrium under Additional Incomplete Information)

Suppose Assumptions IS-2 and SA-1 hold. Suppose also that \( \phi_{1s} > 0, \phi_{1w} > 0, \phi_{2s} \geq 0, \) and \( \phi_{2w} \geq 0 \) (but \( \phi_{2s} \cdot \phi_{2w} \neq 0 \)). Then, there exist PBE for all regions of \((\varepsilon_1, \varepsilon_2)\) given \(X\).

See the Appendix E.2 for the proof. From Figure A3, we see that there exist multiple equilibria in several regions. Similarly to the case of IS, we determine which equilibrium survives the refinement but here we use Banks and Sobel (1987)’s divinity concept instead, which is stronger than Cho and Kreps (1987). The results are summarized in the following theorem (see the Appendix E.3.1).

\( \text{37} \) The idea of “divinity” developed by Banks and Sobel (1987) is that if Player 2 observes a deviation play and if the set of Player 2’s responses that makes type \( t_1 \) willing to deviate from an equilibrium is strictly smaller than such a set for type \( t_1' \), it is reasonable to expect that Player 2 will think such deviation is more likely to be from type \( t_1' \).

\( \text{38} \) We use Banks and Sobel (1987)’s refinement, instead of Cho and Kreps (1987), because it is not feasible to apply Cho and Kreps (1987)’s intuitive criterion to our game with additional incomplete information.
Theorem E.2 Suppose Assumptions IS-2 and SA-1 hold and suppose that $\phi_{1s} > 0$, $\phi_{1w} > 0$, $\phi_{2s} \geq 0$, and $\phi_{2w} \geq 0$ (but $\phi_{2s} \cdot \phi_{2w} \neq 0$). Then, (i) The pooling equilibrium with $(A_{1s}, A_{1w}) = (B, B)$ survives the refinement of Banks and Sobel (1987) under $\varepsilon_1 \leq \mu + X_1' \beta_1 - \frac{\phi_{1s}}{G_1(p)} - \frac{\phi_{1w}}{G_2(p)}$; (ii) The pooling equilibrium $(A_{1s}, A_{1w}) = (Q, Q)$ survives the refinement of Banks and Sobel (1987) under $\varepsilon_1 \geq \mu + X_1' \beta_1 + \frac{\phi_{1s}}{G_1(1) - G_2(p)}$.

Figure A4 depicts the results and we conclude

Theorem E.3 (Uniqueness of Equilibrium)
Suppose Assumptions IS-2 and SA-1 hold. Suppose $\phi_{1s} > 0$, $\phi_{1w} > 0$, $\phi_{2s} \geq 0$, and $\phi_{2w} \geq 0$ (but $\phi_{2s} \cdot \phi_{2w} \neq 0$). Further suppose each Player plays only one equilibrium that survives the refinement of Banks and Sobel (1987), when there exists multiple equilibria. Then, the game $G$ has the unique equilibrium for each region of $(\varepsilon_1, \varepsilon_2)$ given $X$.

Now similarly with the case of IS, from the result of Theorem E.3, we can obtain a well-defined likelihood function under IS-2 and SA-1. Using the conditional probabilities of four possible observed outcomes presented in the Appendix E.3.2, we can estimate the parameters of interest by the conditional ML as in (1). We note that the additional incomplete information considered in this section may help the identification of $\mu$ and other parameters since there are more critical values (four vertical lines in Figure A4 compared to two vertical lines in Figure 4) of $\varepsilon_1$ that switches the kinds of equilibrium and these critical lines vary according to different values of $X_2$. This is an interesting result.

E.2 PBE of the Game G with IS-2

Again we derive equilibria of the game with asymmetric payoffs as in Figure A1. Equilibria of the game with symmetric payoffs ($\mu_s = \mu_w$, $\beta_s = \beta_w$) are easily obtained from the results in this section.
E.2.1 Pooling Equilibria

P-1) Pooling Equilibrium (both \(t_s\) and \(t_w\) choose \(B\)):

In this case, Player 2 does not update her belief and thus the posterior equals to the prior belief: \(\mu_2(t_1 = t_s | B) = p \) and \(\mu_2(t_1 = t_w | B) = 1 - p\). Thus, \(Y_{2|B} = 1 \{X'_2\beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} - \varepsilon_2 \geq 0 | B\}\) on the equilibrium path. Now we define \(G_2^*(a) \equiv G_2(X'_2\beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w})\). The resulting payoffs of each type of Player 1 on-the-equilibrium path will be

| \(Y_{2|B} = 1\) | \(Y_{2|B} = 0\) | Player 1’s expected payoffs |
|-----------------|-----------------|-----------------------------|
| \(u_1(t_s; B, \cdot)\) | \(\mu_s + X'_s\beta_s - \varepsilon_1\) | \(G_2^*(p) (\mu_s + X'_s\beta_s - \varepsilon_1)\) |
| \(u_1(t_w; B, \cdot)\) | \(\mu_w + X'_w\beta_w - \varepsilon_1 - \phi_{1w}\) | \(G_2^*(p) (\mu_w + X'_w\beta_w - \varepsilon_1 - \phi_{1w})\) |

To have the pooling \((A_{1t_s}, A_{1t_w}) = (B, B)\) as an equilibrium:

P1-1 \(t_s\) should have no incentive to deviate. Player 1 of \(t_s\) will note that Player 2 with \(\bar{\pi}_2 \equiv \mu_2( t_1 = t_s | Q)\) as off-the-equilibrium belief will choose \(Y_{2|Q} = 1 \{X'_2\beta_2 + \bar{\pi}_2 (\phi_{2s} + \phi_{2w}) - \phi_{2w} - \varepsilon_2 \geq 0 | Q\}\) which implies \(\text{Pr}(Y_{2|Q} = 1) = G_2(X'_2\beta_2 + \bar{\pi}_2 (\phi_{2s} + \phi_{2w}) - \phi_{2w}) = G_2^*(\bar{\pi}_2)\). Thus, Player 1 of type \(t_s\) will have no incentive to deviate from the equilibrium if

\[
E_1[ u_1(t_s; Q, A_2) ] \leq E_1[ u_1(t_s; B, A_2) ]
\]

which implies

\[
G_2^*(\bar{\pi}_2) \cdot (\mu_s + X'_s\beta_s - \varepsilon_1) - \phi_{1s} \leq G_2^*(p) (\mu_s + X'_s\beta_s - \varepsilon_1). \tag{42}
\]

P1-2 \(t_w\) should have no incentive to deviate. \(t_w\) has an incentive to deviate unless \(E_1[ u_1(t_w; Q, A_2) ] \leq E_1[ u_1(t_w; B, A_2) ]\). This requires

\[
G_2^*(\bar{\pi}_2) \cdot (\mu_w + X'_w\beta_w - \varepsilon_1) \leq G_2^*(p) (\mu_w + X'_w\beta_w - \varepsilon_1) - \phi_{1w}. \tag{43}
\]

Combining P1-1 and P1-2, we conclude that the pooling \((A_{1t_s}, A_{1t_w}) = (B, B)\) with \(E[Y_{2|B}] = G_2^*(p)\) and \(E[Y_{2|Q}] = G_2^*(\bar{\pi}_2)\) can be supported as an equilibrium under the following two cases.

- If \(\bar{\pi}_2 > p\) (noting any \(\bar{\pi}_2 > p\) satisfying (42) and (43) supports the equilibrium\(^{39}\)), we need

\[
\max \left\{ \mu_s + X'_s\beta_s - \frac{\phi_{1s}}{G_2^*(1) - G_2^*(p)}, \mu_w + X'_w\beta_w - \frac{\phi_{1w}}{G_2^*(1) - G_2^*(p)} \right\} \leq \varepsilon_1. \tag{44}
\]

- If \(\bar{\pi}_2 < p\) (noting any \(\bar{\pi}_2 < p\) satisfying (42) and (43) supports the equilibrium similarly as before), we need

\[
\min \left\{ \mu_s + X'_s\beta_s - \frac{\phi_{1s}}{G_2^*(p) - G_2^*(0)}, \mu_w + X'_w\beta_w - \frac{\phi_{1w}}{G_2^*(p) - G_2^*(0)} \right\} \geq \varepsilon_1. \tag{45}
\]

\(^{39}\)This means that there exist values of \(\bar{\pi}_2 = \bar{\pi}_2(p, \mu_s + X'_s\beta_s, \mu_w + X'_w\beta_w, \phi_{1s}, \phi_{1w}, \varepsilon_1), p < \bar{\pi}_2 \leq 1\) such that (42) and (43) hold.
P-2) Pooling Equilibrium (both \(t_s\) and \(t_w\) choose \(Q\)):

In this case, Player 2 does not update its belief and thus the posterior equals to the prior belief: \(\mu_2(t_1 = t_s | Q) = p\) and \(\mu_2(t_1 = t_w | Q) = 1 - p\). Thus, \(Y_{2|Q} = \{X_2^b \beta_2 + p(\phi_{2s} + \phi_{2w}) - \phi_{2w} - \varepsilon_2 \geq 0\}\) on the equilibrium path. The resulting payoffs of each type of Player 1 on-the-equilibrium path will be

| \(Y_{2|Q} = 1\) | \(Y_{2|Q} = 0\) | Player 1’s expected payoffs |
|----------------|----------------|--------------------------|
| \(u_1(t_s; Q, \cdot)\) | \(\mu_s + X_1^b \beta_s - \varepsilon_1 - \phi_{1s}\) | \(-\phi_{1s}\) |
| \(u_1(t_w; Q, \cdot)\) | \(\mu_s + X_1^b \beta_s - \varepsilon_1\) | \(G_2^s(p)(\mu_s + X_1^b \beta_s - \varepsilon_1)\) |

To have the pooling \((A_{1t_s}, A_{1t_w}) = (Q, Q)\) as an equilibrium:

P2-1  \(t_w\) should have no incentive to deviate. Player 1 of \(t_w\) will note that Player 2 with \(\tilde{\mu}_2 = \mu_2(t_1 = t_s | B)\) as off-the-equilibrium belief will choose \(Y_{2|B} = \{X_2^b \beta_2 + \tilde{\mu}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w} - \varepsilon_2 \geq 0 | Q\}\) which means \(\Pr(Y_{2|B} = 1) = G_2(X_2^b \beta_2 + \tilde{\mu}_2(\phi_{2s} + \phi_{2w}) - \phi_{2w}) = G_2^s(\tilde{\mu}_2)\). Thus, Player 1 of type \(t_w\) will have no incentive to deviate from the equilibrium if

\[E_1[u_1(t_w; B, A_2)] \leq E_1[u_1(t_w; Q, A_2)] \]

which implies

\[G_2^s(\tilde{\mu}_2) \cdot (\mu_w + X_1^b \beta_w - \varepsilon_1) - \phi_{1w} \leq G_2^s(p)(\mu_w + X_1^b \beta_w - \varepsilon_1). \tag{46}\]

P2-2  \(t_s\) should have no incentive to deviate. \(t_s\) has an incentive to deviate unless \(E_1[u_1(t_s, B, A_2)] \leq E_1[u_1(t_s, Q, A_2)]\). This requires

\[G_2^s(\tilde{\mu}_2) \cdot (\mu_s + X_1^b \beta_s - \varepsilon_1) \leq G_2^s(p)(\mu_s + X_1^b \beta_s - \varepsilon_1) - \phi_{1s}. \tag{47}\]

Combining P1-1 and P1-2, we conclude that the pooling \((A_{1t_s}, A_{1t_w}) = (Q, Q)\) with \(E[Y_{2|Q}] = G_2^s(p)\) and \(E[Y_{2|B}] = G_2^s(\tilde{\mu}_2)\) can be supported as an equilibrium under the following two cases:

- If \(\tilde{\mu}_2 > p\) (noting any \(\tilde{\mu}_2 > p\) satisfying (46) and (47) supports the equilibrium), we need

\[
\max \left\{ \frac{\phi_{1s}}{G_2^s(1) - G_2^s(p)}, \frac{\phi_{1w}}{G_2^s(1) - G_2^s(p)} \right\} \leq \varepsilon_1. \tag{48}\]

- If \(\tilde{\mu}_2 < p\) (noting any \(\tilde{\mu}_2 < p\) satisfying (46) and (47) supports the equilibrium), we need

\[
\min \left\{ \frac{\phi_{1s}}{G_2^s(p) - G_2^s(0)}, \frac{\phi_{1w}}{G_2^s(p) - G_2^s(0)} \right\} \geq \varepsilon_1. \tag{49}\]

E.2.2  Separating Equilibria

S-1) Separating Equilibrium (\(t_s\) choose \(B\) and \(t_w\) choose \(Q\)):

In the separating equilibrium, Player 2 has complete information once it observes the signal. Thus, \(\mu_2(t_1 = t_s | B) = 1\) and \(\mu_2(t_1 = t_s | Q) = 0\). This implies \(Y_{2|B} = \{X_2^b \beta_2 + \phi_{2s} - \varepsilon_2 \geq 0\}\). Similarly upon seeing \(Q\) Player 2 will have \(Y_{2|Q} = \{X_2^b \beta_2 - \varepsilon_2 \geq \phi_{2w}\}\). The resulting payoffs of each type of Player 1 on-the-equilibrium path will be

\[
\begin{align*}
E_1[u_1(t_s; B, \cdot)] & : G_2(X_2^b \beta_2 + \phi_{2s})(\mu_s + X_1^b \beta_s - \varepsilon_1) \\
E_1[u_1(t_w; Q, \cdot)] & : G_2(X_2^b \beta_2 - \phi_{2w})(\mu_w + X_1^b \beta_w - \varepsilon_1)
\end{align*}
\]
S1-1 $t_s$ should have no incentive to deviate. It requires

\[ E_1[u_1(t_s;B,\cdot)] = G_2(X'_s\beta_2 + \phi_{2w})(\mu_s + X'_s\beta_s - \varepsilon_1) \]
\[ \geq E_1[u_1(t_s;Q,\cdot)] = G_2(X'_s\beta_2 - \phi_{2w})(\mu_s + X'_s\beta_s - \varepsilon_1) - \phi_{1s}. \]

S1-2 $t_w$ should have no incentive to deviate. It requires

\[ E_1[u_1(t_w;Q,\cdot)] = G_2(X'_s\beta_2 + \phi_{2w})(\mu_w + X'_s\beta_s - \varepsilon_1) \]
\[ \geq E_1[u_1(t_w;B,\cdot)] = G_2(X'_s\beta_2 - \phi_{2w})(\mu_w + X'_s\beta_s - \varepsilon_1) - \phi_{1w}. \]

Combining S1-1 and S1-2, we conclude a separating equilibrium $(A_{1t_s}, A_{1w}) = (B, Q)$ can be supported as an equilibrium under\(^\text{40}\)

\[ \mu_w + X'_s\beta_w - \frac{\phi_{1w}}{G_2'(1) - G_2'(0)} \leq \varepsilon_1 \leq \mu_s + X'_s\beta_1 + \frac{\phi_{1s}}{G_2'(1) - G_2'(0)}. \]

S-2) Separating Equilibrium ($t_s$ choose $Q$ and $t_w$ choose $B$):

In the separating equilibrium, Player 2 has complete information once it observes the signal. Thus, \(\mu_2(t_1 = t_s|B) = 0\) and \(\mu_2(t_1 = t_w|B) = 1\). This implies \(Y_{2|B} = \{X'_s\beta_s - \varepsilon_2 \geq \phi_{2w}\}\). Similarly upon seeing $Q$ Player 2 will have \(Y_{2|Q} = \{X'_s\beta_s + \phi_{2s} \geq 0\}\). The resulting payoffs of each type of Player 1 on-the-equilibrium path will be

\[ \begin{align*}
E_1[u_1(t_s;Q,\cdot)] &= G_2(X'_s\beta_2 + \phi_{2w})(\mu_s + X'_s\beta_s - \varepsilon_1) - \phi_{1s} \\
E_1[u_1(t_w;B,\cdot)] &= G_2(X'_s\beta_2 - \phi_{2w})(\mu_w + X'_s\beta_w - \varepsilon_1) - \phi_{1w}
\end{align*} \]

S2-1 $t_s$ should have no incentive to deviate. It requires

\[ E_1[u_1(t_s;Q,\cdot)] = G_2(X'_s\beta_2 + \phi_{2w})(\mu_s + X'_s\beta_s - \varepsilon_1) - \phi_{1s} \]
\[ \geq E_1[u_1(t_s;B,\cdot)] = G_2(X'_s\beta_2 - \phi_{2w})(\mu_s + X'_s\beta_s - \varepsilon_1). \]

S2-2 $t_w$ should have no incentive to deviate. It requires

\[ E_1[u_1(t_w;B,\cdot)] = G_2(X'_s\beta_2 - \phi_{2w})(\mu_w + X'_s\beta_w - \varepsilon_1) - \phi_{1w} \]
\[ \geq E_1[u_1(t_w;Q,\cdot)] = G_2(X'_s\beta_2 + \phi_{2w})(\mu_w + X'_s\beta_w - \varepsilon_1). \]

Combining S2-1 and S2-2, we conclude a separating equilibrium $(A_{1t_s}, A_{1w}) = (Q, B)$ can be supported as an equilibrium under\(^\text{41}\)

\[ \mu_w + X'_s\beta_w + \frac{\phi_{1w}}{G_2'(1) - G_2'(0)} \leq \varepsilon_1 \text{ and } \mu_s + X'_s\beta_s - \frac{\phi_{1s}}{G_2'(1) - G_2'(0)} \geq \varepsilon_1. \]

E.2.3 Semi-separating Equilibrium

There exist other possible equilibria named “semi-separating” where one type of Player 1 randomize between two possible actions while the other type of Player 1 plays a separating equilibrium. One of conditions for the existence of such an equilibrium is that the type of Player 1 who randomizes should be indifferent between two actions.

\(^\text{40}\)If \(\mu_w + X'_s\beta_w - \frac{\phi_{1w}}{G_2'(1) - G_2'(0)} > \mu_s + X'_s\beta_s + \frac{\phi_{1s}}{G_2'(1) - G_2'(0)}\), this equilibrium does not exist.

\(^\text{41}\)No such \(\varepsilon_1\) exists when \(\mu_w + X'_s\beta_w + \frac{\phi_{1w}}{G_2'(1) - G_2'(0)} > \mu_s + X'_s\beta_s - \frac{\phi_{1s}}{G_2'(1) - G_2'(0)}\).
SS1: The weak type of Player 1 plays the separating equilibrium with $Q$ and the strong type randomizes with $\sigma_s(B) = \mu_B^w$:

Then,

$$\mu_2(t_1 = t_s|B) = \frac{\mu_B^w \cdot p}{\mu_B^w \cdot p} = 1$$

and

$$\mu_2^s \equiv \mu_2(t_1 = t_s|Q) = \frac{(1 - \mu_B^w) p}{(1 - \mu_B^w) p + (1 - p)} \in (0, p)$$

which implies that

$$E_1 [u_1 (t_s; B, \cdot)] = G_2(X_2' \beta_s + \phi_{2s}) (\mu_s + X_1' \beta_s - \varepsilon_1) \text{ and } E_1 [u_1 (t_s; Q, \cdot)] = G_2(X_2' \beta_s + \mu_2^s \phi_{2s} - \phi_{2w}) (\mu_s + X_1' \beta_s - \varepsilon_1) - \phi_{1w}.$$ 

For the existence of such a semi-separating equilibrium, it is required that the strong type of Player 1 is indifferent between choosing $B$ or $Q$, $E_1 [u_1 (t_s; B, \cdot)] \geq E_1 [u_1 (t_s; Q, \cdot)]$ and hence

$$G_2(X_2' \beta_s + \phi_{2s}) (\mu_s + X_1' \beta_s - \varepsilon_1) \geq G_2(X_2' \beta_s + \mu_2^s \phi_{2s} - \phi_{2w}) (\mu_s + X_1' \beta_s - \varepsilon_1) - \phi_{1w}$$

which holds as long as

$$\varepsilon_1 \geq \mu_w + X_1' \beta_w - \frac{G_2(X_2' \beta_s + \phi_{2s}) - G_2(X_2' \beta_s + \mu_2^s \phi_{2s} - \phi_{2w})}{\phi_{1w}}.$$ 

Note that by construction, we have $0 < \mu_2^s < p$ and hence (52) implies that under the region of $M_2$, such semi-separating equilibria exist.

SS2: The strong type of Player 1 plays the separating equilibrium with $B$ and the weak type randomizes with $\sigma_s(B) = \mu_B^w$:

Then,

$$\tilde{\mu}_2^s = \mu_2(t_1 = t_s|B) = \frac{p}{p + \mu_B^w \cdot (1 - p)} \in (p, 1)$$

and

$$\mu_2(t_1 = t_s|Q) = \frac{0}{(1 - \mu_B^w) \cdot (1 - p)} = 0$$

which implies that

$$E_1 [u_1 (t_w; B, \cdot)] = G_2(X_2' \beta_s + \tilde{\mu}_2^s \phi_{2s} - \phi_{2w}) (\mu_w + X_1' \beta_w - \varepsilon_1) - \phi_{1w} \text{ and } E_1 [u_1 (t_w; Q, \cdot)] = G_2(X_2' \beta_s - \phi_{2w}) (\mu_w + X_1' \beta_w - \varepsilon_1).$$

For the existence of such a semi-separating equilibrium, it is required that the weak type of Player 1 is indifferent between choosing $B$ or $Q$, $E_1 [u_1 (t_w; B, \cdot)] = E_1 [u_1 (t_w; Q, \cdot)]$ and hence

$$G_2(X_2' \beta_s + \tilde{\mu}_2^s \phi_{2s} - \phi_{2w}) (\mu_w + X_1' \beta_w - \varepsilon_1) - \phi_{1w} = G_2(X_2' \beta_s - \phi_{2w}) (\mu_w + X_1' \beta_w - \varepsilon_1)$$

which implies

$$\mu_w + X_1' \beta_w = \frac{\phi_{1w}}{G_2(X_2' \beta_s + \tilde{\mu}_2^s \phi_{2s} - \phi_{2w}) - G_2(X_2' \beta_s - \phi_{2w})} = \varepsilon_1.$$ 

(53)
Now we check whether the strong type of Player 1 has an incentive to deviate. The strong type has no incentive deviate if \( E_1 \left[ \tau_s (B, \cdot) \right] \geq E_1 \left[ \tau_s (Q, \cdot) \right] \),

\[
G_2 (X'_2 \beta_2 + \mu^*_2 2 (\phi_{2s} + \phi_{2w}) - \phi_{2w}) (\mu_s + X'_1 \beta_s - \epsilon_1) \geq G_2 (X'_2 \beta_2 - \phi_{2w}) (\mu_s + X'_1 \beta_s - \epsilon_1) - \phi_{1s}
\]

which holds as long as \( \epsilon_1 \leq \mu_s + X'_1 \beta_s + \frac{\phi_{1s}}{G_2 (X'_2 \beta_2 + \mu^*_2 2 (\phi_{2s} + \phi_{2w}) - \phi_{2w}) - G_2 (X'_2 \beta_2 - \phi_{2w})} \). Note that by construction, we have \( p < \mu^*_2 < 1 \) and hence (53) implies that under the region of \( M_1 \), such semi-separating equilibria exist.

### E.2.4 Existence of Well-defined Likelihood for Semi-Seperating

Under the equilibrium of (SS1), four observed outcomes can arise with following probabilities denoted by

\[
\Pr(Y_1 = 1, Y_2 = 1; M_2, \mu^*_2) = \Pr(Y_1 = 1; M_2, \mu^*_2) \Pr(Y_2 = 1|Y_1 = 1; M_2, \mu^*_2)
\]

\[
= \int_{\epsilon_1 \in M_2} p \cdot \mu^*_2 2 (\epsilon_1) \cdot G_2 (X'_2 \beta_2 + \phi_{2s}) g_1 (\epsilon_1) d\epsilon_1
\]

\[
\Pr(Y_1 = 1, Y_2 = 0; M_2, \mu^*_2) = \Pr(Y_1 = 1; M_2, \mu^*_2) \Pr(Y_2 = 0|Y_1 = 1; M_2, \mu^*_2)
\]

\[
= \int_{\epsilon_1 \in M_2} p \cdot \mu^*_2 2 (\epsilon_1) \cdot (1 - G_2 (X'_2 \beta_2 + \phi_{2s})) g_1 (\epsilon_1) d\epsilon_1
\]

\[
\Pr(Y_1 = 0, Y_2 = 1; M_2, \mu^*_2) = \Pr(Y_1 = 0; M_2, \mu^*_2) \Pr(Y_2 = 1|Y_1 = 0; M_2, \mu^*_2)
\]

\[
= \int_{\epsilon_1 \in M_2} (p (1 - \mu^*_2 2 (\epsilon_1)) + (1 - p)) \cdot G_2 (X'_2 \beta_2 + \mu^*_2 2 (\epsilon_1)(\phi_{2s} + \phi_{2w}) - \phi_{2w}) g_1 (\epsilon_1) d\epsilon_1
\]

\[
\Pr(Y_1 = 0, Y_2 = 0; M_2, \mu^*_2) = \Pr(Y_1 = 0; M_2, \mu^*_2) \Pr(Y_2 = 0|Y_1 = 0; M_2, \mu^*_2)
\]

\[
= \int_{\epsilon_1 \in M_2} (p (1 - \mu^*_2 2 (\epsilon_1)) + (1 - p)) \cdot (1 - G_2 (X'_2 \beta_2 + \mu^*_2 2 (\epsilon_1)(\phi_{2s} + \phi_{2w}) - \phi_{2w})) g_1 (\epsilon_1) d\epsilon_1
\]

where \( \mu^*_2 2 (\epsilon_1) = \frac{(1 - \mu^*_2 2 (\epsilon_1)) p}{(1 - \mu^*_2 2 (\epsilon_1)) p + (1 - p)} \). For example, \( \Pr(Y_1 = 0, Y_2 = 0; M_2, \mu^*_2) \) is obtained using the following facts. Conditional on \((\epsilon_1, \epsilon_2), (Q, F)\) is observed in two cases: 1) the nature draw the strong type with the probability \( p \), the strong type plays \( Q \) with the probability \((1 - \mu^*_2 (\epsilon_1))\), and Player 2 plays \( F \) with probability \((1 - G_2 (X'_2 \beta_2 + \mu^*_2 2 (\epsilon_1)(\phi_{2s} + \phi_{2w}) - \phi_{2w}))\) after observing \( Q \) and 2) the nature draw the weak type with the probability \( 1 - p \), the weak type plays \( Q \) with the probability 1 (separating), and Player 2 plays \( F \) with probability \((1 - G_2 (X'_2 \beta_2 + \mu^*_2 2 (\epsilon_1)(\phi_{2s} + \phi_{2w}) - \phi_{2w}))\) after observing \( Q \). Thus, the conditional probability of observing \((Q, F)\) conditional on \((\epsilon_1, \epsilon_2)\) is given by

\[
(p (1 - \mu^*_2 2 (\epsilon_1)) + (1 - p)) \cdot (1 - G_2 (X'_2 \beta_2 + \mu^*_2 2 (\epsilon_1)(\phi_{2s} + \phi_{2w}) - \phi_{2w}))
\]

and we obtain the unconditional probability by taking expectation of (54) with respect to \( \epsilon_1 \in M_2 \). Other probabilities can be interpreted similarly. Using a change of variables from the relationship of \( \mu^*_2 \) and \( \epsilon_1 \):

\[
\epsilon_1 (\mu^*_2, p) \equiv \mu_s + X'_1 \beta_s + \frac{\phi_{1s}}{G_2 (X'_2 \beta_2 + \phi_{2s}) - G_2 (X'_2 \beta_2 + \mu^*_2 2 (\phi_{2s} + \phi_{2w}) - \phi_{2w})}
\]

as given in (52), we find

\[
\Pr(Y_1 = 1, Y_2 = 1; M_2, \mu^*_2)
\]

\[
= \int_0^1 p \cdot \mu^*_2 \cdot G_2 (X'_2 \beta_2 + \phi_{2s}) g_1 (\epsilon_1 (\mu^*_2, p)) D(\mu^*_2, p) d\mu^*_2
\]

\[
\Pr(Y_1 = 1, Y_2 = 0; M_2, \mu^*_2)
\]

\[
= \int_0^1 p \cdot \mu^*_2 \cdot (1 - G_2 (X'_2 \beta_2 + \phi_{2s})) g_1 (\epsilon_1 (\mu^*_2, p)) D(\mu^*_2, p) d\mu^*_2
\]

\[
\Pr(Y_1 = 0, Y_2 = 1; M_2, \mu^*_2)
\]

\[
= \int_0^1 (p (1 - \mu^*_2) + (1 - p)) \cdot G_2 (X'_2 \beta_2 + \mu^*_2 2 (\mu^*_2, p)(\phi_{2s} + \phi_{2w}) - \phi_{2w}) g_1 (\epsilon_1 (\mu^*_2, p)) D(\mu^*_2, p) d\mu^*_2
\]

\[
\Pr(Y_1 = 0, Y_2 = 0; M_2, \mu^*_2)
\]

\[
= \int_0^1 (p (1 - \mu^*_2) + (1 - p)) \cdot (1 - G_2 (X'_2 \beta_2 + \mu^*_2 2 (\mu^*_2, p)(\phi_{2s} + \phi_{2w}) - \phi_{2w})) g_1 (\epsilon_1 (\mu^*_2, p)) D(\mu^*_2, p) d\mu^*_2
\]
where \( \mu^*_w(\mu_B^*, p) = \frac{(1-\mu_B^*)p}{(1-\mu_B^*)p + (1-p)}, \quad \varepsilon_1(\mu_B^*, p) \) as given in (55), and

\[
D(\mu_B^*, p) \equiv \left| \frac{d\varepsilon_1(\mu_B^*, p)}{d\mu_B^*} \right| = \frac{\phi_1(\phi_2, + \phi_2w)}{(G_2(X_2^*\beta_2 + \phi_2w, - G_2(X_2^*\beta_2 + \mu^*_w(\mu_B^*, p)(\phi_2, + \phi_2w) - \phi_2w))^2}
\times g_2(X_2^*\beta_2 + \mu^*_w(\mu_B^*, p)(\phi_2, + \phi_2w) - \phi_2w) \frac{p(1-p)}{((1-\mu_B^*)p + (1-p))^2}.
\]

Similarly, under this equilibrium of (SS2), four observed outcomes can arise with following probabilities.

\[
\begin{align*}
\Pr(Y_1 = 1, Y_2 = 1; M_1, \mu_B^w) &= \Pr(Y_1 = 1; M_1, \mu_B^w) \Pr(Y_2 = 1|Y_1 = 1; M_1, \mu_B^w) \\
&= \int_{\varepsilon_1 \in M_1} (p + \mu_B^w(\varepsilon_1) (1-p)) G_2(X_2^*\beta_2 + \mu^*_w(\varepsilon_1)(\phi_2, + \phi_2w) - \phi_2w) g_1(\varepsilon_1) d\varepsilon_1 \\
\Pr(Y_1 = 1, Y_2 = 0; M_1, \mu_B^w) &= \Pr(Y_1 = 1; M_1, \mu_B^w) \Pr(Y_2 = 0|Y_1 = 1; M_1, \mu_B^w) \\
&= \int_{\varepsilon_1 \in M_1} (p + \mu_B^w(\varepsilon_1) (1-p)) (1 - G_2(X_2^*\beta_2 + \mu^*_w(\varepsilon_1)(\phi_2, + \phi_2w) - \phi_2w)) g_1(\varepsilon_1) d\varepsilon_1 \\
\Pr(Y_1 = 0, Y_2 = 1; M_1, \mu_B^w) &= \Pr(Y_1 = 0; M_1, \mu_B^w) \Pr(Y_2 = 1|Y_1 = 0; M_1, \mu_B^w) \\
&= \int_{\varepsilon_1 \in M_1} (1 - \mu_B^w(\varepsilon_1) (1-p)) G_2(X_2^*\beta_2 - \phi_2w) g_1(\varepsilon_1) d\varepsilon_1 \\
\Pr(Y_1 = 0, Y_2 = 0; M_1, \mu_B^w) &= \Pr(Y_1 = 0; M_1, \mu_B^w) \Pr(Y_2 = 0|Y_1 = 0; M_1, \mu_B^w) \\
&= \int_{\varepsilon_1 \in M_1} (1 - \mu_B^w(\varepsilon_1) (1-p)) (1 - G_2(X_2^*\beta_2 - \phi_2w)) g_1(\varepsilon_1) d\varepsilon_1
\end{align*}
\]

where \( \tilde{\mu}^*_w(\varepsilon_1) = \frac{p}{p + \mu_B^w(\varepsilon_1) (1-p)}. \) A change of variables using the relationship of \( \mu^*_w \) and \( \varepsilon_1: \)

\[
\varepsilon_1(\mu_B^*, p) = \mu_w + X_1^*\beta_w - \frac{\phi_1}{G_2(X_2^*\beta_2 + \mu^*_w(\mu_B^*, p)(\phi_2, + \phi_2w) - \phi_2w) - G_2(X_2^*\beta_2 - \phi_2w)}
\]

as given in (53) gives us

\[
\begin{align*}
\Pr(Y_1 = 1, Y_2 = 1; M_1, \mu_B^w) &= \int_0^1 (p + \mu_B^w(\varepsilon_0) (1-p)) G_2(X_2^*\beta_2 + \tilde{\mu}^*_w(\varepsilon_0)(\phi_2, + \phi_2w) - \phi_2w) g_1(\varepsilon_0(\mu_B^*, p)) D(\mu_B^*, p) d\mu_B^w \\
\Pr(Y_1 = 1, Y_2 = 0; M_1, \mu_B^w) &= \int_0^1 (p + \mu_B^w(\varepsilon_0) (1-p)) (1 - G_2(X_2^*\beta_2 + \tilde{\mu}^*_w(\varepsilon_0)(\phi_2, + \phi_2w) - \phi_2w)) g_1(\varepsilon_0(\mu_B^*, p)) D(\mu_B^*, p) d\mu_B^w \\
\Pr(Y_1 = 0, Y_2 = 1; M_1, \mu_B^w) &= \int_0^1 (1 - \mu_B^w(\varepsilon_0) (1-p)) G_2(X_2^*\beta_2 - \phi_2w) g_1(\varepsilon_0(\mu_B^*, p)) D(\mu_B^*, p) d\mu_B^w \\
\Pr(Y_1 = 0, Y_2 = 0; M_1, \mu_B^w) &= \int_0^1 (1 - \mu_B^w(\varepsilon_0) (1-p)) (1 - G_2(X_2^*\beta_2 - \phi_2w)) g_1(\varepsilon_0(\mu_B^*, p)) D(\mu_B^*, p) d\mu_B^w
\end{align*}
\]

where \( \tilde{\mu}^*_w(\mu_B^*, p) = \frac{p}{p + \mu_B^w(1-p)}, \varepsilon_1(\mu_B^*, p) \) as given in (56), and

\[
D(\mu_B^w, p) \equiv \left| \frac{d\varepsilon_1(\mu_B^w, p)}{d\mu_B^w} \right| = \frac{\phi_1}{G_2(X_2^*\beta_2 + \tilde{\mu}^*_w(\mu_B^*, p)(\phi_2, + \phi_2w) - \phi_2w) - G_2(X_2^*\beta_2 - \phi_2w))^2}
\times g_2(X_2^*\beta_2 + \tilde{\mu}^*_w(\mu_B^*, p)(\phi_2, + \phi_2w) - \phi_2w) \frac{p(1-p)}{(p + \mu_B^w(1-p))^2}.
\]

### E.3 Existence of Well-defined Likelihood for the Game with IS-2

We proceed the following discussion for the game with public signals. All the discussions hold true for the game without public signals also by replacing \( p(Z) \) with \( p \). We add the public signal assumption to IS-2.

**Assumption E.2 (IS-2.A)**

1. Assumption IS-2 holds.
2. The public signal \( Z \) about the types of Player 1 is perfectly known to both Player 1 and Player 2.
E.3.1 Equilibrium Refinement and Uniqueness of Equilibrium (Proof of Theorem E.2)

Again we proceed our discussion with asymmetric payoffs case. From the result we obtain here, uniqueness of equilibrium for the symmetric payoffs case immediately follows. We make the following assumptions to obtain uniqueness of equilibrium.

**Condition 5**  
\[ \mu_w + X_1^\beta_w - \frac{\phi_{1w}}{G_2(X_2^\beta_2 + \phi_{2w}) - G_2(X_2^\beta_2 - \phi_{2w})} < \mu_s + X_1^\beta_s + \frac{\phi_{1s}}{G_2(X_2^\beta_2 + \phi_{2s}) - G_2(X_2^\beta_2 - \phi_{2s})} \]  
for all \( X_1 \times X_2 \in \mathcal{S}(X_1) \times \mathcal{S}(X_2) \).

**Condition 6**  
For all \( W \in \mathcal{S}(W) \),  
\[ \mu_w + X_1^\beta_w - \frac{\phi_{1w}}{G_2(X_2^\beta_2 + \phi_{2w}) - G_2(X_2^\beta_2 - \phi_{2w})} < \mu_s + X_1^\beta_s + \frac{\phi_{1s}}{G_2(X_2^\beta_2 + \phi_{2s}) - G_2(X_2^\beta_2 - \phi_{2s})} \]

**Condition 7**  
For all \( W \in \mathcal{S}(W) \),  
\[ \mu_w + X_1^\beta_w - \frac{\phi_{1w}}{G_2(X_2^\beta_2 + \phi_{2w}) - G_2(X_2^\beta_2 + p(Z)(\phi_{2s} + \phi_{2w}) - \phi_{2w})} < \mu_s + X_1^\beta_s + \frac{\phi_{1s}}{G_2(X_2^\beta_2 + \phi_{2s}) - G_2(X_2^\beta_2 + p(Z)(\phi_{2s} + \phi_{2w}) - \phi_{2w})} \]

**Condition 8**  
\[ \mu_s + X_1^\beta_s - \frac{\phi_{1s}}{G_2(X_2^\beta_2 + \phi_{2s}) - G_2(X_2^\beta_2 - \phi_{2s})} < \mu_w + X_1^\beta_w + \frac{\phi_{1w}}{G_2(X_2^\beta_2 + \phi_{2w}) - G_2(X_2^\beta_2 - \phi_{2w})} \]  
for all \( X_1 \times X_2 \in \mathcal{S}(X_1) \times \mathcal{S}(X_2) \).

These conditions are sufficient for uniqueness of equilibria together with the refinement of Banks and Sobel (1987), namely, \emph{divine equilibrium}. Note that Conditions 5-8 hold immediately when \( \mu_s = \mu_w \) and \( \beta_w = \beta_s \) since \( \phi_{1s}, \phi_{1w} > 0 \). Condition 5 makes the separating \((B, Q)\) supported (see (50)) and thus prevents two different semi-separating equilibria from overlapping each other (see Figure A4). Condition 8 eliminates the separating equilibrium \((Q, B)\) (see (51)). Now we eliminate some of pooling equilibria using the refinement.

- Pooling with \((A_{1t_s}, A_{1t_w}) = (B, B)\)

Under this pooling equilibrium, Player 2 knows the weak type is more willing to deviate (at any given Player 2’s action) and hence Player 1 knows that if Player 2 observes a deviation play of \( Q \), she will assign \( \mu_2(t_1 = t_s|Q) < p \). In other words, it is reasonable to expect that the relative probability of \( t_s \) should decrease when Player 2 observes \( Q \) according to Banks and Sobel (1987)’s \emph{divinity} concept. This rules out the case of (44) and only the region defined by (45) supports this equilibrium.

This refinement and Condition 6 make the pooling equilibrium with \((A_{1t_s}, A_{1t_w}) = (B, B)\) supported only under
\[ \varepsilon_1 \leq \mu_w + X_1^\beta_w - \frac{\phi_{1w}}{G_2(X_2^\beta_2 + p(Z)(\phi_{2s} + \phi_{2w}) - \phi_{2w}) - G_2(X_2^\beta_2 - \phi_{2w})} \]

- Pooling with \((A_{1t_s}, A_{1t_w}) = (Q, Q)\)

Under this pooling equilibrium, Player 2 knows the strong type is more willing to deviate (at any given Player 2’s action) and hence Player 1 knows that if Player 2 observes a deviation play of \( B \), she will assign
\( \mu_{2}(t_{1} = t_{s}|B) > p \). Again Banks and Sobel (1987)'s *divinity* requires that we should expect that the relative probability of \( t_{s} \) should increase when Player 2 observes a deviation play with \( B \). This rules out the case of (49) and only the region defined by (48) supports this equilibrium.

This refinement and Condition 7 make the pooling with equilibrium \((A_{1s}, A_{1w}) = (Q, Q)\) supported only under
\[
\varepsilon_{1} \geq \mu_{s} + X_{t}^{1} \beta_{s} + \frac{\phi_{1s}}{G_{2}(X_{2}^{t} \beta_{2} + \phi_{2s}) - G_{2}(X_{2}^{t} \beta_{2} + \phi_{2s} + \phi_{2w} - \phi_{2w})}.
\]

Finally, we note that the estimation strategies considered in the paper are still valid based on the conditional probabilities derived in the next section provided that Assumption **SA-Asym-IS-2** holds.

**Assumption E.3 (SA-Asym-IS-2)**

(i) \( \mu_{w} + X_{1}^{1} \beta_{w} - \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w})}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w})} < \mu_{s} + X_{1}^{1} \beta_{s} - \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s})}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s}) - G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2s})} \)
for all \( X_{1} \times X_{2} \in \mathbb{S}(X_{1}) \times \mathbb{S}(X_{2}) \) and for all \( \mu_{s}, \phi_{1w}, \phi_{2s}, \phi_{2w}, \beta_{s}, \beta_{w}, \beta_{2} \) in the parameter space;

(ii) \( \mu_{w} + X_{1}^{1} \beta_{w} - \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w})}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w})} < \mu_{s} + X_{1}^{1} \beta_{s} + \frac{\phi_{2w}}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})} \)
for all \( W \in \mathbb{W}(W) \) and for all \( \mu_{s}, \phi_{1w}, \phi_{1w}, \phi_{2s}, \phi_{2w}, \beta_{s}, \beta_{w}, \beta_{2} \) and \( h(\cdot) \) in the parameter space;

(iii) \( \mu_{w} + X_{1}^{1} \beta_{w} - \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w})}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w})} < \mu_{s} + X_{1}^{1} \beta_{s} + \frac{\phi_{1w}}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})} \)
for all \( W \in \mathbb{W}(W) \) and for all \( \mu_{s}, \mu_{w}, \phi_{1w}, \phi_{1w}, \phi_{2s}, \phi_{2w}, \beta_{s}, \beta_{w}, \beta_{2} \) and \( h(\cdot) \) in the parameter space;

(iv) \( \mu_{w} + X_{1}^{1} \beta_{w} - \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w})}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w})} < \mu_{s} + X_{1}^{1} \beta_{s} + \frac{\phi_{1w}}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})} \)
for all \( X_{1} \times X_{2} \in \mathbb{S}(X_{1}) \times \mathbb{S}(X_{2}) \) and for all \( \mu_{s}, \mu_{w}, \phi_{1w}, \phi_{1w}, \phi_{2s}, \phi_{2w}, \beta_{s}, \beta_{w}, \beta_{2} \) in the parameter space.

Note that Assumption **SA-Asym-IS-2** holds immediately when \( \mu_{s} = \mu_{w} \) and \( \beta_{s} = \beta_{w} \).

**E.3.2 Conditional Probabilities of Four Observed Outcomes**

From the result of previous sections, here we present the conditional probabilities of four observed outcomes in the game with **IS-2A** and **SA-2** allowing for asymmetric payoffs. For the game with **IS-2** and **SA-1**, we obtain the same conditional probabilities in replace of \( p(Z) \) with \( p \) and \( W \) with \( X \), respectively. Corresponding conditional probabilities of the game with symmetric payoffs are easily obtained by replacing \( \beta_{s} \) and \( \beta_{w} \) with \( \beta_{1} \) and \( \mu_{s} \) and \( \mu_{w} \) with \( \mu_{i} \), respectively.

We will use some simplifying notations:

\[
G_{2}^{s}(a) = G_{2}(X_{2}^{s} \beta_{2} + a(\phi_{2s} + \phi_{2w}) - \phi_{2w}),
\]
\[
\mu_{2}^{s}(\mu_{B}^{s}, p(Z)) = \frac{1}{(1-\mu_{B}^{s})(p(Z)+(1-p(Z)))},
\]
\[
\varepsilon_{1}(\mu_{B}^{s}, p(Z)) = \mu_{w} + X_{1}^{1} \beta_{w} + \frac{\phi_{1w}}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})},
\]
\[
D(\mu_{B}^{s}, p(Z)) = \left| \frac{d\varepsilon_{1}(\mu_{B}^{s}, p(Z))}{d\mu_{B}^{s}} \right| = \frac{1}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})} \times \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2s} - \phi_{2w})}{p(Z)(1-p(Z))},
\]
\[
\tilde{\varepsilon}_{1}(\mu_{B}^{w}, p(Z)) = \mu_{w} + X_{1}^{1} \beta_{w} - \frac{\phi_{1w}}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})},
\]
\[
\tilde{D}(\mu_{B}^{w}, p(Z)) = \left| \frac{d\tilde{\varepsilon}_{1}(\mu_{B}^{w}, p(Z))}{d\mu_{B}^{w}} \right| = \frac{1}{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w}) - G_{2}(X_{2}^{1} \beta_{2} - \phi_{2w})} \times \frac{G_{2}(X_{2}^{1} \beta_{2} + \phi_{2w} - \phi_{2w})}{p(Z)(1-p(Z))}.\]
Now we present the conditional probabilities of four possible observed outcomes.

1 \((Y_1 = 1, Y_2 = 1) : (B, NF)\)

It happens under region of \(L_2\) with probability one (pooling), under region of \(C_2 \cup C_3 \cup C_4\) (separating) with probability \(p(Z)\), and under region of \(M_{12} \cup M_{13} \cup M_{14} \cup M_{22} \cup M_{23} \cup M_{24}\) (semi-separating):

\[
\Pr(Y_1 = 1, Y_2 = 1|W, \alpha) = G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) G_2^* (p(Z)) = G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) G_2^* (p(Z)) + \int_0^1 p(Z) G_1 \left( \mu_s + X'_1\beta_s + \frac{\phi_{1s}}{G_2^*, p(Z)} \right) - G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) G_2^* (1) + \int_0^1 (p(Z) + \mu_B^w (1 - p(Z))) G_2^* (\mu_B^w, p(Z)) g_1(\epsilon_1(\mu_B^w, p(Z))) d\mu_B^w + \int_0^1 p(Z) G_2^* (1) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w.
\]

2 \((Y_1 = 1, Y_2 = 0) : (B, F)\)

It happens under region of \(L_1\) with probability one (pooling) and under region of \(C_1\) (separating) with probability \(p(Z)\), and under region of \(M_{11} \cup M_{12} \cup M_{21}\) (semi-separating):

\[
\Pr(Y_1 = 1, Y_2 = 0|W, \alpha) = G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) (1 - G_2^* (p(Z))) + \int_0^1 p(Z) G_1 \left( \mu_s + X'_1\beta_s + \frac{\phi_{1s}}{G_2^*, p(Z)} \right) - G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) G_2^* (1) + \int_0^1 (1 - \mu_B^w) (1 - p(Z)) G_2^* (0) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w + \int_0^1 (1 - p(Z)) G_2^* (\mu_B^w, p(Z)) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w.
\]

3 \((Y_1 = 0, Y_2 = 1) : (Q, NF)\)

It happens under region of \(R_2\) (pooling) with probability one and under region of \(C_4\) (separating) with probability \(1 - p(Z)\), and under region of \(M_{14} \cup M_{23} \cup M_{24}\) (semi-separating):

\[
\Pr(Y_1 = 0, Y_2 = 1|W, \alpha) = G_1 \left( \mu_w + X'_1\beta_w + \frac{\phi_{1w}}{G_2^*, p(Z)} \right) G_2^* (p(Z)) + (1 - p(Z)) G_1 \left( \mu_s + X'_1\beta_s + \frac{\phi_{1s}}{G_2^*, p(Z)} \right) - G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) G_2^* (0) + \int_0^1 (1 - \mu_B^w) (1 - p(Z)) G_2^* (0) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w + \int_0^1 (1 - p(Z)) G_2^* (\mu_B^w, p(Z)) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w.
\]

4 \((Y_1 = 0, Y_2 = 0) : (Q, F)\)

It happens under region of \(R_1\) (pooling) with probability one and under region of \(C_1 \cup C_2 \cup C_3\) (separating) with probability \(1 - p(Z)\), and under region of \(M_{11} \cup M_{12} \cup M_{13} \cup M_{21} \cup M_{22} \cup M_{23}\) (semi-separating):

\[
\Pr(Y_1 = 0, Y_2 = 0|W, \alpha) = G_1 \left( \mu_w + X'_1\beta_w + \frac{\phi_{1w}}{G_2^*, p(Z)} \right) (1 - G_2^* (p(Z))) + (1 - p(Z)) G_1 \left( \mu_s + X'_1\beta_s + \frac{\phi_{1s}}{G_2^*, p(Z)} \right) - G_1 \left( \mu_w + X'_1\beta_w - \frac{\phi_{1w}}{G_2^*, p(Z)} \right) (1 - G_2^* (0)) + \int_0^1 (1 - \mu_B^w) (1 - p(Z)) G_2^* (0) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w + \int_0^1 (1 - p(Z)) G_2^* (\mu_B^w, p(Z)) g_1(\epsilon_1(\mu_B^w, p(Z))) D(\mu_B^w, p(Z)) d\mu_B^w.
\]
F Mathematical Proofs for the Sieve Conditional ML

All the discussions of this section can be applied to both game models under IS-A and IS-2A based on appropriate conditional probabilities presented in Sections C and E.3.2, respectively.

F.1 Identification (Proof of Lemma 5.1)

Proof. We let \( \mathcal{L}(Y, W, \alpha) = \exp((Y|W, \alpha)) \) and note

\[
\frac{\mathcal{L}(Y, W, \alpha)}{\mathcal{L}(Y, W, \alpha_0)} = \left\{ \frac{\Pr(Y = y|W, \alpha)}{\Pr(Y = y|W, \alpha_0)} : y \in \{(1, 1), (1, 0), (0, 1), (0, 0)\} \right\}
\]

where \( \Pr(Y|W, \alpha) \) denotes the conditional probability of \( Y \) given \( W = X \cup Z \) when the parameter equals to \( \alpha \). Applying Jensen’s inequality, we have

\[
- \ln \left\{ E \left[ \frac{\mathcal{L}(Y, W, \alpha)}{\mathcal{L}(Y, W, \alpha_0)} \right] \right\} < -E \left[ \ln \left\{ \frac{\mathcal{L}(Y, W, \alpha)}{\mathcal{L}(Y, W, \alpha_0)} \right\} \right]
\]

noting \( \frac{\mathcal{L}(Y, W, \alpha)}{\mathcal{L}(Y, W, \alpha_0)} \) is always positive and not constant whenever \( \alpha \neq \alpha_0 \) by Assumption SA-3. We also have

\[
\Pr \left( \frac{\mathcal{L}(Y, W, \alpha)}{\mathcal{L}(Y, W, \alpha_0)} = \frac{\Pr(Y = y|W, \alpha)}{\Pr(Y = y|W, \alpha_0)} \right) = \Pr(Y = y|W, \alpha_0),
\]

for each \( y \in \{(1, 1), (1, 0), (0, 1), (0, 0)\} \) under Assumptions IS-A and SA-2 or Assumptions IS-2A and SA-2. It follows that

\[
E \left[ \frac{\mathcal{L}(Y, W, \alpha)}{\mathcal{L}(Y, W, \alpha_0)} \right] = \int \left\{ \sum_y \Pr(Y = y|W, \alpha) \cdot \Pr(Y = y|W, \alpha_0) \right\} f_W(W) dW = \int \left\{ \sum_y \Pr(Y = y|W, \alpha) \right\} f_W(W) dW = 1
\]

where the last equality holds since \( \sum_y \Pr(Y = y|W, \alpha) = 1 \) for all \( \alpha \in A \). Therefore, combining (57) and (58), we conclude that for all \( \alpha \neq \alpha_0 \in A, 0 < E[\ln \mathcal{L}(Y, W, \alpha_0)] - E[\ln \mathcal{L}(Y, W, \alpha)] \). This implies \( Q(\alpha_0) > Q(\alpha) \) and thus proves the claim.

F.2 Consistency

To prove the consistency, we need an additional condition. Recall that \( \mathcal{L}_{ij}(W, \alpha) \) denotes the conditional probabilities of observed outcomes such that \( \mathcal{L}_{ij}(W, \alpha) = \Pr(Y_1 = i, Y_2 = j|W, \alpha) \) for \( i, j = 0, 1 \).

Condition 9 (Lipschitz Condition)

(i) For \( \mathbf{\pi} \) some convex combination of \( \alpha_1, \alpha_2 \in A_n \), there exists functions \( M_{ij}^{(\phi_{11})}(), M_{ij}^{(\phi_{1w})}(), M_{ij}^{(\phi_{21})}(), M_{ij}^{(\phi_{2w})}(), M_{ij}^{(\mu)}(), M_{ij}^{(\beta_1)}(), M_{ij}^{(\beta_2)}(), \) and \( M_{ij}^{(h)}() \) such that

\[
\begin{align*}
\frac{d\mathcal{L}_{ij}(W, \mathbf{\pi})}{d\alpha} &\left( \alpha_1 - \alpha_2 \right) = \\
M_{ij}^{(\phi_{11})}(W, \mathbf{\pi})(\phi_{1x1} - \phi_{1x2}) &+ M_{ij}^{(\phi_{1w})}(W, \mathbf{\pi})(\phi_{1w1} - \phi_{1w2}) + M_{ij}^{(\phi_{21})}(W, \mathbf{\pi})(\phi_{2s1} - \phi_{2s2}) \\
+ M_{ij}^{(\phi_{2w})}(W, \mathbf{\pi})(\phi_{2w1} - \phi_{2w2}) &+ M_{ij}^{(\mu)}(W, \mathbf{\pi})(\mu_1 - \mu_2) \\
+ M_{ij}^{(\beta_1)}(W, \mathbf{\pi})X_1(\beta_{11} - \beta_{12}) &+ M_{ij}^{(\beta_2)}(W, \mathbf{\pi})X_2(\beta_{21} - \beta_{22}) + M_{ij}^{(h)}(W, \mathbf{\pi})(h_1 - h_2)
\end{align*}
\]

for all \( i, j = 0, 1 \);

(ii) \( \sup_{\alpha \in A_n} \left| M_{ij}^{(t)}(W, \alpha) \right| \leq C_t(W) < \infty, \forall t \in \{\phi_{1x}, \phi_{1w}, \phi_{2s}, \mu, \beta_{1}, \beta_{2}, h\} \), \( \forall i, j = 0, 1, n \geq \exists N \).
This condition is not difficult to verify from the arguments in Section J.

The following theorem is borrowed from Theorem 3.1 in Chen (2005) for the sieve conditional ML estimator defined in (11).

**Theorem F.1 (Theorem 3.1 in Chen (2005))**

Suppose (C1) $Q(\alpha)$ is uniquely maximized on $A$ at $\alpha_0 \in A$, and $Q(\alpha_0) > -\infty$;
(C2) $A_n \subset A_{n+1} \subset A$ for all $n \geq 1$, and for any $\alpha \in A$ there exists $\pi_n \alpha \in A_n$ such that $\|\Pi_n \alpha - \alpha\|_s \to 0$ as $n \to \infty$;
(C3) The criterion function, $Q(\alpha)$, is continuous in $\alpha \in A$ with respect to $\|\|_s$;
(C4) The sieve spaces, $A_n$, are compact under $\|\|_s$;
(C5) $\lim_{n \to \infty} \sup_{\alpha \in A_n} \left| \hat{Q}_n(\alpha) - Q(\alpha) \right| = 0$ holds and let $\hat{\alpha}_n$ be the approximate sieve ML estimator defined by (11), then $\|\hat{\alpha}_n - \alpha\|_s = o_p(1)$.

Remarks (1)-(4) after Theorem 3.1 in Chen (2005) are applied here. Note that Condition (C1) is satisfied by Lemma 5.1. Condition (C2) holds for the sieve space $A_n = \Theta \times \mathcal{H}_n$ with $\mathcal{H}_n$ defined in (8) (see Section 5.3.2 of Timan (1963)). Condition (C3) is satisfied since each $L_{ij}$, $i, j = 1, 0$ is (pointwise) Lipschitz continuous by Condition 9 for the game models with IS-A or IS-2A. Condition (C4) holds for the sieve space (8). Now let $\mathcal{F}_n = \{l(y|w, \theta, h) : (\theta, h) \in A_n\}$ denote the class of measurable functions indexed by $(\theta, h)$. Condition (C5) will be satisfied, for example, if $\mathcal{F}_n$ is $P$-Glivenko-Cantelli as presented in van der Vaart and Wellner (1996). The following lemma establishes the uniform convergence result.

**Lemma F.1 (Uniform convergence over sieves)**

Suppose Assumptions SA-2 and SA-4 hold. Then, for $\hat{Q}_n(\alpha)$ and $Q(\alpha)$ defined in (10) and (12), respectively and for $A_n = \Theta \times \mathcal{H}_n$ with $\mathcal{H}_n$ defined in (8), we have $\lim_{n \to \infty} \sup_{\alpha \in A_n} \left| \hat{Q}_n(\alpha) - Q(\alpha) \right| = 0$.

**Proof.** We prove this lemma by showing that all the conditions (i), (ii), and (iii) of Lemma A2 in Newey and Powell (2003) are satisfied. The condition (i) is satisfied for $A_n = \Theta \times \mathcal{H}_n$ with $\mathcal{H}_n$ defined in (8) and for the metric $\|\|_s$. The condition (ii) will be satisfied if $E[|l(y|w, \theta, h(z))|] < \infty$, for all $(\theta, h) \in A_n$ by the law of large numbers. Note that this condition is satisfied since $L_{ij}(W, \theta, h)$, $\forall i, j = 0, 1$ is uniformly bounded between 0 and 1 over $A_n$ and since under Assumptions SA-2,

$\Pr(L_{ij}(W, \theta, h) = 0 \text{ or } L_{ij}(W, \theta, h) = 1 \text{ for } W \in S(W)) = 0 \text{ for } \forall i, j = 0, 1 \text{ and all } (\theta, h) \in A_n,$

where $\Pr(\cdot)$ is a probability measure over $W$. Therefore, the condition (ii) holds. The condition (iii) is also satisfied since $\hat{Q}_n(\alpha)$ is Lipschitz with respect to $\alpha$ by Condition 9. This completes the proof. \blacksquare

Therefore, under Assumptions SA-2, SA-3, and SA-4 and Condition 9, all the conditions in Theorem F.1 are satisfied for $\hat{\alpha}_n$. This establishes the consistency result of the sieve ML estimator defined in (11).

**F.3 Convergence Rate**

In this section and the next section, we will use the following notations. For any $\alpha_1, \alpha_2, \alpha_3 \in A$, the pathwise first derivatives are defined as

$$\frac{dl(y|w, \alpha_1)}{d\alpha} [\alpha_1 - \alpha_2] \equiv \frac{dl(y|w, \alpha_0)}{d\alpha} [\alpha_1 - \alpha_0] - \frac{dl(y|w, \alpha_0)}{d\alpha} [\alpha_2 - \alpha_0] \text{ and}$$

$$\frac{dl(y|w, \alpha_1)}{d\alpha} [\alpha_1 - \alpha_0] \equiv \lim_{\tau \to 0} \frac{dl(y|w, \alpha_0 + \tau(\alpha_1 - \alpha_0))}{d\tau}.$$
The pathwise second derivatives are defined in a similar way such that

\[
\frac{d^2 l(y_i|w_i, \alpha_3)}{d\alpha^2}[\alpha_1 - \alpha_0, \alpha_2 - \alpha_0] \equiv \lim_{r \to 0} \frac{d^2 l(y_i|w_i, \alpha_3 + r(\alpha_2 - \alpha_0))}{d\alpha} [\alpha_1 - \alpha_0].
\]

In particular, the pathwise second derivative at the direction \([\alpha - \alpha_0]\) evaluated at \(\alpha_0\) is denoted by

\[
\frac{d^2 l(y_i|w_i, \alpha_0)}{d\alpha^2}[\alpha - \alpha_0, \alpha - \alpha_0] \equiv \lim_{r \to 0} \frac{d^2 l(y_i|w_i, \alpha_0 + r(\alpha - \alpha_0))}{d\alpha^2}. \tag{59}
\]

Note that we have

\[
E \left[ \frac{d l(y_i|w_i, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_0] \right] = \lim_{r \to 0} E \left[ \frac{d^2 l(y_i|w_i, \alpha_0)}{d\alpha^2}[\alpha_1 - \alpha_0, \alpha_2 - \alpha_0] \right] - E \left[ \frac{d^2 l(y_i|w_i, \alpha_0)}{d\alpha^2}[\alpha_1 - \alpha_0, \alpha_2 - \alpha_0] \right]
\]


Here we present a convergence rate of the sieve ML estimator defined in (11) under the metric \(\|\cdot\|_2\). We use Theorem 3.2 of Chen (2005) which is a version of Chen and Shen (1996) for iid data. Before stating Theorem 3.2 of Chen (2005) and proving Proposition 5.1, we need to introduce some notations.

Now let \(K(\alpha_0, \alpha) = n^{-1} \sum_{i=1}^n E[l(y_i|w_i, \alpha_0) - l(y_i|w_i, \alpha)]\) denote the Kullback-Leibler information (divergence measure) on \(n\) observations. Let \(\|\cdot\|_a\) be a metric on \(\mathcal{A}\) which is equivalent to \(K(\cdot, \cdot)^{1/2}\). The equivalence means there exist constants \(C_1\) and \(C_2 > 0\) such that \(C_1 K(\alpha_0, \alpha)^{1/2} \leq \|\alpha - \alpha_0\|_a \leq C_2 K(\alpha_0, \alpha)^{1/2}\) for all \(\alpha \in \mathcal{A}\). It is well known in the literature that the convergence rate depends on two factors. One is how fast the sieve approximation error rate, \(\|\alpha_0 - \Pi_n \alpha_0\|_a\), goes to zero. The other is the complexity of the sieve space \(\mathcal{H}_n\). Let \(L_r(P_0), r \in [1, \infty)\) denote the space of real-valued random variables with finite \(r\)-th moments and \(\|\cdot\|_{L_r}\) denote the \(L_r(P_0)\)-norm. Let \(\mathcal{F}_n = \{g(\alpha, \cdot): \alpha \in \mathcal{A}_n\}\) be a class of real-valued, \(L_r(P_0)\)-measurable functions indexed by \(\alpha \in \mathcal{A}_n\). We let \(N(\varepsilon, \mathcal{F}_n, \|\cdot\|_{L_r})\) denote the covering numbers without bracketing, which is the minimal number of \(\varepsilon\)-balls \(\left\{ f : \|f - g\|_{L_r} \leq \varepsilon \right\} \) of the form \(\|g_j\|_{L_r} < \infty, j = 1, \ldots, N \right\}\) that covers \(\mathcal{F}_n\). We often use the notion of \(L_r(P_0)\)-metric entropy, \(H(\varepsilon, \mathcal{F}_n, \|\cdot\|_{L_r}) \equiv \log(N(\varepsilon, \mathcal{F}_n, \|\cdot\|_{L_r}))\) as a measure of the complexity of \(\mathcal{F}_n\), since the covering numbers can grow very fast. The second notion of complexity of the class \(\mathcal{F}_n\) is the covering numbers with bracketing. Let \(\mathcal{C}_{L_r}\) be the completion of \(\mathcal{F}_n\) under the norm \(\|\cdot\|_{L_r}\). The \(L_r(P_0)\)-covering numbers with bracketing, denoted by \(N(\varepsilon, \mathcal{C}_{L_r}, \|\cdot\|_{L_r})\), is the minimal number of \(\varepsilon\)-brackets \(\left\{ [l_j, u_j] : l_j, u_j \in \mathcal{C}_{L_r}, \max_{1 \leq j \leq N} \|l_j - u_j\|_{L_r} \leq \varepsilon, \|l_j\|_{L_r}, \|u_j\|_{L_r} < \infty, j = 1, \ldots, N \right\}\) to cover \(\mathcal{F}_n\). Similarly we let \(H(\varepsilon, \mathcal{C}_{L_r}, \|\cdot\|_{L_r}) \equiv \log(N(\varepsilon, \mathcal{C}_{L_r}, \|\cdot\|_{L_r}))\) which is called the \(L_r(P_0)\)-metric entropy with bracketing of the class \(\mathcal{F}_n\). Pollard (1984), Andrews (1994), van der Vaart and Wellner (1996), and van der Geer (2000) provide more detailed discussions of metric entropy. We will use a simplified notation \(b_{1n} \sim b_{2n}\) for two sequences of positive numbers \(b_{1n}\) and \(b_{2n}\), when the ratio of these two \(b_{1n}/b_{2n}\) is bounded below and above by some positive constants.

Now let \(\mathcal{F}_n = \{l(y_i|w_i, \alpha) - l(y_i|w_i, \alpha_0) : \|\alpha - \alpha_0\|_a \leq \delta, \alpha \in \mathcal{A}_n\}\) and for some constant \(b > 0\), let

\[
\delta_n = \inf \left\{ \delta \in (0, 1) : \frac{1}{\sqrt{n\delta^2}} \int_{b\delta^2} \sqrt{H(\varepsilon, \mathcal{C}_{L_r}, \|\cdot\|_{L_r})} \, dz \leq \text{const} \right\}. \tag{60}
\]

Note that \(\delta_n\) does not only depend on the smoothness of \(l(\cdot|\cdot, \cdot)\) but also on the complexity of the sieve \(\mathcal{A}_n\). Now we present Theorem 3.2 in Chen (2005) tailored to our estimator.

**Theorem F.2** (Theorem 3.2 in Chen (2005))

Suppose that (CC1) The data \(\{Y_1, Y_2, W_i\}\) are iid;
(CC2) There is a $C_1 > 0$ such that for all small $\varepsilon > 0$, $\sup_{\alpha \in A_n, ||\alpha - \alpha_0||_s \leq \varepsilon} \Var (l(y_i | w_i, \alpha) - l(y_i | w_i, \alpha_0)) \leq C_1 \varepsilon^2$;

(CC3) For any $\delta > 0$, there exists a constant $s \in (0,2)$ such that

$$\sup_{\alpha \in A_n, ||\alpha - \alpha_0||_s \leq \delta} |l(y_i | w_i, \alpha) - l(y_i | w_i, \alpha_0)| \leq \delta \varUpsilon(U(w_i))$$

with $E \left[ \left| U(w_i) \right|^t \right] \leq C_2$ for some $t \geq 2$. Let $\tilde{\alpha}_n$ be the approximate sieve ML defined in (11). Then, $\| \tilde{\alpha}_n - \alpha_0 \|_2 = O_p(\varepsilon_n)$ with $\varepsilon_n = \max \{ \delta_n, \| \Pi_n \alpha_0 - \alpha_0 \|_2 \}$.

Note that $\tilde{\alpha}_n$ increases with the complexity of the sieve $A_n$, which can be interpreted as a measure of the standard deviation form, while we interpret the deterministic approximation error $\| \Pi_n \alpha_0 - \alpha_0 \|_2$ as a measure of the bias since it decreases with the complexity of the sieve $A_n$. Now we prove Proposition 5.1 by showing the conditions in Theorem F.2 hold under Assumptions SA-2, SA-3, and SA-4. We impose the following three conditions

**Condition 10** For $\overline{\alpha}$ some convex combination of $\alpha$ and $\alpha_0$ and for $L_{ij}$, there exists functions $M_{ij}^{(\phi_{1w})}()$, $M_{ij}^{(\phi_{2w})}()$, $M_{ij}^{(\phi_{1w})}()$, $M_{ij}^{(\phi_{2w})}()$, $M_{ij}^{()}$, $M_{ij}^{()}$, and $M_{ij}^{()}$ such that

$$\frac{d\ell_{ij}(W, \overline{\alpha})}{d\alpha} =$$

$$M_{ij}^{(\phi_{1w})}(W, \overline{\alpha}) (\phi_{1s} - \phi_{1s0}) + M_{ij}^{(\phi_{2w})}(W, \overline{\alpha}) (\phi_{1w0} - \phi_{1w0}) + M_{ij}^{(\phi_{2w})}(W, \overline{\alpha}) (\phi_{2s} - \phi_{2s0})$$

$$+ M_{ij}^{(\phi_{2w})}(W, \overline{\alpha}) (\phi_{2w} - \phi_{2w0}) + M_{ij}^{()}(W, \overline{\alpha}) (\mu - \mu_0) + M_{ij}^{()}(W, \overline{\alpha}) X_1 (\beta_1 - \beta_10)$$

$$+ M_{ij}^{()}(W, \overline{\alpha}) X_2 (\beta_2 - \beta_20) + M_{ij}^{()}(W, \overline{\alpha}) (h - h_0)$$

for all $i, j = 0, 1$.

**Condition 11**

(i) $\sup_{\alpha \in A_n, ||\alpha - \alpha_0||_s = o(1)} |M_{ij}^{()}(W, \alpha)| \leq C_{11}(W) < \infty$, $\forall t \in \{ \phi_{1s}, \phi_{1w}, \phi_{2w}, \phi_{2w}, \mu, \beta_1, \beta_2, h \}$,

(ii) $\inf_{\alpha \in A_n, ||\alpha - \alpha_0||_s = o(1)} |M_{ij}^{()}(W, \alpha)| \geq C_{12}(W) > 0$, $\forall t \in \{ \phi_{1s}, \phi_{1w}, \phi_{2w}, \phi_{2w}, \mu, \beta_1, \beta_2, h \}$, $\forall i, j = 0, 1$;

**Condition 12**

$$\sup_{\alpha \in A_n, ||\alpha - \alpha_0||_s = o(1)} |M_{ij}^{()}(W, \alpha) - M_{ij}^{()}(W, \alpha_0)| = C_{2}(W) \| \alpha - \alpha_0 \|_s$$

$\forall t \in \{ \phi_{1s}, \phi_{1w}, \phi_{2w}, \phi_{2w}, \mu, \beta_1, \beta_2, h \}$, $\forall i, j = 0, 1$.

These three conditions are not difficult to verify for the game models with IS-A and IS-2A from the arguments in Section J. These three conditions are sufficient to verify the conditions (CC2) and (CC3).

### F.3.1 Proof of Proposition 5.1

The condition (CC1) is directly assumed by Assumption SA-4 (i). Condition 10 implies that the pathwise derivative of $l(\cdot, \cdot)$ is well defined. Condition 10-11 implies that $l(\cdot, \cdot)$ satisfies a Lipschitz condition in $\alpha \in A$. Condition 12 is useful to provide some regularities on the difference of the pathwise derivatives of $l(\cdot, \cdot)$. We first verify the condition (CC2). Using the mean value theorem, we have $l(y_i | w_i, \alpha) - l(y_i | w_i, \alpha_0) = \frac{dl(y_i | w_i, \alpha)}{da} [\alpha - \alpha_0]$ where $\tilde{\alpha}$ lies between $\alpha$ and $\alpha_0$. It follows that

$$E \left[ (l(y_i | w_i, \alpha) - l(y_i | w_i, \alpha_0))^2 \right] = E \left[ \left( \frac{dl(y_i | w_i, \alpha)}{da} [\alpha - \alpha_0] \right)^2 \right]$$
from which we conclude \( E \left[ (\ell(y_i|w_i, \alpha) - \ell(y_i|w_i, \alpha_0))^2 \right] \geq \|\alpha - \alpha_0\|_2^2 \) since Condition 10-12 implies that
\[
\left| \frac{dL(y_{ij}|w_{ij}, \tilde{\alpha})}{d\alpha} [\alpha - \alpha_0] - \frac{dL(y_{ij}|w_{ij}, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right| = O(\|\alpha - \alpha_0\|_2^2) \text{ with } \|\alpha - \alpha_0\|_s = o(1).
\]

Thus, the condition (CC2) holds immediately.

Now we verify the condition (CC3). Note if \( 0 < C \leq a \leq b \), then \( |\ln a - \ln b| \leq |a - b|/C \). Because \( L_i(W, \alpha) \), \( \forall i, j = 0, 1 \) is bounded away from zero and one for all \( \alpha \in \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_2 \leq \varepsilon\} \) for all small \( \varepsilon > 0 \), we have
\[
|\ln L_i(W, \alpha) - \ln L_i(W, \alpha_0)| = |L_i(W, \alpha) - L_i(W, \alpha_0)| / C \text{ for all } i, j = 0, 1.
\]

This fact and the fact that \( L_i(W, \alpha) \) satisfies the Lipschitz condition for all \( i, j = 0, 1 \) (Condition 10-11) implies that
\[
|l(Y|W, \alpha) - l(Y|W, \alpha_0)| \leq C(W) \|\alpha - \alpha_0\|_s \quad (61)
\]
with \( E \left[ C(W)^2 \right] < \infty \). By Theorem 1 of Gabushin (1967) (for integer \( \nu_1 > 0 \)) or Lemma 2 in Chen and Shen (1998) (for \( \nu_1 > 0 \), any positive number), we have
\[
\|\alpha - \alpha_0\|_s \leq C_1 \|\alpha - \alpha_0\|_2^{2\nu_1/d_x + 1}. \quad (62)
\]

From (61) and (62), we see that the condition (CC3) is satisfied with \( s = 2\nu_1/d_x / (2\nu_1 + 1) \) and \( U(w_i) = C_1 C(W) \). We have verified all the conditions in Theorem F.2.

The next step is to derive the convergence rate depending on the choice of sieves. For the sieve \( \mathcal{A}_n = \Theta \times \mathcal{H}_n \) with \( \mathcal{H}_n \) defined in (8), we have
\[
\|\Pi_n \alpha_0 - \alpha_0\|_s = O \left( K_n^{-\nu_1/d_x} \right) \quad (63)
\]
by Lorentz (1986). Now we need to calculate \( \delta_n \) that solves (60). Now let \( C = \sqrt{E \left[ (U(w_i))^2 \right]} \) and \( u_h \equiv \sup_{h \in \mathcal{H}_n} \|h\|_\infty \), then for all \( 0 < \delta \leq \delta < 1 \), we have
\[
H[\varepsilon, \mathcal{F}_n, \|\cdot\|_{L_2}] \leq \log N \left( \frac{\varepsilon}{C}, \mathcal{H}_n, \|\cdot\|_\infty \right) \leq \text{const} \times K_n \times \log \left( 1 + \frac{4u_h}{\varepsilon} \right)
\]
by Lemma 2.5 in van der Geer (2000). Using this result, we obtain
\[
\frac{1}{n^\beta \delta_n^{2\beta}} \int_{|h|^{2\beta}} H[\varepsilon, \mathcal{F}_n, \|\cdot\|_{L_2}] d\varepsilon \leq C \frac{1}{n^{\beta \delta_n^2}} \int_{|h|^{2\beta}} K_n \times \log \left( 1 + \frac{4u_h}{\varepsilon} \right) d\varepsilon \leq C \frac{1}{n^{\beta \delta_n^2}} \sqrt{K_n} \delta_n \leq \text{const.} \quad (64)
\]
From the last inequality of (64), we conclude \( \delta_n \asymp \sqrt{K_n/n} \). We complete the proof by combining this result with (63). Finally, by letting \( \delta_n \asymp \|\Pi_n \alpha_0 - \alpha_0\|_s \) and \( K_n = n^\kappa \), we obtain the optimal rate with \( K_n = n^{1/(2\nu_1/d_x + 1)} \).
F.4 Asymptotic Normality

We derive the asymptotic normality of the structural parameters estimates using Theorem 4.3 in Chen (2005).

First, we let the sieve conditional ML estimator \( \hat{\alpha}_n \) converges to \( \alpha_0 \) at a rate faster than \( \eta_n \). Also let \( \varepsilon_n \) denote any sequence satisfying \( \varepsilon_n = o(n^{-1/2}) \) and \( \mu_n(g(Y,W)) = n^{-1} \sum_{i=1}^{n} \{ g(y_i, w_i) - E[g(y_i, w_i)] \} \) denote the empirical process indexed by the function \( g \). The following theorem show that the plug-in sieve conditional ML estimator \( f(\hat{\alpha}_n) \) achieves the \( \sqrt{n} \)-asymptotic normality.

**Theorem F.3** (Theorem 4.3 in Chen (2005))

Suppose that (AN1) (i) there is an \( \omega > 0 \) such that \( \left| f(\alpha) - f(\alpha_0) - \frac{df(\alpha_0)}{d\alpha} [\alpha - \alpha_0] \right| = O(\|\alpha - \alpha_0\|^\omega) \) uniformly over \( \alpha \in \mathcal{A} \) with \( \|\alpha - \alpha_0\|_2 = o(1) \); (ii) \( \left\| \frac{df(\alpha_0)}{d\alpha} \right\|_2 < \infty \); (iii) there is a \( \Pi_n v^* \in \mathcal{A} \) such that \( \|\Pi_n v^* - v^*\|_2 \times \|\hat{\alpha}_n - \alpha_0\|_2 = o_p(n^{-1/2}) \); (AN2) \( \sup_{\alpha \in \mathcal{A}, \|\alpha - \alpha_0\|_2 \leq \eta_n} \mu_n \left( l(Y|W, \alpha) - l(Y|W, \alpha + \varepsilon_n \Pi_n v^*) - \frac{d(\Pi_n v^*)}{da} \right) = o_p(\varepsilon_n^2) \); (AN3) \( K(\alpha_0, \hat{\alpha}_n) - K(\alpha_0, \alpha_0 + \varepsilon_n \Pi_n v^*) = o(\|\varepsilon_n \Pi_n v^*\|) + o(n^{-1/2}) \); (AN4) \( \mu_n \left( \frac{d(\Pi_n v^*)}{da} \right) = o_p(n^{-1/2}) \); (ii) \( \text{Var} \left( \frac{d(\Pi_n v^*)}{da} \right) = o(n^{-1/2}) \); (AN5) \( n^{1/2} \mu_n \left( \frac{d(\Pi_n v^*)}{da} \right) \rightarrow_d N(0, \sigma^2_{v^*}) \) with \( \sigma^2_{v^*} \equiv \text{Var} \left( \frac{d(\Pi_n v^*)}{da} \right) > 0 \) for iid data holds and \( \|\hat{\alpha}_n - \alpha_0\|^2 = o_p(n^{-1/2}) \). Then, for the sieve ML estimate \( \hat{\alpha}_n \) given in (14), we have \( \sqrt{n} (f(\hat{\alpha}_n) - f(\alpha_0)) \rightarrow_d N(0, \sigma^2_{v^*}) \).

Note that for statistical inference of the sieve plug-in estimate \( f(\hat{\alpha}_n) \), one needs a consistent estimator for \( \sigma^2_{v^*} \). For example, such estimators can be found in Andrews (1994b), Newey (1994), and Ai and Chen (2003). Now we are ready to prove Theorem 5.1 based on Theorem F.3.

F.4.1 Proof of Theorem 5.1

Now note that \( \frac{df(\alpha_0)}{d\alpha} = (\theta - \theta_0)' \lambda \) which implies that \( f(\alpha) - f(\alpha_0) - \frac{df(\alpha_0)}{d\alpha} [\alpha - \alpha_0] = 0 \) and hence the condition (AN1) (i) holds with \( \omega = \infty \). In addition, note

\[
\sup_{0 \neq \alpha - \alpha_0 \in \mathbf{V}} \frac{|f(\alpha) - f(\alpha_0)|^2}{\|\alpha - \alpha_0\|^2} = \sup_{0 \neq \alpha - \alpha_0 \in \mathbf{V}} \text{E} \left[ \frac{\left| \lambda'(\theta - \theta_0) \right|^2}{\left( \frac{d(\Pi_n v^*)}{da} \right)^2 \left( \frac{d(\Pi_n v^*)}{da} \right)^2} \right]
\]

\[
= \sup_{0 \neq (\theta - \theta_0, b) \in \mathbf{V}} \text{E} \left[ \frac{\lambda'(\theta - \theta_0)' \lambda}{\left( \frac{d(\Pi_n v^*)}{da} \right)^2 \left( \frac{d(\Pi_n v^*)}{da} \right)^2} \right]
\]

\[
= \lambda' \left( \text{E} \left[ D_{b^*}(Y, W)' D_{b^*}(Y, W) \right] \right)^{-1} \lambda
\]

which implies \( f(\alpha) = \lambda' \theta \) is bounded if and only if \( \text{E} \left[ D_{b^*}(Y, W)' D_{b^*}(Y, W) \right] \) is finite positive-definite, in which case we have \( v^* \in \mathbf{V} \) such that

\[
f(\alpha) - f(\alpha_0) \equiv \lambda' (\theta - \theta_0) = (v^*, \alpha - \alpha_0) \text{ for all } \alpha \in \mathcal{A}
\]

(65)

by (18) and the Riesz representation theorem. \( v^* \equiv (v^*_b, v^*_h) \in \mathbf{V} \) satisfies (65) with

\[
v^*_b = \left( \text{E} \left[ D_{b^*}(Y, W)' D_{b^*}(Y, W) \right] \right)^{-1} \lambda \text{ and } v^*_h = -b^* \times v^*_b.
\]
Thus, the condition (AN1) (ii) is satisfied under Assumption \textbf{SA-5} (ii). Assumption \textbf{SA-5} (iii) implies that we can find \( \Pi_n v^* \in \mathcal{A}_n \equiv \Theta \times \mathcal{H}_n \) with \( \mathcal{H}_n \) defined in (8) such that \( \| \Pi_n v^* - v^* \|_s = O\left(n^{-1/4}\right) \). Combining this with the condition \( \| \tilde{\alpha}_n - \alpha_0 \|_2 = o_p(n^{-1/4}) \) supported by Proposition 5.1, we obtain \( \| \Pi_n v^* - v^* \|_2 \times \| \tilde{\alpha}_n - \alpha_0 \|_2 = o_p(n^{-1/2}) \) with \( \nu_1/d > 1/2 \). This satisfies the condition (AN1) (iii). Next, we verify the condition (AN3).

Note that we have
\[
E \left[ \frac{d l(Y|W, \alpha_0)}{d \alpha} \right] = 0
\]
for any \( \alpha - \alpha_0 \) (it does not need to be in \( \nabla \)) because (i) the directional derivative of \( l(Y|W, \alpha) \) at \( \alpha_0 \) is well-defined and (ii) it is unconstrained maximization (see Shen (1997)). This is the zero expectation of score function like in a parametric ML. We can show (66) as follows. Denote \( \mathcal{L}(Y, W, \alpha) = \exp(l(Y|W, \alpha)) \).

\[
\begin{align*}
E \left[ \frac{d l(Y|W, \alpha_0)}{d \alpha} \right] &= E \left[ \frac{\partial \mathcal{L}(Y, W, \alpha_0)}{\partial (\alpha - \alpha_0)} \right] \\
&= \int \sum_y \frac{d \mathcal{L}(Y, W, \alpha_0)}{d (Y, W, \alpha_0)} \mathcal{L}(Y = y, W, \alpha_0) f_W(W) dW \\
&= \lim_{T \to 0} \frac{d}{dT} \int \sum_y \mathcal{L}(Y = y, W, \alpha_0 + \tau(\alpha - \alpha_0)) f_W(W) dW = \lim_{T \to 0} \frac{d}{dT} \int f_W(W) dW = 0
\end{align*}
\]

where the second equality holds since \( \mathcal{L}(Y = y, W, \alpha_0) = \Pr(Y = y|W, \alpha_0) \), by construction, for all \( y \in \{1, 1\}, (1,0), (0,1), (0,0) \}, the third equality holds by the interchangeability of integral and derivative and by definition of directional derivative, the fourth equality holds since \( \sum_y \mathcal{L}(Y = y, W, \alpha_0 + \tau(\alpha - \alpha_0)) = 1 \) regardless of \( \tau \) and \( \alpha \), and the last result holds since \( \int f_W(W) dW = 1 \).

From (66), it follows that
\[
E \left[ \frac{d l(Y|W, \alpha_0)}{d \alpha} \right] = 0
\]
and
\[
E \left[ \frac{d l(Y|W, \alpha_0)}{d \alpha} \right] = 0.
\]

We also need the following results to verify the condition (AN3).

For \( \alpha_3 \in \mathcal{A}_n \) such that \( \| \alpha_3 - \alpha_0 \|_s \leq \delta_n \) and for \( \alpha_1, \alpha_2 \in \mathcal{A}_n - \alpha_0 \), note that
\[
\begin{align*}
&\leq \left| E \left[ \frac{d l(Y|W, \alpha_0)}{d \alpha} \right] \right| \| \alpha_1 \|_s \| \alpha_2 \|_s \| \alpha_1 \|_s \leq C_1 \left( \| \alpha_3 - \alpha_0 \|_s \| \alpha_2 \|_s \| \alpha_1 \|_s \right\| \| \alpha_2 \|_s \| \alpha_1 \|_s \right)^{2\nu_1/d \nu_2 + 1}
\end{align*}
\]

where the first inequality uses the triangle inequality, the second inequality holds by Conditions 10-12, and the last result holds by Theorem 1 of Gabushin (1967) (when \( \nu_1/d > 0 \)) or by Lemma 2 in Chen and Shen (1998) for any \( \nu_1/d_2 > 1/2 \). Thus, the condition A4 (i) in Wong and Severini (1991) holds with \( \nu_1/d > 1 \). Note also that
\[
\begin{align*}
&\leq C_1 \left( \| \alpha_3 - \alpha_0 \|_s \| \alpha_2 \|_s \| \alpha_1 \|_s \right)^{2\nu_1/d \nu_2 + 1}
\end{align*}
\]
and hence the condition A4 (ii) in Wong and Severini (1991) holds with \( \nu_1/d_z > 1 \). Now consider

\[
K(\alpha_0, \hat{\alpha}_n) - K(\alpha_0, \hat{\alpha}_n + \varepsilon_n \Pi_n v^*) = E \left[ l(Y|W, \hat{\alpha}_n + \varepsilon_n \Pi_n v^*) - l(Y|W, \hat{\alpha}_n) \right] = E \left[ \frac{dl(Y|W, \hat{\alpha}_n + \varepsilon_n \Pi_n v^*)}{d\alpha} [\hat{\alpha}_n - \alpha_0, \pm \varepsilon_n \Pi_n v^*] \right] + o_p(n^{-1})
\]

by Assumption

\[
E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [\hat{\alpha}_n - \alpha_0, \pm \varepsilon_n \Pi_n v^*] \right] + o(n^{-1}) = \mp \varepsilon_n \times (\hat{\alpha}_n - \alpha_0, \Pi_n v^*) + o(n^{-1})
\]

where the first equality holds by definition of \( K(\cdot, \cdot) \), the second equality is using the mean value theorem with \( \tilde{\varepsilon}_n = o(n^{-1/2}) \), the third equality is obtained using the Taylor expansion, the fourth equality is obtained using

(i) \( E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [\pm \varepsilon_n \Pi_n v^*] \right] = \pm \varepsilon_n E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [v_n^*] \right] = 0 \) by (68) and using (69) with \( \varepsilon_n = o(n^{-1/2}) \),

the last result is from the fact that \( \alpha_1, \alpha_2 = -E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [\alpha_1, \alpha_2] \right] \) from (59) (see Wong and Severini (1991)). Therefore, the condition (AN3) holds since \( \varepsilon_n \) is arbitrary.

Note that the condition (AN4) (ii) immediately holds since \( E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [\Pi_n v^*] \right] = 0 \) by (68). Define

\[
M^{(h)}(Y, W, \alpha_0) = \left\{ \begin{array}{l}
Y_1 Y_2 \frac{M^{(h)}(W, \alpha_0)}{L_1} + Y_1 (1 - Y_2) \frac{M^{(h)}(W, \alpha_0)}{L_2} \\
+ (1 - Y_1) Y_2 \frac{M^{(h)}(W, \alpha_0)}{L_3} + (1 - Y_1) Y_2 \frac{M^{(h)}(W, \alpha_0)}{L_4}
\end{array} \right\}.
\]

From this, it follows that

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{dl(y_i|w_i, \alpha_0)}{d\alpha} [\Pi_n v^* - v^*_n] = \frac{1}{n} \sum_{i=1}^{n} M^{(h)}(y_i, w_i, \alpha_0) (\Pi_n v^*_n(z_i) - v^*_h(z_i))
\]

and thus the condition (AN4) (i) follows using the Chebyshev inequality from \( 0 = E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [\Pi_n v^* - v^*] \right] \) by (67), (68), and from \( \|\Pi_n v^*_n - v^*_h\|_\infty = o(n^{-1/4}) \) by Assumption SA-5 (iii). Now note that

\[
E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*_n] \right] = \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*_h] + \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*_h]
\]

\[
= \frac{dl(Y|W, \alpha_0)}{d\alpha} - M^{(h)}(Y, W, \alpha_0) \left( E [D_{b^r}(Y, W) D_{b^r}(Y, W)] \right)^{-1} \lambda
\]

\[
= D_{b^r}(Y, W) \left( E [D_{b^r}(Y, W) D_{b^r}(Y, W)] \right)^{-1} \lambda
\]

where the second equality is obtained using (70) and the definitions of \( v^*_n \) and \( v^*_h \). This implies that the condition (AN5) is satisfied by the Lindberg and Levy CLT since \( E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*] \right] = 0 \) and

\[
E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*] \right] = E \left[ \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*] \right] = \lambda \left( [D_{b^r}(Y, W) D_{b^r}(Y, W)] \right)^{-1} \lambda
\]

by Assumption SA-5 (ii). Now note \( \sigma^2_{\varepsilon^*} = \text{Var} \left( \frac{dl(Y|W, \alpha_0)}{d\alpha} [v^*] \right) = \lambda \left( [D_{b^r}(Y, W) D_{b^r}(Y, W)] \right)^{-1} \lambda' \). Therefore, the conclusion

\[
\sqrt{n} (\hat{\theta}_n - \theta_0) \rightarrow^d N \left( 0, (E [D_{b^r}(Y, W) D_{b^r}(Y, W)])^{-1} \right)
\]
follows since $\lambda$ is arbitrary with $\lambda \neq 0$.

Now the condition (AN2) remains to be proved. Note that the condition (AN2) is implied by

$$\sup_{\alpha \in A_n, ||\alpha - \alpha_0|| \leq \delta_n} \left( \frac{d l(y_i^1, w_i, \alpha)}{da} [\Pi_n v^*] - \frac{d l(y_i^1, w_i, \alpha_0)}{da} [\Pi_n v^*] \right) = o_p(n^{-1/2}).$$

(71)

Let $F = \{ \frac{d l(y_i^1, w_i, \alpha)}{da} [\Pi_n v^*] : \alpha \in A \}$. Condition 12 implies that $\frac{d l(y_i^1, w_i, \alpha)}{da} [\Pi_n v^*]$ satisfies the Lipschitz condition with respect to $\alpha$ and the metric $|| \cdot ||_s$. Thus, $\frac{d l(y_i^1, w_i, \alpha)}{da} [\Pi_n v^*]$ satisfies the condition (3.1) of Theorem 3 in Chen, Linton, and van Keilegom (2003). Note also that $\Theta$ is compact and $\mathcal{H}$ is a subset of a Hölder space. Thus, from the Lipschitz condition and the remark 3 (ii) of Chen, Linton, and van Keilegom (2003), it follows that

$$\int_0^\infty \sqrt{\log N} \left( \varepsilon, F, \||\cdot\|_{L_2(P_0)} \right) dz < \infty$$

by the proof of Theorem 3 in Chen, Linton, and van Keilegom (2003). Now note

$$E \left[ \left( \frac{d l(y_i^1, w_i, \alpha)}{da} [\Pi_n v^*] - \frac{d l(y_i^1, w_i, \alpha_0)}{da} [\Pi_n v^*] \right)^2 \right]_E 
\leq C \cdot E \left[ \left( \frac{d l(y_i^1, w_i, \alpha)}{da} [\Pi_n v^*] - \frac{d l(y_i^1, w_i, \alpha_0)}{da} [\Pi_n v^*] \right)_E \right] 
\times \sup_{w \times y \in S(W) \times S(Y), ||\alpha - \alpha_0|| \leq \delta_n, \alpha \in A_n} \left\| \frac{d l(y_i^1, w, \alpha)}{da} [\Pi_n v^*] - \frac{d l(y_i^1, w, \alpha_0)}{da} [\Pi_n v^*] \right\|_E \to 0$$

as $||\alpha - \alpha_0||_s \to 0$

(73)

where the last result holds by Condition 12. Therefore, applying Lemma 1 in Chen, Linton, and van Keilegom (2003), we find that (71) is true by (72) and (73). This completes the proof.

**F.4.2 Proof of Proposition 5.2**

Similarly with the proof of Theorem 5.1 in Ai and Chen (2003). We can prove Proposition 5.2. We note that $\sum_{i=1}^n \left( \frac{d l(y_i^1, w_i, \alpha)}{da} - \frac{d l(y_i^1, w_i, \alpha_0)}{da} \right) [b_j]_s$ is globally convex in $b_j$ and hence the solution of (20), $\hat{b}_j^*$, must be bounded by $\left\| \hat{b}_j^* \right\|_s \leq C$. Thus, we only care about the subset $\{ b \in \mathbb{B} : ||b||_s \leq C \}$ in the following discussion.

Note that uniformly over $b_j \in \mathcal{H}_n, ||b_j||_s < C$, we have

$$\frac{1}{n} \sum_{i=1}^n (D_{b_j} (Y_i, W_i, \alpha_0))^2 = \frac{1}{n} \sum_{i=1}^n (D_{b_j} (Y_i, W_i, \alpha_0))^2 + o_p(1)$$

(74)

since $D_{b_j} (Y, W, \alpha) = D_{b_j} (Y_i, W_i, \alpha_0) + o_p(1)$ uniformly over $||\alpha - \alpha_0||_s = o(1)$ by Condition 12 and $||b_j||_s < C$. Thus, it suffices to show that $\left\| \hat{b}_j^*(\cdot) - b_j^*(\cdot) \right\|_s = o_p(1)$ which implies

$$D_{b_j} (Y, W, \alpha_0) = D_{b_j} (Y_i, W_i, \alpha_0) + o_p(1).$$

(75)

Combining (74) and (75), we have

$$\frac{1}{n} \sum_{i=1}^n \left( D_{b_j} (Y_i, W_i, \alpha_0) \right)^2 = \frac{1}{n} \sum_{i=1}^n \left( D_{b_j} (Y_i, W_i, \alpha_0) \right)^2 + o_p(1)$$

from which the claim follows.

Now it remains to show that $\left\| \hat{b}_j^*(\cdot) - b_j^*(\cdot) \right\|_s = o_p(1)$ which is satisfied by Condition 12, Assumption SA-5, $||\alpha_n - \alpha_0||_s = o(1)$, and Lemma A.1 in Newey and Powell (2003).
G  Utilizing the Mixing Distribution of Public Signals

Until now, we have ignored the possibility that we may estimate the finite mixing distribution of $Z$ directly from the data or we may combine this with the estimation procedures we have considered in previous sections (for example, we may estimate the sieve ML estimator that maximizes (13)). Recall that $Z$ follows a mixing distribution with the density

$$f_Z(z) = p f_{(st)}(z) + (1 - p) f_{(we)}(z).$$

Suppose $f_{(st)}(\cdot)$ and $f_{(we)}(\cdot)$ belong to some parametric family $\{ f_{(i)}(\cdot, \pi) \}$. Then, we can estimate $p$ and $\pi$ consistently using the EM algorithm as suggested in the literature (Everitt and Hand (1981), Titterington, Smith, and Makov (1985)). For a comprehensive treatment of mixture models, one can refer to Lindsay (1995). First, we briefly review the EM procedure tailored to our problem, following Arcidiacono and Bailey Jones (2003). Denote the unconditional likelihood of $z_i$ as $f_Z(z; \pi, p) = p f_{(st)}(z; \pi) + (1 - p) f_{(we)}(z; \pi)$. From Bayes’ theorem, $P(t_s|z_i; \pi, p)$, the probability that $i$ is from the strong type conditional on $z_i$ will be

$$P(t_s|z_i; \pi, p) = \frac{p f_{(st)}(z_i; \pi)}{f_Z(z_i; \pi, p)}.$$  \hspace{1cm} (76)

If we maximize the sample average of the log of unconditionally-type-averaged likelihood given by $L(\pi, p) = \frac{1}{n} \sum_{i=1}^{n} \log(f_Z(z_i; \pi, p))$, we will obtain

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} P(t_s|z_i; \hat{\pi}, \hat{p}).$$ \hspace{1cm} (77)

The maximum likelihood estimator $\hat{\pi}$ must solve

$$\frac{1}{n} \sum_{i=1}^{n} \left( P(t_s|z_i; \hat{\pi}, \hat{p}) \frac{\partial \log f_{(st)}(z_i; \pi)}{\partial \pi} + (1 - P(t_s|z_i; \hat{\pi}, \hat{p})) \frac{\partial \log f_{(we)}(z_i; \pi)}{\partial \pi} \right) = 0$$

which is the first order condition of

$$\hat{\pi} = \text{argmax} \frac{1}{n} \sum_{i=1}^{n} \left( P(t_s|z_i; \hat{\pi}, \hat{p}) \log f_{(st)}(z_i; \pi) + (1 - P(t_s|z_i; \hat{\pi}, \hat{p})) \log f_{(we)}(z_i; \pi) \right).$$  \hspace{1cm} (78)

This means that $\hat{\pi}$ maximizes both the sample average of the log of unconditionally- and conditionally-type averaged log-likelihood. In the EM algorithm, starting from an initial value of $\pi^{(1)}$ and $p^{(1)}$, we update the conditional probability using (76) as $P(t_s|z_i; \pi^{(1)}, p^{(1)})$ in the “E” step. In the “M” step, using (77) and (78), we obtain $\pi^{(2)}$ and $p^{(2)}$. Asymptotic properties of this estimator are well-known in the literature.

However, in our problem, the mixing distribution is nonparametrically specified where $f_{(st)}(\cdot)$ and $f_{(we)}(\cdot)$ are infinite dimensional parameters. The identification of the mixture with nonparametric component functions for types has not been studied well with few exceptions (Kitamura (2004)) while Heckman and Singer (1994) provides an important study for the semiparametric models treating the mixing probability nonparametrically. Here we propose a pseudo EM algorithm that combines (76), (77), and the estimation procedures considered in the previous sections. The key idea is that for updating the mixing probability (like in step (77)), other than the prior mixing probability, we only need the ratio of densities for two component distributions, not individual densities.

In the first step, from (10), obtain an initial estimator of $\hat{h}_{(st)}^{(1)}$. Using these, obtain

$$P(t_s|z_i; \hat{h}_{(st)}^{(1)}) = \frac{\exp(\hat{h}_{(st)}^{(1)}(z_i))}{1 + \exp(\hat{h}_{(st)}^{(1)}(z_i))} \text{ and } \hat{p}_{(st)}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} P(t_s|z_i; \hat{h}_{(st)}^{(1)})$$ \hspace{1cm} (79)
from (5) and (24). These replace the steps (76) and (77), respectively. Now note that a directional derivative of the log likelihood function in (13) with respect to \((\theta, h^o)\) at the direction of \((\theta, h^o)\) equals to

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{p \exp(h^o(z_i))}{p \exp(h^o(z_i)) + (1 - p)(\bar{h}^o - h^o)} \frac{d l(y_i, \theta, \log \left( \frac{p}{1-p} \right) + h^o(z_i))}{d(\theta, h^o)} \left[ (\hat{\theta}, \bar{h}^o) - (\theta, h^o) \right]
\]

and note that

\[
\frac{p \exp(h^o(z_i))}{p \exp(h^o(z_i)) + (1 - p)(\bar{h}^o + h^o(x_i))} = \frac{\exp(h(z_i))}{1 + \exp(p/(1-p) + h^o(x_i))}. \quad \text{Similarly with (78), this suggests that we can update \(\theta\) and \(h^o\) by solving}
\]

\[
\left( \hat{\theta}^{(2)}_n, \bar{h}^{(2)}_n \right) = \arg \max_{(\theta, h^o) \in \Theta \times \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(h^o(z_i))}{1 + \exp(h^o(z_i))} h^o(z_i) + \frac{1}{n} \sum_{i=1}^{n} l(y_i, \theta, \log \left( \frac{\hat{p}^o_n}{1 - \hat{p}^o_n} \right) + h^o(z_i)).
\]

Let \(\bar{h}^{(2)}_n(z) = \log \left( \hat{p}^o_n/(1 - \hat{p}^o_n) \right) + \hat{h}^{(2)}_n\). Using this \(\hat{h}^{(2)}_n\), obtain \(\hat{p}^{(2)}_n\) from (79). Iterate this procedure until convergence is obtained.

### H Asymptotic Normality of \(p_0\): Type Distribution

We may prove Proposition 5.3 by showing all the conditions in Theorem 2 of Chen, Linton, and van Keilegom (2003) hold. Most of conditions will be satisfied trivially. Here we directly prove Proposition 5.3 since it is a simple case of Chen, Linton, and van Keilegom (2003).

#### H.1 Proof of Proposition 5.3

We let \(M(h) = \int_{S(Z)} L(h(z)) f_Z(z) dz\) and \(M_n(h) = \frac{1}{n} \sum_{i=1}^{n} L(h(Z_i))\). We have

\[
\sqrt{n} (\hat{p}_n - p_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (L(\bar{h}_n(Z_i)) - L(h_0(Z_i))) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (L(h_0(Z_i)) - E(L(h_0(Z_i)))) = \sqrt{n} \left( M_n(\bar{h}_n) - M_n(h_0) \right) + \sqrt{n} \left( M_n(h_0) - M(h_0) \right)
\]  

(80)

Now let \(\xi(h) = M_n(h) - M(h)\) be a stochastic process indexed by \(h \in \mathcal{H}\). Then, we obtain the following stochastic equicontinuity such that for any positive sequence \(\delta_n = o(1)\),

\[
\sup_{\|h - h_0\| \leq \delta_n} |\xi(h) - \xi(h_0)| = o_p(n^{-1/2})
\]  

(81)

by applying Lemma 1 of Chen, Linton, and van Keilegom (2003) after establishing that (a) \(\{v(h) \equiv L(h) - M(h_0) : h \in \mathcal{H}\}\) is a Donsker class and that (b) \(E \left[ (v(h_1) - v(h_2))^2 \right] \to 0 \) as \(\|h_1 - h_2\| \to 0\) noting \(E [v(h_0)] = 0\) by construction. Consider that \(\{v(h) \equiv L(h) - M(h_0) : h \in \mathcal{H}\}\) is a subset of \(\mathcal{L}_C^2(S(Z))\) and \(\mathcal{L}_C^2(S(Z))\) is a Donsker class by Theorem 2.5.6 of van der Vaart and Wellner (1996). Thus, the condition (a) is satisfied. Now note

\[
E \left[ (v(h_1) - v(h_2))^2 \right] = E \left[ (L(h_1) - L(h_2))^2 \right] \leq E \left[ \|L(h_1) - L(h_2)\| \sup_z L'(\bar{h}(z)) \|h_1 - h_2\| \right] \leq \frac{1}{2} E \left[ \|L(h_1) - L(h_2)\| \|h_1 - h_2\| \right].
\]
since \( L' = L(1 - L) \leq 1/4 \) uniformly where the second equality is obtained by applying the mean value theorem and thus, the condition (b) is satisfied. Therefore, \((81)\) holds. Now consider

\[
\sqrt{n} \left( M_n(h_n) - M_n(h_0) \right) = \sqrt{n} \left( M(h_n) - M(h_0) \right) + \sqrt{n} \left( M_n(h_n) - M(h_n) - M_n(h_0) + M(h_0) \right) = \sqrt{n} \left( M(h_n) - M(h_0) \right) + o_p(1) \tag{82}
\]

where the last result is obtained by \((81)\) and \(\left\| \hat{h}_n - h_0 \right\|_\infty = o_p(1)\). Now applying the mean value theorem, we have

\[
\sqrt{n} \left( M(h_n) - M(h_0) \right) = \sqrt{n} \int_{S(Z)} L'(\hat{h}_n(z)) \left( \hat{h}_n(z) - h_0(z) \right) f_Z(z) dz \\
= \sqrt{n} \int_{S(Z)} L'(h_0(z)) \left( \hat{h}_n(z) - h_0(z) \right) f_Z(z) dz + \sqrt{n} \int_{S(Z)} \left( L'(\hat{h}_n(z)) - L'(h_0(z)) \right) \left( \hat{h}_n(z) - h_0(z) \right) f_Z(z) dz \tag{83}
\]

where \(\hat{h}_n\) lies between \(\hat{h}\) and \(h_0\). Applying the mean value theorem again, we obtain (noting \(L' = L(1 - L)\))

\[
\left| L'(\hat{h}_n) - L'(h_0) \right| = \left| L' \left( \hat{h}_n \right) \left( 1 - L \left( \hat{h}_n \right) \right) - L \left( h_0 \right) \left( 1 - L \left( h_0 \right) \right) \middle| \right| \\
= \left| \left( 1 - L \left( \hat{h}_n \right) - L \left( h_0 \right) \right) L' \left( \hat{h}_n \right) \left( \hat{h}_n - h_0 \right) \middle| \right| \leq \frac{1}{4} \left| \hat{h}_n - h_0 \right|
\]

where \(\tilde{h}_n\) lies between \(\hat{h}_n\) and \(h_0\). It follows that

\[
\sqrt{n} \int_{S(Z)} \left( L'(\tilde{h}_n(z)) - L'(h_0(z)) \right) \left( \tilde{h}_n(z) - h_0(z) \right) f_Z(z) dz \leq \frac{1}{4} \sqrt{n} \left\| \tilde{h}_n - h_0 \right\|_\infty \left\| \hat{h}_n - h_0 \right\|_\infty = o_p(1) \tag{84}
\]

by the condition (i). Thus, we find

\[
\sqrt{n} \left( M(h_n) - M(h_0) \right) = \sqrt{n} \int_{S(Z)} L'(h_0(z)) \left( \hat{h}_n(z) - h_0(z) \right) f_Z(z) dz + o_p(1) \tag{85}
\]

from \((83)\) and \((84)\). From \((80)\), \((82)\), and \((85)\), the claim follows by the condition (ii).

**H.2 Proof of Proposition 5.4**

We let \(M_n^*(h) = \frac{1}{n} \sum_{i=1}^n L(h(Z_i))\). First, we note that the following condition holds by Giné and Zinn (1990),

\[
\sup_{\left\| h - h_0 \right\|_\infty \leq \delta_n} |\xi^*(h) - \xi^*(h_0)| = o_P(1) \tag{86}
\]

where \(\xi^*(h) = M_n^*(h) - M_n(h)\). Note

\[
\sqrt{n} (\hat{p}_n - \hat{p}_n) = \sqrt{n} \left( M_n^*(\hat{h}_n) - M_n^*(\hat{h}_n) \right) + \sqrt{n} \left( M_n^*(\hat{h}_n) - M_n(h_0) \right) \tag{87}
\]

and we have,

\[
\sqrt{n} \left( M_n^*(\hat{h}_n) - M_n(h_0) \right) = \sqrt{n} \left( M_n(h_0) - M(h_0) \right) + o_P(1) \tag{88}
\]
and let stochastic expansion of
To check the condition (ii) in Proposition 5.3 (or the condition (iv) of Proposition 5.4), we need to derive the
Combining (87), (88), (89), (90), and (91), we note the claim follows by the condition (iv).
by the conditions (i), (ii), and (iii). Now similarly with (85), we can show that
where the last equality is obtained using (86) and by the conditions (i) and (ii). Now consider
by the conditions (i), (ii), and (iii). Now similarly with (85), we can show that
Combining (87), (88), (89), (90), and (91), we note the claim follows by the condition (iv).

H.2.1 Stochastic Expansion of \( \hat{h}_n - h_0 \)

To check the condition (ii) in Proposition 5.3 (or the condition (iv) of Proposition 5.4), we need to derive the stochastic expansion of \( \hat{h}_n - h_0 \) for a particular estimator. Here we provide a sketch of such an expansion for the sieve conditional ML estimator. Deriving the stochastic expansion of a nonparametric estimator obtained from a highly nonlinear objective function is often difficult.

To facilitate this task, we define a pseudo true value of \( \theta \) and \( h \) such that
\[
\alpha_{0K} \equiv (\theta_{0K}, h_{0K}) = \arg \max_{\theta \in \Theta, h = R^K(\cdot) \pi \in \mathcal{H}_n} Q_K(\theta, h) \equiv E[(Y_i | W_i, \theta, h(Z_i))]
\] (92)
and let \( h_{0K}(\cdot) = R^K(\cdot)\pi_{0K} \). Similarly, we let \( \hat{h}_n(\cdot) = R^K(\cdot)\hat{\pi}_n \). Then,
\[
\hat{h}_n(\cdot) - h_{0K}(\cdot) = R^K(\cdot)'(\hat{\pi}_n - \pi_{0K}) \text{ with } K = K(n).
\] (93)

Define
\[
\Psi(h) = \int_{S(Z)} L(h(z))(1 - L(h(z)))R^K(z)dF_Z(z).
\] (94)

Now suppose that
\[
\Psi(h_0)'(\hat{\pi}_n - \pi_{0K}) = \frac{1}{n} \sum_{i=1}^n \psi(\alpha_{0K}, Y_i, W_i) + o_p(n^{-1/2}),
\] (95)

\[
E[\psi(\alpha_{0K}, Y_i, W_i)] = o(1), \quad E[\|\psi(\alpha_{0K}, Y_i, W_i)\|_2^2] < \infty. \quad \text{Further suppose,}
\]
\[
\|h_{0K} - h_0\|_\infty = o \left( n^{-1/2} \right).
\] (96)
Then, we have
\[
\sqrt{n} \int_{S(Z)} L(h_0)(1 - L(h_0)) \left( \hat{h}_n - h_0 \right) dF_Z \\
= \sqrt{n} \int_{S(Z)} L(h_0)(1 - L(h_0)) \left( \hat{h}_n - h_{0K} \right) dF_Z + \sqrt{n} \int_{S(Z)} L(h_0)(1 - L(h_0)) (h_{0K} - h_0) dF_Z \\
= \sqrt{n} \Psi(h_0)' \left( \hat{\pi}_n - \pi_{0K} \right) + \sqrt{n} \int_{S(Z)} L(h_0)(1 - L(h_0)) (h_{0K} - h_0) dF_Z = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\alpha_{0K}, Y_i, W_i) + o_p(1)
\]
where the second equality is obtained from (93) and (94) and the third equality is obtained from (95) and (96) noting \(L(\cdot)(1 - L(\cdot)) \leq 1/4\) uniformly. From this, it follows
\[
V_p = \lim_{K \to \infty} E \left[ \left( \Psi(\alpha_{0K}, Y_i, W_i) + \varphi(Z_i) \right) \left( \Psi(\alpha_{0K}, Y_i, W_i) + \varphi(Z_i) \right)' \right]
\]
where \(\varphi(Z_i) = L(h_0(Z_i)) - E[L(h_0(Z_i))]\). Therefore, to verify the condition (ii) of Proposition 5.3 holds, it suffices to show (95) and (96) are true. (95) can be shown using the first order conditions of (10) and (92). For (96), we note we often find \(\|h_{0K} - h_0\|_\infty = \zeta(K) K^{-\nu_1/2d_Z}\) in the literature (Hirano, Imbens, and Ridder (2003)) where \(\zeta(K) = \sup_{z \in S(Z)} \|R^K(z)\|_E\).

I Set Estimation of the Type Distribution (Proposition 6.1)

Here we derive the consistency for the set estimator of the type distribution parameter. We employ the Hausdorff metric measuring the distance between two sets whose elements are \(\dim(\cdot)\)-vectors. For two such sets \(A\) and \(B\), let the maximum distance between any points in \(A\) and \(B\) be given by \(\rho(A|B) = \sup_{a \in A} \rho(a|B)\) where \(\rho(a|B) = \inf \{\|a - b\|_E : b \in B\}\). Then, the Hausdorff metric is defined by \(d(A, B) = \max\{\rho(A|B), \rho(B|A)\}\).

We prove Proposition 6.1 similarly with the proof of Theorem 2 in Andrews, Berry, and Jia (2004).

**Proof.** We let \(p_n(\theta) = \frac{1}{n} \sum_{i=1}^n L(h(z_i, \delta))\) where \(\theta = (\hat{\theta}, \delta) \in \Theta_n\) and let \(p(\theta) = E[L(h(Z, \delta))\) where \(\theta = (\hat{\theta}, \delta) \in \Theta_+\). Now note that for any \(\varepsilon > 0\),
\[
P(d(\hat{\Theta}_n, \Theta_+) > \varepsilon) \leq P(p(\hat{\Theta}_n|\mathcal{P}_+) > \varepsilon) + P(p(\mathcal{P}_+|\hat{\Theta}_n) > \varepsilon) \leq \varepsilon + \zeta
\]
from the definition of \(d(\cdot, \cdot)\). Let \(\zeta = P(d(\hat{\Theta}_n, \Theta_+) > \varepsilon)\) for any \(\varepsilon > 0\) and consider
\[
P(\rho(\hat{\Theta}_n|\mathcal{P}_+) > \varepsilon) \leq P(\rho(\mathcal{P}_+|\hat{\theta}_+) > \varepsilon, d(\hat{\Theta}_n, \Theta_+) \leq \varepsilon) + \zeta
\]
where the equality holds by definitions of \(\rho(\cdot, \cdot), \hat{\Theta}_n,\) and \(\mathcal{P}_+.\) Now consider that if \(d(\hat{\Theta}_n, \Theta_+) \leq \varepsilon\), then \(\rho(\hat{\Theta}_n|\Theta_+) \leq \varepsilon\) and for any \(\hat{\theta} \in \hat{\Theta}_n\), there exists \(\hat{\theta}_+ \in \Theta_+\) such that \(\|\hat{\theta} - \hat{\theta}_+\|_\ast \leq \varepsilon\). It follows that for any \(\hat{\theta} \in \hat{\Theta}_n\) such that \(\|\hat{\theta} - \hat{\theta}_+\|_\ast \leq \varepsilon\), we have
\[
p_n(\hat{\theta} - \hat{\theta}_+) \leq \frac{1}{n} \sum_{i=1}^n \left( L(h(Z_i, \delta)) - L(h(Z_i, \delta_+)) \right) + \frac{1}{n} \sum_{i=1}^n \left( L(h(Z_i, \delta_+)) - E[L(h(Z_i, \delta_+))] \right)
\]
\[
\leq \frac{1}{n} \sum_{i=1}^n \left( L(h(Z_i, \delta)) - L(h(Z_i, \delta_+)) \right) + o_p(1) \leq C \cdot \varepsilon
\]
for sufficiently large \(n \geq \exists N\),
where the second equality is obtained applying the mean value theorem ($\delta$ lies between $\delta$ and $\delta_{++}$), the first inequality holds since $L(1 - L) \leq 1/4$ uniformly and since we bound the second term of (99) by $o_p(1)$ applying the LLN ($\{Z_i\}_{i=1}^n$ are iid and $|L(h)| < 1$ uniformly), the last result holds applying the mean value theorem since $\sup_{\delta \in \mathcal{D}, z \in \mathbb{S}(Z)} \left\| \frac{\partial h(z, \delta)}{\partial \delta} \right\|_E < \infty$. From this result, we have

$$
\sup_{\theta \in \Theta_n} \inf \{ |p_n(\theta) - p(\theta_+) : \theta_+ \in \Theta_+ \} \leq \sup_{\theta \in \Theta_n, \|\theta - \theta_+\|_e \leq \epsilon} |p_n(\theta) - p(\theta_+)| \leq C \cdot \epsilon
$$

for all sufficiently large $n \geq \exists N$. From this, it follows that for sufficient large $n$,

$$
P(\sup_{\theta \in \Theta_n} \inf \{ |p_n(\theta) - p(\theta_+) : \theta_+ \in \Theta_+ \} > \epsilon, d(\hat{\Theta}_n, \Theta_+) \leq \epsilon) \leq \epsilon
$$

and thus

$$
P(\rho(\hat{P}_n|P_+) > \epsilon) \leq \epsilon + P(d(\hat{\Theta}_n, \Theta_+) > \epsilon)
$$

from (98) and (100). An analogous argument provides

$$
P(\rho(P_+|\hat{P}_n) > \epsilon) \leq \epsilon + P(d(\hat{\Theta}_n, \Theta_+) > \epsilon).
$$

Combining (97), (101), and (102), we have $P(d(\hat{P}_n, P_+) > \epsilon) \leq 2\epsilon + 2P(d(\hat{\Theta}_n, \Theta_+) > \epsilon)$. This proves Proposition 6.1 since $\epsilon > 0$ is arbitrary and $d(\hat{\Theta}_n, \Theta_+) = o_p(1)$. ■

### J Smoothness of Conditional Probabilities

Here we note that for the conditional probabilities presented in Appendix C or E.3.2, the pathwise first and second derivatives are well-defined. This result is useful to verify Conditions 10-12 for the sieve conditional ML. It is easy to see that the pathwise derivatives are well-defined as long as $G_1(\cdot)$ and $G_2(\cdot)$ are continuously differentiable since the function $h(\cdot)$ appears only in $p(Z) = p(p, h) = \frac{\exp(\log(p/(1-p)) + h)}{1 + \exp(\log(p/(1-p)) + h)}$ and we have $\frac{dp(p, h)}{dh}[h_1 - h_2] = (1 - p(p, h))p(p, h)(h_1 - h_2)$. Therefore, for the conditional probabilities given in Appendix C or E.3.2, we have

$$
\frac{dP_{ij}(Y|W, \theta, p(p, h))}{dh}[h_1 - h_2] = M^{(h)}_{ij}(h_1 - h_2), \forall i, j = 0, 1,
$$

$$
\frac{d^2P_{ij}(Y|W, \theta, p(p, h))}{dh^2}[h_1 - h_0, h_2 - h_0] = M^{(h)}_{ij}(h_1 - h_0)(h_2 - h_0), \forall i, j = 0, 1, \text{ and}
$$

$$
\frac{d^2P_{ij}(Y|W, \theta, p(p, h))}{dhdt}[h_1 - h_2] = M^{(h)}_{ij}(t)(h_1 - h_0), \forall i, j = 0, 1 \text{ and any element } t \text{ of } \theta,
$$

where $M^{(h)}_{ij}$, $M^{(h)}_{ij}(h)$, and $M^{(h)}_{ij}(t)$ are some well-defined ordinary derivatives. The second thing to note is that those derivatives and other derivatives with respect to finite dimensional parameters are uniformly bounded by some constant since (i) $G_1(\cdot)$ and $G_2(\cdot)$ are continuously differentiable, (ii) the parameter space $\Theta$ is compact, (iii) $0 < p < 1$ and $0 < p(Z) < 1$, and (iv) $h(Z)$ appears only in $p(p, h)$. Therefore, the Lipschitz conditions for the conditional probabilities and the Lipschitz conditions for the pathwise first derivatives of the conditional probabilities are well satisfied. For example, in the case of the sieve conditional ML estimation, this implies that the Lipschitz conditions for the log likelihood and the Lipschitz conditions for the pathwise derivatives of the log likelihood are also well-defined.

Specific forms of derivatives for each model can be provided upon request.
References


Set Inference for Semiparametric Discrete Games

Kyoo il Kim*

School of Economics and Social Sciences
Singapore Management University

November 10, 2006

Abstract

We consider estimation and inference of parameters in discrete games allowing for multiple equilibria, without using an equilibrium selection rule. We do a set inference while a game model can contain infinite dimensional parameters. Examples can include signaling games with discrete types where the type distribution is nonparametrically specified and entry-exit games with partially linear payoffs functions. A consistent set estimator and a confidence interval of a function of parameters are provided in this paper. We note that achieving a consistent point estimation often requires an information reduction. Due to this less use of information, we may end up a point estimator with a larger variance and have a wider confidence interval than those of the set estimator using the full information in the model. This finding justifies the use of the set inference even though we can achieve a consistent point estimation. It is an interesting future research to compare these two alternatives: CI from the point estimation with the usage of less information vs. CI from the set estimation with the usage of the full information.

Keywords: Semiparametric Estimation, Set Inference, Infinite Dimensional Parameters, Inequality Moment Conditions, Signaling Game with Discrete Types

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

1 Introduction

The econometric modeling of game theories has been of significant interest over the last decade including simultaneous games with complete information (Bjorn and Vuong (1984, 1985), Bres-
nahan and Reiss (1990, 1991), Tamer (2003), Bajari, Hong and Ryan (2004)) or with incomplete information (Brock and Durlauf (2001, 2003), Seim (2002), Sweeting (2004), Aradillas-Lopez (2005)), dynamic games (Aguirregabiria and Mira (2003), Bajari, Benkard, and Levin (2003), Berry, Ovstrovsky, and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003)), and signaling games (Kim (2006)). Here we focus on static discrete games. For these games, depending on the equilibrium properties, a researcher can face with the issue of multiple equilibria. Several resolutions have been proposed such as imposing equilibrium selection rules\(^1\) and redefining the space of outcomes in a game\(^2\).

Alternatively, inspired by important work by Manski and co-authors (Manski (1990), Horowitz and Manski (1995), and Manski and Tamer (2002)) on bound analysis, some researchers have started to develop set inferences rather than a point estimation, without attempting to resolve the equilibrium selection (Sutton (2000), Ciliberto and Tamer (2003), and Andrews, Berry, Jia (2004) [ABJ]). In particular, we consider the model where some asymptotic inequalities may define a region of parameters rather than a single point in the parameter space. By definition, when there are multiple equilibria, there exist regions of unobservables that are consistent with the necessary conditions for more than one equilibrium. Therefore, the probability implied by the necessary condition for a given event is greater than or equal to the true probability of the event and a set inference including this paper utilizes these inequality conditions.

Another thing we note in the literature of the set inference is that parameters allowed in game models are only finite dimensional even though infinite dimensional parameter is naturally included in the model (For example, see Kim (2006)) or misspecification of a fully parametric model is concerned. This paper considers a set inference with infinite dimensional parameters. A consistent set estimator and a confidence interval are provided in the paper.

Our proposed set estimation and inference requires a consistent profiled estimator for the infinite dimensional parameters. An interesting case we note in this paper is that sometimes we can achieve a consistent point estimation of all the parameters including finite and infinite dimensional ones by losing some information in the model. For example, in Bresnahan and Reiss (1990, 1991), we disregard the information about which firms enter the market since we redefine the outcome space in terms of how many firms in the market. Due to this omitted information, we may end up point estimators whose variances are larger and thus have wider confidence intervals than those of the set estimator using the full information in the model. Comparison of these two will be also interesting.

The organization of this paper is as follows. Section 2 introduces the model we study. In Section

---


\(^2\)See Bresnahan and Reiss (1990, 1991).
3, we extend the set estimator of ABJ to the semiparametric case and provide two examples of such models. In Section 4, we show the consistency of the set estimator. In Section 5, we propose a set inference. We conclude in Section 6. Technical details and mathematical proofs are presented in Appendix.

2 Model

Let $Y_p$ be player $p$’s action (or strategy) and $X_p$ be a vector of player $p$’s characteristics for $p = 1, \ldots, p$. Let $\varepsilon_p$ be player $p$’s unobservable to econometricians, which is a part of player $p$’s payoff functions. We let $Y = (Y_1, Y_2, \ldots, Y_p)^t \in \mathbb{R}^{1+\cdots+b}$, $X = (X_1, X_2, \ldots, X_p)^t \in \mathbb{R}^k$ and let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^t \in \mathbb{R}^p$. Also let $\alpha_0 \equiv (\theta_0, h_0)$ be the parameters of interest in a game. Then, payoffs of the game are given as functions of $(Y, X, \varepsilon, \alpha_0)$ and an equilibrium of the game can be characterized comparing those payoffs for different actions or strategies. $\alpha_0$ consists of a vector of finite dimensional unknown parameters, denoted by $\theta_0$ and a vector of infinite dimensional unknown functions, denoted by $h_0$. Here we let $h_0$ be functions of $X$ or subset of $X$ alone without loss of generality since $Y$ is discrete. We allow $h_0$ can depend on the parameters $\theta$. We also let $S(W)$ denote the support of the distribution of random variable $W$.

Now denote $\Omega(Y = y, X = x, \alpha)$ to be the region of $\varepsilon$ under which $Y$ takes the value $y$ given $X = x$ and $\alpha$. To be precise, $\Omega(y, x, \alpha) \equiv \Omega(Y = y, X = x, \alpha) = \{\varepsilon|Y_1 = y_1, Y_2 = y_2, \ldots, Y_p = y_p\}$ given $X = x$ and $\alpha$. Then, the probability that the necessary conditions for $Y = y$ holds, denoted by $P(y|x, \alpha)$, will equal the probability that $\varepsilon$ belongs to $\Omega(y, x, \alpha)$ given $X = x$ and $\alpha$. Depending on the game of interest and the equilibrium property of the game, a researcher can construct $\Omega(y, x, \alpha)$ accordingly. Examples can be found in ABJ for entry games and Kim (2006) for signaling games. Note that

$$P(y|x, \alpha) = \Pr(\varepsilon \in \Omega(y, x, \alpha)) \quad (1)$$

and thus the analytic form of $P(y|x, \alpha)$ will be given as long as the distribution of $\varepsilon$ is assumed to be known. When $\alpha = \alpha_0$, this is a simple $\varepsilon$-orthant probability. Note that at the true parameter value $\alpha = \alpha_0$, the probabilities of the necessary conditions must be at least as large as the true probabilities of the events $y \in S(Y)$ given $X = x$, denoted by $P_0(y|x)$:

$$P(y|x, \alpha_0) \geq P_0(y|x), \forall (y, x) \in S(Y) \times S(X). \quad (2)$$

Notice that this inequality follows from the fact that the outcome $y$ implies the necessary conditions for $y$ but the necessary condition need not imply the outcome $y$. Note that the inequalities in (2) are satisfied for the true $\alpha_0$ and possibly for other values. It is possible that only one $\alpha$ satisfies the
inequalities and that if the necessary conditions are derived from an incorrect model, then perhaps no \( \alpha \) will satisfy the inequalities.

Now let \( A_0 \equiv \Theta_0 \times \mathcal{H}_0 \) denote the asymptotically identified set of parameters that satisfy the inequality restrictions in (2). This \( A_0 \) is the object we are trying to estimate from the model\(^3\). We may estimate both the finite dimensional parameters and the infinite dimensional parameters simultaneously. Alternatively, we may obtain consistent profiled estimates of \( h_0(\cdot, \theta) \) from the model or an auxiliary model and then estimate \( \theta_0 \) (thus \( \alpha_0 = (\theta_0, h_0(\cdot, \theta_0)) \)) in the main estimation. Here we adopt the second approach where consistent estimates of \( h(\cdot, \theta) \) for all \( \theta \in \Theta \) are available. We will suppress the arguments of \( h_0 \) for notational convenience such that \( (\theta, h_0) \equiv (\theta, h(\cdot, \theta)) \), \((\theta, h_0) \equiv (\theta, h_0(\cdot, \theta))\), and \((\theta_0, h_0) \equiv (\theta_0, h_0(\cdot, \theta_0))\).

3 Set Estimator

Here we take the approach by ABJ. Noting ABJ only allows for finite dimensional parameters by construction\(^4\), we adopt the second step estimation for infinite dimensional parameters where profiled estimates for infinite dimensional parameters are available in the pre-step and thus in the main estimation, we only deal with finite dimensional parameters. To focus on the treatment of the infinite dimensional parameters in this paper, we simplify discussions regarding the construction of estimators and related issues. Such issues can be found in ABJ. Here we assume that the model probabilities \( \{ P(y|X_i, \alpha) : i = 1, \ldots, n \} \) have analytic closed form solutions.\(^5\) Now we briefly review the data-dependent construction of \( X \) cells following ABJ\(^6\). Noting the data-dependent selection of \( X \) cells will affect the asymptotic distribution of the statistics, we account for this dependency in the determination of the critical values later. Consider a set \( \{ q_\gamma : \gamma \in \Gamma \} \) of real-valued weight functions on \( \mathbb{S}(X) \), where \( \gamma \) is a subset of \( \mathbb{S}(X) \) and \( \Gamma \) is a collection of subsets of \( \mathbb{S}(X) \). In particular, for each \( y^{(j)} \in \mathbb{S}(Y) = \{ y^{(1)}, \ldots, y^{(J)} \} \), we consider such \( \mathcal{M}_j \) subsets of \( \mathbb{S}(X) \) indexed by \( \gamma_{j,m}, m = 1, \ldots, \mathcal{M}_j \). We let \( \Gamma = \{ \gamma_{j,m} \subset \mathbb{S}(X) : (j, m) \in \mathcal{I}_{J, \mathcal{M}} \} \), where \( \mathcal{I}_{J, \mathcal{M}} = \{ (j, m) : m = 1, \ldots, \mathcal{M}_j, j = 1, \ldots, J \} \) with \( J = l_1 \times \ldots \times l_p \). The functions \( \{ q_\gamma : \gamma \in \Gamma \} \)

\(^3\)Note that \( A_0 \) could be (i) the null set, (ii) a single point, (iii) a strict subset of the parameter space consisting of more than one point, or (iv) the entire parameter space. ABJ refers that the model is (i) rejected, (ii) point identified, (iii) set identified, or (iv) completely uninformative.

\(^4\)It is because ABJ utilizes finite numbers of cells to facilitate the estimation, which is not compatible with infinite dimensional parameters. Simply it violates the order condition for identification.

\(^5\)The model probabilities induced by the games may not have analytic closed form expressions. In that case we need to consider the simulated version of the probabilities which are not hard to construct in many cases. The analysis here can easily adopt the simulated version of model probabilities.

\(^6\)We may also need to consider such cells for \( Y \) when the dimension of \( Y \) is high but here we implicitly assume that we do not have such a problem.

\(^7\)Examples of constructing these cells and some efficiency issue can be found in ABJ.
aggregate and/or weight the necessary condition for an equilibrium over different values of \( x \). Now let \( \hat{\Gamma}_n = \{ \hat{\gamma}_{n,j,m} \subset \mathbb{S}(X) : (j,m) \in \mathcal{I}_{j,M} \} \), where \( \hat{\gamma}_{n,j,m} \) is a random subset of \( \mathbb{S}(X) \). For the consistency of the set estimator, we require \( \hat{\Gamma}_n \rightarrow \Gamma_0 \) under certain metric described later where \( \Gamma_0 = \{ \gamma_{0,j,m} \subset \mathbb{S}(X) : (j,m) \in \mathcal{I}_{j,M} \} \). Now we extend ABJ to the semiparametric case where a profiled consistent estimator of \( h_0(\cdot, \theta) \), denoted by \( \hat{h}(\cdot, \theta) \), is available for all \( \theta \in \Theta \). Define

\[
  c_0(j, \gamma, \theta, h) = \int \left( P(y^{(j)}|x, \theta, h(\cdot, \theta)) - P_0(y^{(j)}|x) \right) q_\gamma(x) dF_X(x) \quad \text{and} \quad (3)
\]

\[
  \tilde{c}_n(j, \gamma, \theta, h) = n^{-1} \sum_{i=1}^{n} \left( P(y^{(j)}|X_i, \theta, h(\cdot, \theta)) - 1[Y_i = y^{(j)}] \right) q_\gamma(X_i).
\]

Note that \( E[\tilde{c}_n(j, \gamma, \theta, h)] = c_0(j, \gamma, \theta, h) \) for all \( (j, \gamma, \theta, h) \) by construction. Hence, with iid observations, we have \( \tilde{c}_n(j, \gamma, \theta, h) \rightarrow p_0 c_0(j, \gamma, \theta, h) \) provided that \( c_0(j, \gamma, \theta, h) \) is well-defined. Necessary conditions for \( \theta \) to be the true parameters are

\[
  P(y|x, \theta, h_0(\cdot, \theta)) - P_0(y|x) \geq 0, \ \forall (y, x) \in \mathbb{S}(Y) \times \mathbb{S}(X) \quad (4)
\]

which implies that

\[
  c_0(j, \gamma_{0,k,m}, \theta, h_0(\cdot, \theta)) \geq 0, \ \forall (j, m) \in \mathcal{I}_{j,M}. \quad (5)
\]

Define

\[
  \Theta_0 = \{ \theta \in \Theta : (4) \ \text{holds} \} \quad \text{and} \quad \Theta_+ = \{ \theta \in \Theta : (5) \ \text{holds} \}.
\]

By definition, the set \( \Theta_0 \) is the smallest set of parameter values that necessarily includes the true value \( \theta_0 \) (and thus \( c_0 \in \Theta_0 \times \mathcal{H}_0 \)). By construction, \( \Theta_+ \supset \Theta_0 \) since (4) implies (5). Now suppose that we have an initial nonparametric estimator \( \hat{h}(\cdot, \theta) \) of \( h_0(\cdot, \theta) \) for each \( \theta \). Then we define a set estimator \( \hat{\Theta}_n \) of \( \Theta_+ \) in the spirit of ABJ. To do that, we first define the estimator criterion function

\[
  Q_n(\theta, h) = \sum_{(j,m)\in\mathcal{I}_{j,M}} [\tilde{c}_n(j, \hat{\gamma}_{n,j,m}, \theta, h) - 1[\tilde{c}_n(j, \hat{\gamma}_{n,j,m}, \theta, h) \leq 0]]
\]

whose population version of the criterion function is given by

\[
  Q(\theta, h) = \sum_{(j,m)\in\mathcal{I}_{j,M}} [c_0(j, \gamma_{0,j,m}, \theta, h) - 1[c_0(j, \gamma_{0,j,m}, \theta, h) \leq 0]]. \quad (7)
\]

Note that the function \( Q(\theta, h) \) is minimized and equals zero for all values \( (\theta, h) \) which satisfy the necessary conditions \( c_0(j, \gamma_{0,j,m}, \theta, h) \geq 0 \) for all \( (j, m) \in \mathcal{I}_{j,M} \), which implies that

\[
  \Theta_+ = \{ \theta \in \Theta : \theta \text{ minimizes } Q(\theta, h_0(\cdot, \theta)) \text{ over } \Theta \}.
\]
This justifies the construction of the set estimator \( \hat{\Theta}_n \) to be

\[
\hat{\Theta}_n = \left\{ \theta \in \Theta : \theta \text{ minimizes } Q_n(\theta, \hat{h}(:, \theta)) \text{ over } \Theta \right\},
\]

where \( \hat{h}(:, \theta) \in \mathcal{H}_n \) and \( \mathcal{H}_n \) is a space of sieves that approximates \( \mathcal{H} \) satisfying \( \mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H} \) for all \( n \geq 1 \). Note that if there exists a value of \( (\theta, \hat{h}(:, \theta)) \) for which \( \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \theta, \hat{h}) \geq 0 \) for all \( (j, m) \in I_{JM} \), then \( \hat{\Theta}_n \) equals to

\[
\left\{ \theta \in \Theta : \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \theta, \hat{h}(:, \theta)) \geq 0, \forall (j, m) \in I_{JM} \right\}.
\]

It is possible that the set defined in (10) is empty by the randomness in the estimator of \( \hat{c}_n(\cdot) \). We will let \( A_0 = \Theta_0 \times \mathcal{H}_0, A_+ = \Theta_+ \times \mathcal{H}_+, \) and \( \hat{A}_n = \hat{\Theta}_n \times \hat{\mathcal{H}}_n \) where \( \mathcal{H}_0 = \{ h \in \mathcal{H} : h = h_0(:, \theta) \} \) and \( \theta \in \Theta_0 \}, \mathcal{H}_+ = \{ h \in \mathcal{H} : h = h_0(:, \theta) \text{ and } \theta \in \Theta_+ \} \) and \( \hat{\mathcal{H}}_n = \{ h \in \mathcal{H}_n : h = \hat{h}(:, \theta) \text{ and } \theta \in \hat{\Theta}_n \}, \) respectively.

The existence of \( \hat{\Theta}_n \) (i.e. the existence of \( \hat{A}_n \)) is guaranteed since \( \hat{\Theta}_n \) is defined as the collection of arguments that minimize a continuous function on a compact set\(^8\) as in (9). In Section 4, we establish the convergence of \( \hat{A}_n \) to \( A_+ \) in probability under a certain metric with suitable assumptions. We develop our discussion under some higher level assumptions which should be justified for each example on hand. In the following, before going into the asymptotics of our estimator, we present semiparametric versions of two discrete games providing the inequality conditions of (2) which we are based on for our set estimation and inference.

### 3.1 Example 1: Two Firms Entry-Exist Game

Suppose there are two potential entrants in a market whose profits depend on the existence of its rival. Let \( \Pi_j \) denote the profit of the firm \( j = 1, 2 \) as

\[
\Pi_j(y_1, y_2|x) = \pi_j(Y_1 = y_1, Y_2 = y_2|X = (x_1', x_2')) + \varepsilon_j
\]

where we let \( \pi_1(1, 0|x) = a + x_1' - c_1 \theta_1 + h_1(x_{1c}), \pi_2(0, 1) = a + x_2' - c_2 \theta_2 + h_2(x_{2c}), \pi_1(1, 1) = b + x_1' - c_1 \theta_1 + h_1(x_{1c}), \pi_2(1, 1) = b + x_2' - c_2 \theta_2 + h_2(x_{2c}), \) and \( \pi_1(0, \cdot) = \pi_2(\cdot, 0) = 0 \) with \( a > b \) since a monopoly profit tends to be higher than a duopoly profit. \( \varepsilon = (\varepsilon_1, \varepsilon_2) \) are known to each firm but not to econometricians. Here the payoffs functions of players are given as partially linear forms. We let \( X_{jc} \cap X_{j-c} = \emptyset \) and \( X_{jc} \cup X_{j-c} = X_j \) for \( j = 1, 2 \) and let \( h_1(0) = h_2(0) \) for the identification of parameters. Now assume that \( \varepsilon_j \) follows a normal distribution and \( \varepsilon_1 \perp \varepsilon_2 \). Then, the probability of being a monopolist will be \( \Phi(\mu_{a_j} \equiv a + x_{j-c}' \theta_j + h_j(x_{jc})) = \Pr(\varepsilon_j > -\mu_{a_j}) \) and that of being

\(^8\) \( Q_n(\theta, h) \) is continuous in \( \theta \) as long as \( h \) is continuous in \( \theta \) and \( \hat{c}_n(\cdot) \) is continuous in \( (\theta, h) \), which is also continuous as long as \( F(y|x, \theta, h) \) is continuous in \( (\theta, h) \).
a duopolist will be \( \Phi(\mu_{b_j}) = b + x'_{j-c}\theta_j + h_j(x_{jc}) = \Pr(\varepsilon_j > -\mu_{b_j}) \). For this game, we note that multiple Nash equilibria exist depending on the realization of \((\varepsilon_1, \varepsilon_2)\) and that the necessary conditions of the Nash equilibria give us the following four inequality conditions comparing the true probabilities of events and the model probabilities as the form of (2):

\[
\begin{align*}
P(0,0|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) &= (1 - \Phi(\mu_{a_1}))(1 - \Phi(\mu_{a_2})) \geq P_0(0,0|x), \\
P(0,1|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) &= \Phi(\mu_{a_2})(1 - \Phi(\mu_{b_1})) \geq P_0(0,1|x), \\
P(1,0|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) &= \Phi(\mu_{a_1})(1 - \Phi(\mu_{b_2})) \geq P_0(1,0|x), \text{ and} \\
P(1,1|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) &= \Phi(\mu_{b_1})\Phi(\mu_{b_2}) \geq P_0(1,1|x).
\end{align*}
\]

This game model includes both finite dimensional parameters \((a, b, \theta_j, j = 1, 2)\) and infinite dimensional parameters \((h_j(\cdot), j = 1, 2)\) of interest.

### 3.1.1 Construction of the Profiled Estimator

Let \( \theta = (a, b, \theta_1', \theta_2')' \) and \( h = (h_1, h_2) \) and rewrite \( P(Y_1, Y_2|X, \theta, h) = P(Y_1, Y_2|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) \).

Under the correct model specification with true parameters of \((\theta_0, h_0)\), we have \( P(0,0|X, \theta_0, h_0) = P_0(0,0|X), 1 - P(0,0|X, \theta_0, h_0) - P(1,1|X, \theta_0, h_0) = P_0(0,1|X) + P_0(1,0|X) \), and \( P(1,1|X, \theta_0, h_0) = P_0(1,1|X) \) regardless of the multiplicity of the Nash equilibria\(^9\). Using this fact and the method of sieve MLE\(^10\) similarly with Kim (2006), we estimate \( h_0 \) as a profiled estimate of the form \( \hat{h}(\cdot, \theta) \) such that

\[
\hat{h}(\cdot, \theta) = \operatorname{argmax}_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} l(y_i, x_i; \theta, h)
\]

where \( l(y_i, x_i; \theta, h) \equiv 1[y_{1i} + y_{2i} = 0] \log P(0,0|x_i, \theta, h) + 1[y_{1i} + y_{2i} = 1] \log(1 - P(0,0|x_i, \theta, h) - P(1,1|x_i, \theta, h)) + 1[y_{1i} + y_{2i} = 2] \log P(1,1|x_i, \theta, h) \). Under some regularity conditions similar with those in Kim (2006), we can show that \( \sup_{\theta \in \Theta} \sup_{x_{jc} \in \mathcal{S}(X_{jc})} |\hat{h}_j(\cdot, \theta) - h_j(\cdot, \theta)| = o_p(1) \) for \( j = 1, 2 \).

Interestingly, here we note that we may estimate the parameters simultaneously as \( (\hat{\theta}, \hat{h}) \) such that

\[
(\hat{\theta}, \hat{h}) = \operatorname{argmax}_{\theta \in \Theta, h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} l(y_i, x_i; \theta, h).
\]

The consistency and the asymptotic normality of functional of parameters for such estimators can be also found in Kim (2006). However, we note that to achieve this consistent point estimation, we have to disregard the information about which firm enters the market since we redefine the outcome space in terms of how many firms in the market. Due to this omitted information, we

---

\(^9\)Thus, we have redefined the outcome space in terms of the number of firms in the market.

\(^10\)Alternatively, we can obtain a profiled estimator using the Sieve Minimum Distance estimator proposed by Ai and Chen (2003) noting that the model can be characterized in terms of the moment conditions: \( 0 = E[(1 - Y_1)(1 - Y_2) - P(0,0|X, \theta_0, h_0)|X] \) and \( 0 = E[Y_1 Y_2 - P(1,1|X, \theta_0, h_0)|X] \).
may end up a point estimator whose variance is larger and thus we may have wider confidence interval for a parameter of interest than that of the set estimator using the full information in the model. It will be an interesting future research to compare these two alternatives: CI from the point estimation with the usage of less information vs. CI from the set estimation with the usage of the full information.

3.2 Example 2: Signaling Game with Two Discrete Types

The following example is a simple version of Kim (2006). Consider the beer-quiche game in Cho and Kreps (1987) where we have two players. Player 1 has either of two types \{strong, weak\} with the probability of being the strong type equal to \(p_u\), and knows her type. After observing her type, Player 1 moves first sending one of two messages \{Beer, Quiche\} to Player 2. Then, Player 2 chooses an action “Fight” or “No Fight” after observing the signal sent by Player 1. After the play, a payoff is realized depending on actions chosen by two players and Player 1’s type. The structure and the payoffs of this game is given in Figure 1. Here \(\varphi_1\) denotes the cost of mimicking the other type of Player 1 (cost of signalling falsely) and \(\varphi_2\) measures Player 2’s incentive to single out a particular type of Player 1. In the payoffs functions, Players can observe each other’s \(\varepsilon\) but econometricians only know the distribution of \(\varepsilon\) such as normal distribution.

In this game we have four possible observable outcomes: Player 1 chooses beer but Player 2 decides to fight, Player 1 chooses beer and Player 2 decides not to fight, Player 1 chooses quiche and Player 2 decides to fight, or Player 1 chooses quiche but Player 2 decides not to fight. We let

\[ p_u \]

\[ 1 - p_u \]

Note that \(p_u\) is the unconditional probability while \(p(\cdot)\) denotes the conditional probability.
$Y_1 = 1$ if Player 1 chooses Beer and $Y_1 = 0$ otherwise and let $Y_2 = 1$ if Player 2 chooses No Fight and $Y_2 = 0$ otherwise. From the result of Kim (2006)\textsuperscript{12}, using the Perfect Bayesian Equilibrium (PBE), we can characterize the equilibrium of the game as summarized in Figure 2. As illustrated in the figure, depending on the realizations of $(\varepsilon_1, \varepsilon_2)$, we can have Pooling equilibria, Separating equilibria, or Semi-separating equilibria. Then, we obtain the model probabilities for each four possible outcome by integrating regions of $\varepsilon$ for each particular observable outcome as below.\textsuperscript{13}

We let $\mu_1 \equiv X_1^1 \theta_1$, $\mu_2 \equiv X_2^2 \theta_2$, and $p \equiv p(Z)$ where $p(Z)$ is the conditional probability of being a strong type conditional on the public signal $Z$ regarding Player 1’s type, which are available to both Player 2 and econometricians.

\[
P (1, 1|p, \mu_1, \mu_2, \varphi_1, \varphi_2) = \\
\Phi (\mu_1 - \varphi_1) \Phi (\mu_2 + (2p - 1) \varphi_2) - \Phi (\mu_2 - \varphi_2) + p \Phi (\mu_2 - \varphi_2) \\
+ p \Phi (\mu_1 + \varphi_1) \Phi (\mu_2 + \varphi_2) - \Phi (\mu_2 - \varphi_2) \\
+ \int_0^1 \phi \left( \mu_2 + \left( \frac{p - \omega (1-p)}{p + \omega (1-p)} \right) \varphi_2 \right) 2p (1-p) \varphi_2 d\omega \\
+ \int_0^1 \phi \left( \mu_2 + \left( \frac{1-p}{1-\omega} p + (1-p) \right) \varphi_2 \right) 2p (1-p) \varphi_2 d\omega \\
+ \int_0^1 \phi \left( \mu_1 + \varphi_1 \right) \frac{\varphi_1}{1-\omega} d\omega \\

P (1, 0|p, \mu_1, \mu_2, \varphi_1, \varphi_2) = \\
(1 - \Phi (\mu_1 + \varphi_1)) \Phi (\mu_2 + (2p - 1) \varphi_2) - \Phi (\mu_2 - \varphi_2) + p (1 - \Phi (\mu_2 + \varphi_2)) \\
+ \int_0^1 \phi \left( \mu_2 + \left( \frac{p - \omega (1-p)}{p + \omega (1-p)} \right) \varphi_2 \right) 2p (1-p) \varphi_2 d\omega \\
+ \int_0^1 \phi \left( \mu_2 + \left( \frac{1-p}{1-\omega} p + (1-p) \right) \varphi_2 \right) 2p (1-p) \varphi_2 d\omega \\
+ \int_0^1 \phi \left( \mu_1 + \varphi_1 \right) \frac{\varphi_1}{1-\omega} d\omega \\

P (0, 1|p, \mu_1, \mu_2, \varphi_1, \varphi_2) = \\
\Phi (\mu_1 - \varphi_1) \Phi (\mu_2 + (2p - 1) \varphi_2) - \Phi (\mu_2 - \varphi_2) + (1 - p) \Phi (\mu_2 - \varphi_2) \\
+ \int_0^1 \phi \left( \mu_2 + \left( \frac{p - \omega (1-p)}{p + \omega (1-p)} \right) \varphi_2 \right) 2p (1-p) \varphi_2 d\omega \\
+ \int_0^1 \phi \left( \mu_2 + \left( \frac{1-p}{1-\omega} p + (1-p) \right) \varphi_2 \right) 2p (1-p) \varphi_2 d\omega \\
+ \int_0^1 \phi \left( \mu_1 + \varphi_1 \right) \frac{\varphi_1}{1-\omega} d\omega \\

P (0, 0|p, \mu_1, \mu_2, \varphi_1, \varphi_2) = \\
1 - P (1, 1|p, \mu_1, \mu_2, \varphi_1, \varphi_2) - P (1, 0|p, \mu_1, \mu_2, \varphi_1, \varphi_2) - P (0, 1|p, \mu_1, \mu_2, \varphi_1, \varphi_2) - P (0, 0|p, \mu_1, \mu_2, \varphi_1, \varphi_2)
\]

\textsuperscript{12}When PBE is adopted as the equilibrium concept, Kim (2006) shows that this signaling game has multiple equilibria depending on the realizations of $(\varepsilon_1, \varepsilon_2)$. In the region $E_1 \equiv \{(\varepsilon_1, \varepsilon_2): \varepsilon_1 \geq \mu + X_1^1 \theta_1 + \varphi_1 \& X_2^2 \theta_2 + (2p - 1) \varphi_2 \leq \varepsilon_2 \leq X_2^2 \theta_2 + \varphi_2\}$, we can have two equilibria: Pooling equilibrium with Beer & Fight or Pooling equilibrium with Quiche & Fight while in $E_2 \equiv \{(\varepsilon_1, \varepsilon_2): \varepsilon_1 \leq \mu + X_1^1 \theta_1 - \varphi_1 \& X_2^2 \theta_2 - \varphi_2 \leq \varepsilon_2 \leq X_2^2 \theta_2 + (2p - 1) \varphi_2\}$, we can have two equilibria: Pooling with Beer & No Fight or Pooling with Quiche & No Fight. Kim (2006) also shows that we can still achieve the uniqueness of equilibrim by strengthening the concept of equilibrim such as Cho and Kreps (1987)’s Intuitive Criterion. Only allowing equilibrium that survives this Intuitive Criterion, Kim (2006) shows that only (Quiche, Fight) survives in $E_1$ and only (Beer, No Fight) survives in $E_2$.

\textsuperscript{13}For details how to derive the equilibria of the game and the resulting model probabilities, see Kim (2006).
This game model also includes both finite dimensional parameters \((\varphi_j, p_u, \beta_j, j = 1, 2)\) and infinite dimensional parameters \((p(\cdot))\) of interest. The estimation and inference of this signaling game model will be based on the four inequality conditions as the form of (2): \(P(y_1, y_2|p, \mu_1, \mu_2, \varphi_1, \varphi_2) \geq P_0(y_1, y_2|X = x)\) where \(X = X_1 \cup X_2 \cup Z\) and \(y_1, y_2 = \{0, 1\}\).

### 3.2.1 Construction of the Profiled Estimator

Now let\(^{14}\) \(\theta = (\varphi_1, \varphi_2, \theta_1', \theta_2')'\) and \(h = L^{-1}(p(Z))\) with \(L(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))\) under the logit specification of \(p(\cdot)\). Rewrite \(P(Y_1, Y_2|X, \theta, h) = P(Y_1, Y_2|p(\cdot), \mu_1, \mu_2, \varphi_1, \varphi_2)\). To obtain a consistent estimate of \(h_0(\cdot, \theta)\), we again consider a redefinition of outcome space. From Figure 2, we note that regardless of the multiple equilibria in the regions of \(\mathcal{E}_1 \equiv \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 \geq \mu + X_1' \theta_1 + \varphi_1 \& \ X_2' \theta_2 + (2p - 1) \varphi_2 \leq \varepsilon_2 \leq X_2' \theta_2 + \varphi_2\}\) and \(\mathcal{E}_2 \equiv \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 \leq \mu + X_1' \theta_1 - \varphi_1 \& \ X_2' \theta_2 - \varphi_2 \leq \varepsilon_2 \leq X_2' \theta_2 + (2p - 1) \varphi_2\}\), we have a well-defined likelihood function when we redefine the outcomes of the game in terms of whether a fight is raised or not. Figure 3\(^{15}\) shows the resulting redefinition of outcome space. From this observation, we have

\[
P(Y_2 = 0|X, \theta, h) = P(0, 0|X, \theta, h) + P(1, 0|X, \theta, h) - \Pr(\varepsilon \in \mathcal{E}_1)
\]

\[
P(Y_2 = 1|X, \theta, h) = P(0, 1|X, \theta, h) + P(1, 1|X, \theta, h) - \Pr(\varepsilon \in \mathcal{E}_2)
\]

where we let \(\Pr(\varepsilon \in \mathcal{E}_1) = (1 - \Phi(\mu_1 + \varphi_1))(\Phi(\mu_2 + \varphi_2) - \Phi(\mu_2 + (2p - 1) \varphi_2))\) and \(\Pr(\varepsilon \in \mathcal{E}_2) = \Phi(\mu_1 - \varphi_1)(\Phi(\mu_2 + (2p - 1) \varphi_2) - \Phi(\mu_2 - \varphi_2))\). Then, we obtain a consistent profiled estimator, \(\widehat{h}(\cdot, \theta)\), using the sieve MLE similarly with Example 1 such that

\[
\widehat{h}(\cdot, \theta) = \arg\max_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} l(y_i, x_i; \theta, h) = \{1 \; [y_{2i} = 1] \log P(0|x_i, \theta, h) + 1 \; [y_{2i} = 0] \log P(1|x_i, \theta, h)\}
\]

Under some regularity conditions similar with those in Kim (2006), we can also show that \(\sup_{\theta \in \Theta} \sup_{\varepsilon \in S(\varepsilon)} \left[ \hat{h} - h_0 \right] = o_p(1)\). As in Example 1, here we can also estimate the parameters simultaneously as \((\hat{\theta}, \hat{h})\) such that \((\hat{\theta}, \hat{h}) = \arg\max_{\theta \in \Theta, h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} l(y_i, x_i; \theta, h)\). However, this consistent point estimation again requires the information reduction. We do not use the information regarding Player 1’s actions. Because of this, we may obtain a point estimator with larger variance and thus we may have wider confidence interval than that of the set estimator considered in this paper. This concern justifies the use of the set inference even though we can achieve a consistent point estimation.

\(^{14}\) We note that in the conditional probabilities of observed outcomes, implied by the model, \(p_u\) does not appear separately from \(p(Z)\). Nonetheless, we can still estimate \(p_u\) as a functional of \(p(Z)\). We will discuss this in Section 4.1.

\(^{15}\) The dotted area is for “Fight” and the solid area is for “No Fight”. “\(\sigma\)” denotes the Semi-Separating Equilibria.
4 Probability Limit of the Set Estimator

We use a version of the Hausdorff metric measuring the distance between two sets whose elements are \((e_1, e_2)\)’s such that \(e_1 \in \Theta\) and \(e_2 = (e_{21}, \ldots, e_{2n})\) belongs to a class of vector of continuous functions defined on \(S(Z_1) \times \ldots \times S(Z_n)\) where \(Z_v\)’s are subsets of \(X\). For two such sets \(A\) and \(B\) whose elements are \(a = (a_1, a_2)\) and \(b = (b_1, b_2)\) respectively, let the maximum distance between any points in \(A\) and \(B\) be given by \(\rho(A|B) = \sup_{a \in A} \rho(a|B)\), where \(\rho(a|B) = \inf \{\|a - b\|_s = \|a_1 - b_1\|_E + \|a_2 - b_2\|_{\mathcal{H}} : b \in B\}\) for \(a \in A\) and a pseudo metric \(\|\cdot\|_{\mathcal{H}}\) is defined by \(\|\cdot\|_{\mathcal{H}} \equiv \sup_{a_1, b_1} \sum_{v=1}^{n} \sup_{z_v \in S(Z_v)} |a_{2v}(z_v) - b_{2v}(z_v)|\). Notice that \(\rho(a|B)\) is the distance from an element \(a\) to the set \(B\). By definition, if \(A = \emptyset\) and \(B \neq \emptyset\), then \(\rho(A|B) = 0\) while \(\rho(A|B) = \infty\) if \(B = \emptyset\). The Hausdorff metric distance between \(A\) and \(B\) is given by \(d(A, B) = \max \{\rho(A|B), \rho(B|A)\}\). For completeness, we also let \(\|a - b\|_s = \|a_1 - b_1\|_E\) and define the Hausdorff metric accordingly when we have only the finite dimensional parameters. Also note that \(\|\cdot\|_{\mathcal{H}} = \sup_{a_1, b_1} \sum_{v=1}^{n} \sup_{z_v \in S(Z_v)} |a_{2v}(z_v) - b_{2v}(z_v)|\) when \(a_1 = b_1\). Here we provide some conditions under which

\[
\begin{align*}
(i) \quad &\rho(\hat{A}_n|A_+) \xrightarrow{p} 0 \quad \text{and} \quad (ii) \quad \rho(A_+|\hat{A}_n) \xrightarrow{p} 0
\end{align*}
\]

and thus we establish the conditions for \(d(\hat{A}_n, A_+) \xrightarrow{p} 0\). Note that (11)(i) ensures that \(\hat{A}_n\) is not larger than \(A_+\) asymptotically while (11)(ii) ensures that \(\hat{A}_n\) is not smaller than \(A_+\) asymptotically. Also note that (11)(ii) alone ensures that the distance of the true value \(\alpha_0 (\in A_0 \subset A_+)\) from the set estimator \(\hat{A}_n\) satisfies \(\rho(\alpha_0|\hat{A}_n) \xrightarrow{p} 0\).
Let $\Gamma_{all}$ be a class of subsets of $S(X)$ that includes all possible realizations of $\tilde{\gamma}_{n,j,m}$ for all $(j,m) \in I_{J,M}$ and $n \geq 1$. Also let $I_J = \{1, \ldots, J\}$. We use “wpl” to denote “with probability that approaches to one”. The following assumptions are sufficient to establish the result (11)(i).

**Assumption 4.1** \(\{(Y_i, X_i)\}_{i=1}^n\) are iid.

**Assumption 4.2** The true parameter $\alpha_0$ satisfies $P(y|x, \alpha_0) - P_0(y|x) \geq 0, \forall (y, x) \in S(Y) \times S(X)$.

**Assumption 4.3** (i) $\Theta \times H$ is compact under the metric $\|\cdot\|_S$; (ii) $H_n \subseteq H_{n+1} \subseteq H$ for all $n \geq 1$ and for any $h \in H$, there exists $\Pi_n h \in H_n$ such that $\|\Pi_n h - h\|_H = o(1)$; (iii) $\|h - h_0\|_H = o_p(1)$; (iv) $h(\cdot, \theta)$ is continuous in $\theta$ for all $h$ s.t. $\|h - h_0\|_H = o(1)$.

**Assumption 4.4** (i) $P(y|x, \alpha)$ is Hölder continuous in $\alpha$ on $\mathcal{A}$; (ii) $|q_{\gamma}(x)| < \infty$ for all $\gamma \in \Gamma_{all}$ uniformly over $x \in S(X)$.

To establish the second result (11) (ii), we need additional assumptions. The following conditions are in the line with ABJ. Let $\text{int}(A)$ and $\text{cl}(A)$ denote the interior and closure of a set $A$, respectively.

**Assumption 4.5** Either (i) $\Theta_+ = \{\theta_0\}$ or (ii) (a) $\Theta_+ = \text{cl}(\text{int}(\Theta_+))$ and (b) $\forall \theta \in \text{int}(\Theta_+), \inf_{(j,m)\in I_{J,M}} c_0(j, \gamma_{0,j,m}, \theta, h_0(\cdot, \theta)) > 0$.

Assumption 4.2 states that the model (conditional probabilities implied by the game) is correctly specified, which ensures that $\mathcal{A}_0$ and $\mathcal{A}_+$ are not empty. Note that Assumptions 4.3 and 4.4 are standard assumptions in the semi-nonparametric literature. Note that Assumption 4.5 (i) holds when the necessary conditions (5) are strong enough that $\mathcal{A}_+$ only contains the true parameter $\alpha_0 = (\theta_0, h_0)$. Assumption 4.5 (ii) (a) implies that $\Theta_+$ has a non-empty interior and does not contain isolated points. Assumption 4.5 (ii) (b) requires that for any $\theta \in \text{int}(\Theta_+)$, the necessary conditions (5) hold with a strict inequality. We also need some consistency results for $\tilde{c}_n(\cdot)$ and $\tilde{\Gamma}_n$ under certain metrics. First, define $\mathcal{H}_{\delta_n} \equiv \{h \in H : \|h(\cdot, \theta) - h_0(\cdot, \theta)\|_H \leq \delta_n, \theta \in \Theta\}$, $\mathcal{H}_{n,\delta_n} \equiv \{h \in H_n : \|h(\cdot, \theta) - h_0(\cdot, \theta)\|_H \leq \delta_n, \theta \in \Theta\}$ with $\delta_n = o(1)$, and for any two real functions $c_1$ and $c_2$ on $I_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$, define $\|c_1 - c_2\|_{U_n} \equiv \sup_{(j,\gamma,\theta,h)\in I_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}} |c_1(j, \gamma, \theta, h) - c_2(j, \gamma, \theta, h)|$.

The following is the semiparametric version of Assumption 5 in ABJ.

**Assumption 4.6** $\|\tilde{c}_n(\cdot) - c_0(\cdot)\|_{U_n} \to 0.$

Let $\mathcal{F}_\xi = \{\xi(y, x, j, \gamma, \theta, h) = (P(y^{(j)}|x, \theta, h) - 1[y = y^{(j)}]) \cdot q_\gamma(x) : (j, \gamma, \alpha) \in I_J \times \Gamma_{all} \times \Theta \times \mathcal{H}\}$ denote the class of measurable functions indexed by $(j, \gamma, \theta, h)$. Assumption 4.6 will hold when
\( \mathcal{F}_\xi \) is a P- Glivenko-Cantelli class as presented in van der Vaart and Wellner (1996). Now define a semi-norm \( \| \cdot \| \) as follows. For \( \gamma_1 \) and \( \gamma_2 \in \Gamma_{\text{all}} \), we let

\[
\| \gamma_1 - \gamma_2 \| = \left( \int |q_{\gamma_1}(x) - q_{\gamma_2}(x)|^2 dF_X(x) \right)^{1/2}
\]

and

\[
\| \Gamma_1 - \Gamma_2 \| = \max_{(j,m) \in I_{J,M}} \| \gamma_{1,j,m} - \gamma_{2,j,m} \|
\]

where \( \Gamma_1 = \{ \gamma_{1,j,m} \in \Gamma_{\text{all}} : (j,m) \in I_{J,M} \} \) and \( \Gamma_2 = \{ \gamma_{2,j,m} \in \Gamma_{\text{all}} : (j,m) \in I_{J,M} \} \). We assume

**Assumption 4.7** \( \| \hat{\Gamma}_n - \Gamma_0 \| \to 0 \) (Assumption 6 of ABJ).

Now we are ready to present the consistency result of the set estimator \( \hat{A}_n \).

**Theorem 4.1** (i) Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.6, and 4.7 hold. Then, \( \rho(\hat{\Theta}_N | \Theta_+) \to 0 \).

(ii) Under Assumptions 4.3, 4.4, 4.5, 4.6, and 4.7, \( \rho(\Theta_+ | \hat{\Theta}_n) \to 0 \). Thus, we have \( d(\hat{A}_n, A_+) \to 0 \).

In Section 5, we discuss how to construct the confidence interval of a real functional \( \beta_n(\alpha) \) of \( \alpha \). In particular, we may restrict our interests to real functions such that \( \beta_n(\alpha) = r_n(\theta) \) where \( r_n(\theta) \) is a real function of \( \theta \). For example, we can have \( \beta_n(\alpha) = \theta_{(k)} \) where \( \theta_{(k)} \) is the \( k \)-th element of \( \theta \). When constructing the confidence interval of the real functional \( \beta_n(\alpha) \), its largest and smallest values are of interest. The largest and smallest values of \( \beta_n(\alpha) \) across all \( \alpha \in A_+ \) defined by

\[
\beta_{n,U} = \sup \{ \beta_n(\alpha) : \alpha \in A_+ \} \quad \text{and} \quad \beta_{n,L} = \inf \{ \beta_n(\alpha) : \alpha \in A_+ \},
\]

respectively. We estimate these values, respectively, by

\[
\hat{\beta}_{n,U} = \sup \{ \hat{\beta}_n(\alpha) : \alpha \in \hat{A}_n \} \quad \text{and} \quad \hat{\beta}_{n,L} = \inf \{ \hat{\beta}_n(\alpha) : \alpha \in \hat{A}_n \}.
\]

The consistency of \( \hat{\beta}_{n,U} \) for \( \beta_{n,U} \) and \( \hat{\beta}_{n,L} \) for \( \beta_{n,L} \) is obtained from Corollary 4.1 as long as \( \beta_n(\cdot) \) has some continuity property (with respect to the metric \( \| \cdot \|_s \)). We note that if \( \{ \beta_n(\cdot) : n \geq 1 \} \) is stochastically equicontinuous on \( A \), the consistency results hold.

**Assumption 4.8** \( \{ \beta_n(\cdot) : n \geq 1 \} \) is stochastically equicontinuous on \( A \), i.e. for any given \( \varepsilon > 0 \) and any \( \alpha \in A \), there exists \( \delta > 0 \) such that

\[
\lim_{n \to \infty} P(\sup_{\|\tilde{\alpha} - \alpha\|_s \leq \delta} | \beta_n(\tilde{\alpha}) - \beta_n(\alpha) | > \varepsilon ) < \varepsilon.
\]

In particular, when \( \beta_n(\cdot) \) is (pointwise) Lipschitz continuous with respect to \( \alpha \), primitive sufficient conditions for stochastic equicontinuity can be found in Andrews (1994) or Newey and McFadden (1994). Even when \( \beta_n(\cdot) \) is (pointwise) Lipschitz continuous with respect to \( h \) but not in \( \theta \), we can still apply the results in Andrews (1994) for certain cases. Chen, Linton, and van Keilegom (2003) also provide some stochastic equicontinuity results even when \( \beta_n(\cdot) \) is not (pointwise) continuous with respect to \( h \) and \( \theta \). From the result of Theorem 4.1, \( d(\hat{A}_n, A_+) \to 0 \), we obtain

**Corollary 4.1** Under Assumptions 4.1-4.8, \( \hat{\beta}_{n,U} - \beta_{n,U} \to 0 \) and \( \hat{\beta}_{n,L} - \beta_{n,L} \to 0 \).
4.1 Example: Set Estimation of the Type Distribution

We note that in the conditional probabilities of observed outcomes, implied by the model, $p_u$ does not appear separately from $p(Z)$ in Section 3.2 as originally noted in Kim (2006). However, using the law of iterated expectation $p_u = E[p(Z)]$, we can still identify the type distribution parameter $p_u$. Recalling $p(Z) = L(h(Z)) \equiv \exp(h(Z))/(1 + \exp(h(Z)))$, we obtain a set estimator of $p_u$ such that

$$
\hat{P}_n = \left\{ p_u : p_u = \frac{1}{n} \sum_{i=1}^{n} L(h(z_i)) = \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(h(z_i))}{1+\exp(h(z_i))} \text{ for each } h \in \mathcal{R}_n \right\}.
$$

We note that as long as $d(\mathcal{A}_n, \mathcal{A}_+) = o_p(1)$, $\hat{P}_n$ converges to its population counterpart $\mathcal{P}_+$ defined by $\mathcal{P}_+ = \{ p_u : p_u = E[L(h)] = E\left[\frac{\exp(h)}{1+\exp(h)}\right] \text{ for each } h \in \mathcal{H}_+ \}$.

Proposition 4.1 Suppose $\{Z_i\}_{i=1}^{n}$ are iid and $d(\mathcal{A}_n, \mathcal{A}_+) = o_p(1)$. Then, $d(\hat{P}_n, \mathcal{P}_+) = o_p(1)$.

5 Confidence Intervals

In this section, we construct a CI\textsuperscript{16} for the true value $\beta_0 = \beta_n(\alpha_0)$, where $\beta_n(\cdot)$ is a known real functional of $\alpha_0$. Note that we suppress the potential dependence of $\beta_0$ on $n$ for notational simplicity. In the spirit of Imbens and Manski (2003), the CI we consider is for the true value $\beta_0$, not for the set values $\beta_n(\alpha)$ for $\alpha \in \mathcal{A}_+$. The CI provided here is an extension of ABJ to the semiparametric case. Now let

$$
\mathcal{A}_{n,U} = \{ \alpha \in \mathcal{A}_n : \beta_n(\alpha) = \beta_{n,U} \}. \tag{15}
$$

Note that $\mathcal{A}_{n,U}$ is not empty since $\mathcal{A}_n$ is compact under a metric $\| \cdot \|$. The compactness of $\mathcal{A}_n$ under $\| \cdot \|$ comes from (i) $\mathcal{A}_n$ is compact under $\| \cdot \|$ and (ii) $\mathcal{A}_n$ is defined using the non-strict inequality. We choose a unique value $\alpha_{n,U}$ such that

$$
\hat{\alpha}_{n,U} = \text{argmin} \left\{ \|\alpha\| : \alpha \in \mathcal{A}_{n,U} \right\}. \tag{16}
$$

Again note that the existence of $\alpha_{n,U}$ is guaranteed since $\mathcal{A}_{n,U}$ is compact under $\| \cdot \|$. The solution of (16) may not be unique. In that case, a researcher can choose a particular value of $\alpha_{n,U}$ according to certain criterion. We define $\mathcal{A}_{n,L}$ and $\alpha_{n,L}$ analogously replacing $\beta_{n,U}$ with $\beta_{n,L}$. Note that by construction, we have $\beta_n(\alpha_{n,U}) = \hat{\beta}_{n,U}$ and $\beta_n(\alpha_{n,L}) = \hat{\beta}_{n,L}$. Now let $\mathcal{B}_{n,U}$ and $\mathcal{B}_{n,L}$ be collections

---

\textsuperscript{16}We may consider the CI for identified set of $\beta_n(\cdot)$ alternatively. Such CI can be found in ABJ. The asymptotic justification of the semiparametric version of this can be given similarly with Theorem 5.1. In this paper, we provide the CI for the true value only since our focus here is in extending ABJ to the semiparametric models.
of \((j, m) \in \mathcal{I}_{J, M}\) such that the corresponding constraints bind at \(\tilde{\alpha}_{n,U}\) and \(\tilde{\alpha}_{n,L}\), respectively, which are the boundary points of \(\tilde{A}_n\). Thus, we have \(\tilde{B}_{n,U} = \{(j, m) \in \mathcal{I}_{J, M} : \tilde{c}(n, j, \tilde{\gamma}_{n,j,m}, \tilde{\alpha}_{n,U}) = 0\}\) and \(\tilde{B}_{n,L} = \{(j, m) \in \mathcal{I}_{J, M} : \tilde{c}(n, j, \tilde{\gamma}_{n,j,m}, \tilde{\alpha}_{n,L}) = 0\}\).

Now we are ready to present the \((1 - a)\)-CI of the true value \(\beta_0\). We consider several versions of CI’s depending on the choice of the sieve spaces that are used for constructing upper and lower bounds of CI’s. We have not considered a specific function space and a sieve space for the estimation stage but in the inference stage, we need to do so since the construction of CI’s critically depend on the choice of the function space for \(h\). A Hölder space, denoted by \(\Lambda^\nu(S(Z))\), is a space of functions \(g : S(Z) \to \mathbb{R}\) such that the first \(\nu\) derivatives are bounded, and the \(\nu\)-th derivatives are Hölder continuous with the exponent \(\nu - \frac{1}{2} \in (0, 1]\), where \(\nu\) is the largest integer smaller than \(\nu\). The Hölder space becomes a Banach space when endowed with the Hölder norm: 
\[
\|g\|_{\Lambda^\nu} = \sup\limits_{z \neq z'} |g(z) - g(z')| / |z - z'|^{\nu - \frac{1}{2}} < \infty,
\]
where \(\frac{1}{2} < \nu \leq 1\). The Hölder ball (with radius \(c\)) \(\Lambda^\nu_c(S(Z))\) is defined accordingly as \(\Lambda^\nu_c(S(Z)) = \{g \in \Lambda^\nu(S(Z)) : \|g\|_{\Lambda^\nu} \leq c < \infty\}\).

Now let \(\mathcal{H}^1 \equiv \mathcal{H}^1 \times \ldots \times \mathcal{H}^n = \Lambda^\nu_1(S(Z_1)) \times \ldots \times \Lambda^\nu_n(S(Z_n))\). Then, it is well known that functions in \(\mathcal{H}\) can be approximated well by power series, Fourier series, splines, and wavelets. For example, we may let \(\mathcal{H}^1 = \{h_1 : h_1(z_1) = \sum_{k=1}^{\infty} (a_k \cos(kz_1) + b_k \sin(kz_1)), \|h_1\|_{\Lambda^\nu_1} \leq c_1\}\) where \(h_1(z_1)\) is given as an infinite Fourier series and its derivative with a fractional power is also defined in terms of Fourier series.

The \((1 - a)\)-CI of the true value \(\beta_0\) is given by
\[
CI_n(1 - a) = [\tilde{\beta}_{n,L}, \tilde{\beta}_{n,U}]
\]
for some upper and lower bounds, \(\tilde{\beta}_{n,U}\) and \(\tilde{\beta}_{n,L}\) such that
\[
\liminf_{n \to \infty} P(\beta_0 \subset CI_n(1 - a)) = \liminf_{n \to \infty} P(\tilde{\beta}_{n,L} \leq \beta_0 \leq \tilde{\beta}_{n,U}) \geq 1 - a.
\]

We will consider three alternatives. First, define \(\tilde{\mathcal{H}}_{\delta_n} \equiv \{h \in \mathcal{H} : \|h - \tilde{h}(\cdot, \theta)\|_{\mathcal{H}} \leq \delta_n, \theta \in \Theta\}\) and \(\tilde{\mathcal{H}}_{l,\delta_n} \equiv \{h \in \mathcal{H}_l : \|h - \tilde{h}(\cdot, \theta)\|_{\mathcal{H}_l} \leq \delta_n, \theta \in \Theta\}\) with \(\delta_n = o(1)\) where \(\mathcal{H}_l\) is a finite dimensional sieve space such that \(\mathcal{H}_l \subseteq \mathcal{H}_{l+1} \subseteq \mathcal{H}\) for all \(l \geq 1\). Then, the upper and lower bounds, \(\tilde{\beta}_{n,U}\) and \(\tilde{\beta}_{n,L}\) for three alternative CI’s are given as the following form:

- Alternative CI1: CI over the whole infinite dimensional space \((\Theta \times \mathcal{H})\): \(\tilde{\beta}_{n,U} \equiv \tilde{\beta}_{n,U}^{(1)}\) and

---

\(^{17}\)For detailed discussions regarding finite dimensional or infinite dimensional sieve spaces, see Chen (2005) and Shen (1997, 1998).

\(^{18}\)Note that with some abuse of notation, when we consider \(\tilde{\mathcal{H}}_{l,\delta_n}\) as a sequence of sets indexed by \(l\), we treat \(\tilde{h}(\cdot, \theta)\) is fixed. In other words, the degree of approximation of the sieve space for the estimation stage does not have to agree with that of the sieve space for the inference, which we can let arbitrary large regardless of the sample size.
such as the bounds do not involve the optimization. Moreover, we do not require some slackness variable note that this is not too demanding compared to the estimation stage since the construction of since we construct the bounds over the whole inﬁnite dimensional parameter space. However, we critical values that are constructed by the bootstrap procedure described in the following section.

where $b(j; n; \Lambda)$ so that while reducing the functional space for such construction from $\Lambda \times H_0$ to $\Lambda \times H_0$, we have comparable distributions across different $j; m$ have been non-negative $B_n, B_m$ are non-negative $B_n, B_m$ are non-negative

Alternative CI2: CI over the inﬁnite dimensional space around the true value $h_0 (\Theta \times H_{\hat{\delta}_n})$: $\hat{\beta}_{n,U} = \sup \{ \beta_n(\alpha) : \alpha \in \Theta \times H_{\hat{\delta}_n}, \hat{c}_{n,U}(j, \hat{\gamma}_{n,j,m}, \alpha) \geq 0, \forall (j, m) \in \hat{B}_{n,U} \}$ $\hat{\beta}_{n,L} = \inf \{ \beta_n(\alpha) : \alpha \in \Theta \times H_{\hat{\delta}_n}, \hat{c}_{n,L}(j, \hat{\gamma}_{n,j,m}, \alpha) \geq 0, \forall (j, m) \in \hat{B}_{n,L} \}$, (19)

Alternative CI3: CI over the inﬁnite dimensional sieve space around the true value $h_0 (\Theta \times H_{\hat{H}_{\delta_n}})$: $\hat{\beta}_{n,U} = \sup \{ \beta_n(\alpha) : \alpha \in \Theta \times H_{\hat{\delta}_n}, \hat{c}_{n,U}(j, \hat{\gamma}_{n,j,m}, \alpha) \geq 0, \forall (j, m) \in \hat{B}_{n,U} \}$ $\hat{\beta}_{n,L} = \inf \{ \beta_n(\alpha) : \alpha \in \Theta \times H_{\hat{\delta}_n}, \hat{c}_{n,L}(j, \hat{\gamma}_{n,j,m}, \alpha) \geq 0, \forall (j, m) \in \hat{B}_{n,L} \}$, (20)

Alternative CI3: CI over the inﬁnite dimensional sieve space around the true value $h_0 (\Theta \times H_{\hat{H}_{\delta_n}})$: $\hat{\beta}_{n,U} = \sup \{ \beta_n(\alpha) : \alpha \in \Theta \times H_{\hat{\delta}_n}, \hat{c}_{n,U}(j, \hat{\gamma}_{n,j,m}, \alpha) \geq 0, \forall (j, m) \in \hat{B}_{n,U} \}$ $\hat{\beta}_{n,L} = \inf \{ \beta_n(\alpha) : \alpha \in \Theta \times H_{\hat{\delta}_n}, \hat{c}_{n,L}(j, \hat{\gamma}_{n,j,m}, \alpha) \geq 0, \forall (j, m) \in \hat{B}_{n,L} \}$, (21)

For Alternative CI3, we require $l \geq n$ so that $H_{l,\delta_n} \supseteq \hat{H}$. This guarantees that $\hat{\beta}_{l,U} \geq \hat{\beta}_{n,U}$ and $\hat{\beta}_{l,L} \leq \hat{\beta}_{n,L}$, which are necessary to justify the proposed CI asymptotically. Here $\hat{c}_{n,U}$ is an upper bound on $c_0$ for those $(y(j), \hat{\gamma}_{n,j,m})$ sets for which $(j, m)$ belongs to $\hat{B}_{n,U}$ at a particular value $\alpha = \hat{\alpha}_{n,U}$ and, analogously, $\hat{c}_{n,L}$ is an upper bound on $c_0$ for $(j, m) \in \hat{B}_{n,L}$ at $\alpha = \hat{\alpha}_{n,L}$. Thus, $\hat{c}_{n,U}$ and $\hat{c}_{n,L}$ are given as real random functions on $A$:

$$\hat{c}_{n,U}(j, \hat{\gamma}_{n,j,m}, \alpha) = \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \alpha) + \hat{w}_n(j, \hat{\gamma}_{n,j,m}, \alpha) \lambda^*_n(j, m, a) / \sqrt{n}$$

$$\hat{c}_{n,L}(j, \hat{\gamma}_{n,j,m}, \alpha) = \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \alpha) + \hat{w}_n(j, \hat{\gamma}_{n,j,m}, \alpha) \lambda^*_n(j, m, a) / \sqrt{n},$$ (22)

where $\hat{w}_n(j, \gamma, \alpha)$ is a positive weight function and $\lambda^*_n(j, m, a) \& \lambda^*_n(j, m, a)$ are non-negative critical values that are constructed by the bootstrap procedure described in the following section.

Now we compare the three alternative CIs. Alternative CI1 is most computationally demanding since we construct the bounds over the whole inﬁnite dimensional parameter space. However, we note that this is not too demanding compared to the estimation stage since the construction of the bounds do not involve the optimization. Moreover, we do not require some slackness variable such as $\delta_n$. For Alternative CI2, we include the slackness variable $\delta_n$ in constructing $\hat{\beta}_{n,L}$ and $\hat{\beta}_{n,U}$ so that while reducing the functional space for such construction from $\Theta \times H$ to $\Theta \times H_{\hat{\delta}_n}$, we

---

19The weight function is used to make $\hat{c}_n(j, \hat{\gamma}_{n,j,m}, \alpha) / \hat{w}_n(j, \hat{\gamma}_{n,j,m}, \alpha)$ have comparable distributions across different $(j, m)$. Examples of such weight functions can be found in ABJ.
make sure \( \Theta_+ \times \mathcal{H}_+ \) is included in \( \Theta \times \hat{\mathcal{H}}_{\delta_n} \) with probability approaching to one. This is critical in justifying the proposed CI asymptotically. In practice, we can let \( \delta_n \) be some fixed number since fixed \( \delta_n \) does not affect the asymptotics for the CI. We still have a valid CI with a fixed \( \delta_n \). However, the choice of \( \delta_n \) will affect the cost of computation and thus we want to let \( \delta_n \) be small as long as the sample size is relatively large. Alternative CI3 requires the least computation among three alternatives since the CI is constructed over the finite dimensional sieve space but we need to admit the possibility that the coverage probability is smaller than \( 1 - a \) with a finite \( l \). However, in practice, we can let \( l \) be arbitrary large noting the sieve space for the construction of the bounds can be larger than that for the estimation stage and thus the size of data does not restrict the smoothness of the sieve space for the inference stage. In consequence, we can make the smallest value of coverage probability arbitrary close to \( 1 - a \).

Now we consider how to obtain the critical values \( \lambda_{n,U}^*(j, m, a) \) and \( \lambda_{n,L}^*(j, m, a) \) using the standard nonparametric bootstrap.

### 5.1 Bootstrap Critical Values

Here we briefly review the bootstrap procedure to obtain the critical values following ABJ. Most of their discussions hold by replacing their \( \theta \) and \( \hat{\theta} \) with the infinite dimensional parameter \( \alpha \) and \( \hat{\alpha} \). The purpose of our paper is to provide some conditions under which the bootstrap critical values can be justified asymptotically for the semiparametric case.

Let \( \{(Y_i^*, X_i^*) : i = 1, \ldots, n\} \) denote a standard nonparametric bootstrap sample conditional on the original sample \( \{(Y_i, X_i) : i = 1, \ldots, n\} \). First, we obtain the bootstrap conditional probabilities implied by the model using the bootstrap sample such that \( P^*(g^{(j)}|X_i^*, \theta, h(\tau)) = P(g^{(j)}|X_i^*, \theta, h(\tau)) \). Similarly, we define \( \hat{c}_n(j, \gamma, \alpha), \hat{\gamma}_{n,j,m}, \hat{\Gamma}_n, \) and \( \hat{w}_n(j, \gamma, \alpha) \) using the bootstrap sample as we define \( \hat{c}(j, \gamma, \alpha), \hat{\gamma}_{n,j,m}, \hat{\Gamma}_n, \) and \( \hat{w}_n(j, \gamma, \alpha) \), respectively. Define

\[
D_{n,U}^*(j, m) = \sqrt{n} \left( \frac{\hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_n,U) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \alpha_{n,U})}{\hat{w}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_n,U)} \right)
\]

\[
D_{n,L}^*(j, m) = \sqrt{n} \left( \frac{\hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_n,L) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \alpha_{n,L})}{\hat{w}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_n,L)} \right).
\]

Note that when we construct \( D_{n,U}^*(j, m) \) and \( D_{n,L}^*(j, m) \), in the arguments of \( \hat{c}_n(\cdot) \), we use \( \hat{\alpha}_{n,U} \) and \( \hat{\alpha}_{n,L} \) not \( \alpha_{n,U} \) and \( \alpha_{n,L} \), respectively. This ensures \( E^* \left[ c_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_n,U) - c_n(j, \hat{\gamma}_{n,j,m}, \alpha_{n,U}) \right] = 0 \) and \( E^* \left[ c_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_n,L) - c_n(j, \hat{\gamma}_{n,j,m}, \alpha_{n,L}) \right] = 0 \) where \( E^* \) is the expectation operator taken conditional on the original sample. We denote \( P^*(\cdot) \) to be the probability with respect to the bootstrap sample conditional on the original sample. Now we obtain the critical values \( \lambda_{n,U}^*(j, m, a) \) for \( (j, m) \in \hat{B}_{n,U} \) and \( \lambda_{n,L}^*(j, m, a) \) for \( (j, m) \in \hat{B}_{n,L} \) as non-negative constants satisfying the condition

\[
P^* \left( D_{n,U}^*(j, m) + \lambda_{n,U}^*(j, m, a) \geq 0 \text{ for } (j, m) \in \hat{B}_{n,U} \right) \quad \text{and} \quad D_{n,U}^*(j, m) + \lambda_{n,U}^*(j, m, a) \geq 0 \text{ for } (j, m) \in \hat{B}_{n,L} \right) \quad = 1 - a
\] (23)
and the same condition with $U$ and $L$ interchanged.\textsuperscript{20} We summarize the procedure to construct the CIs.

1. Obtain the following objects in the order that they are presented:
   \[ \hat{\Gamma}_n, \tilde{c}_n(j, \tilde{\gamma}_{n,j,m}, \alpha), \hat{A}_n, \hat{\beta}_{n,U}, \hat{\beta}_{n,L}, \hat{\alpha}_{n,U}, \hat{\alpha}_{n,L}, \hat{B}_n, \tilde{c}_n(j, \tilde{\gamma}_{n,j,m}, \alpha), \tilde{w}_n(j, \tilde{\gamma}_{n,j,m}, \alpha) \]

2. Obtain the bootstrap critical values as described in this section:
   \[ \lambda_{n,U}^*(j, m, a) \text{ for } (j, m) \in \hat{B}_n \text{ and } \lambda_{n,L}^*(j, m, a) \text{ for } (j, m) \in \hat{B}_n \]

3. Construct the confidence intervals defined in (17):
   \[ \hat{c}_n(j, \hat{\gamma}_{n,k,m}, \hat{\alpha}_n), \hat{c}_n(j, \hat{\gamma}_{n,k,m}, \hat{\alpha}_n), \hat{\beta}_{n,U}, \text{ and } \hat{\beta}_{n,L} \text{ from (22) and (19), (20), or (21), respectively.} \]

Under some higher level assumptions presented in the Appendix, the following theorem justifies the confidence intervals suggested in (17) asymptotically for all the three alternatives.

**Theorem 5.1** Suppose Assumptions B.1-B.6 in the Appendix are satisfied. Further suppose Assumptions 4.1, 4.2, 4.3, and 4.7 hold. Then, Alternative CI1 satisfies (18) with \( \hat{\beta}_{n,U} = \bar{\beta}_{n,U}^{(1)} \) and \( \hat{\beta}_{n,L} = \bar{\beta}_{n,L}^{(1)} \). Further suppose \( d(\hat{A}_n, A_+) \rightarrow 0 \). Then, (18) with \( \bar{\beta}_{n,U} = \beta_{n,U}^{(2)} \) and \( \bar{\beta}_{n,L} = \beta_{n,L}^{(2)} \) holds for Alternative CI2. Further suppose \( \mathcal{H}_l \subseteq \mathcal{H}_{l+1} \subseteq \mathcal{H} \) for all \( l \geq 1 \). Then, we have
\[ \liminf_{l,n \rightarrow \infty, l \geq n} P(\beta_0 \in CI_n(1 - a)) = \liminf_{l,n \rightarrow \infty, l \geq n} P(\beta_{t,n}^{(3)} \leq \beta_0 \leq \beta_{t,n,U}^{(3)}) \geq 1 - a \text{ for Alternative CI3.} \]

6 **Concluding Remarks**

This paper considers estimation and inference of parameters in discrete games with multiple equilibria, without using an equilibrium selection rule, while the game model can contain infinite dimensional parameters. In particular, we adopt a set inference approach popularized recently. Noting the literature only allows for finite dimensional parameters in the model even though infinite dimensional parameter is naturally included in the model or misspecification of a fully parametric model is concerned, this paper extends a current literature to a set inference with infinite dimensional parameters where a consistent profiled estimator of infinite dimensional parameters is available. A consistent set estimator and confidence intervals are provided. Examples of signaling games with discrete types where the type distribution is nonparametrically specified and entry-exit games with partially linear payoffs functions are considered.

\[ \text{\textsuperscript{20}} \text{Though the requirement of (23) for } \lambda_{n,L}^*(\cdot, \cdot, \cdot) \text{ and } \lambda_{n,U}^*(\cdot, \cdot, \cdot) \text{ is enough to justify the CI asymptotically, it does not uniquely determine these values. Also in principle, these bootstrapped critical values can be obtained analytically but in practice, they need to be simulated. See some related issues in ABJ. We omit these discussions since our focus here is to provide an asymptotic justification for the semiparametric version of the CI proposed in ABJ.} \]
In this paper, we note that achieving a consistent point estimation often requires some information reduction (for example, redefinition of outcome spaces). Due to this less use of information than available in the model, we may end up a point estimator with larger variances and have wider confidence intervals than those of the set estimator using the full information in the model. This finding justifies the use of the set inference even though we can achieve a consistent point estimation in some cases. It is also an interesting future research to compare these two alternatives: CI from the point estimation with the usage of less information vs. CI from the set estimation with the usage of the full information.
Appendix

A Consistency Proofs

For any real functional $c$ on $\mathcal{I}_f \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$ and any collection of $\sum_{j=1}^{J} \mathcal{M}_j$ subsets of $\mathbb{S}(X)$, define $\Theta(c, \Gamma, h) = \{ \theta \in \Theta : c(j, \gamma_{j,m}, \theta, h(\cdot), \theta) \geq 0, \forall (j, m) \in \mathcal{I}_f, \mathcal{M}_j \}$ and note that $\Theta(c_0, \Gamma_0, h_0) = \Theta_+$. To prove Theorem 4.1, we need the following lemma which extends Lemma 4 in ABJ to the semiparametric case.

Lemma A.1 Under Assumptions 4.4 and 4.5 (ii), $\rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) \rightarrow 0$ as $c \rightarrow c_0, \Gamma \rightarrow \Gamma_0, h \rightarrow h_0$ under $\| \cdot \|_{\mathcal{U}_n}, \| \cdot \|_{\mathcal{H}}$.

Proof. For any $\theta \in \Theta$, we have

$$\limsup_{(c, \Gamma, h) \rightarrow (c_0, \Gamma_0, h_0)} \left| \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c(j, \gamma_{j,m}, \theta, h) - \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c_0(j, \gamma_{j,m}, \theta, h) \right|$$

$$\leq \limsup_{c \rightarrow c_0} \sup_{\Gamma \in \Gamma_{all}} \sup_{h \in \mathcal{H}_{\delta_n}} \left| \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c(j, \gamma_{j,m}, \theta, h) - c_0(j, \gamma_{j,m}, \theta, h) \right| = 0,$$

because $c \rightarrow c_0$ with respect to the uniform metric over $\mathcal{I}_f \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$. It follows that for any $\theta \in \Theta$,

$$\liminf_{(c, \Gamma, h) \rightarrow (c_0, \Gamma_0, h_0)} \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c(j, \gamma_{j,m}, \theta, h) = \liminf_{\Gamma \rightarrow \Gamma_0, h \rightarrow h_0} \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c_0(j, \gamma_{j,m}, \theta, h).$$

Next, consider that for any $\alpha \in \Theta \times \mathcal{H}_{\delta_n}$, we have

$$\limsup_{\Gamma \rightarrow \Gamma_0} \left| \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c_0(j, \gamma_{j,m}, \alpha) - \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c_0(j, \gamma_{0,j,m}, \alpha) \right|$$

$$\leq \limsup_{\Gamma \rightarrow \Gamma_0} \max_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} 2 \int q_{\gamma_{j,m}}(x) - q_{\gamma_{0,j,m}}(x) \ dF_X(x) = 0,$$

where the inequality holds by the definition of $c_0$ in (3) and the equality holds by the definition of $\Gamma \rightarrow \Gamma_0$ in (12) and the Cauchy-Schwarz inequality. Also note for any $\Gamma \in \Gamma_{all}$ and $\theta \in \Theta$, we have

$$\limsup_{h \rightarrow h_0} \left( \sup_{x \in \mathbb{S}(X)} \max_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} q_{\gamma_{j,m}}(x) \right) \| h - h_0 \|_{\mathcal{H}} = 0,$$

where the inequality holds by the construction of $c_0(\cdot)$ and Assumption 4.4 (i) and the equality holds by the definition of the metric $\| \cdot \|_{\mathcal{H}}$ and Assumption 4.4 (ii). It follows that for any $\theta \in \text{int}(\Theta_+)$,

$$\liminf_{\Gamma \rightarrow \Gamma_0, h \rightarrow h_0} \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c_0(j, \gamma_{j,m}, \theta, h) = \min_{(j,m) \in \mathcal{I}_f, \mathcal{M}_j} c_0(j, \gamma_{0,j,m}, \theta, h_0) > 0,$$

where the last result holds by Assumption 4.5 (ii). From (24) and (25), we conclude that for any $\theta \in \text{int}(\Theta_+)$, it also holds that $\theta \in \Theta(c, \Gamma, h)$ for $(c, \Gamma, h)$ sufficiently close to $(c_0, \Gamma_0, h_0)$. 

20
Now suppose \( \rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) \to 0 \) as \((c, \Gamma, h) \to (c_0, \Gamma_0, h_0)\). Then, by definition of
\[
\rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) = \sup_{\theta \in \Theta(c_0, \Gamma_0, h_0)} \rho(\Theta(c, \Gamma, h)),
\]
there exists (i) a constant \( \varepsilon > 0 \), (ii) a sequence of functions on \( \mathcal{I}_j \times \mathcal{M}_j \times \Theta \times \mathcal{H}_\delta_n \), \( \{c_j : j \geq 1\} \), and a sequence of collections \( \sum_{j=1}^J \mathcal{M}_j \) in \( \mathcal{M}_j \), \( \{j : j \geq 1\} \), s.t. \( (c_j, \Gamma_j, h_j) \to (c_0, \Gamma_0, h_0) \), and (iii) a sequence of parameters \( \{\theta_j \in \Theta(c_0, \Gamma_0, h_0) : j \geq 1\} \) s.t. \( \rho(\theta_j | \Theta(c_j, \Gamma_j, h_j)) \geq \varepsilon \) for all \( j \geq 1 \). The sequence \( \{\theta_j \in \Theta(c_0, \Gamma_0, h_0) : j \geq 1\} \) has a subsequence, say \( \{\theta_{j_l} : l \geq 1\} \), that converges to a point \( \Theta(c_0, \Gamma_0, h_0) \) because \( \Theta(c_0, \Gamma_0, h_0) \) is compact (This is because \( \Theta \) is compact and \( \Theta(c_0, \Gamma_0, h_0) \) is defined from the non-strict inequality.). That is, \( \rho(\theta_{j_l} | \Theta(c_{j_l}, \Gamma_{j_l}, h_{j_l}))-\varepsilon/2 \)
by the triangle inequality. Thus, for all \( l \) sufficiently large,
\[
\rho(\theta_{j_l} | \Theta(c_{j_l}, \Gamma_{j_l}, h_{j_l})) \geq \varepsilon/2.
\]
If \( \theta_{j_l} \in \Theta(c_{j_l}, \Gamma_{j_l}, h_{j_l}) \), (26) contradicts to the fact that for any \( \theta \in \Theta(c_{j_l}, \Gamma_{j_l}, h_{j_l}) \) sufficiently close to \( (c_0, \Gamma_0, h_0) \), \( \rho(\theta | \Theta(c_0, \Gamma_0, h_0)) \) is a contradiction. We conclude that \( \rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) \to 0 \) as \((c, \Gamma, h) \to (c_0, \Gamma_0, h_0)\). \( \blacksquare \)

### A.1 Proof of Theorem 4.1

We first prove part (i) by extending a consistency result of a class of extremum estimator. Under Assumption 4.4, \( Q(\theta, h) \) defined in (7) is continuous (with respect to the metric \( \|\cdot\|_s \)). Note that this holds even though \( Q(\theta, h) \) contains an indicator function because \( \|b(\theta, h)\| \cdot 1[b(\theta, h) \leq 0] \) is continuous as long as \( b(\theta, h) \) is continuous (with respect to the metric \( \|\cdot\|_s \)), which follows from the fact that \( |b(\theta_1, h_1)| \cdot 1[b(\theta_1, h_1) \leq 0] - |b(\theta_2, h_1)| \cdot 1[b(\theta_2, h_1) \leq 0] \leq |b(\theta_1, h_1)| - |b(\theta_2, h_1)| \) for all \( (\theta_1, h_1), (\theta_2, h_2) \). Because \( Q(\theta, h) \) is continuous and \( \Theta \times \mathcal{H} \) is compact under \( \|\cdot\|_s \), \( Q(\cdot, \cdot) \) attains its minimum value zero at points in the set \( \mathcal{A}_+ \) by definition in (8). Now we claim that for all \( \varepsilon > 0 \), there exists a \( \delta > 0 \) such that
\[
\inf_{\theta \in \mathcal{E}(\Theta_+, \varepsilon), \theta \in \Theta} Q(\theta, h) \geq \delta > 0 \tag{27}
\]
where \( \mathcal{E}(\Theta_+, \varepsilon) = \{\theta \in \Theta : \rho(\theta | \Theta_+) \leq \varepsilon\} \). Suppose not. Then, for some \( \varepsilon > 0 \) and \( h \) s.t. \( \|h - h_0\|_\mathcal{H} \leq \delta_n \), there is a sequence \( \{\theta_l \in \Theta \setminus \mathcal{E}(\Theta_+, \varepsilon) : l \geq 1\} \) for which \( \lim_{l \to \infty} Q(\theta_l, h(\cdot, \theta_l)) = 0 \). Because \( \Theta \) is compact and \( \mathcal{E}(\Theta_+, \varepsilon) \) is open, the set \( \Theta \setminus \mathcal{E}(\Theta_+, \varepsilon) \) is compact. Hence, \( \{\theta_l : l \geq 1\} \) has a convergent subsequence, say \( \{\theta_{l_j} : j \geq 1\} \), that converges to a point \( \Theta \setminus \mathcal{E}(\Theta_+, \varepsilon) \), say \( \theta_\infty \). Continuity of \( Q(\cdot, \cdot) \) and \( (\cdot, \cdot) \in \mathcal{H}_\delta_n \) in \( \theta \) imply that \( Q(\theta_\infty, h(\cdot, \theta_\infty)) = \lim_{j \to \infty} Q(\theta_{l_j}, h(\cdot, \theta_{l_j})) = 0 \). This implies that \( \theta_\infty \in \Theta_+ \), which is a contradiction. This proves (27). From \( \mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H} \) by Assumption 4.3 (ii) (note also \( \mathcal{H}_n \) is compact) and the fact that \( Q(\theta, h) \) and \( h(\cdot, \theta) \) are continuous in \((\theta, h)\) and \( \theta \), respectively, we note that there is an \( N \geq 1 \) such that \( \inf_{\theta \in \Theta} Q(\theta) \geq \inf_{\theta \in \Theta} Q(\theta) \geq \inf_{\theta \in \Theta} Q(\theta) \geq \inf_{\theta \in \Theta} Q(\theta) \).
\[ \mathcal{H}_{\delta_n} \equiv \{ h \in \mathcal{H} : ||h(\cdot, \theta) - h_0(\cdot, \theta)||_{\mathcal{H}} \leq \delta_n, \theta \in \Theta \} \] and \[ \mathcal{H}_{n, \delta_n} \equiv \{ h \in \mathcal{H}_n : ||h(\cdot, \theta) - h_0(\cdot, \theta)||_{\mathcal{H}} \leq \delta_n, \theta \in \Theta \} \] with \( \delta_n = o(1) \). From this result and (27), it follows that

\[ \inf_{g \in \mathcal{E}(\Theta, e), \theta \in \Theta, h \in \mathcal{H}_{n, \delta_n}} Q(\alpha) \geq \delta > 0 \text{ for all } n \geq N. \] (28)

This is a version of the identification condition. Now we derive the uniform convergence of \( Q_n(\theta, h) \) defined in (6) to \( Q(\theta, h) \) uniformly over \( \Theta \times \mathcal{H}_{\delta_n} \). Consider

\[ \sup_{\theta \in \Theta, h \in \mathcal{H}_{\delta_n}} \left| Q_n(\theta, h) - Q(\theta, h) \right| \leq \sup_{\theta \in \Theta, h \in \mathcal{H}_{\delta_n}} \left\{ \max_{(j,m) \in I_{j,M}} \left| \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \theta, h) - c_0(j, \gamma_{0,j,m}, \theta, h) \right| \right\} \]

\[ \leq \sup_{\theta \in \Theta, h \in \mathcal{H}_{\delta_n}} \left\{ \max_{(j,m) \in I_{j,M}} \left| \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \theta, h) - c_0(j, \gamma_{0,j,m}, \theta, h) \right| \right\} \]

\[ + \sup_{\theta \in \Theta, h \in \mathcal{H}_{\delta_n}} \left\{ \max_{(j,m) \in I_{j,M}} \left| c_0(j, \gamma_{0,j,m}, \theta, h) - c_0(j, \gamma_{0,j,m}, \theta, h) \right| \right\} \]

\[ \leq \sup_{(j, \gamma, \theta, h) \in I_{j,M} \times \Theta} \left| \hat{c}_n(j, \gamma, \theta, h) - c_0(j, \gamma, \theta, h) \right| + 2 \max_{(j,m) \in I_{j,M}} \int_{x} q_{\gamma_{n,j,m}}(x) - q_{\gamma_{0,j,m}}(x) \right| dF_X(x) \to 0 \] (29)

where the first inequality holds by the definitions of \( Q_n(\theta, h) \) and \( Q(\theta, h) \), the second inequality is from the triangle inequality, the third inequality is obtained using the definition of \( c_0(j, \gamma, \theta, h) \) in (3), and last result holds by Assumptions 4.6 (i) and 4.7 using the definitions of the metric \( \| \cdot \|_{\mathcal{U}_n} \) and \( \| \cdot \| \). Now we are ready to prove Theorem 4.1 (i). Note that the set \( \Theta_+ \) is not empty by Assumption 4.2 and \( \tilde{\Theta}_n \) is not empty by construction. Let \( \alpha_{n+} = (\theta_n, h_{n+}) \) denote some element of \( \Theta_+ \times \{ h \in \mathcal{H}_n : \|h(\cdot, \theta) - h_0(\cdot, \theta)\|_{\mathcal{H}} \leq \delta_n, \theta \in \Theta_+ \} \). Then, there exist \( \alpha_+ \in \mathcal{A}_+ \) such that \( \|\alpha_{n+} - \alpha_+\| = o(1) \) and thus \( Q(\alpha_{n+}) - \delta/2 \leq Q(\alpha_+) \) for \( n \geq 3N \) by the continuity of \( Q(\cdot, \cdot) \). It follows that

\[ -Q(\alpha_{n+}) \geq -\delta/2 \text{ for all } n \geq 3N \] (30)

since \( Q(\alpha_+) = 0 \) for every \( \alpha_+ \in \mathcal{A}_+ \). (28) and the fact that \( \Theta_+ \) and \( \tilde{\Theta}_n \) are not empty imply that for all \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that

\[ P \left( \rho \left( \Theta_n \cap \Theta_+ \right) > \varepsilon \right) = P \left( \Theta_n \cap \Theta_+ \right) \neq 0 \]

\[ \leq P \left( \sup_{\theta \in \Theta_n, h \in \mathcal{H}_{\delta_n}} Q(\alpha) \geq \delta \right) = P \left( \sup_{\theta \in \Theta_n, h \in \mathcal{H}_{\delta_n}} (Q(\alpha) - Q_n(\alpha) + Q_n(\alpha)) \geq \delta \right) \]

\[ \leq P \left( \sup_{\theta \in \Theta_n, h \in \mathcal{H}_{\delta_n}} (Q(\alpha) - Q_n(\alpha) + Q_n(\alpha + \alpha(1)) \geq \delta/2 \right) \]

\[ \leq P \left( \sup_{\theta \in \Theta_n, h \in \mathcal{H}_{\delta_n}} |Q(\alpha) - Q_n(\alpha)| \geq \delta/2 \right) \to 0 \]

where the first inequality holds by (28), the second inequality holds because \( i \) \( Q_n(\alpha) \leq Q_n(\alpha_{n+}) \) for each \( \alpha \in \mathcal{A}_n \) since each \( \alpha \in \mathcal{A}_n \) minimizes \( Q_n(\alpha) \) over \( \Theta \times \mathcal{H}_{\delta_n} \) and (ii) \( -Q(\alpha_{n+}) + \alpha(1) \geq -\delta/2 \) as noted in (30). The last result comes from (29) and \( \mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H} \) for all \( n \). This completes the proof of part (i).

Now we turn to part (ii). Suppose Assumption 4.5 (ii) holds. Then, by Lemma A.1, given \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that \( \|c - c_0\|_{\mathcal{U}_n} < \delta, \|\Gamma - \Gamma_0\| < \delta, \) and \( \|h - h_0\|_{\mathcal{H}} < \delta \) imply \( \rho(\Theta(c_0, \Gamma_0, h_0)|\Theta(c, \Gamma, h)) < \varepsilon \) as \( n \to \infty \). It follows that

\[ P \left( \rho \left( \Theta(c_0, \Gamma_0, h_0)|\Theta(\hat{c}_n, \hat{\Gamma}_n, \hat{h}) \right) < \varepsilon \right) \geq P \left( \|\hat{c}_n - c_0\|_{\mathcal{U}_n} < \delta, \|\hat{\Gamma}_n - \Gamma_0\| < \delta, \|\hat{h} - h_0\|_{\mathcal{H}} \right) \to 1, \] (31)
where the convergence holds by Assumption 4.3 (iii), 4.6, and 4.7. From $\Theta\left(\tilde{c}_n, \Gamma_n, \tilde{h}\right) \subset \tilde{\Theta}_n$, it follows that

$$P\left(\rho\left(\Theta(c_0, \Gamma_0, h_0) | \tilde{\Theta}_n\right) < \varepsilon\right) \geq P\left(\rho\left(\Theta(c_0, \Gamma_0, h_0) | \Theta_n(\tilde{c}_n, \Gamma_n, \tilde{h})\right) < \varepsilon\right).$$  

(32)

Combining (31) and (32), we establish the part (ii) of Theorem 4.1 under Assumption 4.5 (ii).

Next, suppose Assumption 4.5 (i) holds. Then, $\rho(\Theta | \tilde{\Theta}_n) = \rho(\{ \theta_0 \} | \tilde{\Theta}_n) \leq \rho(\tilde{\Theta}_n | \{ \theta_0 \}) \rightarrow 0$ where the inequality holds since (i) the distance from a point to a non-empty set is less than or equal to the distance from the set to the point by definition of $\rho(\cdot | \cdot)$ and (ii) the set $\tilde{\Theta}_n$ is not empty. The convergence result holds by the proof of part (i).

Now we show that Assumption 4.3 (iv) and $d(\tilde{\Theta}_n, \Theta_+) \rightarrow 0$ imply $d(\tilde{A}_n, A_+) \rightarrow 0$. Note for any $\alpha = (\theta, h(\cdot, \theta)) \in \tilde{A}_n$, we can find $\exists \alpha' = (\theta', h_0(h, \theta')) \in A_+$ s.t. $\|\alpha - \alpha\|_s = \|\theta - \theta\|_E + \|h(\cdot, \theta) - h_0(\cdot, \theta')\|_{\mathcal{H}_t} \rightarrow 0$ since $\|\theta - \theta\|_E \rightarrow 0$ by $d(\tilde{\Theta}_n, \Theta_+) \rightarrow 0$ and since

$$\|h(\cdot, \theta) - h_0(\cdot, \theta')\|_{\mathcal{H}_t} \leq \|h(\cdot, \theta) - h_0(\cdot, \theta)\|_{\mathcal{H}_t} + \|h_0(\cdot, \theta') - h_0(\cdot, \theta)\|_{\mathcal{H}_t} \rightarrow 0$$

because $h(\cdot, \theta) \in \mathcal{H}_n \subset \mathcal{H}_{\delta_{n}}$, by definition and because $h_0(\cdot, \theta)$ is continuous in $\theta$ and $d(\tilde{\Theta}_n, \Theta_+) \rightarrow 0$. It follows that $\rho(\tilde{A}_n | A_+) \rightarrow 0$. Similarly for any $\alpha = (\theta, h(\cdot, \theta)) \in A_+$, we can find $\exists \alpha' = (\theta', h(\cdot, \theta')) \in \tilde{A}_n$ such that $\|\alpha - \alpha\|_s \rightarrow 0$, which implies $\rho(A_+ | \tilde{A}_n) \rightarrow 0$. This completes the proof of Theorem 4.1.

A.2 Consistency of $\hat{\beta}_{n,U}$ and $\hat{\beta}_{n,L}$ (Proof of Corollary 4.1)

The following proof is essentially the same with that of Theorem 2 in ABJ and can be omitted.

For any two sets of real numbers $B_1$ and $B_2$, let $b_j^* = \sup \{b \in B_j\}$ for $j = 1, 2$ and note that $|b_1^* - b_2^*| \leq d(B_1, B_2)$. Now define $\hat{B}_n = \{\beta_n(\alpha) : \alpha \in \tilde{A}_n\}$ and $B_{n,+} = \{\beta_n(\alpha) : \alpha \in A_+\}$. Then, from the result above, it follows that

$$\|\hat{\beta}_{n,U} - \beta_{n,U}\| \leq d(\hat{B}_n, B_{n,+})$$  

(33)

by definitions of $\hat{\beta}_{n,U}$ and $\beta_{n,U}$ given in (13) and (14), respectively. Now we note that

$$P(d(\hat{B}_n, B_{n,+}) > \varepsilon) \leq P(\rho(\hat{B}_n | B_{n,+}) > \varepsilon) + P(\rho(B_{n,+} | \hat{B}_n) > \varepsilon)$$  

(34)

from the definition of $d(\cdot, \cdot)$. Let $p_\delta = P(d(\tilde{A}_n, A_+) > \delta)$ and consider

$$P(\rho(\hat{B}_n | B_{n,+}) > \varepsilon) \leq P(\rho(\hat{B}_n | B_{n,+}) > \varepsilon, d(\tilde{A}_n, A_+) \leq \delta) + p_\delta = P(\sup_{\alpha \in \tilde{A}_n} \inf \{\|\beta_n(\alpha) - \beta_n(\alpha_+)\| : \alpha_+ \in A_+\} > \varepsilon, d(\tilde{A}_n, A_+) \leq \delta) + p_\delta$$  

(35)

where the equality holds by definitions of $\rho(\cdot, \cdot)$, $\hat{B}_n$, and $B_{n,+}$. Now consider that if $d(\tilde{A}_n, A_+) \leq \delta$, then $\rho(\tilde{A}_n | A_+) \leq \delta$ and for any $\alpha \in \tilde{A}_n$, there exists $\alpha_+ \in A_+$ such that $\|\alpha - \alpha_+\|_s \leq \delta$. It follows that

$$\sup_{\alpha \in \tilde{A}_n} \inf \{\|\beta_n(\alpha) - \beta_n(\alpha_+)\| : \alpha_+ \in A_+\} \leq \sup_{\alpha \in \tilde{A}_n, \|\alpha - \alpha_+\|_s \leq \delta} \|\beta_n(\alpha) - \beta_n(\alpha_+)\| \leq \sup_{\|\alpha_1 - \alpha_2\|_s \leq \delta} \|\beta_n(\alpha_1) - \beta_n(\alpha_2)\|$$

(36)

$^{21}$To prove this, suppose $b_1^* > b_2^*$. Then, $|b_1^* - b_2^*| = \rho(b_1^* | B_2) \leq \sup_{b \in B_1} \rho(b | B_2) = \rho(B_1 | B_2) \leq d(B_1, B_2)$. Analogously, we can show this is true for $b_1^* < b_2^*$. 

23
where the second inequality holds since \(\|\alpha - \alpha_+\|_\infty \leq \delta\) for any \(\alpha \in \tilde{A}_n\). From (35), (36), and Assumption 4.8, it follows that for sufficient large \(n\), \(P(\sup_{\alpha \in \tilde{A}_n} \inf \{|\beta_n(\alpha) - \beta_n(\alpha_+)| : \alpha_+ \in A_+\} > \varepsilon, d(\tilde{A}_n, A_+) \leq \delta) \leq \varepsilon\) and thus
\[
P(\rho(\tilde{B}_n|B_{n,+}) > \varepsilon) \leq \varepsilon + P(d(\tilde{A}_n, A_+) > \delta). \tag{37}
\]

An analogous argument provides the same result as (37) but replacing \(\rho(\tilde{B}_n|B_{n,+})\) with \(\rho(B_{n,+}|\tilde{B}_n)\) and \(\rho(\tilde{A}_n|A_+)\), respectively. To be precise, \(P(\rho(B_{n,+}|\tilde{B}_n) > \varepsilon) \leq \varepsilon + P(d(\tilde{A}_n, A_+) > \delta)\). Combining this with (33), (34), and (37), we obtain \(P(|\tilde{\beta}_{n,U} - \beta_{n,U}| > \varepsilon) \leq P(d(\tilde{B}_n, B_{n,+}) > \varepsilon) \leq 2\varepsilon + 2P(d(\tilde{A}_n, A_+) > \delta)\). This proves Corollary 4.1 since \(\varepsilon > 0\) is arbitrary and \(P(d(\tilde{A}_n, A_+) > \delta) \rightarrow 0\) by Theorem 4.1.

### A.3 Estimation of the Type Distribution: Proof of Proposition 4.1

We derive the consistency for the set estimator of the type distribution parameter. We let \(p_{n,n}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} L(h(z_i))\) with \(\alpha = (\theta, h) \in \hat{A}_n\) and let \(p_{n}(\alpha) = E[L(h)]\) with \(\alpha = (\theta, h) \in A_+\). Now note that for any \(\varepsilon > 0\),
\[
P(d(\hat{P}_n, P_+) > \varepsilon) \leq P(\rho(\hat{P}_n|P_+) > \varepsilon) + P(\rho(P_+|\hat{P}_n) > \varepsilon) \tag{38}
\]
from the definition of \(d(\cdot, \cdot)\). Let \(\epsilon_\delta = P(d(\tilde{A}_n, A_+) > \delta)\) and consider
\[
P(\rho(\hat{P}_n|P_+) > \varepsilon) \leq P(\rho(\hat{P}_n|P_+) > \varepsilon, d(\hat{A}_n, A_+) \leq \delta) + \epsilon_\delta
\]
\[
= P(\sup_{\alpha \in \tilde{A}_n} \inf \{|p_{n,n}(\alpha) - p_{n}(\alpha_+)| : \alpha_+ \in A_+\} > \varepsilon, d(\tilde{A}_n, A_+) \leq \delta) + \epsilon_\delta \tag{39}
\]
where the equality holds by definitions of \(\rho(\cdot, \cdot)\), \(\hat{P}_n\), and \(P_+\). Now consider that if \(d(\hat{A}_n, A_+) \leq \delta\), then \(\rho(\hat{A}_n|A_+) \leq \delta\) and for any \(\alpha \in \hat{A}_n\), there exists \(\alpha_+ \in A_+\) such that \(\|\alpha - \alpha_+\|_\infty \leq \delta\). We let \(\alpha = (\theta, h)\) and \(\alpha_+ = (\theta_+ + h_+\)). It follows that for any \(\alpha \in \hat{A}_n\) such that \(\|\alpha - \alpha_+\|_\infty \leq \delta\), we have
\[
p_{n,n}(\alpha) - p_{n}(\alpha_+)
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} (L(h(Z_i)) - L(h_+(Z_i))) + \frac{1}{n} \sum_{i=1}^{n} (L(h_+(Z_i)) - E[L(h_+(Z_i))])
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} (L(h(Z_i)) - L(h_+(Z_i)))(h(Z_i) - h_+(Z_i)) + \frac{1}{n} \sum_{i=1}^{n} (L(h_+(Z_i)) - E[L(h_+(Z_i))])
\]
\[
\leq \frac{1}{4} \|h - h_+\|_L^2 + o_p(1) \leq \frac{1}{4} \|\alpha - \alpha_+\|_\infty + o_p(1) \leq \delta
\]
for sufficiently large \(n \geq 3N\), where the second equality is obtained applying the mean value theorem and the first inequality holds since \(L(1 - L) \leq 1/4\) uniformly and since we bound the second right-hand side (RHS) term of the second equality by \(o_p(1)\) applying the LLN \(\{Z_i\}_{i=1}^n\) are iid and \(|L(h)| < 1\) uniformly. From this result, we have
\[
\sup_{\alpha \in \hat{A}_n} \inf \{|p_{n,n}(\alpha) - p_{n}(\alpha_+)| : \alpha_+ \in A_+\} \leq \sup_{\alpha \in \hat{A}_n, \|\alpha - \alpha_+\|_\infty \leq \delta} |p_{n,n}(\alpha) - p_{n}(\alpha_+)| \leq \delta
\]
for all sufficiently large \(n \geq 3N\). From this, it follows that for sufficient large \(n\),
\[
P(\sup_{\alpha \in \hat{A}_n} \inf \{|p_{n,n}(\alpha) - p_{n}(\alpha_+)| : \alpha_+ \in A_+\} > \varepsilon, d(\hat{A}_n, A_+) \leq \delta) \leq \varepsilon \tag{40}
\]
and thus from (39) and (40), we obtain \(P(\rho(\hat{P}_n|P_+) > \varepsilon) \leq \varepsilon + P(d(\hat{A}_n, A_+) > \delta)\). An analogous argument provides \(P(\rho(P_+|\hat{P}_n) > \varepsilon) \leq \varepsilon + P(d(\hat{A}_n, A_+) > \delta)\). Combining these two results with (38), we conclude \(P(d(\hat{P}_n, P_+) > \varepsilon) \leq 2\varepsilon + 2P(d(\hat{A}_n, A_+) > \delta)\). This proves Proposition 4.1 since \(\varepsilon > 0\) is arbitrary and \(d(\hat{A}_n, A_+) = o_p(1)\).
B Large Sample Theory for CI

B.1 High-level Assumptions and Primitive Conditions

Here we provide a semiparametric version of high-level assumptions given in ABJ to justify the CI’s asymptotically. We also provide sets of sufficient conditions that satisfy some of such high-level assumptions. Define $\mathcal{H}_\delta \equiv \{ h \in \mathcal{H} : \| h(\cdot, \theta) - h_0(\cdot, \theta) \|_\mathcal{H} \leq \delta, \theta \in \Theta \}$ for some small $\delta > 0$ and let $\mathcal{A}_\delta \equiv \Theta \times \mathcal{H}_\delta$. This section provides the high-level assumptions that justify the CI’s introduced in Section 5. Let

$$\hat{\nu}_n(j, \gamma, \theta, h) = \sqrt{n}(\hat{e}_n(j, \gamma, \theta, h) - c_0(j, \gamma, \theta, h))$$

and

$$\hat{Z}_n(j, m, \theta, h) = \sqrt{n}(c_0(j, \gamma_0, \theta, h) - c_0(j, \gamma_0, \theta, h))$$

(41)

(42)

Viewed as a function of $(j, \gamma, \theta, h)$, $\hat{\nu}_n(j, \gamma, \theta, h)$ is a stochastic process on $\mathcal{J}_J \times \Gamma_\text{all} \times \mathcal{A}_\delta$. Under suitable conditions, $\hat{\nu}_n(j, \gamma, \theta, h)$ converges weakly to a mean zero Gaussian process $\nu_0(j, \gamma, \theta, h)$ on $\mathcal{J}_J \times \Gamma_\text{all} \times \mathcal{A}_\delta$. The covariance function of $\nu_0(\cdot, \cdot, \cdot, \cdot)$ is given by

$$V_0((j_1, \gamma_1, \theta_1, h_1), (j_2, \gamma_2, \theta_2, h_2)) \equiv \text{Cov}(\nu_0(j_1, \gamma_1, \theta_1, h_1), \nu_0(j_2, \gamma_2, \theta_2, h_2))$$

$$= E \left[ ((P(y_j)|X, \theta_1, h_1) - 1|Y = y(j)|) q_{\gamma_1}(X) - c_0(j_1, \gamma_1, \theta_1, h_1)) \right]$$

$$\times ((P(y_j)|X, \theta_2, h_2) - 1|Y = y(j)|) q_{\gamma_2}(X) - c_0(j_2, \gamma_2, \theta_2, h_2)) \right]$$

(43)

for $(\theta_1, h_1), (\theta_2, h_2) \in \mathcal{A}_\delta$. Now let $\hat{Z}_n(\theta, h)$ denote the $\sum_{j=1}^d \mathcal{M}_j \times 1$ column vector whose elements are $\{\hat{Z}_n(j, \theta, h) : (j, \theta) \in \mathcal{J}_J, \mathcal{A}_\delta \}$ such that $\hat{Z}_n(1, 1, \theta, h)$ is the first element and $\hat{Z}_n(1, 2, \theta, h)$ is the second element, etc. At last, let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes. The following assumptions extend the assumptions in ABJ allowing for infinite dimensional parameters.

Assumption B.1 $\hat{\nu}_n(\cdot, \cdot, \cdot, \cdot) \Rightarrow \nu_0(\cdot, \cdot, \cdot, \cdot)$, where $\nu_0(\cdot, \cdot, \cdot, \cdot)$ is a mean zero Gaussian process indexed by $(j, \gamma, \theta, h) \in \mathcal{J}_J \times \Gamma_\text{all} \times \mathcal{A}_\delta$ with bounded and continuous sample paths a.s. (with respect to \| \cdot \| on $\Gamma_\text{all}$ and the metric \| \|_{\mathcal{A}_\delta}$ on $\mathcal{A}_\delta$) with covariance function $V_0(\cdot)$ defined in (43).

We note that the following stochastic equicontinuity condition is sufficient for Assumption B.1:

$$\sup_{\| (\theta', h') - (\theta, h) \|_{\mathcal{A}_\delta} \leq \delta_n, \| \gamma' - \gamma \| \leq \delta_n} \| \hat{\nu}_n(j, \gamma', \theta', h') - \hat{\nu}_n(j, \gamma, \theta, h) \| = o_p(1), \text{ for any given } (\gamma, \theta, h) \in \Gamma_\text{all} \times \mathcal{A}_\delta$$

(44)

for any positive sequence $\delta_n$ tending to zero.

Recall that $\mathcal{A}_\delta$ is compact (with respect to $\| \cdot \|_{\mathcal{A}_\delta}$). When $X$ is discrete, then it is obvious $\Gamma_\text{all}$ is finite. When $X$ is continuous, we construct $\gamma$’s such that they have non-empty interior. Then, $\Gamma_\text{all}$ is still a finite set since the number of all subsets (with nonempty interior) of $\mathcal{S}(X)$ is finite due to the compactness of $\mathcal{S}(X)$ (any compact set is totally bounded). Thus, $\Gamma_\text{all}$ is totally bounded with respect to $\| \cdot \|$. Therefore, $(\mathcal{A}_\delta, \Gamma_\text{all}, (\| \cdot \|_{\mathcal{A}_\delta}, \| \cdot \|))$ is a totally bounded pseudometric space. It is not difficult to show the finite dimensional (fidii) convergence holds, i.e. all the finite subsets $((\theta_1, h_1, \gamma_1), \ldots, (\theta_j, h_j, \gamma_j))$ of $\mathcal{A}_\delta$, $(\Gamma_\text{all}), (\hat{\nu}_n(j, \gamma_1, \theta_1, h_1), \ldots, \hat{\nu}_n(j, \gamma_j, \theta_j, h_j))$ converge in distribution. Therefore, as long as the condition (44)
Lemma B.1 Suppose Assumptions 4.1 and 4.4 hold. Further suppose that
(a) $\Theta$ is a compact subset of $\mathbb{R}^d$ and $\int^\infty_0 \sqrt{\log N(\varepsilon, \mathcal{H}, \|\cdot\|)} d\varepsilon < \infty$; and that
(b) (Lipschitz Condition) (i) For $P(Y = y^{(j)}|X, \theta, h)$, $j = 1, \ldots, J$, the pathwise derivative at the direction $[(\tilde{\theta}, \tilde{h}) - (\theta, h)]$ exists for all $(\tilde{\theta}, \tilde{h}), (\theta, h) \in \Theta \times \mathcal{H}_d$ and hence for $\exists M_1(j, X, \theta, h) \equiv dP(Y = y^{(j)}|X, \theta, h)$ and $\exists m_2(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot)$, we have $dP(Y = y^{(j)}|X, \theta, h) = M_1(j, X, \theta, h) - \theta + \sum_{v=1}^v m_2(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot)$.
(ii) $sup_{(\theta, h)} |M_1(j, X, \theta, h)| \leq C_1(X) < \infty$ and $sup_{(\theta, h)} |m_2(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot)| \leq C_2(X) < \infty$ for all $j = 1, \ldots, J$ and $v = 1, \ldots, v$.
Then, Condition \((44)\) holds.

Assumption B.2 $\tilde{Z}_n(\theta, h) \rightarrow Z_0(\theta, h)$, where $Z_0(\cdot, \cdot)$ is a mean zero Gaussian process indexed by $(\theta, h) \in \mathcal{A}_d$ with bounded and continuous sample paths a.s. (with respect to the metric $\|\cdot\|_s$) and the convergence holds jointly with that in Assumption B.1 with the joint limit being Gaussian.

Assumption B.3 (i) $sup_{(j, \gamma, \theta, h) \in \mathcal{I}_J \times \Gamma_{all} \times \mathcal{A}_d} |\tilde{w}_n(j, \gamma, \theta, h) - w_0(j, \gamma, \theta, h)| \rightarrow 0$ for some non-random positive functional $w_0$ on $\mathcal{I}_J \times \Gamma_{all} \times \mathcal{A}_d$ that is bounded and bounded away from zero. (ii) $w_0(j, \gamma, \theta, h)$ is continuous in $(\gamma, \theta, h)$ (with respect to the product of $\|\cdot\|$ on $\Gamma_{all}$ and the metric $\|\cdot\|_s$ on $\mathcal{A}_d$) at $(\gamma_0, j, m; \theta, h)$, \forall$(\theta, h) \in \mathcal{A}_d$, \forall$(j, m) \in \mathcal{I}_{J,M}$.

To develop the asymptotics, we need to define the population analogues of $\hat{\alpha}_{n,U}$ and $\hat{\alpha}_{n,L}$ defined in (16) replacing $\hat{\mathcal{A}}_{n,U}$ and $\hat{\mathcal{A}}_{n,U}$ with $\mathcal{A}_+$ and $\mathcal{A}_+$, respectively. We let\(^{22}\)
\[
\alpha_{n,U} = \arg\min\{\|\alpha\|_s : \alpha \in \mathcal{A}_+, \beta_{n,U}(\alpha) = \beta_{n,U}\}
\]
and define $\alpha_{n,L}$ analogously with $\beta_{n,L}$ in place of $\beta_{n,U}$.

Assumption B.4 (i) $\beta_{n}(\cdot) \rightarrow \beta_{0}(\cdot)$ uniformly over $\alpha \in \mathcal{A}_d$ for some non-random continuous functional $\beta_{0}(\cdot)$ on $\mathcal{A}_d$. (ii) $\|\alpha_{n,U} - \alpha_{+U}\|_s \rightarrow 0$ where $\alpha_{+,U} = \arg\min\{\|\alpha\|_s : \alpha \in \mathcal{A}_+, \beta_0(\alpha) = \beta_{+U}\}$ and $\beta_{+,U} = sup\{\beta_{0}(\alpha) : \alpha \in \mathcal{A}_+\}$. Also $\|\alpha_{n,L} - \alpha_{+L}\|_s \rightarrow 0$ where $\beta_{+,L}$ is defined analogously with $sup$ replaced by $inf$ and $\alpha_{+,L}$ is defined with $\beta_{+,L}$ replaced by $\beta_{+,L}$.

Note that if $\beta_{n}(\cdot)$ is non-random and does not depend on $n$, Assumption B.4 immediately holds by construction. Now define
\[
\mathcal{B}_{n,U} = \{(j, m) \in \mathcal{I}_{J,M} : c_0(j, \gamma_{0,j,m}; \alpha_{n,U}) = 0\}, \mathcal{B}_{+,U} = \{(j, m) \in \mathcal{I}_{J,M} : c_0(j, \gamma_{0,j,m}; \alpha_{+,U}) = 0\},
\]
and define $\mathcal{B}_{n,L}$ & $\mathcal{B}_{+,L}$ analogously replacing $\alpha_{n,U}$ and $\alpha_{+,U}$ with $\alpha_{n,L}$ and $\alpha_{+,L}$, respectively. We assume
\(^{22}\)Here we assume that $\alpha_{n,U}$ is unique. If it is not unique, then we select one of those functions that satisfy (45) according to a certain criterion.
Assumption B.5 (i) $\widehat{\beta}_{n,U} - \beta_{n,U} \xrightarrow{p} 0$ and $\widehat{\beta}_{n,L} - \beta_{n,L} \xrightarrow{p} 0$; (ii) $P(\widehat{B}_{n,U} \subseteq B_{n,U} \subseteq B_{+,U}) \xrightarrow{} 1$ and $P(\widehat{B}_{n,L} \subseteq B_{n,L} \subseteq B_{+,L}) \xrightarrow{} 1$.

Theorem 4.1 provides sufficient conditions for Assumption B.5 (i). ABJ note that by allowing the estimated binding constraints sets $\widehat{B}_{n,U}$ and $\widehat{B}_{n,L}$ to be smaller than the population versions $B_{n,U}$ and $B_{n,L}$, a researcher can consider more constraints in the estimation stage than in the CI construction stage. Lastly, we assume that when employing the CI for $\beta_0$, the critical values $\lambda_{n,U}^*(j, m, a)$ and $\lambda_{n,L}^*(j, m, a)$ defined in (23) converge in probability to non-negative constants $\lambda_{0,U}^*(j, m, a)$ and $\lambda_{0,L}^*(j, m, a)$ that satisfy

$$P \left( \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) + Z_0(j, m, \alpha_{+,U}) + w_0(j, \gamma_{0,j,m}, \alpha_{+,U}) \lambda_{0,U}^*(j, m, a) \geq 0 \text{ for } (j, m) \in B_{+,U} \right. \\
& \quad \left. \& \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) + Z_0(j, m, \alpha_{+,U}) + w_0(j, \gamma_{0,j,m}, \alpha_{+,U}) \lambda_{0,L}^*(j, m, a) \geq 0 \text{ for } (j, m) \in B_{+,L} \right) = 1 - a$$

(47)

and the same condition holds with $U$ and $L$ interchanged.

Assumption B.6 For the CI of the true value $\beta_0$, $\lambda_{n,U}^*(j, m, a) \xrightarrow{p} \lambda_{0,U}^*(j, m, a) \geq 0$ for all $(j, m) \in B_{+,U}$ and $\lambda_{n,L}^*(j, m, a) \xrightarrow{p} \lambda_{0,L}^*(j, m, a) \geq 0$ for all $(j, m) \in B_{+,L}$ where $\lambda_{0,U}^*(j, m, a)$ and $\lambda_{0,L}^*(j, m, a)$ satisfy (47).

Suppose that

$$\sqrt{n} \left( \frac{\tilde{c}_n(j, \tilde{\gamma}_{n,j,m}, \tilde{\alpha}_n, U) - c_0(j, \gamma_{0,j,m}, \alpha_{+,U})}{\tilde{w}_n(j, \tilde{\gamma}_{n,j,m}, \tilde{\alpha}_n, U)} \right) \xrightarrow{d} \frac{\nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) + Z_0(j, m, \alpha_{+,U})}{w_0(j, \gamma_{0,j,m}, \alpha_{+,U})}$$

(48)

and suppose the same condition holds with $U$ replaced by $L$. Then, Assumption B.6 will be satisfied if

$$\sqrt{n} \left( \frac{\tilde{c}_n^*(j, \tilde{\gamma}_{n,j,m}^*, \tilde{\alpha}_n, U) - \tilde{c}_n(j, \tilde{\gamma}_{n,j,m}, \tilde{\alpha}_n, U)}{\tilde{w}_n^*(j, \tilde{\gamma}_{n,j,m}^*, \tilde{\alpha}_n, U)} \right) \xrightarrow{d} \frac{\nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) + Z_0(j, m, \alpha_{+,U})}{w_0(j, \gamma_{0,j,m}, \alpha_{+,U})}$$

(49)

in $P^*$-probability and the same condition holds with $U$ replaced by $L$. Lemma B.2 provides sufficient conditions for (48) and Lemma B.3 provides sufficient conditions for (49). Analogous sufficient conditions for (48) and (49) with $U$ replaced by $L$ can be found in Lemma B.2 and Lemma B.3 with $L$ in replace of $U$.

Lemma B.2 Suppose that $A_+$ satisfies the condition (5), that $d(\hat{A}_n, A_+) \xrightarrow{p} 0$, and that

(i) $\|\hat{\alpha}_{n,U} - \alpha_{n,U}\|_s = o_P(1)$; (ii) (44) holds; (ii) Assumptions B.2, B.3, B.4, and 4.7 hold. Then, (48) is satisfied.

Note that Assumptions B.2, B.3 and 4.7 can be verified for a particular choice of $\tilde{Y}_n$ and a weight function $\tilde{w}_n(\cdot)$. Thus, they are directly assumed in this paper. To prove Lemma B.3, we need the following condition:

$$\sup_{\|\alpha' - \alpha\| \leq \delta_n, \|\gamma' - \gamma\| \leq \delta_n} |\tilde{c}_n^*(j, \gamma', \alpha') - \tilde{c}_n(j, \gamma', \alpha') - (\tilde{c}_n^*(j, \gamma, \alpha) - \tilde{c}_n(j, \gamma, \alpha))| = o_{P^*}(n^{-1/2})$$

(50)

for any given $(\gamma, \alpha) \in \Gamma_{all} \times A_3$. This is a bootstrap version of the stochastically equicontinuity condition of (44) and will be satisfied under the same sufficient conditions for (44).

Note that the use of fewer constraints cannot reduce the coverage probability of the CI.
Lemma B.3 Suppose that $A_+$ satisfies the condition (5), that $d(\widehat{A}_n, A_+) \to 0$ a.s., and that

(i) $||\widehat{A}_n, - \alpha_n, U||_s = o(1)$ a.s.;
(ii) $\sup_{(j, \gamma, \alpha) \in I_j \times \Gamma_{all} \times A_+} |\hat{\gamma}_n(j, \gamma, \alpha) - \hat{\gamma}_n(j, \gamma, \alpha)| = o_p(1)$; (iii) (50) holds;
(iv) (44) and Assumptions B.3, B.4, and 4.7 hold with “in probability” replaced by “almost surely”.

Then, (49) holds in $P^*$-probability.

B.2 Asymptotics for Confidence Interval

The proof of Theorem 5.1 and the proofs of lemmas in Appendix B use the following lemma, which extends Lemma 5 of ABJ to the semiparametric case.

Lemma B.4 Suppose Assumptions B.1-B.3, B.6, 4.2, 4.3 (i), 4.4 (i), and 4.7 for all $(j, m) \in I_{J, M}$ hold. Then, for any $\alpha \in A_\delta$ such that \( \| \alpha - \alpha_{+U} \|_s \to 0 \), we have

(i) \( \bar{\nu}_n(j, \bar{\gamma}_{n,j,m}, \alpha) \to \nu_0(j, \gamma_{0,j,m}, \alpha_{+U}) \); (ii) \( \bar{\nu}_n(j, m, \alpha) \to Z_0(j, m, \alpha_{+U}) \);

(iii) \( \hat{\gamma}_n(j, \gamma_{n,j,m}, \alpha) \to w_0(j, \gamma_{0,j,m}, \alpha_{+U}) \); (iv) \( \bar{\nu}_n(j, \bar{\gamma}_{n,j,m}, \alpha) + \hat{\gamma}_n(j, \gamma_{n,j,m}, \alpha) \alpha_{n,U} \to \nu_0(j, \gamma_{0,j,m}, \alpha_{+U}) + Z_0(j, m, \alpha_{+U}) + w_0(j, \gamma_{0,j,m}, \alpha_{+U}) \lambda_{n,U} \); and (v) the results of parts (i)-(iv) hold with $U$ replaced by $L$ and all the convergence results of the lemma hold jointly.

Proof. Combining Assumptions B.1 and 4.7, we obtain for any $\alpha \in A_\delta$ such that \( \| \alpha - \alpha_{+U} \|_s \to 0 \), $(\bar{\nu}_n(\cdot, \cdot, \cdot), \bar{\gamma}_n, \alpha)$ weakly converges to $(\nu_0(\cdot, \cdot, \cdot), \Gamma_0, \alpha_{+U})$ as processes indexed by $(j, \gamma, \alpha) \in I_j \times \Gamma_{all} \times A_\delta$ and this convergence is joint with that in Assumption B.2. Note that the function $g(\nu(\cdot, \cdot, \cdot), \Gamma, \alpha) = \nu(j, \gamma_{j,m}, \alpha)$ is continuous at $(\nu_0(\cdot, \cdot, \cdot), \Gamma_0, \alpha_{+U})$ because $\nu_0(\cdot, \cdot, \cdot)$ has continuous sample paths a.s. with respect to the product of the $\| \cdot \|$ norm and the $\| \cdot \|_s$ metric. Thus, applying the continuous mapping theorem (e.g., see Pollard (1984)), for any $\alpha \in A_\delta$ such that \( \| \alpha - \alpha_{+U} \|_s \to 0 \), we find $\bar{\nu}_n(j, \bar{\gamma}_{n,j,m}, \alpha) \to \nu_0(j, \gamma_{0,j,m}, \alpha_{+U})$, which proves part (i).

Similarly with part (i), we see part (ii) holds by Assumptions B.2 from the continuous mapping theorem.

Next, we prove part (iii). Using the triangle inequality and Assumption B.3, we have for any $\alpha \in A_\delta$ such that \( \| \alpha - \alpha_{+U} \|_s \to 0 \),

\[
\frac{1}{\alpha} \left| \frac{1}{\alpha} \hat{\gamma}_n(j, \gamma_{n,j,m}, \alpha) - w_0(j, \gamma_{0,j,m}, \alpha_{+U}) \right|
\leq \frac{1}{\alpha} \left| \hat{\gamma}_n(j, \gamma_{n,j,m}, \alpha) - w_0(j, \gamma_{n,j,m}, \alpha) \right| + \frac{1}{\alpha} \left| w_0(j, \gamma_{n,j,m}, \alpha) - w_0(j, \gamma_{0,j,m}, \alpha_{+U}) \right|
\leq \frac{1}{\alpha} \sup_{(j, \gamma, \alpha) \in I_j \times \Gamma_{all} \times A_\delta} \left| \hat{\gamma}_n(j, \gamma, \alpha) - w_0(j, \gamma, \alpha) \right| + \frac{1}{\alpha} \left| w_0(j, \gamma_{n,j,m}, \alpha) - w_0(j, \gamma_{0,j,m}, \alpha_{+U}) \right| \to 0
\]

where the first RHS term in the second inequality goes to zero by Assumption B.3 (i) and the second term goes to zero by the continuity assumed in Assumption B.3 (ii). Combining parts (i)-(iii) of the lemma and Assumption B.6 proves part (iv). ■

B.2.1 Proof of Theorem 5.1

By Assumption B.4 (i), $\beta_0 = \beta_n(\alpha_0) \to \beta_0(\alpha_0)$. We let $\beta_{0,0} = \beta_0(\alpha_0)$ denote the asymptotic true value. The following cases are considered separately: (i) $\beta_{+,L} < \beta_{0,0} < \beta_{+,U}$, (ii) $\beta_{+,L} < \beta_{0,0} = \beta_{+,U}$, (iii)
\( \beta_{+,L} = \beta_{0,0} < \beta_{+,U} \), and (iv) \( \beta_{+,L} = \beta_{0,0} = \beta_{+,U} \). The proofs are given for all three alternative CI's. We let \( \bar{\beta}_{n,U} \) and \( \bar{\beta}_{n,U} \) denote the generic lower bound and the generic upper bound for three alternative CI's, which suppress \( f \) in Alternative CI3 for notational convenience.

**CASE (i):** We have \( \bar{c}_{n,U}(j;\bar{\gamma}_{n,j,m},\alpha) \geq \hat{c}_{n,U}(j;\hat{\gamma}_{n,j,m},\alpha) \) for all \((j,m,\alpha)\), since \( \hat{w}_{n}(j;\gamma,\alpha) > 0 \) and \( \lambda^{*}_{n,U}(j,m,\alpha) \geq 0 \). For all the three alternative CI's, this implies that \( \bar{\beta}_{n,U} \geq \bar{\beta}_{n,U} \) by constructions of \( \bar{\beta}_{n,U} \) and \( \bar{\beta}_{n,U} \) in (14) and (19), (20), or (21) (note \( \tilde{A}_{n} \subseteq \Theta \times \tilde{\mathcal{T}}_{U} \subseteq \Theta \times \tilde{\mathcal{H}} \) by construction), respectively. Combining this with Theorem 4.1 and Assumption B.4 (i) gives \( \bar{\beta}_{n,U} - \beta_{0} \geq \bar{\beta}_{n,U} - \beta_{0} \rightarrow 0 \) and \( P(\beta_{0} \leq \bar{\beta}_{n,U}) \rightarrow 1 \). By an analogous argument, we can show that \( P(\beta_{0} \geq \bar{\beta}_{n,U}) \rightarrow 1 \), which establishes the result of Theorem 5.1 for case (i).

**CASE (ii):** From \( \beta_{0,0} > \beta_{+,L} \) and the same argument as above, it follows that \( P(\beta_{0} \geq \bar{\beta}_{n,U}) \rightarrow 1 \). It remains to show that \( \liminf_{n \to \infty} P(\beta_{0} \leq \bar{\beta}_{n,U}) \geq 1 - a \) for Alternative CI1, \( \liminf_{n \to \infty} P(\beta_{0} \leq \bar{\beta}_{n,U}) \geq 1 - a \) for Alternative CI2, and \( \liminf_{n \to \infty} P(\beta_{0} \leq \bar{\beta}_{n,U}) \geq 1 - a \) for Alternative CI3. We start with Alternative CI1. From definition of \( \alpha_{n,U} \) in (45), we have \( \beta_{0}(\alpha_{n,U}) = \beta_{n,U} \). Also note that if \( \bar{c}_{n,U}(j;\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \) \( \forall (j,m) \in \tilde{b}_{n,U} \), then \( \bar{\beta}_{n,U} \) cannot be smaller than \( \beta_{n,U} \) by constructions of \( \bar{\beta}_{n,U} \) and \( \beta_{n,U} \) (\( \alpha_{n,U} \) becomes an element of the set to which we take the sup operator to obtain \( \bar{\beta}_{n,U} \)). Moreover, \( \beta_{0} \leq \beta_{n,U} \) by definition of \( \beta_{n,U} \) in (13) and the fact that \( \alpha_{0} \in A_{+} \). Combining these results, we obtain

\[
P(\beta_{0} \leq \bar{\beta}_{n,U}) \geq P(\beta_{n,U} \leq \bar{\beta}_{n,U}) \geq P(\bar{c}_{n,U}(j;\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \forall (j,m) \in \tilde{b}_{n,U})
\]

for all \( n \geq \exists N \). Now note that

\[
\lim_{n \to \infty} \inf_{n \to \infty} P(\bar{c}_{n,U}(j;\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \forall (j,m) \in \tilde{b}_{n,U}) = 0 \forall (j,m) \in \tilde{b}_{n,U}) = 0 \forall (j,m) \in \tilde{b}_{n,U} \cap B_{n,U}
\]

where the equality holds because \( \beta_{0}(\alpha_{0,j,m},\alpha_{n,U}) = 0 \forall (j,m) \in B_{n,U} \) by definition of \( B_{n,U} \) in (46), the inequality holds by Assumption B.5 (ii) and the fact that a set can not be larger when it is defined using more restrictions.

Let \( Q_{n,U} = P(\bar{c}_{n,U}(j;\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \forall (j,m) \in B_{n,U}) \). This can be rewritten as

\[
Q_{n,U} = P(\bar{v}_{n}(j;\bar{\gamma}_{n,j,m},\alpha_{n,U}) + \tilde{Z}_{n}(j,m,\alpha_{n,U}) + \hat{w}_{n}(j;\hat{\gamma}_{n,j,m},\alpha_{n,U})\lambda^{*}_{n,U}(j,m,\alpha) \geq 0 \forall (j,m) \in B_{n,U})
\]

using the definitions of \( \bar{c}_{n,U}(j;\gamma,\alpha), \bar{v}_{n}(j;\gamma,\alpha) \), and \( \tilde{Z}_{n}(j,m,\alpha) \). From Lemma B.4 (iv) (note \( \alpha_{n,U} \in A_{+} \subseteq A_{b} \)) and Assumption B.5 (ii), it follows that

\[
\liminf_{n \to \infty} Q_{n,U} \geq P(\nu_{0}(j;\gamma_{0,j,m},\alpha_{+,U}) + Z_{0}(j,m,\alpha_{+,U}) + \nu_{0}(j;\gamma_{0,j,m},\alpha_{+,U})\lambda_{0,U}(j,m,\alpha) \geq 0 \forall (j,m) \in B_{+,U})
\]
Note that the strict inequality in (53) is allowed since Assumption B.5 (ii) allows \( B_{n,U} \) to be a strict subset of \( \mathcal{B}_{+} \) wpa1. If \( \beta_n(\cdot) \) is non-random and does not depend on \( n \), we have \( B_{n,U} = \mathcal{B}_{+} \) by definition and hence, (53) holds with equality. Now note that by definition of \( \lambda_{0,U}(j,m,a) \) in (47), the RHS of (53) is greater than or equal to \( 1 - a \). This completes the proof of Theorem 5.1 for case (ii) with Alternative CI1.

Now we turn to Alternative CI2. Note that \( \Theta_+ \times \hat{\mathcal{H}}_+ \) is included in \( \Theta \times \hat{\mathcal{H}}_+ \) wpa1 because \( d(\hat{\mathcal{A}}_n, \mathcal{A}_n) \to 0 \) and \( \hat{\mathcal{A}}_n \subseteq \Theta \times \hat{\mathcal{H}}_+ \). Thus, if \( \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \bar{B}_{n,U} \), then \( \beta_{n,U}^{(2)} \) cannot be smaller than \( \beta_{n,U} \) by constructions of \( \beta_{n,U}^{(2)} \) and \( \beta_{n,U} \) wpa1 since wpa1, \( \alpha_{n,U} \in \Theta \times \hat{\mathcal{H}}_{\delta_n} \). Similarly with Alternative CI1, we obtain

\[
P\left( \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \bar{B}_{n,U} \right) = P\left( \{ \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \bar{B}_{n,U} \} \text{ and } \{ \alpha_{n,U} \in \Theta \times \hat{\mathcal{H}}_{\delta_n} \} \right) \\
+ P\left( \{ \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \bar{B}_{n,U} \} \text{ and } \{ \alpha_{n,U} \notin \Theta \times \hat{\mathcal{H}}_{\delta_n} \} \right) \\
\leq P\left( \beta_{n,U} \leq \beta_{n,U}^{(2)} \right) + P\left( \alpha_{n,U} \notin \Theta \times \hat{\mathcal{H}}_{\delta_n} \right) = P\left( \beta_{0} \leq \beta_{n,U}^{(2)} \right) + o(1)
\]

where the last inequality holds by the same reason with Alternative CI1 and \( \alpha_{n,U} \in \Theta \times \hat{\mathcal{H}}_{\delta_n} \) wpa1. The remaining proof exactly follows that of Alternative CI1 and thus this completes the proof of Theorem 5.1 for case (ii) with Alternative CI2.

Now we turn to Alternative CI3. Recall that \( \bar{\beta}_{l,n,U}^{(3)} = \sup \{ \beta_{n}(\alpha) : \alpha \in \Theta \times \hat{\mathcal{H}}_{l,\delta_n}, \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha) \geq 0, \forall (j,m) \in \bar{B}_{n,U} \} \) and \( \bar{\beta}_{\infty,n,U}^{(3)} = \sup \{ \beta_{n}(\alpha) : \alpha \in \Theta \times \hat{\mathcal{H}}_{\delta_n}, \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha) \geq 0, \forall (j,m) \in \bar{B}_{n,U} \} \). Then, by construction of \( \bar{\beta}_{l,n,U} \), we have \( \bar{\beta}_{l,n,U}^{(3)} \leq \bar{\beta}_{l+1,n,U}^{(3)} \leq \bar{\beta}_{\infty,n,U}^{(3)} \) for all \( l \geq 1 \) since \( \Theta \times \hat{\mathcal{H}}_{l,\delta_n} \subseteq \Theta \times \hat{\mathcal{H}}_{l+1,\delta_n} \subseteq \Theta \times \hat{\mathcal{H}}_{\delta_n} \) for all \( l \). Similarly with Alternatives CI1 and CI2, it follows that if \( \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \bar{B}_{n,U} \), then \( \bar{\beta}_{\infty,n,U}^{(3)} \) cannot be smaller than \( \beta_{n,U} \) wpa1 by definition of \( \bar{\beta}_{\infty,n,U}^{(3)} \) since wpa1, \( \alpha_{n,U} \in \Theta \times \hat{\mathcal{H}}_{\delta_n} \). Also note \( \bar{\beta}_{l,n,U}^{(3)} \geq \bar{\beta}_{\infty,n,U}^{(3)} - \epsilon \) for arbitrary small number \( \epsilon > 0 \) for all large \( l \) since \( \bar{\beta}_{l,n,U}^{(3)} \to \bar{\beta}_{\infty,n,U}^{(3)} \). We will let \( \kappa_{l,U} = \bar{\beta}_{\infty,n,U}^{(3)} - \bar{\beta}_{l,n,U}^{(3)} \). Then, we have wpa1, \( \beta_{l,n,U}^{(3)} \geq \beta_{n,U} - \kappa_{l,U} \). Combining these results and noting \( \kappa_{l,U} = o_p(l) \) by construction, similarly with CI1 and CI2, we obtain

\[
P\left( \beta_{0} \leq \bar{\beta}_{l,n,U}^{(3)} \right) + o_l(1) + o(1) \geq P\left( \beta_{n,U} \leq \bar{\beta}_{l,n,U}^{(3)} \right) + P\left( \bar{\beta}_{l,n,U}^{(3)} < \beta_{n,U} \leq \beta_{l,n,U} + \kappa_{l,U} \right) + o(1) = P\left( \beta_{n,U} \leq \beta_{l,n,U} + \kappa_{l,U} \right) + o(1) \geq P\left( \bar{c}_{n,U}(j,\bar{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \bar{B}_{n,U} \right)
\]

for all \( n \geq 3N \) and \( o_l(1) \) denotes some non-random sequence that goes to zero as \( l \to \infty \).

The remaining proof exactly follows that of Alternative CI1&CI2 and thus this completes the proof of Theorem 5.1 for case (ii) with Alternative CI3.

**CASE (iii):** First we define \( \kappa_{L,U} = \bar{\beta}_{l,n,U}^{(3)} - \bar{\beta}_{\infty,n,L}^{(3)} \) where \( \bar{\beta}_{l,n,U}^{(3)} \) and \( \bar{\beta}_{\infty,n,L}^{(3)} \) are defined by replacing
with $L$ and sup with $\inf$ for the definitions of $\tilde{\beta}_{1,n,U}$ and $\tilde{\beta}_{\infty,n,U}$, respectively. Then, it can be proved analogously to case (ii).

**CASE (iv):** Note that analogous results of each (51), (54), and (55) for $L$ in replace of $U$ throughout holds such that

$$
P \left( \tilde{\beta}_{n,L} \leq \beta_0 \right) \geq P \left( \tilde{\beta}_{n,L} \leq \beta_{n,L} \right) \geq \rho \left( c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L} \right),
$$

$$
P \left( \tilde{\beta}_{n,L} \leq \beta_0 \right) + o(1) \geq P \left( \tilde{\beta}_{n,L} \leq \beta_{n,L} \right) + o(1) \geq \rho \left( c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L} \right),
$$

$$
P \left( \tilde{\beta}_{1,n,L} \leq \beta_0 \right) + o(1) + o(1) \geq P \left( \tilde{\beta}_{1,n,L} \leq \beta_{n,L} \right) + P \left( \tilde{\beta}_{1,n,L} - \kappa_{n,L} \leq \beta_{n,L} \leq \tilde{\beta}_{1,n,L} \right) + o(1)
$$

(56)

by the same argument as in (51), (54), and (55), alternatively. Combining (51), (54), or (55) with (56), we obtain

$$
P \left( \tilde{\beta}_{n,L} \leq \beta_0 \leq \tilde{\beta}_{n,U} \right) \geq P \left( \tilde{\beta}_{n,L} \leq \beta_{n,L} \leq \beta_{n,U} \leq \tilde{\beta}_{n,U} \right)
$$

$$
\geq P \left( c_{n,U}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,U}) \geq 0 \forall (j,m) \in \tilde{B}_{n,U}, c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L} \right),
$$

$$
P \left( \tilde{\beta}_{n,L} \leq \beta_0 \leq \tilde{\beta}_{n,U} \right) + o(1) \geq P \left( \tilde{\beta}_{n,L} \leq \beta_{n,L} \leq \beta_{n,U} \leq \tilde{\beta}_{n,U} \right) + o(1)
$$

(57)

$$
\geq P \left( c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L}, c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L} \right),
$$

$$
P \left( \tilde{\beta}_{1,n,L} \leq \beta_0 \leq \tilde{\beta}_{1,n,U} \right) + o(1) + o(1) \geq P \left( \tilde{\beta}_{1,n,L} \leq \beta_{n,L} \leq \beta_{n,U} \leq \tilde{\beta}_{1,n,U} \right) + o(1) + o(1)
$$

$$
\geq P \left( c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L}, c_{n,L}(j, \tilde{\gamma}_{n,j,m}, \alpha_{n,L}) \geq 0 \forall (j,m) \in \tilde{B}_{n,L} \right).
$$

Analogous results to those of (52)-(53) holds with $L$ in place of $U$ throughout. Note that the limit inf of the RHS of the second inequality in (57) for each alternative CI is at least as large as the RHS of (53) because $\beta_{+,U} = \beta_{+,L}$ implies that $\alpha_{+,U} = \alpha_{+,L}$ and $B_{+,U} = B_{+,L}$ by the definitions of $\beta_{+,U}$, $\alpha_{+,U}$, and $B_{+,U}$ in Assumption B.4 and (46), respectively. We have shown Theorem 5.1 holds for all four cases.

**B.2.2 Proof of Lemma B.1**

For each $j \in \mathcal{I}_J$, let $\mathcal{F}_{\xi_j} = \{ \xi_j(y,x,\gamma, \theta, h) : (P(y(j)|x, \theta, h) - 1[y = y(j)]) q_j(x) - c_0(j, \gamma, \theta, h) : (\gamma, \theta, h) \in \Gamma_{all} \times \Theta \times \mathcal{H} \}$ denote the class of measurable functions indexed by $(\gamma, \theta, h)$. Note that $\tilde{\nu}_n(j, \gamma, \theta, h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_j(y_i, x_i, \gamma, \theta, h)$. Also note that $E[\xi_j(Y, X, \gamma, \theta, h)] = 0$ for all $(j, \gamma, \theta, h) \in \mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}$ by construction. To prove this lemma, first, we extend Lemma 1 of Chen, Linton, and van Keilegom (2003) to fit our case.

**Lemma B.5** Let $\{ Y_i, X_i \}_{i=1}^n$ be iid with $E[\xi_j(Y, X, \gamma, \theta, h)] = 0$ for all $(j, \gamma, \theta, h) \in \mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_\delta$. Suppose that $\mathcal{F}_{\xi_j} = \{ \xi_j(y, x, \gamma, \theta, h) : (\gamma, \theta, h) \in \Gamma_{all} \times \Theta \times \mathcal{H} \}$ is $P$-Donsker (i.e. $\int_0^\infty \left[ \log N \left( \epsilon, \mathcal{F}_{\xi_j}, \| \cdot \|_{L_2(P)} \right) \right] \, d\epsilon < \infty$); and that $\xi_j(y, x, \gamma, \theta, h)$ is $L_2(P)$-continuous at all $(\gamma, \theta, h) \in \Gamma_{all} \times \Theta \times \mathcal{H}_\delta$. Then, (44) and (50) hold.
Proof. Noting \( \tilde{d}_t(j, \gamma, \theta, h) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_j(y_i, x_i, \gamma, \theta, h) \) by definition, (44) is obtained by extending Pakes and Pollard (1989)'s Lemma 2.17 from the case \( m(\cdot, \theta) \) to our case \( \xi_j(Y, X, \gamma, \theta, h) \). Its proof is essentially the same with theirs. Now (50) is obtained from Giné and Zinn (1990). 

Now we prove Lemma B.1. From Lemma B.5, it suffices to show \( \int_{0}^{\infty} \log N_{\| ||} \left( \epsilon, \mathcal{F}_{\xi_j}, || \cdot ||_{L_2(P)} \right) d\epsilon < \infty \). Note

\[
\left| \xi_j(y, x, \gamma_1, \theta_1, h_1) - \xi_j(y, x, \gamma_2, \theta_2, h_2) \right| \\
\leq \left| \xi_j(y, x, \gamma_1, \theta_1, h_1) - \xi_j(y, x, \gamma_1, \theta_2, h_2) \right| + \left| \xi_j(y, x, \gamma_1, \theta_2, h_2) - \xi_j(y, x, \gamma_2, \theta_2, h_2) \right| \\
\leq \left| P(y^{(j)}|x, \theta_1, h_1) - P(y^{(j)}|x, \theta_2, h_2) \right| q_{\gamma_1}(x) + \left| \int P(y^{(j)}|x, \theta_1, h_1) - P(y^{(j)}|x, \theta_2, h_2) \right| q_{\gamma_1}(x) dF_X(x) \\
+ \left| P(y^{(j)}|x, \theta_2, h_2) - 1[y = y^{(j)}] \right| q_{\gamma_2}(x) \\
and hence we have
\]

\[
\left| \xi_j(y, x, \gamma_1, \theta_1, h_1) - \xi_j(y, x, \gamma_2, \theta_2, h_2) \right|^2 \\
\leq 3 \left( P(y^{(j)}|x, \theta_1, h_1) - P(y^{(j)}|x, \theta_2, h_2) \right)^2 q_{\gamma_1}(x)^2 \\
+ 3 \int P(y^{(j)}|x, \theta_1, h_1) - P(y^{(j)}|x, \theta_2, h_2) \right)^2 q_{\gamma_1}(x) dF_X(x) \\
+ 3 \left( P(y^{(j)}|x, \theta_2, h_2) - 1[y = y^{(j)}] \right)^2 q_{\gamma_2}(x) \\
\leq (C_1(X) + C_2) \left( ||\theta_1 - \theta_2||_E^2 + ||h_1 - h_2||_{\mathcal{H}_\delta}^2 \right) + 3 \left( q_{\gamma_1}(x) - q_{\gamma_2}(x) \right)^2
\]

for some \( C_1(X) < \infty \) and \( C_2 < \infty \) by Assumption 4.4 and the condition (b). Thus, it follows that

\[
\left( E \sup_{\| \gamma_1 - \gamma_2 \|_{\mathcal{H}} < \delta, \| \theta_1 - \theta_2 \|_E < \delta, \| h_1 - h_2 \|_{\mathcal{H}_\delta} < \delta} \left| \xi_j(Y, X, \gamma_1, \theta_1, h_1) - \xi_j(Y, X, \gamma_2, \theta_2, h_2) \right|^2 \right)^{1/2} \leq C \delta
\]

from the definition of the semi-norm \( \| \gamma_1 - \gamma_2 \| \). Therefore, \( \xi_j(y, x, \gamma, \theta, h) \) is locally uniformly \( L_2(P) \)-continuous with respect to \( (\gamma, \theta, h) \in \Gamma_{all} \times \Theta \times \mathcal{H}_\delta \) by Theorem 6 in Andrews (1994a). The remaining proof is obtained similarly with the proof of the theorem 3 in Chen, Linton, and van Keilegom (2003). Now let \( \{\gamma_k : k = 1, \ldots, N_1\} \) be a \( \delta \)-cover for \( (\Gamma_{all}, \| \cdot \|) \), \( \{\theta_k : k = 1, \ldots, N_2\} \) be a \( \delta \)-cover for \( (\Theta, \| \cdot \|_E) \), \( \{h_k : k = 1, \ldots, N_3\} \) be a \( \delta \)-cover for \( (\mathcal{H}, \| \cdot \|_{\mathcal{H}}) \). Also let \( N_1 \equiv \{1, \ldots, N_1\} \), \( N_2 \equiv \{1, \ldots, N_2\} \), and \( N_3 \equiv \{1, \ldots, N_3\} \). Then, by (58), for any \( \xi_j(y, x, \gamma, \theta, h) \), there exist \( k_1 \in N_1 \), \( k_2 \in N_2 \), and \( k_3 \in N_3 \) such that

\[
\left| \xi_j(y, x, \gamma, \theta, h) - \xi_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}) \right| \leq b_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}, \delta)
\]

It follows that

\[
\xi_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}) - b_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}, \delta) \leq \xi_j(y, x, \gamma, \theta, h) \leq \xi_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}) + b_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}, \delta) \quad \text{and that} \quad \left( E \left[ b_j(Y, X, \gamma_{k_1}, \theta_{k_2}, h_{k_3}, \delta) \right]^2 \right)^{1/2} \leq C \delta \quad \text{for all } (\gamma_{k_1}, \theta_{k_2}, h_{k_3}) \quad \text{and all positive sequence tending to zero } \delta = o(1). \]

Therefore, \( \epsilon = 2C\delta \)-bracket for \( (\mathcal{F}_{\xi_j}, \| \cdot \|_{L_2(P)}) \) is formed as

\[
\left\{ \xi_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}) - b_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}, \delta), \xi_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}) + b_j(y, x, \gamma_{k_1}, \theta_{k_2}, h_{k_3}, \delta) \right\}
\]

It follows that \( N_{\| \epsilon, \mathcal{F}_{\xi_j}, \| \cdot \|_{L_2(P)} \|} \leq \sum N \left( \Theta, \| \cdot \|_E \right) \times \sum N \left( \mathcal{H}, \| \cdot \|_{\mathcal{H}_\delta} \right) \times N \left( \Gamma_{all}, \| \cdot \| \right). \)

Combining this result, the condition (a), and the arguments following (44), we complete the proof.
B.2.3 Proof of Lemma B.2

We can rewrite
\[
\sqrt{n} \left( \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \gamma_{0,j,m}, \hat{\alpha}_{n,U}) \right) = \sqrt{n} \left( \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) + \sqrt{n} \left( c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \gamma_{0,j,m}, \hat{\alpha}_{n,U}) \right) \\
= \hat{\nu}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) + \hat{Z}_n(j, m, \hat{\alpha}_{n,U})
\]
using the definitions of \( \hat{\nu}_n \) and \( \hat{Z}_n \) in (41) and (42), respectively. From the results of part (i) and part (ii) in Lemma B.4, we have
\[
\hat{\nu}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \to_d \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) \text{ and } \hat{Z}_n(j, m, \hat{\alpha}_{n,U}) \to_d Z_0(j, m, \alpha_{+,U}) \tag{59}
\]
since \( \|\hat{\alpha}_{n,U} - \alpha_{+,U}\|_p \to 0 \) by the condition (i) and Assumption B.4 (ii) and since \( \|\hat{\gamma}_{n,j,m} - \gamma_{0,j,m}\| \to 0 \) by Assumption 4.7. Finally, note Assumption B.3, Assumption 4.7, and the condition (i) imply that
\[
|\hat{\nu}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - w_0(j, \gamma_{0,j,m}, \alpha_{+,U})| \\
\leq |\hat{\nu}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - w_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U})| + |w_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - w_0(j, \gamma_{0,j,m}, \alpha_{+,U})| \tag{60}
\]
Combing (59) and (60), the claim follows.

B.2.4 Proof of Lemma B.3

Consider
\[
\sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) = \sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) - \sqrt{n} \left( \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) + \sqrt{n} \left( c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \gamma_{0,j,m}, \hat{\alpha}_{n,U}) \right) \tag{61}
\]
\[
+ \sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) \tag{62}
\]
\[
+ \sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) \tag{63}
\]
Now note that (61) is \( o_p(1) \) by the condition (i) and (iii). Note that from the definition of \( \hat{\nu}_n \), we have
\[
\sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) - \sqrt{n} \left( \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) + \sqrt{n} \left( c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \gamma_{0,j,m}, \hat{\alpha}_{n,U}) \right)
\]
and hence (62) is \( o(1) \) a.s. by (44) (a.s. version). From Giné and Zinn (1990), we know that
\[
\sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) = \sqrt{n} \left( \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) + o_p(1) \tag{64}
\]
and note also that \( \sqrt{n} \left( c_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) \to_d \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) \) by the part (i) of Lemma B.4 (note that \( \hat{\alpha}_{n,U} \in \mathcal{A}_n \subset \mathcal{A}_S \) wp1) and since \( \|\hat{\alpha}_{n,U} - \alpha_{+,U}\|_p = o(1) \) a.s. by the condition (i) and Assumption B.4 (ii) (a.s. version). From these results, it follows that for the first term
in (63), $\sqrt{n} \left( c_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) \rightarrow d \nu_0(j, \gamma_{0,j,m}, \alpha_{+}, U) + o_P(1)$. Finally, we note that for the second term in (63), $\sqrt{n} \left( c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - c_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) \rightarrow d Z_0(j, m, \alpha_{+}, U) + o_P(1)$ by the part (ii) of Lemma B.4 and since $\|\hat{\alpha}_{n,U} - \alpha_{+}, U\|_s = o(1)$ a.s.. Therefore, we have

$$\sqrt{n} \left( c_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{c}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) \right) \rightarrow d \nu_0(j, \gamma_{0,j,m}, \alpha_{+}, U) + Z_0(j, m, \alpha_{+}, U) + o_P(1).$$ (64)

Now it remains to show that $\hat{\omega}_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) = w_0(j, \gamma_{0,j,m}, \alpha_{+}, U) + o_P(1)$. This holds because

$$|\hat{\omega}_n^*(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - w_0(j, \gamma_{0,j,m}, \alpha_{+}, U)|$$

$$\leq |\hat{\omega}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - \hat{\omega}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U})| + |\hat{\omega}_n(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U}) - w_0(j, \hat{\gamma}_{n,j,m}, \hat{\alpha}_{n,U})|$$

$$+ |w_0(j, \gamma_{0,j,m}, \alpha_{+}, U) - w_0(j, \gamma_{0,j,m}, \alpha_{+}, U)|$$

$$\leq \sup_{(j, \gamma, \alpha) \in \Gamma \times \Gamma_{a(i)} \times \mathcal{A}_i} |\hat{\omega}_n(j, \gamma, \alpha) - \hat{\omega}_n(j, \gamma, \alpha)| + \sup_{(j, \gamma, \alpha) \in \Gamma \times \Gamma_{a(i)} \times \mathcal{A}_i} |\hat{\omega}_n(j, \gamma, \alpha) - w_0(j, \gamma, \alpha)|$$

$$+ |w_0(j, \gamma_{0,j,m}, \alpha_{+}, U) - w_0(j, \gamma_{0,j,m}, \alpha_{+}, U)| = o_P(1)$$

where the last result holds by Assumptions B.3 and 4.7 (a.s. version), by the condition (ii), and by $\|\hat{\alpha}_{n,U} - \alpha_{+}, U\|_s = o(1)$ a.s.. Therefore, from this result and (64), the claim follows.

References


