Essays on the Econometrics of Games

Shuo Jiang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Yanqin Fan, Chair

Quan Wen

Jing Tao

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on the Econometrics of Games

Shuo Jiang

Chair of the Supervisory Committee:
Professor Yanqin Fan
Department of Economics

This dissertation studies identification, estimation and inference for various types of static games of incomplete information, a class of games in which players do not have full information about their opponents. Such games have been widely used in the empirical studies of strategic interactions such as market entry, technology adoption and so on.

Chapter 1 studies sequential estimation and uniform inference in a static game of incomplete information with nonseparable unobserved heterogeneity. We propose a novel method for sequentially estimating payoff function and conducting uniform inference in static games of incomplete information with non-separable unobserved heterogeneity (and multiple equilibria). We tackle the matching-types problem by constructing a new characterization of the payoff function via a minimum distance model with incorrect "moments". For several specifications of the payoff function, we propose to select the correct matching and estimate the payoff function jointly using a minimum distance type criterion function with a rewarding term when needed; we show consistency of the selected matching and the estimator of the payoff function; we construct an asymptotically uniformly valid and easy-to-implement test for the linear hypothesis on the payoff function; and for large state spaces, we introduce a sequential Monte Carlo method to ease computational burden. We report results from a small simulation study and an application to the dataset of Sweeting (2009).

Chapter 2 proposes a simple estimator for static game of incomplete information with action complementarity. Oligopolists often engage in strategic interactions in multiple related businesses or industries. Such phenomenon could be analyzed using game theoretic models with action complementarity (substitutability). In this paper we study the semiparametric identification and estimation of static games of incomplete information with complementary (substitutable) actions. Building on and extending the identifiability result for bundled demand in Fox and Lazzati (2017), we show that structural parameters in this game are identified. A simple closed-form estimator for the structural parameters is proposed based on our identification strategy. The estimator could be implemented easily by running a three-stage least squares, and no numerical optimization is needed. We establish the root-n consistency and asymptotic normality of this estimator. A small Monte Carlo simulation shows the efficacy of our methods in finite samples.

Chapter 3 studies identification and estimation of a binary game of incomplete information under symmetry of the unobservables. We study the semiparametric identification and estimation of a class of binary game of incomplete information under the restriction of conditional symmetry for unobserved private information. We use a two-step identification strategy that is based on the equilibrium condition and the symmetry restriction. We propose a two-step minimum distance estimator, and prove its root-N consistency and asymptotic normality. Compared to existing semiparametric method in the literature, our estimator could adapt arbitrary forms of heteroskedasticity in common knowledge state variables and does not require stringent support and tail conditions. Our method could be extended to allow for multiple equilibria and symmetrically distributed random coefficients. A small Monte Carlo study demonstrates the efficacy and robustness of our estimator compared to the popular two-step pseudo maximum likelihood method.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my greatest gratitude to Professor Yanqin Fan for her continuous guidance, encouragement and support throughout the writing of this dissertation, and for her great patience in nurturing me intellectually. Personally, she is the best thesis advisor one could possibly imagine. The numerous insights that she gives have helped me catch the essence of the problems and have led to exciting discoveries. In the process I learned how to refine a paper like perfecting a piece of fine art. Being a close observer of her dedicated mentorship, I also learned how to cultivate student with great care and compassion. She will always be the best role model in my career, and I believe the experience of working with her will be the most invaluable intellectual treasure I could luckily carry with and continuously gain wisdom and strength from for the rest of my life.

I would also like to thank my committee members Professor Quan Wen and Professor Jing Tao, both gave me tremendous support in the writing of this dissertation, in job search and beyond. They shared with me without reservation state-of-the-art techniques from their own field, and they are always available whenever I have any request. I'm really grateful to the great generosity of their support. I would like to thank Professor Fang Han for deepening my understanding of statistics with the course STAT583, and for his patient guidance during my RAship. I also thank Xuetao Shi for kindly spending time discussing with me about my research, and thank Sunny Huang and Ming He for their help and guidance too.

I'm also thankful to Professor Jin-Chuan Duan, Ruli Xiao, Arthur Lewbel, Xun Tang and Aureo de Paula for helpful discussions, and I'm grateful to the excellent computing support provided by the UW Hyak supercomputer system and the CSDE simulation cluster.

In addition, I would like to thank my friends Thor Dodson, Junwei Mao, Zinan Wang, Kwongyu Wong, Matthew Daniels and Qifan Huang for their company, without which the past few years would be much less joyful.

Finally, I'm particularly grateful to my parents for being so supportive during my pursuit of the Ph.D. degree. I would like to give my sincerest thanks to them for their unconditional love and support.

Chapter 1

GAMES OF INCOMPLETE INFORMATION WITH NON-SEPARABLE UNOBSERVED HETEROGENEITY: ESTIMATION, INFERENCE, AND COMPUTATION [1]

1.1 Introduction

Motivation and Main Contributions The sequential approach to identification and estimation of discrete games of incomplete information is widely used in the literature (see Aguirregabiria and Mira (2019) and references therein). In this approach, the equilibrium conditional choice probabilities (CCPs hereafter) are identified and estimated from the data in the first step. In the second step, the payoff functions are identified and estimated using variations in observed state variables such as exclusion restrictions.[2] By avoiding the computation of equilibrium for every given state and parameter value, the sequential approach to estimation is computationally less costly than the all-solution method such as the nested fixed point algorithm.

The existing sequential approach to estimation relies on two assumptions: no common knowledge payoff relevant unobserved heterogeneity and a single equilibrium being played in the data. Both assumptions may be violated in data. Sweeting (2009) provides empirical evidence for the existence of multiple equilibria, De Paula and Tang (2012) propose a test for the existence of multiple equilibria when there is no unobserved heterogeneity, and Aguirregabiria and Mira (2019) summarize the importance of allowing for multiple equilibria and unobserved heterogeneity. Many recent works focus on developing sequential estimation methods in the presence of multiple equilibria and/or unobserved heterogeneity such as those

[1]This Chapter is a joint work with Yanqin Fan and Xuetao Shi from the University of Washington.

[2]Depending on the contexts/specifications, we will use payoff functions, payoff vectors, and payoff parameters interchangeably throughout this paper.

discussed next. To the best of the authors' knowledge, a unified framework for the sequential estimation and inference for the payoff function in static games of incomplete information with non-separable unobserved heterogeneity and possibly multiple equilibria is yet to be developed. This paper provides one.

As pointed out in Aguirregabiria and Mira (2019), a major impediment for developing valid sequential estimation and inference methods for games of incomplete information in the presence of common unobserved heterogeneity is the *matching-types problem*. This is because in the first step of the sequential approach, the presence of multiple equilibria or unobserved heterogeneity requires identification of a nonparametric finite mixture model, and CCPs are only identified up to a label swapping.[3] In the second step, *when there is payoff relevant unobserved heterogeneity*, identifying and estimating payoff vectors using variations in observed states require correctly tracking the ordering of unobserved states across different values of the observed state variables. This results in the matching-types problem.

To tackle the matching-types problem, we make use of the necessary and sufficient condition for step-2 identification of the payoff vector in Proposition 3 in Aguirregabiria and Mira (2019). We construct a novel characterization of the payoff vector as the unknown parameter in a minimum distance model with both correct and incorrect "moments". The correct set of "moments" corresponds to the correct matching and the true payoff vector is uniquely determined by the set of correct "moments". Based on this new characterization, we adapt the moment selection procedure in Andrews (1999) to address the matching-types problem. First, we define a selection vector, construct selection vector spaces, and propose to select the correct matching and estimate the payoff vector jointly. This is achieved using a minimum distance type criterion function with a rewarding term when needed over the payoff parameter space and selection vector space. Second, we establish asymptotic theory including consistency of the selection vector and estimated payoff vector. Third, we develop an asymptotically uniformly valid test for the linear hypothesis on the payoff vector. Our test

[3]Like majority work in this literature, we assume that the unobserved state variable takes a finite number of values.

is easy to implement with known critical values from chi-squared distribution. Moreover, this formulation allows us to test the null hypothesis of no unobserved heterogeneity against the presence of unobserved heterogeneity using J-test. Fourth, to overcome the computational challenge when the state space is large, we adapt the sequential Monte Carlo (SMC) method developed by Duan (2019) to our set-up. We provide numerical evidence on its superior performance over the exhaustive search for large selection vector spaces.

To demonstrate the broad applicability of the methodology we develop for the nonparametric game[4] studied in Aguirregabiria and Mira (2019), we introduce three of its variants. First, in many empirical applications, there may be some common observed state variable in addition to players' exclusive state variable. The first variant incorporates common observed state variables in a partial linear specification of the payoff function, which allows us to identify partial effects of common observed state variables. Second, we propose a parametric random coefficients specification for the payoff function in which the random coefficients depend on the latent state variables. For the third variant, we construct a modified version of the symmetric game in Sweeting (2009) and Xiao (2018) with random coefficients to account for the presence of unobserved heterogeneity. We show that the estimation and inference methods we develop for the nonparametric game can be easily modified to estimate and conduct inference in these models. Via the parametric game with random coefficients, we demonstrate the finite sample performance of our methods. As an empirical illustration, we apply our methods to the study of coordination incentives of radio stations in the timing choice of commercial breaks using the data from Sweeting (2009).

Related Works Motivated by the analysis of the strategic timing incentives among radio stations, Sweeting (2009) proposes to estimate a parametric game of incomplete information allowing for multiple equilibria but no unobserved heterogeneity. As a robustness check, Sweeting (2009) incorporates normally distributed unobserved heterogeneity and states that

[4] In this paper, we refer to games with parametric, semiparametric, and nonparametric payoff functions as parametric, semiparametric, and nonparametric games respectively.

"estimation of a game with many possible choices, multiple equilibria, and observed and possibly unobserved heterogeneity is well beyond the current literature." Sweeting (2009) has sparked important research such as Grieco (2014), Xiao (2018), and Aguirregabiria and Mira (2019). Grieco (2014) studies identification and estimation in a game with normally distributed private information allowing for the presence of multiple equilibria and normally distributed unobserved heterogeneity. Xiao (2018) studies sequential identification and estimation in a game with only multiple equilibria and no unobserved heterogeneity

This paper builds on Aguirregabiria and Mira (2019) which focuses on sequential and joint identification of a static nonparametric game with both multiple equilibria and common knowledge payoff relevant unobserved heterogeneity. To address the matching-types problem, Aguirregabiria and Mira (2019) suggest two approaches, both of which rely on additional assumptions. The first approach makes use of the ranking independence assumption. If mixing weights satisfy ranking independence, then we can match latent state across observed states based on the ranking of its mixing weight.[5] This assumption requires that the number of mixtures be the same across different values of observed states, which fails when there are potentially multiple equilibria. Even when there does not exist multiple equilibria, the assumption that the ranking of mixing weight is the same across all observed states is a strong restriction. In the second approach, unobserved heterogeneity is assumed to be additively separable and mean independent of the observed states. This allows to integrate out the unobserved heterogeneity and recover the part of payoff function that does not depend on the unobserved heterogeneity. However, this assumption rules out games with non-separable unobserved heterogeneity such as entry games with game-level random coefficients,[6] see Section 1.3 for an example.

[5]Complete statistical independence between latent state and observed states is a special case of ranking independence, and is used to match unobserved types in Kasahara and Shimotsu (2009).

[6]Here game-level has the same meaning as in De Paula (2013), who discusses game-level heterogeneity and game-level shock.

Organization of the Rest of This Paper and Notation We first develop our idea and methodology for a simple game with only unobserved heterogeneity in Section 1.2 referred to as the Simple Game and then extend the results to the general case with both unobserved heterogeneity and multiple equilibria. Section 1.2 reviews sequential identification procedure for the Simple Game building on Xiao (2018) and Aguirregabiria and Mira (2019), and characterizes the payoff function as the unknown parameter in a minimum distance model with incorrect "moments". In Section 1.3, we introduce a generic minimum distance model and three variants of the Simple Game. Section 1.4 proposes a sequential estimator, proves consistency of the estimated payoff, and develops an asymptotically valid test for the linear hypothesis for the generic minimum distance model, the Simple Game, and three variants. To ease the computational burden, we introduces a SMC method in Section 1.4. Section 1.5 introduces a general game with both unobserved heterogeneity and multiple equilibria referred to as the Extended Game. It also develops methods for sequential estimation and inference in the Extended Game and its variants. In Section 1.6, we investigate the finite sample performance of our estimation and inference methods via Monte Carlo simulation, and compares the running time of SMC with the conventional exhaustive search. Section 1.7 applies our methods to the study of radio stations' timing choice of commercial breaks using the data from Sweeting (2009). Section 1.8 concludes. Appendices of this paper are given in Fan et al. (2020). Appendix A.1 presents the approach in Xiao (2018) for the identification of CCPs. Appendix A.2 introduces the SMC method. Technical proofs are collected in Appendix B. Appendix C discusses three variants of the Extended Game.

For any $q \times 1$ vector E, let $\|E\|$ denote its Euclidean norm and $\|E\|_0$ denote its L_0 norm, i.e., the number of non-zero elements in E. For Ω being some $q \times q$ matrix, denote $\|E\|_\Omega \equiv \left(E^\top \Omega E\right)^{1/2}$. For any random variable X, we use S_X to denote its support. For any finite set, $|\cdot|$ denotes its cardinality. For any given q-dimensional vector c of zeros and ones and some $q \times p$ matrix A, let A_c denote the submatrix of A generated by deleting the rows in A corresponding to zeros in c. Let I_q be a $q \times q$ identity matrix. "wp $\to$ 1" denotes "with probability approaching one".

1.2 The Simple Game, Sequential Identification, and Minimum Distance Characterization

We introduce a simple game of incomplete information with only unobserved heterogeneity (referred to as the Simple Game) and use it to illustrate our idea and methods. We review sequential identification of the payoff function in the Simple Game and characterize the payoff function as the unknown parameter in a minimum distance model with incorrect "moments".

1.2.1 The Simple Game and Equilibrium CCPs

The Simple Game is a static game with three players.[7] Each player, denoted as $i = 1, 2, 3$, chooses an action $d_i \in \{0, 1\}$. Before choosing his action, player i draws his private information $\widetilde{\epsilon}_i\left(d_i\right)$ for two actions $d_i = 0$ and $d_i = 1$ from a bivariate distribution. Player i's payoff from choosing action d_i is given by $\widetilde{\pi}_i\left(d_i, \mathbf{d}_{-i}, z_i, k, \widetilde{\epsilon}_i\left(d_i\right)\right)$, where the vector $\mathbf{d}_{-i}$ denotes the joint actions of all the other players, $z_i \in S_{z_i} = \{L, M, H\}$ for all $i = 1, 2, 3$, is an observable exclusive state variable which does not enter the payoffs of other players, and $k \in S_k = \{A, B\}$ is a common knowledge state variable that is known by all players but unobserved by the econometrician.

We assume that player i's payoff is additively separable in his private information $\widetilde{\epsilon}_i\left(d_i\right)$ and write

$$\widetilde{\pi}_i\left(d_i, \mathbf{d}_{-i}, z_i, k, \widetilde{\epsilon}_i\left(d_i\right)\right) = \widetilde{\pi}_i\left(d_i, \mathbf{d}_{-i}, z_i, k\right) - \widetilde{\epsilon}_i\left(d_i\right),$$

where $\widetilde{\pi}_i(d_i, \mathbf{d}_{-i}, z_i, k)$ captures how payoff of player i for choosing d_i changes with respect to his opponents' actions and state variables. Since the optimality of an action is invariant under monotonically increasing transformations, we normalize payoffs using $\widetilde{\pi}_i\left(0, \mathbf{d}_{-i}, z_i, k, \tilde{\epsilon}_i\left(0\right)\right)$ and define the normalized payoff for $d_i = 1$ as

$$\pi_i\left(1, \mathbf{d}_{-i}, z_i, k\right) \equiv \frac{\widetilde{\pi}_i\left(1, \mathbf{d}_{-i}, z_i, k\right) - \widetilde{\pi}_i\left(0, \mathbf{d}_{-i}, z_i, k\right)}{sd\left(\widetilde{\epsilon}_i\left(1\right) - \widetilde{\epsilon}_i\left(0\right)\right)},$$

where $sd\left(\widetilde{\epsilon}_i\left(1\right) - \widetilde{\epsilon}_i\left(0\right)\right)$ denotes the standard deviation of $\widetilde{\epsilon}_i\left(1\right) - \widetilde{\epsilon}_i\left(0\right)$. We refer to $\pi_i\left(d_i, \mathbf{d}_{-i}, z_i, k\right)$

[7] It follows from Allman et al. (2009) and Aguirregabiria and Mira (2019) that when the number of mixtures is 2, the minimum number of players required for identifying CCPs up to a label swapping is 3.

as the *payoff function* for player i hereafter. By normalization, the payoff function $\pi_i(d_i, \mathbf{d}_{-i}, z_i, k)$ for $d_i = 0$ is zero. Stacking the payoff function evaluated at all combinations of opponents' actions, we obtain the following *payoff vector* for player i at observed state z_i and unobserved state k:

$$\pi_{iz_ik} \equiv \begin{bmatrix} \pi_i(1, (1,1), z_i, k) \\ \pi_i(1, (0,1), z_i, k) \\ \pi_i(1, (1,0), z_i, k) \\ \pi_i(1, (0,0), z_i, k) \end{bmatrix}. \tag{1.1}$$

We define the normalized private information for player i as

$$\epsilon_i \equiv \frac{\widetilde{\epsilon}_i(1) - \widetilde{\epsilon}_i(0)}{sd(\widetilde{\epsilon}_i(1) - \widetilde{\epsilon}_i(0))}.$$

Let $\mathbf{z} \equiv (z_1, z_2, z_3) \in S_{\mathbf{z}} \equiv S_{z_1} \times S_{z_2} \times S_{z_3}$, where $S_{\mathbf{z}}$ is the support of $\mathbf{z}$. Following Aguirregabiria and Mira (2019), we impose the following assumption on the normalized private information ϵ_i.

Assumption 1.2.1. *(i)* $\{\epsilon_i\}_{i=1}^{3} \overset{i.i.d}{\sim} F(\cdot)$, *where* $F(\cdot)$ *is known to the econometrician and is absolutely continuous with a probability density function denoted as* $f(\cdot)$*; (ii)* $\epsilon_1, \epsilon_2,$ *and* ϵ_3 *are independent of the state variables* $(\mathbf{z}, k)$*.*

A (pure) strategy in this game is defined as follows:

Definition 1.2.1 (Strategy)**.** *For given* $\mathbf{z}$ *and* k*, a (pure) strategy for player* i *is a mapping* $\sigma_i(\epsilon_i, \mathbf{z}, k) : S_\epsilon \times S_{\mathbf{z}} \times S_k \to \{1, 0\}$*.*

For notational compactness, we use $\boldsymbol{\sigma} \equiv (\sigma_1(\epsilon_1, \mathbf{z}, k), \sigma_2(\epsilon_2, \mathbf{z}, k), \sigma_3(\epsilon_3, \mathbf{z}, k))$ to denote a strategy profile given $(\mathbf{z}, k)$. Let $\mathbb{1}(\cdot)$ denote the indicator function. Any given $\boldsymbol{\sigma}$ is completely characterized by the following CCPs:

$$p_i \equiv \int \mathbb{1}(\sigma_i(\epsilon_i, \mathbf{z}, k) = 1) f(\epsilon_i) d\epsilon_i, \text{ for } i = 1, 2, 3.$$

Denote j and q as the two players other than player i. The *expected payoff function* for player i with $d_i = 1$ for given $(\mathbf{z}, k)$ and $\boldsymbol{\sigma}$ is computed as:[8]

$$\overline{\pi}_i (1, \mathbf{z}, k, \boldsymbol{\sigma}) = \sum_{d_j, d_q \in \{0,1\}} p_j^{d_j} (1 - p_j)^{1-d_j} p_q^{d_q} (1 - p_q)^{1-d_q} \pi_i (1, (d_j, d_q), z_i, k). \tag{1.2}$$

Bayesian Nash Equilibrium is then defined as follows:

Definition 1.2.2 (Equilibrium)**.** *For any given* $(\mathbf{z}, k)$*, a Bayesian Nash Equilibrium (BNE) of the game is a strategy profile* $\boldsymbol{\sigma}^*$ *such that for any player* i *and for any* ϵ_i*,*

$$\sigma_i^* (\epsilon_i, \mathbf{z}, k) = \arg \max_{d_i \in \{0,1\}} \{\overline{\pi}_i (d_i, \mathbf{z}, k, \boldsymbol{\sigma}^*) - \epsilon_i\}.$$

Note that the game typically may have multiple equilibria, but not necessarily all of them are played in the observed data. To simplify the discussion, we assume that the data are rationalized by a single equilibrium in the Simple Game. In Section 1.5, we extend our methods to allow for both unobserved heterogeneity and multiple equilibria.

Assumption 1.2.2. *There is a single equilibrium being played in the data for each* $(\mathbf{z}, k) \in S_{\mathbf{z}} \times S_k$*.*

Under Assumption 1.2.2, there is a unique equilibrium denoted as $\boldsymbol{\sigma}^*$ which rationalizes the data for each $(\mathbf{z}, k)$. Denote the *equilibrium* CCP of choosing action 1 for player i as

$$p_i (\mathbf{z}, k) \equiv \Pr (d_i = 1 \mid \mathbf{z}, k).$$

Then it holds that

$$p_i (\mathbf{z}, k) = \int \mathbb{1} (\sigma_i^* (\epsilon_i, \mathbf{z}, k) = 1) f (\epsilon_i) d\epsilon_i.$$

For any $(\mathbf{z}, k)$, the BNE of the game is equivalently characterized by the equilibrium CCP vector $\mathbf{p} (\mathbf{z}, k) \equiv (p_1 (\mathbf{z}, k), p_2 (\mathbf{z}, k), p_3 (\mathbf{z}, k))$ such that the following system holds:

$$p_i (\mathbf{z}, k) = F (\overline{\pi}_i (1, \mathbf{z}, k, \boldsymbol{\sigma}^*)) \text{ for } i = 1, 2, 3.$$

[8]Because of the normalization, $\overline{\pi}_i (0, \mathbf{z}, k, \boldsymbol{\sigma}) = 0$.

Under Assumption 1.2.2, the equilibrium expected payoff function is unique and only depends on $\mathbf{z}$ and k. For brevity, we denote it as $\overline{\pi}_i(1, \mathbf{z}, k) \equiv \overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma}^*)$.

Remark 1.2.1. *For any given $(\mathbf{z}, k)$ and strategy profile $\boldsymbol{\sigma}$ with corresponding CCP vector (p_1, p_2, p_3), the probability that choice 1 is optimal for player i in the Simple Game is given by $F(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma}))$ for $i = 1, 2, 3$. This defines the best response mapping for players $i = 1, 2, 3$ on $(\mathbf{z}, k)$ as follows:*

$$\Lambda_{i\mathbf{z}k}(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma})) = F(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma})). \tag{1.3}$$

In the Simple Game, the best response mapping is simply the cumulative distribution function. Given the values of payoff functions, by Equation (1.2), these best response mappings only depend on CCPs of other players. Aguirregabiria and Mira (2019) characterizes a BNE as a fixed point of the best response mapping of all players.

The game together with the equilibrium constitute the underlying structure that generates the data. The form of data available to the econometrician is specified in the following assumption.

Assumption 1.2.3. *The econometrician observes a random sample on players' actions and observable state variables $\{(d_{1l}, d_{2l}, d_{3l}, z_{1l}, z_{2l}, z_{3l})\}_{l=1}^{n}$.*

Let $p^A(\mathbf{z}) \equiv \Pr(k = A \mid \mathbf{z})$ and $p^B(\mathbf{z}) \equiv \Pr(k = B \mid \mathbf{z})$. We impose the following assumption on the distribution of the latent state.

Assumption 1.2.4. *(i) For any $\mathbf{z} \in S_{\mathbf{z}}$, $0 < p^A(\mathbf{z}) < 1$; (ii) For $k \in \{A, B\}$, $|S_k|$ is known to the econometrician.*

Assumption 1.2.4 (ii) is introduced for the Simple Game to simplify exposition and will be relaxed in Section 1.5.

1.2.2 Sequential Identification of Payoff Vectors

Consider sequential identification of the payoff vectors for player i corresponding to $d_i = 1$ and exclusive observed state variable z_i.[9] Under Assumption 1.2.4, there are two payoff vectors corresponding to two latent states: π_{iz_iA} and π_{iz_iB} defined in (1.1). The goal is to identify and make inference on them from the data. Since the identification procedure is exactly the same for all i and z_i, we focus on the identification of the payoff vectors for player 1 with $z_1 = L$ below and denote them as $\pi_{0k} \equiv \pi_{1Lk}$ for $k = A, B$. The subscript 0 in π_{0k} is used to emphasize that it is the true payoff.

Step-1: Identification of Equilibrium CCPs (up to a label swapping) The identification of equilibrium CCPs, $p_i(\mathbf{z}, k)$ and mixing weights, $p^A(\mathbf{z})$ makes use of the following system of equations: for $(d_1, d_2, d_3) \in \{0,1\} \times \{0,1\} \times \{0,1\}$, $\mathbf{z} \in S_{\mathbf{z}}$, and $k \in \{A, B\}$,

$$p(d_1, d_2, d_3 \mid \mathbf{z}) = \sum_{k \in \{A,B\}} \left[p^k(\mathbf{z}) \prod_{i=1}^{3} (p_i(\mathbf{z}, k))^{d_i} (1 - p_i(\mathbf{z}, k))^{1-d_i} \right], \tag{1.4}$$

where for any (d_1, d_2, d_3), $p(d_1, d_2, d_3 \mid \mathbf{z})$ denotes the probability of players' joint actions identified from the sample information. It is written as a weighted sum of the product of individual CCPs on each unobserved state. Equation (1.4) is a nonparametric finite mixture model. Sufficient conditions for identifiability of this model can be found in Theorem 4 and Corollary 5 in Allman et al. (2009). In the Simple Game, the following assumption guarantees identification in the first step.

Assumption 1.2.5 (Step-1 Identification)**.** $\mathbf{P}_{i\mathbf{z}}$ *has full rank for any* $\mathbf{z} \in S_{\mathbf{z}}$ *and* $i = 1, 2, 3,$

[9] We consider identification, estimation, and inference for each player and exclusive state pair separately to reduce the complexity of the selection vector space introduced in the next section. While Proposition 3 in Aguirregabiria and Mira (2019) requires that if the matching is wrong (for at least one player), then the system of some player (not necessarily the player whose matching is wrong) has no solution, we work with the condition that whenever the matching is wrong for some player, his own system has no solution. The approach developed in the paper can be easily extended to study a game satisfying the weaker condition specified in Proposition 3 of Aguirregabiria and Mira (2019).

where

$$\mathbf{P}_{i\mathbf{z}} \equiv \begin{bmatrix} p_i(\mathbf{z},k) & p_i(\mathbf{z},k') \\ 1-p_i(\mathbf{z},k) & 1-p_i(\mathbf{z},k') \end{bmatrix} \text{ for } k,k' \in \{A,B\} \text{ and } k \neq k'.$$

Constructive identification results for nonparametric finite mixture models have been established by Bonhomme et al. (2016) and Xiao (2018), among others. The approach in Bonhomme et al. (2016) is based on the simultaneous diagonalization of a set of matrices in the same non-orthogonal basis, while the procedure in Xiao (2018) is based on eigendecomposition of a set of matrices. The latter approach is more convenient in our context, because it avoids using permutation matrices to keep the same order of unobserved states across players. Our paper adopts this method and presents the steps involved in Xiao (2018) in Appendix A.1.

Step-2: Identification of payoff vectors—the matching-types problem It follows from Section 1.2.1 that the equilibrium expected payoff function for player 1 given $(\mathbf{z},k)$ is $\overline{\pi}_1(1,\mathbf{z},k) = F^{-1}(p_1(\mathbf{z},k))$. From now on we treat the expected payoff function as known since F is known and $p_1(\mathbf{z},k)$ is identified from step-1 in Section 1.2.2. For any $\mathbf{z} \in S_{\mathbf{z}}$ and $k \in \{A,B\}$, let $\mathbf{p}_{-1}(\mathbf{z},k)$ denote the following row vector:

$$[p_2(\mathbf{z},k)\,p_3(\mathbf{z},k),\,(1-p_2(\mathbf{z},k))\,p_3(\mathbf{z},k),\,p_2(\mathbf{z},k)\,(1-p_3(\mathbf{z},k)),\,(1-p_2(\mathbf{z},k))\,(1-p_3(\mathbf{z},k))]\,.$$

It consists of joint probabilities of player 1's opponents' actions on latent state k.

Let $\mathbf{z}^1 \equiv (L,L,L)\,,\mathbf{z}^2 \equiv (L,L,M)\,,\ldots,\mathbf{z}^9 \equiv (L,H,H)$ denote the nine possible values that $\mathbf{z}$ can take holding $z_1 = L$. For $\mathbf{z} \in \{\mathbf{z}^1,\ldots,\mathbf{z}^9\}$, player 1's equilibrium expected payoff can be written as:

$$\overline{\pi}_1(1,\mathbf{z},k) = \mathbf{p}_{-1}(\mathbf{z},k)\,\pi_{0k} \quad \text{for } k = A,B.$$

As $\mathbf{z}$ varies in $\{\mathbf{z}^1,\ldots,\mathbf{z}^9\}$, z_1 is held constant at $z_1 = L$. In order to identify π_{0k}, we need to also hold k constant at A and B respectively. If A and B are observable, this would be trivial. Collecting equations corresponding to A and B separately delivers two systems of equations in π:

$$\overline{\pi}_1(1,\mathbf{z},A) = \mathbf{p}_{-1}(\mathbf{z},A)\,\pi \text{ for all } \mathbf{z} \in \left\{\mathbf{z}^1,\ldots,\mathbf{z}^9\right\} \text{ and} \tag{1.5}$$

$$\overline{\pi}_1(1,\mathbf{z},B) = \mathbf{p}_{-1}(\mathbf{z},B)\,\pi \text{ for all } \mathbf{z} \in \left\{\mathbf{z}^1,\ldots,\mathbf{z}^9\right\}. \tag{1.6}$$

As long as the matrices

$$\left(\mathbf{p}_{-1}\left(\mathbf{z}^1,A\right)^\top,\ldots,\mathbf{p}_{-1}\left(\mathbf{z}^9,A\right)^\top\right)^\top \text{ and } \left(\mathbf{p}_{-1}\left(\mathbf{z}^1,B\right)^\top,\ldots,\mathbf{p}_{-1}\left(\mathbf{z}^9,B\right)^\top\right)^\top$$

have full column rank, systems (1.5) and (1.6) both have unique solutions corresponding to π_{0A} and π_{0B}. However, neither A nor B is observable. Systems like (1.5) and (1.6) are not immediately available after step-1. The researcher needs to track the same latent state across different values of observed state $\mathbf{z}$, i.e., solve the matching-types problem. Aguirregabiria and Mira (2019) illustrate the matching-types problem in a numerical study with a three player binary game focusing on identifying the payoff vector for player 1, and show that an incorrectly matched system leads to serious bias for every element in estimated player 1's payoff vectors.

To solve the matching-types problem, Aguirregabiria and Mira (2019) imposes either ranking independence or additive separability on the model structure.[10] For games that do not satisfy either assumption, Proposition 3 in Aguirregabiria and Mira (2019) presents a necessary and sufficient condition for step-2 identification of the payoff vector: a correct matching leads to a unique solution to the system for expected payoffs, and this solution identifies the payoff vector; while an incorrect matching delivers no solution. Building on this result, in Section 1.2.3, we construct a novel minimum distance model with incorrect "moments" to characterize the true payoff vectors. Based on the model we develop a unified framework in Section 1.3 for the estimation and inference on the payoff vectors without

[10] As noted by Aguirregabiria and Mira (2019), the matching-types problem also exists in the sequential identification of single agent discrete choice model. Luo (2020) studies the matching-types problem in dynamic games of incomplete information making use of the special structure of Markov perfect equilibrium, which is not applicable here. Other related works include Luo et al. (2018), who studies identification of the ordering of latent value distributions using restricted stochastic dominance.

assuming either ranking independence or additive separability.

1.2.3 A Minimum Distance Model for the Payoff Vectors

For each observed state, we first stack the expected payoff function evaluated at two latent states and then all observed states to obtain the expected payoff vector $\overline{\pi}$ of dimension 18. In the same fashion, we construct the coefficient matrix Γ of dimension 18×4. They are

$$\begin{aligned}\overline{\pi} &= \left[\overline{\pi}_1\left(1, \mathbf{z}^1, k_1\right), \overline{\pi}_1\left(1, \mathbf{z}^1, k_1'\right), \ldots, \overline{\pi}_1\left(1, \mathbf{z}^9, k_9\right), \overline{\pi}_1\left(1, \mathbf{z}^9, k_9'\right)\right]^\top \text{ and} \\ \Gamma &= \left[\mathbf{p}_{-1}\left(\mathbf{z}^1, k_1\right)^\top, \mathbf{p}_{-1}\left(\mathbf{z}^1, k_1'\right)^\top, \ldots, \mathbf{p}_{-1}\left(\mathbf{z}^9, k_9\right)^\top, \mathbf{p}_{-1}\left(\mathbf{z}^9, k_9'\right)^\top\right]^\top,\end{aligned}$$

where k_t and k_t' are used to denote the pair of latent states on the t-th observed state vector for $t = 1, ..., 9$. This notation makes it clear that for any $t \neq t'$, k_t and $k_{t'}$ (k_t' and $k_{t'}'$) do not necessarily correspond to the same latent state. After we have identified CCPs for each observed state up to label swapping in step-1, we obtain the following system of eighteen "moment" functions in $\pi \in \Pi$:

$$G\left(\pi\right) \equiv \overline{\pi} - \Gamma\pi, \tag{1.7}$$

where Π denotes the space for π. There are nine different observed states in this game. Each observed state has two groups of identified CCPs, but we do not know which group corresponds to latent state A and which group corresponds to latent state B. The aim is to select all components of $G\left(\pi\right)$ that correspond to the same latent state and use them to recover the payoff vector.

We define a selection vector and selection vector space. Given a selected latent state for the first observed state $\mathbf{z}^1$, we select one latent state from each subsequent observed state to match it. There are $2^8 = 256$ different possible matchings. Each matching corresponds to a set of "moments" selected from $G\left(\pi\right)$. It can be represented by a selection vector c of zeros and ones, such that a "one" indicates the corresponding element of $G\left(\pi\right)$ being selected and a "zero" indicates the corresponding element of $G\left(\pi\right)$ not being selected. We denote the selected "moment" functions by a selection vector c as $G_c(\pi)$: $G_c\left(\pi\right) = \overline{\pi}_c - \Gamma_c\pi$. Given a selected latent state on the first observed state, the set of all selection vectors (with the structure of our problem imposed) constitutes the selection vector space:

Definition 1.2.3 (Selection Vector Space)**.** *The two selection vector spaces corresponding to System (1.7) are defined as*

$$\mathscr{C}^1 = \left\{ (c_1, \ldots, c_9)^\top : c_1 = (1,0) \text{ and } c_t \in \{(1,0), (0,1)\} \text{ for } t \in \{2, ..., 9\} \right\} \text{ and}$$
$$\mathscr{C}^2 = \left\{ (c_1, \ldots, c_9)^\top : c_1 = (0,1) \text{ and } c_t \in \{(1,0), (0,1)\} \text{ for } t \in \{2, ..., 9\} \right\}.$$

Selection vectors in $\mathscr{C}^1$ always select the first latent state on the first observed state; and selection vectors in $\mathscr{C}^2$ always select the second latent state on the first observed state. For our analysis below, we focus on selection vector space $\mathscr{C}^1$ and suppress superscripts 1 and 2. Based on Definition 1.2.3, each selection vector $c \in \mathscr{C}$ is of dimension 18 with 9 non-zero elements.

Definition 1.2.4 (Correct Matching)**.** *A correct matching is defined by a selection vector c_0, which selects 9 components of $G(\pi)$ defined in (1.7) with the same underlying latent state.*

Primitive conditions that deliver step-2 identification are specified as the following:

Assumption 1.2.6 (Step-2 Identification)**.** Γ_{c_0} *has full column rank, and it holds that* $\operatorname{rank}([\overline{\pi}_c, \Gamma_c]) > \operatorname{rank}(\Gamma_c)$ *for any $c \neq c_0$.*

Since $\overline{\pi}_{c_0} - \Gamma_{c_0}\pi = \mathbf{0}$ has at least one solution, which is the true payoff vector, the assumption on Γ_{c_0} having full column rank is sufficient for $\overline{\pi}_{c_0} - \Gamma_{c_0}\pi$ to have a unique solution. For any $c \neq c_0$, the system of linear equations: $\overline{\pi}_c - \Gamma_c\pi = \mathbf{0}$ has no solution if $\operatorname{rank}([\overline{\pi}_c, \Gamma_c]) > \operatorname{rank}(\Gamma_c)$. Therefore, any system with a wrong matching has no solution under Assumption 1.2.6.

Lemma 1.2.1. *Assumptions 1.2.1-1.2.6 are sufficient for the identification of* (c_0, π_0).

Note that the true payoff vector is only defined for the system selected by c_0. Lemma 1.2.1 characterizes the correct matching and the true payoff vector via a minimum distance model.

This allows us to address the matching-types problem by selecting the correct "moment" functions from (1.7) and estimating the true payoff using the selected "moments" jointly.

Remark 1.2.2. *Assumption 1.2.6 is sufficient for identifying* (c_0, π_0)*, but not necessary. It is stronger than the necessary and sufficient condition discussed by Aguirregabiria and Mira (2019) in Proposition 3, because we consider identification for each pair of player and exclusive state separately to reduce the complexity of selection vector space. The estimation and inference procedure for the minimum distance model introduced in the next section can be directly applied if the condition in Proposition 3 of Aguirregabiria and Mira (2019) is imposed. The system would stack "moment" functions for three players together, and the selection vectors would select the same latent state on any observed state for all players.*

1.3 A Generic Minimum Distance Model and Three Variants of the Simple Game

Lemma 1.2.1 characterizes the payoff vectors in the Simple Game as parameters in a linear-in-parameters minimum distance model with incorrect "moments". In this section, we introduce three variants of the Simple Game and demonstrate that parameters in these games can be characterized by the same class of linear-in-parameters minimum distance models as that of the Simple Game. We call this class of minimum distance models as the *Generic Model*.

We will use the same notations in the Generic Model as in the Simple Game with the understanding that they differ across models. Let $G(\pi)$ defined below be the "moment" functions which are linear in $\pi \in \Pi \subseteq \mathbb{R}^{l_\pi}$:

$$G(\pi) \equiv \overline{\pi} - \Gamma\pi, \tag{1.8}$$

where $\overline{\pi}$ is of dimension $l_g \times 1$ and Γ is of dimension $l_g \times l_\pi$. For some generic selection vector space $\mathscr{C}$, define

$$\mathscr{C}\mathscr{Z} \equiv \{c \in \mathscr{C} : G_c(\pi) = \mathbf{0} \text{ has a solution}\}. \tag{1.9}$$

The Generic Model is defined by (1.8) and the following assumption.

Theorem 1.3.1. *(i) Both* $\overline{\pi}$ *and* Γ *can be identified; (ii)* Π *is compact; (iii)* $\mathscr{C}$ *is known; (iv)* $\mathscr{C}\mathscr{Z}$ *has a unique element denoted as* c_0 *with known* $\|c_0\|_0$*; (v) The system:* $G_{c_0}(\pi) = 0$ *has a unique solution in* Π *denoted as* π_0*.*

Under Assumption 1.3.1, the unknown vector c_0 is the unique element in $\mathscr{C}\mathscr{Z}$, and the number of non-zero elements in c_0 is known. Assumption 1.3.1 is satisfied for the Simple Game by step-1 and step-2 identification. In particular, Lemma 1.2.1 shows that under Assumption 1.2.6, c_0 is the unique element of $\mathscr{C}\mathscr{Z}$ with $\|c_0\|_0 = 9$, and the system: $G_{c_0}(\pi) = \mathbf{0}$ has a unique solution π_0 with $\pi_0 = \pi_{0A}$ or $\pi_0 = \pi_{0B}$ depending on the selection vector space we are using, $\mathscr{C}^1$ or $\mathscr{C}^2$. This verifies Assumption 1.3.1 (iv) and (v). Assumption 1.3.1 (iv) will be relaxed in Section 1.3 to allow for $\mathscr{C}\mathscr{Z}$ to have more than one element. Typically, for Assumption 1.3.1 (v) to hold, we need the number of "moment" functions to be larger than the dimension of π.

In the following subsections, we present three variants of the Simple Game. Since the focus here is on illustrating the broad applicability of the Generic Model, we emphasize on showing the connection between the three variants and the Generic Model, rather than providing rigorous conditions that verify Assumption 1.3.1.

All three variants adopt the basic structure and notation of the Simple Game: there are three players and two latent states in each game; when applicable, there is an exclusive observed variable z_i which takes three distinct values $\{L, M, H\}$. Further, recall that $\mathbf{z} = (z_1, z_2, z_3) \in S_{\mathbf{z}}$, $\mathbf{z}^1 = (L, L, L)$, $\mathbf{z}^2 = (L, L, M)$,..., and $\mathbf{z}^9 = (L, H, H)$. Let $\mathbf{z}^{10} \equiv (M, L, L)$, $\mathbf{z}^{11} \equiv (M, L, M)$,..., $\mathbf{z}^{18} \equiv (M, H, H)$ denote the nine possible values that $\mathbf{z}$ takes holding $z_1 = M$; and let $\mathbf{z}^{19} \equiv (H, L, L)$, $\mathbf{z}^{20} \equiv (H, L, M)$,..., $\mathbf{z}^{27} \equiv (H, H, H)$ denote the nine possible values that $\mathbf{z}$ takes holding $z_1 = H$.

1.3.1 The Simple Game with Common Observed State

In many empirical applications, there is some common observed state variable x. To simplify the discussion, we assume that x is one dimensional and consider sequential identification of the Simple Game with a common observed state variable x. Extending notations for the Simple Game, let the equilibrium CCP vector be: $\mathbf{p}(\mathbf{z},x,k) \equiv (p_1(\mathbf{z},x,k), p_2(\mathbf{z},x,k), p_3(\mathbf{z},x,k))$ for $\mathbf{z} \in S_{\mathbf{z}}$ and $x \in S_x$, where for $i = 1,2,3$, $p_i(\mathbf{z},x,k) \equiv \Pr(d_i = 1 \mid \mathbf{z},x,k)$.

Step-1 identification of equilibrium CCPs and mixing weights makes use of the following system of equations: for $(d_1,d_2,d_3) \in \{0,1\} \times \{0,1\} \times \{0,1\}$,

$$p(d_1,d_2,d_3 \mid \mathbf{z},x) = \sum_{k\in\{A,B\}} \left[p^k(\mathbf{z},x) \prod_{i=1}^{3} (p_i(\mathbf{z},x,k))^{d_i} (1-p_i(\mathbf{z},x,k))^{d_i-1} \right]$$

and can proceed in exactly the same way as step-1 for the Simple Game under similar assumptions.

Step-2 identification based on Aguirregabiria and Mira (2019) is done conditional on x for each player and exclusive state pair. This implies that the unobserved state is not matched across different values of x. When there is a need to match unobserved state across x such as identifying the partial effect of some common observed state variable, some structure on the payoff function may be imposed to relate the data generating processes across different values of x. One possibility is to parametrize the payoff function with respect to the observed state, so that the effect of the observed state is common across all values of x.

Suppose the payoff function for player 1 when choosing $d_1 = 1$ is given by

$$\pi_1(1,\mathbf{d}_{-1},z_1,x,k) = \beta_k x + \pi_1(1,\mathbf{d}_{-1},z_1,k), \tag{1.10}$$

where β_k characterizes the effect of x and changes with k. Under this parametrization, we can formulate the step-2 identification in the same way as (1.8). To illustrate, suppose x is a discrete random variable taking two values, i.e., $x \in \{x^1,x^2\}$. We obtain the following system for player 1 at $z_1 = L$ via stacking two latent states on each $(\mathbf{z},x)$ for $\mathbf{z} \in \{\mathbf{z}^1,\ldots,\mathbf{z}^9\}$:

$$G(\pi) = \begin{bmatrix} \overline{\pi}_1(x^1) \\ \overline{\pi}_1(x^2) \end{bmatrix} - \begin{bmatrix} \Gamma_1(x^1) \\ \Gamma_1(x^2) \end{bmatrix} \pi, \tag{1.11}$$

where

$$\overline{\pi}_1(x) \equiv \left[\overline{\pi}_1\left(1,\mathbf{z}^1,x,k_1\right), \overline{\pi}_1\left(1,\mathbf{z}^1,x,k_1'\right), \ldots, \overline{\pi}_1\left(1,\mathbf{z}^9,x,k_9\right), \overline{\pi}_1\left(1,\mathbf{z}^9,x,k_9'\right)\right]^\top \text{ and}$$

$$\Gamma_1(x) \equiv \begin{bmatrix} (x, \mathbf{p}_{-1}(\mathbf{z}^1, x, k_1))^\top, (x, \mathbf{p}_{-1}(\mathbf{z}^1, x, k_1'))^\top, \ldots, \\ (x, \mathbf{p}_{-1}(\mathbf{z}^9, x, k_9))^\top, (x, \mathbf{p}_{-1}(\mathbf{z}^9, x, k_9'))^\top \end{bmatrix}^\top.$$

System (1.11) is a special case of (1.8) with

$$\overline{\pi} = \begin{bmatrix} \overline{\pi}_1(x^1) \\ \overline{\pi}_1(x^2) \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} \Gamma_1(x^1) \\ \Gamma_1(x^2) \end{bmatrix}.$$

The dimensions for $\overline{\pi}$ and Γ are 36×1 and 36×5 respectively.

Given some selected latent state for $(\mathbf{z}^1, x^1)$, we could match this latent state across all exclusive and common observed states under similar conditions to those for step-2 identification of the Simple Game in Section 1.2.3 with the selection vector spaces given by

$$\mathscr{C}^1 = \left\{(c_1, ..., c_{18})^\top : c_1 = (1,0) \text{ and } c_t \in \{(1,0), (0,1)\} \text{ for } t = 2, ..., 18\right\} \text{ and}$$
$$\mathscr{C}^2 = \left\{(c_1, ..., c_{18})^\top : c_1 = (0,1) \text{ and } c_t \in \{(1,0), (0,1)\} \text{ for } t = 2, ..., 18\right\}.$$

For c belonging to $\mathscr{C}^1$ or $\mathscr{C}^2$, we have $\|c\|_0 = 18$. Given a selection vector space with the first selected latent state being k, the unique solution to the system $G_{c_0}(\pi) = \mathbf{0}$ gives us $\left(\beta_k, \pi_{1z_1k}^\top\right)^\top$.

This procedure can be generalized straightforwardly to the case that x is a random vector with both discrete and continuous components. For continuous components, we could use a finite number of points in their supports in step-2 identification.

1.3.2 A Parametric Model with Game-Level Random Coefficients

In empirical applications, researchers often fully parametrize the payoff functions. Sequential identification and estimation of a parametric game with latent-state-dependent parameters face the matching-types problem as well.

Consider the variant of the Simple Game with the following parametric payoff function for player i when choosing $d_i = 1$:

$$\pi_i\left(1, \mathbf{d}_{-i}, z_i, k\right) = z_i\theta_{ik} + \delta_{ik}\left(\sum_{j\neq i} d_j\right), \tag{1.12}$$

where $(\theta_{ik}, \delta_{ik})$ are game-level random coefficients that depend on the realization of a latent variable k with support $\{A, B\}$.[11] These coefficients are correlated game-level random coefficients in the sense that the conditional distribution of random coefficients changes with $\mathbf{z}$. Identification of equilibrium CCPs is the same as that of the Simple Game. In contrast to the Simple Game, we can pool all the observed states together and obtain one system for each player because of the parametric payoff function. For example, the system for player 1 is given by

$$G\left(\pi\right) = \begin{bmatrix} \overline{\pi}_1\left(\mathbf{z}^1\right) \\ \vdots \\ \overline{\pi}_1\left(\mathbf{z}^{27}\right) \end{bmatrix} - \begin{bmatrix} \Gamma_1\left(\mathbf{z}^1\right) \\ \vdots \\ \Gamma_1\left(\mathbf{z}^{27}\right) \end{bmatrix} \pi,$$

where

$$\overline{\pi}_1\left(\mathbf{z}^t\right) = \begin{bmatrix} \overline{\pi}_1\left(1, \mathbf{z}^t, k_t\right) \\ \overline{\pi}_1\left(1, \mathbf{z}^t, k_t'\right) \end{bmatrix} \text{ and } \Gamma_1\left(\mathbf{z}^t\right) = \begin{bmatrix} z_1 & p_2\left(\mathbf{z}^t, k_t\right) + p_3\left(\mathbf{z}^t, k_t\right) \\ z_1 & p_2\left(\mathbf{z}^t, k_t'\right) + p_3\left(\mathbf{z}^t, k_t'\right) \end{bmatrix}.$$

The model fits into the Generic Model in (1.8) with

$$\overline{\pi} = \begin{bmatrix} \overline{\pi}_1\left(\mathbf{z}^1\right) \\ \vdots \\ \overline{\pi}_1\left(\mathbf{z}^{27}\right) \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} \Gamma_1\left(\mathbf{z}^1\right) \\ \vdots \\ \Gamma_1\left(\mathbf{z}^{27}\right) \end{bmatrix},$$

where vector $\overline{\pi}$ is of dimension 54×1 and Γ is the coefficient matrix of dimension 54×2. Given some selected latent state for $\mathbf{z}^1$, we could match this latent state across all observed states under similar conditions to those for step-2 identification of the Simple Game in Section 1.2.3

[11]For notational simplicity, we do not consider common observed state variables in this model. It is straightforward to incorporate such variables.

with

$$\mathscr{C}^1 = \left\{(c_1, ..., c_{27})^\top : c_1 = (1,0), c_t \in \{(1,0),(0,1)\} \text{ for } t = 2, ..., 27\right\} \text{ and}$$
$$\mathscr{C}^2 = \left\{(c_1, ..., c_{27})^\top : c_1 = (0,1), c_t \in \{(1,0),(0,1)\} \text{ for } t = 2, ..., 27\right\}.$$

For any $c \in \mathscr{C}^1$ or $c \in \mathscr{C}^2$, $\|c\|_0 = 27$. Let k be the first selected latent state. The unique solution to the system $G_{c_0}(\pi) = \mathbf{0}$ gives us $(\theta_{1k}, \delta_{1k})^\top$.

1.3.3 A Symmetric Game With Random Coefficients

So far we have focused on asymmetric games in which each player has an exclusive state variable. Sometimes such variables might not be available in data, or might be available without enough variation.[12] A symmetric game is studied in both Sweeting (2009) and Xiao (2018). Suppose we have a symmetric game with three identical players and observed common state variable x. The payoff function for $d_i = 1$ is parametrized as:[13]

$$\pi_i(1, \mathbf{d}_{-i}, x, k) = x\beta_k + \frac{1}{2}\delta_k \sum_{j \neq i} d_j \equiv \pi(1, \mathbf{d}_{-i}, x, k). \tag{1.13}$$

By symmetry, given any x and k, the equilibrium CCPs for all three players are the same, denoted as $p(x,k) \equiv \Pr(d_i = 1 \mid x, k)$.

Step-1 identification of the CCP vector is based on

$$p(d_1, d_2, d_3 \mid x) = \sum_{k \in \{A,B\}} \left[p^k(x) \prod_{i=1}^{3} (p(x,k))^{d_i} (1 - p(x,k))^{d_i - 1}\right]$$

and can achieved in the same way as step-1 for the Simple Game under similar assumptions.

To illustrate step-2 identification, suppose $x \in \{x^1, ..., x^4\}$. Because the expected payoff

---

[12]This is the case for the data set in Sweeting (2009) on radio stations' timing decision for commercials.

[13]Following Xiao (2018), the strength of strategic interaction is allowed to change with x so that the second term could be $\frac{1}{2}\delta_k x \sum_{j \neq i} d_j$, which is the specification we use in the empirical illustration.

function at observed state x and latent state k is

$$\overline{\pi}\left(1,x,k\right)=x\beta_{k}+\delta_{k}p\left(x,k\right),$$

we obtain the following system for the whole game via stacking two latent states on each value of x:

$$G\left(\pi\right)=\begin{bmatrix}\overline{\pi}\left(1,x^{1},k_{1}\right)\\ \overline{\pi}\left(1,x^{1},k_{1}'\right)\\ \vdots\\ \overline{\pi}\left(1,x^{4},k_{4}\right)\\ \overline{\pi}\left(1,x^{4},k_{4}'\right)\end{bmatrix}-\begin{bmatrix}x^{1} & p\left(x^{1},k_{1}\right)\\ x^{1} & p\left(x^{1},k_{1}'\right)\\ \vdots & \vdots\\ x^{4} & p\left(x^{4},k_{4}\right)\\ x^{4} & p\left(x^{4},k_{4}'\right)\end{bmatrix}\pi.$$

The system fits into the Generic Model with $\overline{\pi}$ being the first term on the right hand side of the equality and Γ being the coefficient matrix of dimension 8×2.

Given some selected latent state for x^1, we could match this latent state across all common observed states under similar conditions to those for step-2 identification of the Simple Game with

$$\begin{aligned}\mathscr{C}^{1}&=\left\{\left(c_{1},...,c_{4}\right)^{\top}:c_{1}=\left(1,0\right),c_{t}\in\left\{\left(1,0\right),\left(0,1\right)\right\}\text{ for }t=2,...,4\right\}\text{ and}\\ \mathscr{C}^{2}&=\left\{\left(c_{1},...,c_{4}\right)^{\top}:c_{1}=\left(0,1\right),c_{t}\in\left\{\left(1,0\right),\left(0,1\right)\right\}\text{ for }t=2,...,4\right\}.\end{aligned}$$

For any $c\in\mathscr{C}^{1}$ or $c\in\mathscr{C}^{2}$, we have $\left\|c\right\|_{0}=4$. For k being the first selected latent state, the unique solution to the system $G_{c_0}\left(\pi\right)=\mathbf{0}$ delivers $\left(\beta_{k},\delta_{k}\right)^{\top}$.

1.4 Estimation, Inference, and Computation

Section 1.4.1 proposes a method for estimating c_0 and the true parameter π_0, and develops a test for the linear hypothesis on π_0 in the Generic Model. Section 1.4.2 provides primitive conditions for the theorems in Section 1.4.1 to hold for the Simple Game and its three variants. Lastly, we introduce a stochastic optimization algorithm based on sequential MC to ease computational burden when the state space is large.

1.4.1 The Generic Model

Estimation and Consistency

The proposed approach to consistently selecting the set of correct "moment" conditions and estimating parameters in the Generic Model is inspired by Andrews (1999). Recall that for any $c \in \mathscr{C}$ and $\pi \in \Pi$, $G_c(\pi)$ denotes the "moment" functions selected by c. By Assumption 1.3.1, the correct selection vector and true parameter vector satisfies

$$(c_0, \pi_0) = \arg \min_{c \in \mathscr{C}, \pi \in \Pi} \|G_c(\pi)\|^2_{W(c)},$$

where $W(c)$ is a positive definite weighting matrix depending on c.

Define the sample "moment" functions as $G_n(\pi) \equiv \overline{\pi}_n - \Gamma_n \pi$, where $\overline{\pi}_n$ and Γ_n are consistent estimators of $\overline{\pi}$ and Γ. Our estimator is defined as

$$(\widehat{c}, \widehat{\pi}) = \arg \min_{c \in \mathscr{C}, \pi \in \Pi} \|G_{nc}(\pi)\|^2_{W_n(c)}, \tag{1.14}$$

where $G_{nc}(\pi)$ is the sample "moment" functions selected by c and $W_n(c)$ is the corresponding sample weighting matrix.

The following assumptions are sufficient for establishing the consistency of $(\widehat{c}, \widehat{\pi})$.

Theorem 1.4.1. *(i) For any $\pi \in \Pi$, $G_n(\pi) = G(\pi) + o_p(1)$; (ii) For any $c \in \mathscr{C}$, $W_n(c) = W(c) + o_p(1)$ with $W(c)$ being positive definite; and (iii) For any $c \in \mathscr{C}$, $\min_{\pi \in \Pi} \|G_{nc}(\pi)\|^2_{W_n(c)} \xrightarrow{p} \min_{\pi \in \Pi} \|G_c(\pi)\|^2_{W(c)}$.*

Theorem 1.4.1. *Under Assumptions 1.3.1 and 1.4.1, it holds that $\widehat{c} = c_0$ wp $\to 1$ and $\widehat{\pi} \xrightarrow{p} \pi_0$.*

Testing Linear Hypothesis

We now develop a test for the following linear hypothesis on the parameter π_0 in Model (1.8):

$$H_0 : R\pi_0 = r \text{ against } H_1 : R\pi_0 \neq r, \tag{1.15}$$

where R is of dimension $l_R \times l_\pi$ with rank $(R) = l_R$ and r is of dimension l_R. The test statistic is defined as

$$T_n \equiv \min_{c \in \mathscr{C},\, R\pi = r} \left\| \sqrt{n} G_{nc}(\pi) \right\|^2_{W_n(c)}, \tag{1.16}$$

which is expected to be large if the null is incorrect. Notice that this is a post-selection inference problem because of the built-in "moment" selection in our test statistic. Albeit the many challenges faced with general post-selection inference documented in Leeb and Pötscher (2005) and Leeb and Pötscher (2008), we are able to construct an asymptotically uniformly valid and computationally simple test based on T_n.

Assume that Model (1.8) is fully characterized by $\xi \in \Xi$, where Ξ is some compact parameter space and is possibly infinite dimensional. Denote Ξ_R as the parameter space consistent with the null hypothesis and $\Pr_\xi(\cdot)$ as the probability calculated under ξ. The objective is to find a critical value CV that controls the asymptotic size defined as:

$$AsySize \equiv \limsup_{n \to \infty} \sup_{\xi \in \Xi_R} \Pr_\xi(T_n > CV).$$

We consider the drifting model parameter sequence ξ_n and the set of drifting model parameter sequences under H_0 with limit ξ as

$$\Xi_R(\xi) = \{\{\xi_n \in \Xi_R : n \geq 1\} : \xi_n \to \xi \in \Xi_R\}.$$

The important role of the analysis under drifting (sub)sequences has been emphasized in Andrews and Cheng (2012), Cheng (2015), and Andrews et al. (2020). For the system of "moment" functions $G(\pi)$, the selection is correct if $G_c(\pi) = \mathbf{0}$ for some π, and is incorrect if $G_c(\pi) \neq \mathbf{0}$ for any π. Under drifting sequence of model parameters, it is possible that some selected system is not correct, but close to being correct. We call such system a nearly correct system; see Appendix B for a formal definition and characterization of a nearly correct system. Using a simple critical value from chi-squared distribution, the asymptotic size of our test can be controlled even in the presence of nearly correct systems.

Denote $\chi^2_{\left[\|c_0\|_0 - (l_\pi - l_R)\right], 1-\alpha}$ as the $(1-\alpha)$-th quantile of the chi-squared distribution with

$\|c_0\|_0 - (l_\pi - l_R)$ degrees of freedom. Assumption 1.3.1 (v) implies that $\|c_0\|_0 - (l_\pi - l_R) \geq 1$. Asymptotic size control under the null hypothesis requires the following assumption under drifting parameter sequence:

Theorem 1.4.2. *For any parameter sequence $\xi_n \in \Xi_R(\xi)$ with any $\xi \in \Xi_R$: (i) $\sqrt{n}\left(G_n(\pi) - G(\pi)\right) \xrightarrow{d} N(0, \Omega(\pi))$ for some positive definite covariance matrix $\Omega(\pi)$; (ii) For any $c \in \mathscr{C}$, $W_n(c) = W(c) + o_p(1)$ with $W(c)$ being positive definite; and (iii) $W(c_0) = \Omega_0^{-1}$, where Ω_0 is the asymptotic variance of $\sqrt{n}\left(G_{nc_0}(\pi_0) - G_{c_0}(\pi_0)\right)$.*

The following theorem establishes asymptotic validity of the test T_n based on the critical value $\chi^2_{\left[\|c_0\|_0 - (l_\pi - l_R)\right], 1-\alpha}$ and its consistency for the Generic Model.

Theorem 1.4.2. *Let Assumption 1.3.1 hold. (i) If in addition Assumption 1.4.2 is satisfied, then*

$$\limsup_{n\to\infty} \sup_{\xi\in\Xi_R} \Pr_\xi\left(T_n > \chi^2_{\left[\|c_0\|_0 - (l_\pi - l_R)\right], 1-\alpha}\right) = \alpha; \tag{1.17}$$

(ii) If in addition Assumption 1.4.1 holds, then for any $\xi \notin \Xi_R$,

$$\lim_{n\to\infty} \Pr_\xi\left(T_n > \chi^2_{\left[\|c_0\|_0 - (l_\pi - l_R)\right], 1-\alpha}\right) = 1. \tag{1.18}$$

Remark 1.4.1. *Note that although Assumption 1.4.1 (ii) and Assumption 1.4.2 (ii) are both about asymptotic positive definiteness of the weighting matrix, they differ in that the former is under fixed parameter ξ while the latter is under drifting parameter sequence ξ_n. This distinction also applies to subsequent assumptions regarding the Simple Game and its variants.*

1.4.2 The Simple Game and its Variants

Estimation and Consistency

The Simple Game and its variants can be estimated sequentially. We first estimate equilibrium CCPs on all observed and latent states using the sample version of the eigendecomposition procedure presented in Appendix A.1. By replacing the population CCPs with their estimators in $\overline{\pi}$ and Γ respectively, we obtain $\overline{\pi}_n$, Γ_n, and sample "moment" functions for the Simple Game and its variants. The payoff vector π and correct matching can be estimated by solving the minimization problem defined in (1.14). In the remaining part of this section, we first provide primitive assumptions on the Simple Game, under which conditions in Theorem 1.4.1 are satisfied. Then we state results for the variants of the Simple Game.

Assumption 1.4.1. *(i) For* $\forall\ c \in \mathscr{C}$, $W_n(c) \xrightarrow{p} W(c)$ *for some positive definite matrix* $W(c)$*; and (ii)* $f(\cdot)$ *is continuous.*

Assumption 1.4.2. *There exists a known player and an action such that his equilibrium CCP for choosing this action is different across different values of* k*.*

Assumption 1.4.1 (i) imposes a standard assumption on the weighting matrix, and Assumption 1.4.1 (ii) implies that the Jacobian of the "moment" functions with respect to CCPs is bounded. Assumption 1.4.2 is needed for the consistency and asymptotic normality of estimated CCPs using eigendecomposition in Xiao (2018). It requires that the eigenvalues in our eigendecomposition be simple. In the Simple Game, because the CCPs for player 3 choosing 1 are eigenvalues in the eigendecomposition, this assumption requires that player 3 have different probabilities of choosing action 1 on different latent states.

Assumption 1.4.3. *The space* Π *is compact.*

The following Proposition states consistency of the estimators $\widehat{c}$ and $\widehat{\pi}$.

Proposition 1.4.1. *Under Assumptions 1.2.1-1.2.6 and 1.4.1-1.4.3, it holds that* $\widehat{c} = c_0$ *wp* $\to 1$ *and* $\widehat{\pi} \xrightarrow{p} \pi_0$ *for the Simple Game.*

For the variants of the Simple Game, consistency of the estimated payoff vector and selected matching is provided in the following Proposition.

Proposition 1.4.2. *Adapt Assumptions 1.2.1-1.2.6 and 1.4.1-1.4.3 to the variants of the Simple Game. Then* $\widehat{c} = c_0$ *wp* $\to 1$ *and* $\widehat{\pi} \xrightarrow{p} \pi_0$ *hold for each game in Sections 1.3.1-1.3.3.*

Testing Linear Hypothesis

Testing a linear restriction on the payoff vector in the Simple Game and its variants fits into the testing framework developed in Section 1.4.1. Consider the following linear restriction on the payoff vector in the Simple Game and its three variants:

$$H_0 : R\pi_0 = r \text{ against } H_1 : R\pi_0 \neq r.$$

For $l_R = \operatorname{rank}(R)$, we have $l_R \leq 4$ for the Simple Game, $l_R \leq 5$ for the game in Section 1.3.1, and $l_R \leq 2$ for the games in Sections 1.3.2 and 1.3.3. In the following discussion, we provide sufficient conditions and construct critical values for tests that control the asymptotic size as in Theorem 1.4.2 for the Simple Game and its variants.

The Simple Game is fully characterized by the model parameter $\xi \in \Xi$, which includes the distribution of private information, conditional probability of latent state on observed state, probability of observed state, and payoff vectors for each individual. We impose the following assumptions on the model parameters for the Simple Game.

Assumption 1.4.4. *(i) The derivative of $f(\cdot)$ is bounded; (ii) For any $\xi \in \Xi_R$, $\mathbf{z}$ takes each value in $S_{\mathbf{z}}$ with positive probability; (iii) For any $\xi \in \Xi_R$ and the parameter sequence $\{\xi_n\} \in \Xi_R(\xi)$, given each $c \in \mathscr{C}$, $W_n(c) = W(c) + o_p(1)$ with $W(c)$ being positive definite; and (iv) $W(c_0) = \Omega_0^{-1}$ for Ω_0 being the asymptotic variance of $\sqrt{n}\left(G_{nc_0}(\pi_0) - G_{c_0}(\pi_0)\right)$.*

Assumption 1.4.4 (i) is satisfied by commonly used distributions for the private information such as the normal distribution. Assumption 1.4.4 (ii) assumes that the support of $\mathbf{z}$ is the same for all $\xi \in \Xi_R$. Assumption 1.4.4 (iii) and (iv) require that the probability limit of $W_n(c)$ be positive definite for each $c \in \mathscr{C}$ and that the optimal weighting matrix be used for c_0.

The following proposition states asymptotic validity and consistency of the test based upon the test statistic T_n defined in (1.16) and critical value $\chi^2_{[5+l_R],1-\alpha}$ for the Simple Game.

Proposition 1.4.3. *Let Assumptions 1.2.1-1.2.6 hold. (i) If in addition Assumptions 1.4.2-1.4.4 hold, then*

$$\limsup_{n\to\infty} \sup_{\xi\in\Xi_R} \Pr_\xi\left(T_n > \chi^2_{[5+l_R],1-\alpha}\right) = \alpha.$$

(ii) If in addition Assumptions 1.4.1-1.4.3 hold, then for any $\xi \notin \Xi_R$,

$$\lim_{n\to\infty} \Pr_\xi\left(T_n > \chi^2_{[5+l_R],1-\alpha}\right) = 1.$$

Results for the variants of the Simple Game are stated below.

Proposition 1.4.4. *(i) Adapting Assumptions 1.2.1-1.2.6 and 1.4.2-1.4.4, the same test statistic together with $\chi^2_{[13+l_R],1-\alpha}$ for the game in Section 1.3.1, with $\chi^2_{[6+l_R],1-\alpha}$ for the game in Section 1.3.2, and with $\chi^2_{[2+l_R],1-\alpha}$ for the game in Section 1.3.3 achieve asymptotic size control in the sense of (1.17); and (ii) Adapting Assumptions 1.2.1-1.2.6 and 1.4.1-1.4.3 to the games in Sections 1.3.1-1.3.3, the tests are consistent for each game in the sense of (1.18).*

Bootstrap Estimation of the Weighting Matrix

To implement the proposed tests for the Simple Game and its variants, the weighing matrices need to satisfy Assumption 1.4.4 (iii) and (iv), which involves estimating the asymptotic covariance matrix of $\sqrt{n}\left(G_{nc_0}(\pi_0) - G_{c_0}(\pi_0)\right)$. We propose to estimate it using the non-parametric bootstrap.

Any $\pi \in \Pi$ satisfying the system of linear equations $R\pi = r$ can be expressed as $\Psi\pi_f + \mu$, where Ψ is a known $l_\pi \times (l_\pi - l_R)$ matrix, π_f is the free parameter vector of dimension $l_\pi - l_R$, and μ is a known $l_\pi \times 1$ vector. Computation of bootstrap weighting matrix $W_n^b(c)$ includes the following steps.

Step 1: For any given $c \in \mathscr{C}$, if $\arg\min_{\pi\in\Pi} \left\|G_{nc}(\Psi\pi_f + \mu)\right\|^2$ is not unique, then set $W_n^b(c)$ as some known positive definite matrix $W_P(c)$ such as the identify matrix. Otherwise, let $\widehat{\pi}_f(c) = \arg\min_{\pi\in\Pi} \left\|G_{nc}(\Psi\pi_f + \mu)\right\|^2$ and continue to Step 2.

Step 2: Compute the bootstrap variance

$$\Sigma_n^b\left(c, \widehat{\pi}_f(c)\right) = \frac{n}{B}\sum_{b=1}^{B}\left(G_{nc}^{(b)}\left(\Psi\widehat{\pi}_f(c)+\mu\right)-\overline{G}_{nc}\right)\left(G_{nc}^{(b)}\left(\Psi\widehat{\pi}_f(c)+\mu\right)-\overline{G}_{nc}\right)^{\top},$$

where $G_{nc}^{(b)}(\cdot)$ is calculated using the b-th nonparametric bootstrap sample and $\overline{G}_{nc} = \frac{1}{B}\sum_{b=1}^{B} G_{nc}^{(b)}\left(\Psi\widehat{\pi}_f(c)+\mu\right)$. Set $W_n^b(c)$ as $\left(\Sigma_n^b\left(c, \widehat{\pi}_f(c)\right)\right)^{-1}$.

Step 3: Repeat Step 1 and 2 for every $c \in \mathscr{C}$.

By Assumption 1.2.6, $\operatorname{rank}\left(\Gamma_{nc_0}\Psi\right) = l_\pi - l_R$ for n sufficiently large. However, for $c \neq c_0$, it is possible that $\operatorname{rank}\left(\Gamma_c\Psi\right) < l_\pi - l_R$ and $\arg\min_{\pi\in\Pi}\left\|G_{nc}\left(\Psi\pi_f+\mu\right)\right\|^2$ is not unique. Fortunately, for such c, Assumption 1.4.4 (iii) only requires the weighing matrix to be positive definite in the limit, which is satisfied by matrix $W_P(c)$.

For $\Sigma_n^b\left(c_0, \widehat{\pi}_f(c_0)\right)$ to be a consistent estimator of Ω_0, we need to match the labels of latent states across different bootstrap draws. The following assumption provides a sufficient condition.

Assumption 1.4.5. *There is a known function $\psi(\cdot)$, such that for each $\mathbf{z} \in S_{\mathbf{z}}$,*

$$\psi\left(p_1(\mathbf{z},k), p_2(\mathbf{z},k), p_3(\mathbf{z},k), p^k(\mathbf{z})\right) \neq \psi\left(p_1(\mathbf{z},k'), p_2(\mathbf{z},k'), p_3(\mathbf{z},k'), p^{k'}(\mathbf{z})\right).$$

In the Simple Game, this function could be taken as $\psi\left(\cdot,\cdot,\cdot,p^k(\mathbf{z})\right) = p^k(\mathbf{z})$, which allows researchers to track the labels of latent states across bootstrap draws using the distinctive weight corresponding to each latent state. Another example could be

$$\psi\left(p_1(\mathbf{z},k), p_2(\mathbf{z},k), p_3(\mathbf{z},k), p^k(\mathbf{z})\right) = \sum_{i=1}^{3} p_i(\mathbf{z},k).$$

Under Assumption 1.4.5, the labels of latent states are matched across bootstrap draws with probability 1 as n goes to infinity. Conditional on latent states being matched across bootstrap draws, consistency of bootstrap variance at c_0 follows from the standard argument. The property of $W_n^b(c)$ is summarized in the following proposition. The result applies to the Simple Game and its variants.

Proposition 1.4.5. *Suppose Assumptions 1.2.1-1.2.6, 1.4.2, 1.4.3, 1.4.4 (i) and (ii), and 1.4.5 hold. Then $W_n^b(c)$ satisfies Assumption 1.4.4 (iii) and (iv).*

1.4.3 Computation via Sequential Monte Carlo

In empirical studies of games, the number of players is usually small, but the cardinality of the support of state variables can potentially be large. This might be the case, in particular, when the state variable is obtained by discretizing some continuous variable. In the Simple Game, the number of elements in each selection vector space for player 1 on exclusive observed state z_1 is $2^{|S_{z_2}|\times|S_{z_3}|}$, which increases exponentially with $|S_{z_2}| \times |S_{z_3}|$. Solving the optimization problem (1.14) and computing the test statistic defined in (1.16) by exhaustive search over the selection vector space quickly becomes infeasible as the cardinality of each observed state variable increases. To address this issue, we propose a stochastic optimization algorithm based on the sequential Monte Carlo (SMC) method developed by Duan (2019).

Based on the insight from Duan (2019), we note that finding the minimum of

$$J_n(c) \equiv \|G_{nc}(\widehat{\pi}(c))\|^2_{W_n(c)}$$

on $\mathscr{C}$ is equivalent to finding the mode of the following probability mass function:

$$f_n(c) = \frac{\exp(-J_n(c))}{\sum\limits_{c\in\mathscr{C}} \exp(-J_n(c))} \equiv \frac{\exp(-J_n(c))}{D_{n\mathscr{C}}}.$$

The mode of this probability mass function could be estimated based on subsampling. If we could draw a representative sample from this probability mass function, then the empirical mode in our sample will very likely deliver our desired solution $\widehat{c}$.

Notice that each selection vector $c \in \mathscr{C}$ determines a combination of rows to select. Corresponding to each combination of rows, there are 8! number of permutations in the case of the Simple Game (given some selected latent state on the first observed state). Since the SMC sample representing the distribution function over combinations is exactly the same as the one over permutations after collapsing permutations into combinations, we convert the problem of finding the mode of $f_n(c)$ over combinations c into the corresponding one over permutations denoted as u. As in Duan (2019), a distribution function defined over permutations is easier to work with, because permutations are easier to sample and the distribution function defined over permutations can be proportionally scaled up to obtain the target distribution function defined over combinations.

As an illustration, let us fix the first row on the first observed state, and obtain the following selection vector space defined using permutations

$$\mathscr{U} \equiv \left\{(u_1, ..., u_9)^\top : 1 \leq u_r \leq 18 \text{ for } r = 1, \ldots, 9 \text{ and } u_{r_1} \neq u_{r_2} \text{ for } r_1 \neq r_2\right\}.$$

For example, for selection vector $c = (1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0)^\top$, it selects rows $(1,3,5,7,9,11,13,15,17)^\top$, corresponding to $8!$ elements in $\mathscr{U}$ by fixing the first row at 1 and exhausting all permutations of $(3,5,7,9,11,13,15,17)^\top$ for the second to ninth rows. Given a permutation $u \in \mathscr{U}$, let

$$J_n(u) \equiv \|G_{nu}(\widehat{\pi}(u))\|^2_{W_n(u)},$$

where the rows specified by u are selected. The probability mass function defined over permutations is

$$f_n(u) = \frac{\exp(-J_n(u))}{\sum_{u \in \mathscr{U}} \exp(-J_n(u))} \equiv \frac{\exp(-J_n(u))}{D_{n\mathscr{U}}}.$$

For large state space, direct sampling from this distribution is not possible because the normalizing constant in the denominator is not computable when $|\mathscr{U}|$ is huge. The idea for density-tempered SMC is to draw an initial sample (a set of particles) from some known initial distribution $I(u)$ which is easy to sample from, then move the sample points along a controlled smooth bridge towards our final target distribution $f_n(u)$. We present the detailed algorithm in Appendix A.2.

1.5 An Extended Game with Both Multiple Equilibria and Unobserved Heterogeneity

This section extends the Simple Game introduced in Section 1.2 to allow for any finite number of players, any finite number of actions, and more importantly both multiple equilibria and unobserved heterogeneity. We call this game the *Extended Game*. Methods developed for the Simple Game are not directly applicable to the Extended Game because of the presence of multiple equilibria. In Section 1.5.2, we develop sequential estimation and inference procedures for an Extended Generic Model. In Section 1.5.2 we apply these procedures to the Extended Game, and provide an approach to testing the existence of unobserved heterogeneity in the Extended Game. We introduce three variants of the Extended Game and discuss estimation and inference in these games in Appendix C.

1.5.1 The Extended Game and Minimum Distance Characterization

In the Extended Game, there are N players denoted as $i = 1, ..., N$. Player i has $J + 1$ available actions denoted as $d_i \in \{0, ..., J\}$ and an exclusive observed state variable z_i. For notational simplicity, we assume that z_i for $i = 1, 2, ..., N$ have common support S_z.[14] Like in the Simple Game in Section 1.2, $k \in S_k$ is a payoff relevant state variable that is common knowledge for players but unobserved by the econometrician. We impose similar assumptions on the private information and data:

Assumption 1.5.1. *(i)* $\{\epsilon_i\}_{i=1}^{N} \overset{i.i.d}{\sim} F(\cdot)$*, where* $F(\cdot)$ *is known to the econometrician and is absolutely continuous with a probability density function denoted as* $f(\cdot)$*; (ii)* $\epsilon_1, \ldots, \epsilon_N$ *are independent of the state variables* $(\mathbf{z}, k)$*.*

Assumption 1.5.2. *The econometrician observes a random sample on players' actions and observable state variables* $\{(d_{1l}, \ldots, d_{Nl}, z_{1l}, \ldots, z_{Nl})\}_{l=1}^{n}$*.*

In contrast to the Simple Game, we allow for the presence of multiple equilibria in the

[14] All our results can be easily extended to the case where z_i's do not have common support.

Extended Game and use τ to denote the equilibrium indicator. Both the number of latent states and the number of multiple equilibria on each latent state are assumed to be finite.

Let $b(\cdot,\cdot)$ be some bivariate injective function. We define $\omega \equiv b(k,\tau)$ as the scalar composite latent variable for the combination of latent state and equilibrium indicator. Let $\mathbf{z} \equiv (z_1, ..., z_N) \in S_{\mathbf{z}} \equiv S_z \times ... \times S_z$. And denote $p(\omega \mid \mathbf{z})$ as the conditional probability mass function for this composite latent variable given observed state vector $\mathbf{z}$ with support $S_{\omega|\mathbf{z}}$. Without loss of generality, we assume that the function $b(\cdot,\cdot)$ is such that $\omega \in \left\{1, ..., \left|S_{\omega|\mathbf{z}}\right|\right\}$ for any $\mathbf{z} \in S_{\mathbf{z}}$.

In the presence of multiple equilibria, the support of the composite latent variable ω depends on $\mathbf{z}$, because the number of active equilibria might change with $\mathbf{z}$. To simplify the discussion, we assume that the number of active latent states is the same across all $\mathbf{z} \in S_{\mathbf{z}}$. The following assumption is on the distribution of the latent state.

Assumption 1.5.3. $p^k(\mathbf{z}) > 0$ *for any* $k \in S_k$ *and* $\mathbf{z} \in S_{\mathbf{z}}$.

Under Assumption 1.5.3, the total number of parameters to be estimated for player i and his exclusive observed state z_i is equal to $J \times (J+1)^{N-1} \times |S_k|$.

Step-1: Identification of Equilibrium CCPs For identification of equilibrium CCPs in step-1, we make use of the following system of equations:

$$p(d_1, ..., d_N \mid \mathbf{z}) = \sum_{\omega \in S_{\omega|\mathbf{z}}} \left[p(\omega \mid \mathbf{z}) \prod_{i=1}^{N} p(d_i \mid \mathbf{z}, \omega) \right], \text{ for } (d_1, ..., d_N) \in \{0, ..., J\}^N .$$

For a given $\mathbf{z}$, the number of mixtures is $\left|S_{\omega|\mathbf{z}}\right| \geq |S_k|$ because of the possible presence of multiple equilibria. Identification follows from the approach in Xiao (2018) described in Appendix A.1. Below we introduce the assumption needed for identification.

Divide the N players into three groups, such that the third group has exactly one player for odd N and two players for even N, and each of the first two groups has $\widetilde{N}$ players. Thus, $N = 2\widetilde{N} + 1$ when N is odd; and $N = 2\widetilde{N} + 2$ when N is even. Player group i is denoted as g_i for $i = 1, 2, 3$. By definition, $\bigcup_{i=1}^{3} g_i = \{1, ..., N\}$. For each group, we create a group action variable, denoted by d_{g_1}, d_{g_2}, and d_{g_3}. We have $d_{g_1}, d_{g_2} \in \left\{0, ..., (J+1)^{\widetilde{N}} - 1\right\}$, and $d_{g_3} \in \{0, ..., J\}$ if there is one player in group 3 and $d_{g_3} \in \left\{0, ..., (J+1)^2 - 1\right\}$ if there are two players in group 3. We introduce the following notation to denote the matrix composed of CCPs for group action d_{g_i} for $i = 1, 2$ on each latent state:

$$\mathbf{P}_{g_i \mathbf{z}} = \left(\Pr\left(d_{g_i} = j \mid \mathbf{z}, \omega\right)\right)_{j=0,\omega=1}^{(J+1)^{\widetilde{N}}-1, \left|S_{\omega|\mathbf{z}}\right|}.$$

Based on Xiao (2018) and Aguirregabiria and Mira (2019), we impose the following assumption for identification of CCPs (up to a label swapping) in the first step.

Assumption 1.5.4 (Step-1 Identification)**.** *(i)* $(J+1)^{\widetilde{N}} > \left|S_{\omega|\mathbf{z}}\right|$ *for any* $\mathbf{z}$*; (ii)* $N \geq 3$ *and* $(J+1)^{\widetilde{N}} \geq |S_k|$*; (iii) For each* $\mathbf{z}$*, there exists a partition* $(d_{g_1}, d_{g_2}, d_{g_3})$ *of joint actions* $(d_1, ..., d_N)$ *such that* $\mathbf{P}_{g_1\mathbf{z}}$ *and* $\mathbf{P}_{g_2\mathbf{z}}$ *both have full column rank; (iv) There exists some* $\mathbf{z}^*$ *such that a single equilibrium is being played on* $(\mathbf{z}^*, k)$ *for any* $k \in S_k$*.*

Assumption 3.7.3 guarantees the identification of equilibrium CCPs (up to a label swapping) and the identification of the number of latent states. The latter identification is needed to determine the number of different payoff vectors for a player on his exclusive observed state. Under Assumption 3.7.3 (i), the approach in Xiao (2018) applies to identify $\left|S_{\omega|\mathbf{z}}\right|$ for each observed state $\mathbf{z}$. Assumption 3.7.3 (ii) and (iii) are conditions required for nonparametric identification of a finite mixture model given $\left|S_{\omega|\mathbf{z}}\right|$. Finally, Assumption 3.7.3 (iv) implies that it can be identified via $|S_k| = \min_{\mathbf{z}} \left|S_{\omega|\mathbf{z}}\right|$.

Remark 1.5.1. *When Assumption 3.7.3 (iv) is not imposed,* $|S_k|$ *is still identified after the payoff vector corresponding to each* ω *is recovered: it equals to the number of distinct payoff vectors. However, in this case, the estimation procedure for recovering economically meaningful payoff vectors is different (an additional step is needed to consistently collect estimated payoffs corresponding to the same* k *together), we study such case in a subsequent paper.*

Step-2: Identification of Payoffs and Minimum Distance Characterization Without loss of generality, we focus on the identification of player 1's payoff vector when his exclusive observed state z_1 takes the smallest value in S_z. With slight abuse of notation, for $s = 1, \ldots, \left|S_{\omega|\mathbf{z}}\right|$, we define $\omega(s, \mathbf{z})$ as the s-th value for the composite latent variable on observed state vector $\mathbf{z}$, and define $T \equiv |S_z|^{N-1}$ as the total number of different observed state vectors holding constant player 1's exclusive observed state. Let $\left\{\mathbf{z}^1, \ldots, \mathbf{z}^T\right\}$ be the collection of observed state vectors when z_1 is fixed at the smallest value in S_z. Define the following vector as the expected payoffs for all actions on some generic observed state $\mathbf{z}$ and composite latent state ω:

$$\overline{\pi}_1(\mathbf{z}, \omega) \equiv \begin{bmatrix} \overline{\pi}_1(1, \mathbf{z}, \omega) \\ \vdots \\ \overline{\pi}_1(J, \mathbf{z}, \omega) \end{bmatrix}.$$

Let $\mathbf{p} \equiv [\Pr(d_1 = 1 \mid \mathbf{z}, \omega), \ldots, \Pr(d_1 = J \mid \mathbf{z}, \omega)]$ denote the identified CCPs for all actions on $(\mathbf{z}, \omega)$. We define the mapping from equilibrium expected payoffs to CCPs on observed state $\mathbf{z}$ and composite latent state ω as:

$$\Lambda_{1\mathbf{z}\omega}(\overline{\pi}_1(\mathbf{z}, \omega)) = \begin{bmatrix} \Pr\left(\overline{\pi}_1(1, \mathbf{z}, \omega) + \epsilon_1 \geq \overline{\pi}_1(j, \mathbf{z}, \omega) + \epsilon_j \text{ for every } j\right) \\ \vdots \\ \Pr\left(\overline{\pi}_1(J, \mathbf{z}, \omega) + \epsilon_J \geq \overline{\pi}_1(j, \mathbf{z}, \omega) + \epsilon_j \text{ for every } j\right) \end{bmatrix} = \mathbf{p}.$$

The inversion of this mapping $\Lambda_{1\mathbf{z}\omega}^{-1}(\mathbf{p})$ identifies $\overline{\pi}_1(\mathbf{z}, \omega)$ from CCPs (see Hotz and Miller (1993) and Aguirregabiria and Mira (2019)). Although this inversion will no longer be the inverse cumulative distribution function when $J \geq 2$, it has a closed form when private information follows a logistic distribution (see Bajari et al. (2013)). Given the identification of expected payoffs for all $(\mathbf{z}, \omega)$, we can construct player 1's expected payoff vector and coefficient matrix for the fixed value of his exclusive observed state by stacking the corresponding expected payoffs and CCPs for all j, $\mathbf{z}$ and ω:[15]

$$\overline{\pi} = \left[\overline{\pi}_1\left(\mathbf{z}^1\right)^\top, \ldots, \overline{\pi}_1\left(\mathbf{z}^T\right)^\top\right]^\top \text{ and } \Gamma = \left[\Gamma_1\left(\mathbf{z}^1\right)^\top, \ldots, \Gamma_1\left(\mathbf{z}^T\right)^\top\right]^\top,$$

where for $t = 1, ..., T$,

$$\overline{\pi}_1\left(\mathbf{z}^t\right) = \begin{bmatrix} \overline{\pi}_1\left(1, \mathbf{z}^t, \omega\left(1, \mathbf{z}^t\right)\right) \\ \vdots \\ \overline{\pi}_1\left(J, \mathbf{z}^t, \omega\left(1, \mathbf{z}^t\right)\right) \\ \vdots \\ \overline{\pi}_1\left(1, \mathbf{z}^t, \omega\left(\left|S_{\omega|\mathbf{z}^t}\right|, \mathbf{z}^t\right)\right) \\ \vdots \\ \overline{\pi}_1\left(J, \mathbf{z}^t, \omega\left(\left|S_{\omega|\mathbf{z}^t}\right|, \mathbf{z}^t\right)\right) \end{bmatrix} \text{ and } \Gamma_1\left(\mathbf{z}^t\right) = \begin{bmatrix} \iota_1\left(1, \mathbf{z}^t\right) \\ \vdots \\ \iota_1\left(\left|S_{\omega|\mathbf{z}^t}\right|, \mathbf{z}^t\right) \end{bmatrix}.$$

Note that matrix $\iota_1\left(s, \mathbf{z}^t\right)$ with $s \in \left\{1, \ldots, \left|S_{\omega|\mathbf{z}^t}\right|\right\}$ is a block diagonal matrix with J identical blocks defined as:[16]

$$\iota_1\left(s, \mathbf{z}^t\right) = \begin{bmatrix} \mathbf{p}_{-1}\left(\mathbf{z}^t, \omega\left(s, \mathbf{z}^t\right)\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{p}_{-1}\left(\mathbf{z}^t, \omega\left(s, \mathbf{z}^t\right)\right) \end{bmatrix},$$

where $\mathbf{p}_{-1}\left(\mathbf{z}^t, \omega\left(s, \mathbf{z}^t\right)\right)$ is a row vector with $(J+1)^{N-1}$ elements given by the probabilities of joint actions for all other players in the same spirit as the Simple Game in Section 1.2.2.

Step-2 makes use of the following system of "moment" functions to identify the payoff vector:

$$G\left(\pi\right) = \overline{\pi} - \Gamma\pi, \tag{1.19}$$

where $\overline{\pi}$ has dimension $J\sum_{i=1}^{T}\left|S_{\omega|\mathbf{z}^i}\right|$, Γ has dimension $J\sum_{i=1}^{T}\left|S_{\omega|\mathbf{z}^i}\right|$ by $J\left(J+1\right)^{N-1}$, and π has dimension $J\left(J+1\right)^{N-1}$.

[15] Similar to the Simple Game introduced in Section 1.2, we have normalized the payoff function for the first action $d_i = 0$ to 0.

[16] The block diagonal structure is a consequence of $J > 1$, as now each row in $\overline{\pi}_1(\mathbf{z}^t)$ corresponds to a single action for player 1 while the payoff vector stacks other actions as well.

Given (1.19), methods for the identification, estimation, and inference for the Simple Game can be modified to the Extended Game accounting for the presence of multiple equilibria. There are in total $|S_k|$ number of selection vector spaces. When there are multiple equilibria on some latent state, there will be more than one selection vectors (in the selection vector space corresponding to this latent state) that generates a system that has a solution (not necessarily unique). To make this clear, we define the selection vector space for the Extended Game below. Let $\mathbf{e}_1$ be a J-dimensional row vector of ones, and $\mathbf{e}_0$ be a J-dimensional row vector of zeros. Without loss of generality, suppose $\mathbf{z}^1$ has the minimum number of mixtures.[17]

Definition 1.5.1 (Selection Vector Space)**.** *For $s = 1, ..., |S_k|$, the selection vector space $\mathscr{C}^s$ is defined as*

$$\mathscr{C}^s \equiv \left\{ \begin{array}{c} (c_1, ..., c_T)^\top : c_1 = \left[c_{1,1}, \ldots, c_{1,\left|S_{\omega|\mathbf{z}^1}\right|}\right] \textit{ with } c_{1,s} = \mathbf{e}_1 \textit{ and } c_{1,j} = \mathbf{e}_0 \textit{ for } j \neq s; \\ \textit{for } t = 2, \ldots, T,\ c_t = \left[c_{t,1}, \ldots, c_{t,\left|S_{\omega|\mathbf{z}^t}\right|}\right] \textit{ with } c_{t,w} \in \{\mathbf{e}_1, \mathbf{e}_0\}, \\ \textit{where } w \in \left\{1, \ldots, \left|S_{\omega|\mathbf{z}^t}\right|\right\} \end{array} \right\}.$$

Following Andrews (1999), we extend the definition of correct matching in the selection vector space for the Simple Game to the Extended Game as follows. The definition of $\mathscr{C}\mathscr{Z}$ is provided in (1.9).

Definition 1.5.2. *The correct matching in a selection vector space for the Extended Game is defined as the unique element in $\mathscr{C}\mathscr{Z}$ that selects the maximum number of components of $G(\pi)$ corresponding to the same latent state. This correct matching is denoted as c_0.*

By definition, c_0 separates from other elements in $\mathscr{C}\mathscr{Z}$ because $\|c_0\|_0 > \|c\|_0$ for any $c \in \mathscr{C}\mathscr{Z}$ and $c \neq c_0$. For the Extended Game, the maximum number of components in $G(\pi)$ are selected only when all equilibria on a given latent state are selected. The step-2 identification is achieved if c_0 is identified in $\mathscr{C}\mathscr{Z}$ and the true system selected by c_0 from

[17] If the number of mixtures on $\mathbf{z}^1$ is not minimum, we can rename the observed state with minimum number of mixtures as $\mathbf{z}^1$. Under Assumption 3.7.3, this implies $\left|S_{\omega|\mathbf{z}^1}\right| = |S_k|$. This assumption is also adapted to the three variants of the Extended Game defined in Appendix C.

(1.7): $G_{c_0}(\pi) = \mathbf{0}$ has a unique solution π_0. The following assumption guarantees step-2 identification:

Assumption 1.5.5 (Step-2 Identification)**.** Γ_{c_0} *has full column rank, and it holds that* $\operatorname{rank}([\overline{\pi}_c, \Gamma_c]) > \operatorname{rank}(\Gamma_c)$ *for any* $c \notin \mathscr{C}\mathscr{Z}$*.*

A necessary condition for Assumption 1.5.5 to hold is that $\sum_{t=1}^{T} \left|S_{\omega|\mathbf{z}^t}\right| \geq (J+1)^{N-1}$, i.e., the number of equations is no less than the number of unknowns for each action.

Lemma 1.5.1. *Assumptions 1.5.1-1.5.5 are sufficient for the identification of* (c_0, π_0)*.*

Given some selection vector space with the first selected latent state being k, the unique solution π_0 to the system $G_{c_0}(\pi) = \mathbf{0}$ gives us the true payoff vector written as

$$\pi_{1z_1k} = \left[\pi_1(1, z_1, k), \ldots, \pi_1(J, z_1, k)\right]^\top,$$

where for $j = 1, \ldots, J$

$$\pi_1(j, z_1, k) = \left[\pi_1(j, (0, \ldots, 0), z_1, k), \ldots, \pi_1(j, (J, \ldots, J), z_1, k)\right].$$

1.5.2 Estimation and Inference

A Generic Model

Estimation and inference in the presence of multiple equilibria differ from the Simple Game. As we have redefined the correct matching as the element in selection vector space that selects the maximum number of "moments" corresponding to the same latent state, consistency of the selected matching requires to select this vector with probability going to one as the sample size approaches infinity. To account for this new feature, we generalize the Generic Model in Section 1.3 to the Extended Generic Model characterized by (1.8) and the following assumption.

Assumption 1.5.1. *(i) Both* $\overline{\pi}$ *and* Γ *can be identified; (ii)* Π *is compact; (iii)* $\mathscr{C}$ *is known; (iv) There is a unique selection vector denoted as* $c_0 \in \mathscr{C}\mathscr{Z}$ *such that* $\|c_0\|_0 > \|c\|_0$ *for any*

$c \in \mathscr{C}\mathscr{Z}$ and $c \neq c_0$; (v) $cl \equiv \min_{c \in \mathscr{C}\mathscr{Z}} \|c\|_0$ is known; (vi) the system: $G_{c_0}(\pi) = \mathbf{0}$ has a unique solution denoted as π_0.

Assumption 1.5.1 for the Extended Generic Model relaxes Assumption 1.3.1 (iv) for the Generic Model by allowing elements in $\mathscr{C}\mathscr{Z}$ to have different numbers of non-zero elements. Assumption 1.5.1 (v) only requires $\min_{c \in \mathscr{C}\mathscr{Z}} \|c\|_0$ to be known. In the Extended Game, $\min_{c \in \mathscr{C}\mathscr{Z}} \|c\|_0 = JT$.

Following Andrews (1999), we modify the sample criterion function in Section 1.4.1 with a rewarding term. Let $G_n(\pi)$ denote the sample counterpart of $G(\pi)$. Under a generic selection vector space $\mathscr{C}$, the parameter vector can be consistently estimated by

$$(\widehat{c}, \widehat{\pi}) = \arg \min_{c \in \mathscr{C}, \pi \in \Pi} \left[n \|G_{nc}(\pi)\|^2_{W_n(c)} - h(\|c\|_0)\, \kappa_n \right],$$

where $h(\cdot)$ and κ_n are added to reward the selection of more "moments" and $W_n(c)$ converges in probability to some positive definite matrix $W(c)$ for any $c \in \mathscr{C}$. The normalizing factor in front of $\|G_{nc}(\pi)\|^2_{W_n(c)}$ in the above definition is determined by the convergence rate of $G_n(\pi)$ and is set as n to simplify the discussion. For consistency, the following assumption is imposed.

Assumption 1.5.2. *(i) For any π, $G_n(\pi) = G(\pi) + O_p\left(n^{-1/2}\right)$; (ii) For any $c \in \mathscr{C}$, $W_n(c) = W(c) + o_p(1)$ with $W(c)$ being positive definite; (iii) For any $c \in \mathscr{C}$, $\min_{\pi \in \Pi} \|G_{nc}(\pi)\|^2_{W_n(c)} \xrightarrow{p} \min_{\pi \in \Pi} \|G_c(\pi)\|^2_{W(c)}$; (iv) $h(\cdot)$ is a known strictly increasing function and $\kappa_n \to \infty$ with $\kappa_n = o(n)$.*

Note that the condition regarding the order of κ_n can be easily modified to accommodate models with normalizing factors different from n. Consistency of this procedure is given in the following theorem.

Theorem 1.5.1. *Suppose Assumptions 1.5.1 and 1.5.2 hold, then $\widehat{c} = c_0$ wp $\to 1$ and $\widehat{\pi} \xrightarrow{p} \pi_0$.*

For testing the linear hypothesis as (1.15), we still use the test statistic T_n proposed in (1.16). The following theorem shows that using an appropriately chosen chi-squared critical value, the test achieves asymptotic size control and is consistent. Let $\chi^2_{[cl-(l_\pi-l_R)],1-\alpha}$ be the $(1-\alpha)$-th quantile of the chi-squared distribution with $cl-(l_\pi-l_R)$ degrees of freedom. Since $l_R \geq 1$ and $cl \geq l_\pi$ by Assumption 1.5.1 (vi), it holds that $cl-(l_\pi-l_R) \geq 1$.

Assumption 1.5.3. *For any parameter sequence $\xi_n \in \Xi_R(\xi)$ with any $\xi \in \Xi_R$: (i) $\sqrt{n}\left(G_n(\pi)-G(\pi)\right) \xrightarrow{d} N(0,\Omega(\pi))$ for some positive definite covariance matrix $\Omega(\pi)$; (ii) For any $c \in \mathscr{C}$, $W_n(c) = W(c)+o_p(1)$ with $W(c)$ being positive definite, and (iii) $W(c_0) = \Omega_0^{-1}$ for Ω_0 being the asymptotic variance of $\sqrt{n}\left(G_{nc_0}(\pi_0)-G_{c_0}(\pi_0)\right)$.*

Theorem 1.5.2. *Suppose Assumption 1.5.1 hold. (i) If in addition Assumption 1.5.3 is satisfied, then*

$$\limsup_{n\to\infty} \sup_{\xi\in\Xi_R} \Pr_\xi\left(T_n > \chi^2_{[cl-(l_\pi-l_R)],1-\alpha}\right) \leq \alpha. \tag{1.20}$$

(ii) If in addition Assumption 1.5.2 is satisfied, then for any $\xi \notin \Xi_R$,

$$\lim_{n\to\infty} \Pr_\xi\left(T_n > \chi^2_{[cl-(l_\pi-l_R)],1-\alpha}\right) = 1. \tag{1.21}$$

The proof of this theorem is similar to that of Theorem 1.4.2 and is provided in Appendix B. When the model does not have multiple equilibria, we obtain equality instead of inequality in (1.20).

The Extended Game

In this section, we provide primitive assumptions such that the conditions in Theorem 1.5.1 and Theorem 1.5.2 are satisfied for the Extended Game.

Assumption 1.5.6. *(i) $\forall\ c \in \mathscr{C}$, $W_n(c) \xrightarrow{p} W(c)$ for some positive definite matrix $W(c)$; (ii) $f(\cdot)$ is continuous; (iii) $h(\cdot)$ is a known strictly increasing function and $\kappa_n \to \infty$ with $\kappa_n = o(n)$.*

Assumption 1.5.7. *(i) There exists a known group action for player group 3 such that the corresponding equilibrium CCP of choosing this action is different across different values of ω; (ii) Π is compact.*

Proposition 1.5.1. *Under Assumptions 1.5.1-1.5.7, it holds that* $\widehat{c} = c_0 \quad wp \to 1$ *and* $\widehat{\pi} \xrightarrow{p} \pi_0$.

For testing linear hypothesis on the payoff vectors, we have $l_R \leq J(J+1)^{N-1}$. Recall that here we are focusing on player 1 for some fixed value of z_1. For asymptotic size control, we introduce an additional assumption.

Assumption 1.5.8. *(i) On any* $(\mathbf{z}, \omega)$, $\frac{\partial^2 \Lambda^{-1}_{1\mathbf{z}\omega}(\mathbf{p})}{\partial \mathbf{p} \partial \mathbf{p}^\top}$ *is bounded for any* $\mathbf{p}$*; (ii) For any* $\xi \in \Xi_R$, $\mathbf{z}$ *takes each value in* $S_{\mathbf{z}}$ *with positive probability; (iii) For any* $\xi \in \Xi_R$ *and the parameter sequence* $\{\xi_n\} \in \Xi_R(\xi)$*, given each* $c \in \mathscr{C}$, $W_n(c) = W(c) + o_p(1)$*, with* $W(c)$ *being positive definite; and (iv)* $W(c_0) = \Omega_0^{-1}$ *for* Ω_0 *being the asymptotic variance of* $\sqrt{n}\left(G_{nc_0}(\pi_0) - G_{c_0}(\pi_0)\right)$.

The following Proposition applies Theorem 1.5.2 to the Extended Game.

Proposition 1.5.2. *Let Assumptions 1.5.1-1.5.5 hold. (i) If in addition Assumptions 1.5.7 and 1.5.8 are satisfied, then*

$$\limsup_{n\to\infty} \sup_{\xi \in \Xi_R} \Pr_\xi \left(T_n > \chi^2_{[JT - J(J+1)^{N-1} + l_R], 1-\alpha} \right) \leq \alpha.$$

(ii) If in addition Assumptions 1.5.6 and 1.5.7 hold, then for any $\xi \notin \Xi_R$,

$$\lim_{n\to\infty} \Pr_\xi \left(T_n > \chi^2_{[JT - J(J+1)^{N-1} + l_R], 1-\alpha} \right) = 1.$$

The bootstrap weighting matrix $W_n^b(c)$ can be computed in a similar way to that in Section 1.4.2. For implementation of the test using $W_n^b(c)$, we need an assumption similar to Assumption 1.4.5.

Assumption 1.5.9. *For each* $\mathbf{z}$*, there exists a known scalar valued function of the CCPs on each* ω*, such that the values for the function are all different across* ω.

Under Assumptions 1.5.3-1.5.5, 1.5.7, 1.5.8 (i) and (ii), and 1.5.9, the bootstrap weighting matrix satisfies Assumption 1.5.8 (iii) and (iv). Note that the critical value used in the

Extended Game is the same as the one without multiple equilibria. This shows that the testing framework we propose in Section 1.4.2 is robust to the presence of multiple equilibria.

Remark 1.5.2. *For inference, we assume that the number of mixtures is known and ignore the estimation error involved in estimating the number of mixtures. Alternatively, one could adopt sample splitting by using one subsample to estimate the number of mixtures and the other subsample to compute the test statistic.*

Testing for Unobserved Heterogeneity

When there is no unobserved heterogeneity and all mixtures are generated by multiple equilibria, the Extended Game reduces to that studied in Xiao (2018). Whether or not these mixtures are generated by unobserved heterogeneity can be tested empirically using Hansen's J-test. The test statistic is defined as

$$J_n = \min_{\pi \in \Pi} \left\| \sqrt{n} G_n \left(\pi \right) \right\|^2_{W_n},$$

where W_n is the inverse of a consistent estimator of the asymptotic variance of $\sqrt{n} G_n \left(\pi_0 \right)$. Intuitively, under the null hypothesis of no unobserved heterogeneity, a single payoff vector satisfies all "moments", and the asymptotic distribution of the test statistic is tight. On the other hand, if there is unobserved heterogeneity, a single payoff vector cannot satisfy all "moments" and the test statistic will diverge to infinity.

Remark 1.5.3. *Note that although the framework we have developed so far assumes that there are at least two mixtures for each observed state, it could be easily extended to the situation in which some observed states have only one mixture. The observed states with only one mixture are treated as having a degenerate latent state. In this case the number of latent states is equal to the minimum number of mixtures among all observed states with non-degenerate latent states.*

1.6 Monte Carlo Simulation

We conduct a small Monte Carlo simulation to examine the performance of our estimation and inference methods in finite samples. We also compare the running times of the stochastic optimization algorithm based on SMC described in Appendix A.2 and the exhaustive search.

For the estimation and inference, we adopt a data generating process that follows a slight variation of the parametric game with random coefficients introduced in Section 1.3.2. There are three players each having an exclusive state variable taking two values $\{L, M\}$. The payoff function for player i of choosing $d_i = 1$ is given by Equation (1.12) and ϵ_i is drawn independently across players from a logistic distribution. This implies that the equilibrium CCP vector is a fixed point of the following system:

$$\begin{aligned}
p_1 &= \frac{\exp\left(\theta_{1k} z_1 + \delta_{1k}\left(p_2 + p_3\right)\right)}{1 + \exp\left(\theta_{1k} z_1 + \delta_{1k}\left(p_2 + p_3\right)\right)}, \\
p_2 &= \frac{\exp\left(\theta_{2k} z_2 + \delta_{2k}\left(p_1 + p_3\right)\right)}{1 + \exp\left(\theta_{2k} z_2 + \delta_{2k}\left(p_1 + p_3\right)\right)}, \\
p_3 &= \frac{\exp\left(\theta_{3k} z_3 + \delta_{3k}\left(p_1 + p_2\right)\right)}{1 + \exp\left(\theta_{3k} z_3 + \delta_{3k}\left(p_1 + p_2\right)\right)}.
\end{aligned}$$

Let $\mathbf{z}^1, \mathbf{z}^2, ..., \mathbf{z}^8$ denote the eight distinct values that $\mathbf{z}$ takes. The distribution of the latent variable is specified as

$$\Pr\left(k = A \mid \mathbf{z}\right) = \begin{cases} \gamma - \frac{\alpha}{\sum_{i=1}^{3} |z_i|} & \text{for } \mathbf{z} \neq \mathbf{z}^2 \\ 0.5 & \text{for } \mathbf{z} = \mathbf{z}^2 \end{cases}.$$

This distribution does not satisfy ranking independence used in Aguirregabiria and Mira (2019). For example, when $\mathbf{z} = \mathbf{z}^2$, components A and B have equal weights and they cannot be separated using their weights. Parameter values used in simulation are specified in Table 2.1 and the equilibrium CCPs are given in Table 1.2.

Table 1.1: Parameter Values

δ_{1A}	δ_{1B}	δ_{2A}	δ_{2B}	δ_{3A}	δ_{3B}	L	M	θ_{1A}	θ_{1B}	θ_{2A}	θ_{2B}	θ_{3A}	θ_{3B}	γ	α
−0.01	−5	−0.02	−5.5	−0.02	−5.5	0.55	1.2	1.7	0.55	2.5	0.5	2.5	0.4	0.75	0.1

Table 1.2: Equilibrium CCPs

	$(\mathbf{z}^1, A)$	$(\mathbf{z}^1, B)$	$(\mathbf{z}^2, A)$	$(\mathbf{z}^2, B)$	$(\mathbf{z}^3, A)$	$(\mathbf{z}^3, B)$	$(\mathbf{z}^4, A)$	$(\mathbf{z}^4, B)$
Player 1	0.715	0.296	0.715	0.202	0.715	0.117	0.714	0.143
Player 2	0.793	0.124	0.793	0.113	0.951	0.396	0.951	0.285
Player 3	0.793	0.110	0.951	0.222	0.793	0.069	0.951	0.133
	$(\mathbf{z}^5, A)$	$(\mathbf{z}^5, B)$	$(\mathbf{z}^6, A)$	$(\mathbf{z}^6, B)$	$(\mathbf{z}^7, A)$	$(\mathbf{z}^7, B)$	$(\mathbf{z}^8, A)$	$(\mathbf{z}^8, B)$
Player 1	0.883	0.549	0.883	0.521	0.883	0.179	0.883	0.470
Player 2	0.793	0.048	0.792	0.050	0.951	0.352	0.951	0.085
Player 3	0.793	0.045	0.951	0.065	0.792	0.057	0.951	0.071

Results For estimation, we focus on estimating parameters for player 1 corresponding to latent state A. The system of "moment" functions consists of sixteen elements:

$$G(\pi) = \begin{bmatrix} \overline{\pi}(1, \mathbf{z}^1, k_1) \\ \overline{\pi}(1, \mathbf{z}^1, k_1') \\ \overline{\pi}(1, \mathbf{z}^2, k_2) \\ \overline{\pi}(1, \mathbf{z}^2, k_2') \\ \vdots \\ \overline{\pi}(1, \mathbf{z}^8, k_8) \\ \overline{\pi}(1, \mathbf{z}^8, k_8') \end{bmatrix} - \begin{bmatrix} L & p_2(\mathbf{z}^1, k_1) + p_3(\mathbf{z}^1, k_1) \\ L & p_2(\mathbf{z}^1, k_1') + p_3(\mathbf{z}^1, k_1') \\ L & p_2(\mathbf{z}^2, k_2) + p_3(\mathbf{z}^2, k_2) \\ L & p_2(\mathbf{z}^2, k_2') + p_3(\mathbf{z}^2, k_2') \\ \vdots & \vdots \\ M & p_2(\mathbf{z}^8, k_8) + p_3(\mathbf{z}^8, k_8) \\ M & p_2(\mathbf{z}^8, k_8') + p_3(\mathbf{z}^8, k_8') \end{bmatrix} \pi.$$

Table 1.3 presents the percentage of selecting the correct matching and mean squared error (MSE) of the estimated the parameter vector using the identity weighting matrix based on 1000 repetitions. Let n_s denote the number of observations per observed state and the MSE reported in the table is calculated as the sum of MSE of each parameter. It can be seen that the percentage of selecting the correct matching is close to 100% and MSE decreases as n_s increases.

Table 1.3: Estimation Result

n_s	625	750	875	1000	1125	1250
Percentage	99.7%	100%	99.9%	100%	100%	100%
MSE	0.043	0.032	0.032	0.024	0.022	0.020

Table 1.4: Rejection Probabilities under H_0

n_s	625	750	875	1000	1125	1250
$R = I_2$	0.041	0.038	0.041	0.043	0.044	0.045
$R = (0, 1)$	0.036	0.043	0.043	0.042	0.041	0.047

Table 1.5: Rejection Probabilities under H_1 for $n_s = 1000$

Deviation	-0.5	-0.4	-0.3	-0.2	0.2	0.3	0.4	0.5
Rejection Probability	1	1	0.978	0.680	0.444	0.816	0.957	0.994

To study the finite sample performance of our inference procedure, we consider two null hypotheses of the form $H_0 : R\pi_0 = r$ for $R = I_2$, $r = (\theta_{1A}, \delta_{1A})^\top$; and $R = (0, 1)$, $r = \delta_{1A}$. The first hypothesis is on the whole payoff parameter vector and the second one is on the parameter of strategic interaction, which is of great interest in empirical research. We set the nominal size as 5% and use the 95% quantile of χ^2_8 and χ^2_7 as the critical values for $R = I_2$ and $R = (0, 1)$. In both cases, the bootstrap weighting matrices are calculated with 4000 bootstrap samples. Across bootstrap draws, we always put the latent state with larger sum of CCPs in the odd rows. The results in Tables 1.4, 1.5, and 1.6 are based on 2000 Monte Carlo repetitions.

Table 1.4 presents the simulation results on the null rejection probabilities. The size is well controlled and is getting closer to 0.05 as the sample size increases for both hypotheses.

Tables 1.5 and 1.6 provide the power results for testing $H_0 : R\pi_0 = r$ with $R = (0, 1)$ and $r = -0.01$. When the sample size is fixed, Table 1.5 shows that as the hypothesized value deviates further from the true value, the probability of rejecting the null hypothesis

Table 1.6: Rejection Probabilities under H_1 for Different Sample Sizes

n_s	625	750	875	1000	1125	1250
Deviation = 0.3	0.527	0.620	0.731	0.816	0.861	0.901
Deviation = −0.3	0.838	0.919	0.959	0.978	0.989	0.996

Table 1.7: Running Time Comparison (in seconds)

Observed States	Particles	SMC	Exhaustive Search
8	80	0.805	0.149
18	250	21.575	5.999
27	400	222.85	3469.1

increases. When the absolute value of the deviation is 0.5, the rejection probabilities are almost 1. The results in Table 1.6 show that for fixed deviation from the null hypothesis, the rejection probability increases with the sample size.

Comparison of SMC and Exhaustive Search To compare the running time of SMC and exhaustive search, we compute the running time for the above DGP and two additional DGPs: one with 18 observed states, and the other with 27 observed states.[18]

From Table 1.7, we see that the exhaustive search is faster than SMC when the number of observed states is small. However, when the number of observed states gets large, SMC is much faster than the exhaustive search. In the case of 27 observed states, SMC is more than 15 times faster than the exhaustive search.[19]

1.7 An Empirical Illustration

This section applies the methods developed in the previous sections to Sweeting (2009)'s dataset on radio stations' strategic timing decision to air commercials between two time slots. In contrast to Sweeting (2009), De Paula and Tang (2012), and Xiao (2018), we allow

[18]We use more particles in SMC when the number of observed states gets larger.

[19]In this experiment, both SMC and exhaustive search find the correct matching for all DGPs.

the presence of non-separable unobserved heterogeneity in the payoff functions. According to Sweeting (2009) and Xiao (2018), advertisers would like different radio stations to play commercials simultaneously so that drivers could hardly avoid them. On the other hand, due to the way listenership is calculated, radio stations might prefer to differentiate with each other on the timing of commercials. Whether radio stations prefer to air commercials at the same time or not, i.e., whether the incentives of radio stations and advertisers are aligned or not, has been of great interest to economists, see Sweeting (2009), De Paula and Tang (2012), and Xiao (2018).

Similar to Xiao (2018), we specify a game with symmetric players and suppress station-specific covariate. We include one market-level covariate: an index for the ranking of market size denoted as x. There are in total 144 different market ranks, and we classify them into six categories with 24 market ranks in each category. The (unnormalized) expected payoff function for the representative player in market rank x and latent state k when taking action 1 is specified as:

$$\tilde{\overline{\pi}}\left(1,x,k\right)=\beta_{1k}\left(x\right)+\delta_{k}\left(x\right)p\left(x,k\right).$$

Similarly the expected payoff function when taking action 0 is specified as:

$$\tilde{\overline{\pi}}\left(0,x,k\right)=\beta_{0k}\left(x\right)+\delta_{k}\left(x\right)\left(1-p\left(x,k\right)\right).$$

Note that in this specification, the strategic effect is assumed to be invariant across choices,[20] while the effect of market state variable is assumed to be choice dependent. The normalized expected payoff is given by

$$\overline{\pi}\left(1,x,k\right)=\beta_{1k}\left(x\right)-\beta_{0k}\left(x\right)+\delta_{k}\left(x\right)\left(2p\left(x,k\right)-1\right).$$

Following Xiao (2018), we allow the strategic effect to vary with the observed state. We parametrize $\beta_{1k}\left(x\right)=\beta_{1k}x$, $\beta_{0k}\left(x\right)=\beta_{0k}x$, and $\delta_{k}\left(x\right)=\delta_{k}x$ to make use of variation in x. Similar to Sweeting (2009) and Xiao (2018), we impose the normalization: $\beta_{0k}=0$. We

[20] Sweeting (2009) and Xiao (2018) also assume that strategic effect is invariant across choices.

deviate from Sweeting (2009) and Xiao (2018) by allowing the effect of market rank and strategic effect to depend on some non-separable latent state k.

Following Xiao (2018), in the first step, we collect all market-day-hour with more than four players to estimate the number of mixtures using the sequential testing procedure introduced by Robin and Smith (2000). We find that during drivetime (hour of the day equals 16pm or 17pm), market rank 5 and rank 6 have two mixtures and all other market ranks have only one mixture in drivetime. During non-drivetime, all market ranks have only one mixture. These results are consistent with the findings in Xiao (2018) that there are two mixtures for smaller markets. From now on we focus our analysis on drivetime as this is where our method is relevant. Instead of interpreting them as two equilibria, we use the aforementioned J-test to test the null hypothesis that there is no unobserved heterogeneity. We find that J-test rejects this null hypothesis at 5% significance level.[21]

We thus proceed to use our methods which allows for the presence of unobserved heterogeneity to estimate the parameters in the payoff functions. Since there are two latent states, we obtain the point estimates for parameters on the first and second latent state by fixing the first and second row corresponding to the market with smallest rank among the markets with two mixtures. For the remaining four observed states with only one mixture (degenerate latent states), we assume that two of them belong to the first latent state and two of them belong to the second latent state so that each latent state will correspond to an overidentified system with four "moments" and two parameters.[22]

The estimated parameters on two latent states and their significance results based on the test developed in the paper are presented in Table 1.8. Several conclusions emerge. First, the estimated coefficients of strategic interactions are all positive. This is consistent with the

[21] We also perform the J-test using seven categories of market ranks and five categories of market ranks. In the seven-category case, the first to sixth categories contain 20 market ranks each and the seventh category contains the remaining market ranks; in the five-category case, the first to fourth categories contain 30 market ranks each and the fifth category contains the remaining market ranks. Both tests reject the null hypothesis at 5% significance level.

[22] Note that here we need at least three equations for each latent state, and we have made use of degenerate latent states as discussed in Remark 1.5.3.

Table 1.8: Estimated Coefficients and Significance Results

	First Latent State	Second Latent State
Market Rank Effect	0.010	0.106^{*}
Strategic Effect	0.390	0.668^{**}

Notes: ** denotes significance at 5% confidence level; and * denotes significance at 10% confidence level.

finding in Sweeting (2009), De Paula and Tang (2012), and Xiao (2018) that radio stations seem to prefer to play commercials at the same time. Second, similar to Xiao (2018) who does not find one timing choice to be significantly more attractive than the other, we find that the coefficient for market rank effect is small. An application of our test to testing significance of this coefficient on two latent states shows that they are mostly insignificant (except that the coefficient on the second latent state is significant at the 10% level), implying that market rank does not make one timing choice more attractive than the other. Third, the estimated strategic effects on two latent states are quite different: the first latent state corresponds to a smaller strategic effect, while the second latent state corresponds to a larger strategic effect. Our tests suggest that on the second latent state, the estimated strategic coefficient is significant at 5% level, while on the first latent state the estimated strategic coefficient is not significant even at 10% level.[23]

1.8 Conclusion

In this paper, we propose a novel method to sequentially estimate payoff functions and to conduct uniform inference in static games of incomplete information with non-separable unobserved heterogeneity and multiple equilibria. For identification, our approach combines and extends the sequential identification framework of Xiao (2018) and Aguirregabiria and

[23] As a robustness check, we conduct estimation and significance tests with the aforementioned five and seven categories of market ranks respectively. We find that point estimate of strategic coefficient is relatively large on the second latent state for both cases. In both cases all coefficients on the first latent states are not significant, and for the second latent state the estimated strategic coefficient is significant at 5% level with five categories and significant at 10% level with seven categories. In all these tests, we use 4000 bootstrap samples to calculate the weighting matrix.

Mira (2019). Estimation of the payoff in the presence of matching-types problem is formulated as a minimum distance problem with incorrect "moments". We estimate the correct matching and payoff simultaneously using a minimum distance criterion with a rewarding term, if needed, over the payoff space and selection vector space. For inference, we construct an asymptotically uniformly valid test for the linear hypothesis on payoff vectors. To make our methods feasible for large state spaces, we introduce a stochastic optimization algorithm based on SMC method. A Monte Carlo study is carried out to investigate finite sample performance of our estimation and inference procedures. As an empirical illustration, we revisit strategic timing decisions for broadcasting commercial among radio stations originally studied in Sweeting (2009).

Chapter 2

A SIMPLE SEMIPARAMETRIC ESTIMATOR FOR STATIC GAMES OF INCOMPLETE INFORMATION WITH ACTION COMPLEMENTARITY

2.1 Introduction

In this paper we study the semiparametric identification and estimation for a class of static games of incomplete information in which each player faces two complementary (substitutable) actions, and could take one, or both, or neither of them. Such models are important for understanding firm behaviors, as the literature have documented both the strategic motive and the complementarity motive as important driving forces for firms' actions, see Athey and Stern (1998), Bajari et al. (2013) and references therein.[1] As summarized in Berry et al. (2014), "*Firms make complementary choices when determining production inputs, entering related markets, and strategic mergers*".

Recent empirical studies of firm behavior start to combine these two motives together in their models. In an incomplete information setting, Augereau et al. (2006) studies the internet service providers' strategic adoption of two technology standards, and Lenzo (2006) studies the strategic adoption of SPECT and/or PET diagnostic imaging technologies among hospitals. All these studies use fully parametric models, and in particular Augereau et al. (2006) and Lenzo (2006) follows and extends the identification and estimation strategy developed by Gentzkow (2007) for single agent parametric discrete choice model for bundles.[2] Our paper makes a direct contribution to this literature by studying identification and esti-

[1] To shorten terminology we use the single word "complementarity" to represent complementarity or substitutability when there is no confusion throughout this paper, like in Lenzo (2006).

[2] Throughout this paper we use the phrase "error term" and "unobservable private information" interchangeably.

mation of such games without a parametric error distribution. To the best of our knowledge, this is the first paper that gives a semiparametric estimator for such games. Building on and extending the identification results of Fox and Lazzati (2017), we show that the structural parameters in this model are identified and have a nice closed-form. Identification makes use of player-action specific state variables that are additively separable and have large supports. Such player-action specific state variables are also required to be private information among players when the game is played, but available to the econometrician afterwards. A simple three-stage least squares estimator for the structural parameters is proposed based on this identification strategy. Although the game has a complicated structure, this estimator has a very simple closed-form, and doesn't require any numerical optimization. We layout sufficient conditions that guarantee its root-n consistency and normality given knowledge of *near-tail intervals*, which could be consistently estimated [3].

Our paper is closely related to several strands of literature. Firstly, our paper is related to the burgeoning literature on identification and estimation of discrete model for bundles (or discrete choice model with complementarity). Gentzkow (2007) proposes a random utility model for bundles in which the utility for a bundle equals sum of two stand-alone utility for each product plus an extra term which quantifies the degree of substitutability or complementarity between products. A lesson from Gentzkow (2007) for models involving discrete choice on bundles is that reliable estimation and inference on the degree of complementarity shall not rely on strong ad-hoc assumption on the joint distribution of error terms. This has motivated a study of the nonparametric identifiability of such models in Fox and Lazzati (2017). In particular, he proves that it suffices to identify model parameters with product-specific covariates that are additively separable and have large supports. Note that such a discrete choice model for bundles could be rewritten as an ordinary multinomial choice model by redefining each combination of products as a different choice and introducing a new error term for the combination satisfying standard assumptions in multinomial choice models.

[3]The definition of such interval is given in Section 3.3

The attractive feature of Fox and Lazzati (2017)'s identification strategy compared to the traditional nonparametric or semiparametric identification argument for multinomial choice model is that unlike the latter, only product-specific covariate is required and we do not need combination-specific covariate like in Lewbel (2000). This greatly facilitates empirical work as combination-specific covariates are not as easily available. Fox and Lazzati (2017) itself doesn't give an estimator, and our paper adds to this literature by providing a closed-form semiparametric estimator for such models (as when strategic interaction coefficient is equal to zero, our game with action complementarity will reduce to a single-agent discrete choice model for bundles).

Our paper is also related to the literature on semiparametric identification and estimation of discrete game of incomplete information under weak distributional assumptions. Tang (2010) studies identification and estimation of an incomplete information game under the median independence assumption between error term and state variables. He establishes consistency for his estimators without establishing rates of convergence. Lewbel and Tang (2015) studies identification and estimation of binary games of incomplete information using special regressors. Their paper is similar to ours in that we both use a special regressor type identification and estimation strategy, but their paper also differs from ours in several aspects. First, Lewbel and Tang (2015)'s identification strategy and the resulting root-n consistent estimator only works for binary games, while games studied in this paper allow players to take all combinations of two binary actions (there are 4 choices for each player), which are allowed to have complementary or substitutable relationships with each other, and we are also interested in quantifying the degree of complementarity or substitutability. Secondly, the root-n normality in Lewbel and Tang (2015) could allow the error term's support to be the whole real line, as long as the tail of the generated special regressor is also the real line and is thick enough (having infinite variance), but in our setting, if the joint distribution of errors has unbounded support, then root-n rate will not be possible even if our special regressor has unbounded support and thick enough tail. This is because in our problem we need to identify and estimate marginal cdf from a joint cdf, and the marginal cdf is only identified when the other variable of the joint distribution is evaluated at infinity. This feature determines how we construct our estimator, which is not a trivial extension of their approach. Thirdly, in the construction of the special regressor, Lewbel and Tang (2015) uses a generated special regressor for identification, and they require player specific covariate being monotone in generated regressor in order to have root-n normality. In contrast, our special regressor doesn't contain any estimated quantity due to the fact that we assume the special regressor is not observed by the other player when the game is played. As a result, monotonicity assumption like theirs is not needed for our root-n normality.

Our paper is also connected to the literature on the use of "special regressor" for semi-parametric identification and estimation of discrete choice model such as Lewbel (2000), Dong and Lewbel (2015) and Lewbel et al. (2012). When there are more than 2 choices, as noted in Lewbel (2000), the naive special regressor estimator (for the coefficients of each choice) fails to be root-n consistent (or even consistent) because it inevitably involves evaluating unknown regression function at a single point (usually a point on the boundary of support) for the special regressors of other choices. Our problem is similar to this problem in that naive estimator directly based on our identification strategy requires evaluating unknown regression function at a single point, resulting in slower rate of convergence. To solve this problem, the estimator we propose takes averages for these special regressors in their *near-tail intervals*. This solution is also applicable to restore root-n rate for the multinomial discrete choice models studied in Lewbel (2000). The price we pay for this is that the error terms must have a bounded support.

Our paper is related to the static entry game literature, like Seim (2006), Sweeting (2009), Zhu and Singh (2009), Bajari et al. (2013), Grieco (2014), Deng and Picone (2018) among others. Most of them focus on entry into a particular industry or business, and do not model simultaneous entry into multiple businesses, nor do they attempt to quantify the degree of substitutability or complementarity between businesses or actions. The only exceptions are Jia (2008) and Ren (2017). Jia (2008)'s study is under the complete information setting. She considers chain store's simultaneous entry into multiple locations taking into account the chain effects from entering neighboring locations, The chain effects is a quantification of the degree of substitutability between entry actions. Our studies differ in several aspects: firstly, she studies a game of complete information while we are studying a game of incomplete information; secondly, she assumes fully parametric payoff function and error distribution while we do not specify the distribution of the error terms; thirdly, in her study the entry outcome in different markets are correlated while in our study the entry outcome in different markets are independent (but entry outcomes for different products in a single market are dependent). Ren (2017) estimates a parametric entry game of complete information for airline oligopolists allowing the payoff of providing nonstop service being affected by the provision of connecting service in the same market, and she quantifies this effect, which is exactly the effect we try to quantify in this paper. Nevertheless, our paper differs from this paper in the following aspects: Firstly, we study a game of incomplete information; secondly, we do not require the error term to have a known distribution and we establish identification results. Our paper contributes to the empirical entry game literature by providing a framework for analyzing simultaneous entry into two related industries (or the markets for two related products) under incomplete information.

The rest of the paper is organized as follows: Section 2.2 defines the baseline model and the equilibrium; Section 2.3 studies identification; Section 2.4 defines our estimator; Section 2.5 establishes asymptotic theory; Section 2.6 presents simulation results; Section 2.7 concludes.

Notations Throughout the paper, a variable name in capital letter denotes a random variable and the corresponding lowercase denotes its realization. The exception is ϵ, whose realization uses the same notation as the random variable.

2.2 Models Under Consideration

2.2.1 Complementary or Substitutable Actions in Static Games of Incomplete Information

According to Vives (2005), the intuitive meaning of complementarity "is the notion, due to Edgeworth, that the marginal value of an action or variable increases in the level of another action or variable". Substitutability likewise corresponds to the notion that marginal value of an action or variable decreases in the level of another action or variable. Formally, in a single agent binary decision problem, the agent will have 4 choices: take action 1 and action 2, take action 1 but not action 2, take action 2 but not action 1, taking neither action. Thus $d \in \mathcal{D} = \{(1,1), (1,0), (0,1), (0,0)\}$[4]. With some slight abuse of notation, let $\Pi^{\dagger}(d)$ denote the utility for this agent. In the study of consumer bundled demand, Gentzkow (2007) defines the following quantity that captures complementarity or substitutability between choices

$$\Pi^{\dagger}((1,1)) - \Pi^{\dagger}((0,1)) - \Pi^{\dagger}((1,0)) - \Pi^{\dagger}((0,0)) = \alpha,$$

the discrete version of cross partial derivatives. He shows that with payoff that are quasilinear in money, this quantity exactly captures the magnitude of complementarity or substitutability in the Hicksian sense. Similarly in the study of firm's combinations of conduct in Athey and Stern (1998), a pair of actions is defined as complementary if $\alpha > 0$.

These definitions naturally extend to a static game of incomplete information. Formally, we consider a very simple static game with two players. Each player faces two complementary (substitutable) actions, and could take one, or both, or neither of them. Let $d_k^i = 1$ if player i chooses action k and 0 otherwise. Let $d^i = (d_1^i, d_2^i)$ denote the choice player i takes out of the combinations of actions, so we have $d^i \in \mathcal{D} = \{(1,1),(1,0),(0,1),(0,0)\}$. Player i's primitive payoff of choosing d^i when the other player is choosing d^{-i} is given by payoff function $\Pi^{\dagger i}(d^i, d^{-i}, \boldsymbol{\epsilon}^i)$, where $\boldsymbol{\epsilon}^i = (\epsilon_1^i, \epsilon_2^i)$ denotes private information of each action for player i. The primitive payoff function of player i for choosing each of these options are

[4] Note that we could rename these choices as $1, 2, 3, 4$, but that will conceal the intuitive meaning for the choices.

specified as follows:

$$\Pi^{\dagger i}((0,0), d^{-i}, \boldsymbol{\epsilon}^i) = 0,$$

$$\Pi^{\dagger i}((1,0), d^{-i}, \boldsymbol{\epsilon}^i) = b_1^i + \delta_1^i d_1^{-i} + \epsilon_1^i,$$

$$\Pi^{\dagger i}((0,1), d^{-i}, \boldsymbol{\epsilon}^i) = b_2^i + \delta_2^i d_2^{-i} + \epsilon_2^i,$$

$$\Pi^{\dagger i}((1,1), d^{-i}, \boldsymbol{\epsilon}^i) = \sum_{k=1}^{2}(b_k^i + \delta_k^i d_k^{-i} + \epsilon_k^i) + \alpha^i,$$

where b_k^i is the payoff to action k that do not depend on strategic interaction. Note that here, again

$$\big(\Pi^{\dagger i}((1,1), d^{-i}, \boldsymbol{\epsilon}^i) - \Pi^{\dagger i}((0,1), d^{-i}, \boldsymbol{\epsilon}^i)\big) - \big(\Pi^{\dagger i}((1,0), d^{-i}, \boldsymbol{\epsilon}^i) - \Pi^{\dagger i}((0,0), d^{-i}, \boldsymbol{\epsilon}^i)\big) = \alpha^i,$$

which is similar to Gentzkow (2007).

Remark 2.2.1. *In the baseline model, player i's payoff from taking action 1 is only affected by action 1 of player $-i$, and player i's payoff from taking action 2 is only affected by action 2 of player $-i$. A real world example for this could be oligopolists' strategic entry into markets that are not related on the demand side. For example, Airbus and Boeing could both enter the passenger airliner market and the fighter market. The fact that Airbus enters the fighter market only makes the fighter market less profitable for Boeing, but not the passenger airliner market, as the demand side of these two markets are completely separated. The baseline model could be easily extended to allow for the situation in which both action 1 and action 2 of player $-i$ affects player i's payoff of action 1 (and action 2):*

$$\Pi^{\dagger i}((0,0), d^{-i}, \boldsymbol{\epsilon}^i) = 0,$$

$$\Pi^{\dagger i}((1,0), d^{-i}, \boldsymbol{\epsilon}^i) = b_1^i + \delta_1^i (d_1^{-i} + d_2^{-i}) + \epsilon_1^i,$$

$$\Pi^{\dagger i}((0,1), d^{-i}, \boldsymbol{\epsilon}^i) = b_2^i + \delta_2^i (d_1^{-i} + d_2^{-i}) + \epsilon_2^i,$$

$$\Pi^{\dagger i}((1,1), d^{-i}, \boldsymbol{\epsilon}^i) = \sum_{k=1}^{2}(b_k^i + \delta_k^i (d_1^{-i} + d_2^{-i}) + \epsilon_k^i) + \alpha^i,$$

A real world example for this could be oligopolists' strategic entry into markets that are related on the demand side. Foe example, whether Delta wants to provide a direct flight between Seattle and Portland will not only affect Alaska's payoff from direct flight between these two cities, but also affect its payoff from connecting flight between these two cities, as direct flight and connecting flight are usually viewed by consumers as substitutes. Although this is a different model, all subsequent analysis on the baseline model trvially extends to this model.

Definition 2.2.1. *We define action 1 and action 2 as complementary if* $\alpha^i > 0$*; as substitutable if* $\alpha^i < 0$*; as independent if* $\alpha^i = 0$*.*

A continuous example in which there are complementary actions and strategic actions could be the example of Bertrand oligopolists strategically choosing both price and advertising efforts in Vives (2005).

2.2.2 The Econometric Model and the Equilibrium

In this section, we define the econometric model for the game we introduced in the previous section. In particular, we will introduce state variables and define the equilibrium of the game. Let $b_k^i = z_k^i + \beta_k^i x$ for $i = 1, 2$ and $k = 1, 2$. Let $\mathbf{z}^i = (z_1^i, z_2^i)$. Note that here z_k^i is a player-action specific state variable, and x is a state variable common to all players and actions.

Assumption 2.2.1 (Information)**.** *(i). The realization of* $(\mathbf{Z}^i, \boldsymbol{\epsilon}^i)$ *is player* i*'s private information when the game is played; (ii). The realization of* $\mathbf{Z}^i$ *is observable to the econometrician (after the game is played).*

Note that Aradillas-Lopez (2010) also allows some state variables to be private information among players when the game is played, although it doesn't provide identification power in his paper.

Assumption 2.2.2 (Independence Relationships)**.** *The following independence conditions hold: (i).* $(\mathbf{Z}^1, \boldsymbol{\epsilon}^1) \perp (\mathbf{Z}^2, \boldsymbol{\epsilon}^2)|X$*; (ii). For* $i = 1, 2$, $\mathbf{Z}^i \perp \boldsymbol{\epsilon}^i|X$*; (iii).* $\mathbb{E}(\epsilon^i|X) = \mathbf{0}$.

Remark 2.2.2. *Under Assumption 2.2.2, $\boldsymbol{\epsilon}^i$ and X cannot be correlated, so that X is exogenous. $\mathbf{Z}^i$ and X are allowed to be correlated. Conditional on X, Z_1^i and Z_2^i are allowed to be correlated, and ϵ_1^i and ϵ_2^i are allowed to be correlated.*

Assumption 2.2.3 (Distribution of $\boldsymbol{\epsilon}^i$ and $\mathbf{Z}^i$)**.** *Conditional on $X = x$, and for $i = 1, 2$, the distributions of $\boldsymbol{\epsilon}^i$, $\boldsymbol{Z}^i$ and $\boldsymbol{\epsilon}^i + \boldsymbol{Z}^i$ are absolutely continuous with respect to the Lebesgue measure and have bounded positive Radon-Nikodym densities almost everywhere on their supports, and the density of $\boldsymbol{\epsilon}^i$ is positive at the boundary of the support. The support of $\boldsymbol{\epsilon}^i$ and $\mathbf{Z}^i$ are both rectangular.*

Assumption 2.2.4 (Data)**.** *The econometrician observes a random sample on players' actions and state variables $\{\mathbf{Z}_m^1, \mathbf{Z}_m^2, X_m, D_m^1, D_m^2\}_{m=1}^M$.*

Without loss of generality, let $[L, U]$ be the support of z_k^i for $i, k = 1, 2$[5]. Also, we focus on the case when X is a discrete scalar. For a generic random variable Q, let Ω_Q denote its support. We focus on pure strategies defined as follows:

Definition 2.2.2 (Pure Strategy)**.** *A pure strategy for player i in this game is a mapping $g^i : \Omega_{\mathbf{Z}^i} \times \Omega_X \times \Omega_{\boldsymbol{\epsilon}^i} \to \mathcal{D}$.*

By definition $D^i = g^i(\mathbf{Z}^i, X, \boldsymbol{\epsilon}^i)$. Based on his available information, player i forms a belief about player $-i$'s choice being a certain value, for example, his belief for $D^{-i} = (1, 1)$ is:

$$Pr\left(g^{-i}\left(\mathbf{Z}^{-i}, X, \boldsymbol{\epsilon}^{-i}\right) = (1,1) \,|\mathbf{Z}^i, X, \boldsymbol{\epsilon}^i\right) = Pr\left(g^{-i}\left(\mathbf{Z}^{-i}, X, \boldsymbol{\epsilon}^{-i}\right) = (1,1) \,|X\right), \tag{2.1}$$

where the equality holds by Assumption 2.2.1 (i) and Assumption 2.2.2 (i). Note that the belief only depends on commonly observed state variable X. Intuitively, under conditional independence, once X is known, $\mathbf{Z}^i$ and $\boldsymbol{\epsilon}^i$ doesn't contain any additional information about

[5]Note that we could allow this support to depend on x, and be different for each i and k, but that only complicates our notation without providing additional insights.

opponent's choice.[6] Given some belief and realized private information as well as state variables, the expected payoff from action 1 is $\pi_1^i(z_1^i, x) = z_1^i + \beta_1^i x + \delta_1^i P_1^{-i}(x)$, and the expected payoff from action 2 is $\pi_2^i(z_2^i, x) = z_2^i + \beta_2^i x + \delta_2^i P_2^{-i}(x)$, where $P_k^{-i}(x)$ represent the belief that player $-i$ also takes action k. Note that there are two (non-overlapping) ways of taking action k: taking both actions or taking only action k. The expected payoff from choosing d^i is $d_1^i(\pi_1^i(z_1^i, x) + \epsilon_1^i) + d_2^i(\pi_2^i(z_2^i, x) + \epsilon_2^i) + \alpha^i d_1^i d_2^i$. Both players choose the combination of actions that maximizes their expected payoff. As an illustration, we plot the optimal choices for player 1 in Figure 2.1 ($\alpha^1 < 0$) and Figure 2.2 ($\alpha^1 > 0$) .[7]

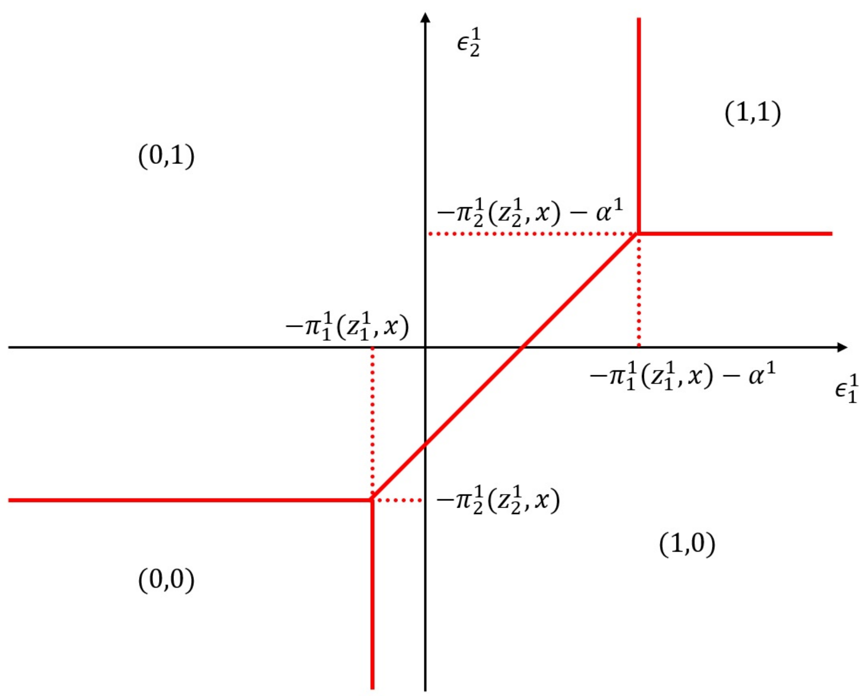

Figure 2.1: Characterization of optimal choices for player 1 ($\alpha^1 < 0$)

[6]When the private information is correlated across players conditional on X, belief will still depend on X, $\mathbf{Z}^i$ and $\boldsymbol{\epsilon}^i$.

[7]Note that Gentzkow (2007) gives a graphical characterization of optimal choice in the utility plane. Our graph is completely different as we draw it in the plane for unobservables, although the shape for the boundaries between choices look similar.

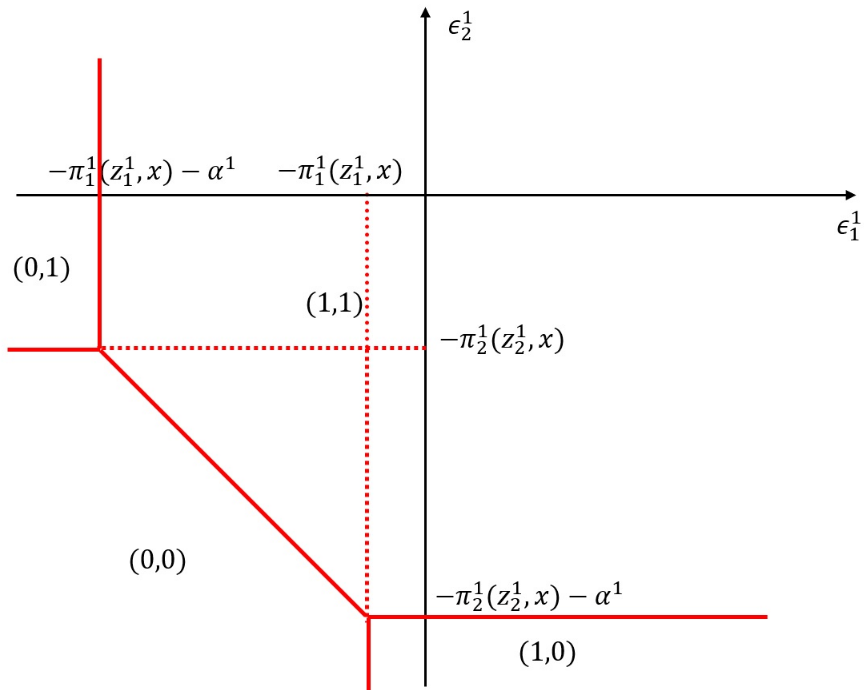

Figure 2.2: Characterization of optimal choices for player 1 ($\alpha^1 > 0$)

In equilibrium, beliefs are consistent with actual choice probabilities. For example, if $\alpha^1 < 0$ and $\alpha^2 < 0$, then the conditional choice probabilities for both players are:[8]

$$P_{00}^1\left(z_1^1, z_2^1, x\right) = Pr\left(\epsilon_1^1 \leq -\pi_1^1\left(z_1^1, x\right), \epsilon_2^1 \leq -\pi_2^1\left(z_2^1, x\right)\right),$$

$$P_{11}^1\left(z_1^1, z_2^1, x\right) = Pr\left(\epsilon_1^1 \geq -\pi_1^1\left(z_1^1, x\right) - \alpha^1, \epsilon_2^1 \geq -\pi_2^1\left(z_2^1, x\right) - \alpha^1\right),$$

$$\begin{aligned} P_{01}^1\left(z_1^1, z_2^1, x\right) &= Pr\left(\epsilon_1^1 \leq -\pi_1^1\left(z_1^1, x\right), \epsilon_2^1 \geq -\pi_2^1\left(z_2^1, x\right)\right) \\ &+ Pr\left(-\pi_1^1\left(z_1^1, x\right) \leq \epsilon_1^1 \leq -\pi_1^1\left(z_1^1, x\right) - \alpha^1, \epsilon_2^1 \geq \epsilon_1^1 + \pi_1^1\left(z_1^1, x\right) - \pi_2^1\left(z_2^1, x\right)\right), \end{aligned}$$

[8]Note that because of Assumption 2.2.1, belief and conditional choice probabilities are different in our game. Belief only depends on x, while conditional choice probabilities depend on all observed state variables.

$$\begin{aligned}P_{10}^{1}\left(z_{1}^{1},z_{2}^{1},x\right) &= Pr\left(\epsilon_{1}^{1}\geq-\pi_{1}^{1}\left(z_{1}^{1},x\right)-\alpha^{1},\epsilon_{2}^{1}\leq-\pi_{2}^{1}\left(z_{2}^{1},x\right)-\alpha^{1}\right)\\ &+Pr\left(-\pi_{1}^{1}\left(z_{1}^{1},x\right)\leq\epsilon_{1}^{1}\leq-\pi_{1}^{1}\left(z_{1}^{1},x\right)-\alpha^{i},\epsilon_{2}^{1}\leq\epsilon_{1}^{1}+\pi_{1}^{1}\left(z_{1}^{1},x\right)-\pi_{2}^{1}\left(z_{2}^{1},x\right)\right),\end{aligned}$$

$$P_{00}^{2}\left(z_{1}^{2},z_{2}^{2},x\right) = Pr\left(\epsilon_{1}^{2}\leq-\pi_{1}^{2}\left(z_{1}^{2},x\right),\epsilon_{2}^{2}\leq-\pi_{2}^{2}\left(z_{2}^{2},x\right)\right),$$

$$P_{11}^{2}\left(z_{1}^{2},z_{2}^{2},x\right) = Pr\left(\epsilon_{1}^{2}\geq-\pi_{1}^{2}\left(z_{1}^{2},x\right)-\alpha^{2},\epsilon_{2}^{2}\geq-\pi_{2}^{2}\left(z_{2}^{2},x\right)-\alpha^{2}\right),$$

$$\begin{aligned}P_{01}^{2}\left(z_{1}^{2},z_{2}^{2},x\right) &= Pr\left(\epsilon_{1}^{2}\leq-\pi_{1}^{2}\left(z_{1}^{2},x\right),\epsilon_{2}^{2}\geq-\pi_{2}^{2}\left(z_{2}^{2},x\right)\right)\\ &+Pr\left(-\pi_{1}^{2}\left(z_{1}^{2},x\right)\leq\epsilon_{1}^{2}\leq-\pi_{1}^{2}\left(z_{1}^{2},x\right)-\alpha^{2},\epsilon_{2}^{2}\geq\epsilon_{1}^{2}+\pi_{1}^{2}\left(z_{1}^{2},x\right)-\pi_{2}^{2}\left(z_{2}^{2},x\right)\right),\end{aligned}$$

$$\begin{aligned}P_{10}^{2}(z_{1}^{2},z_{2}^{2},x) &= Pr(\epsilon_{1}^{2}\geq-\pi_{1}^{2}\left(z_{1}^{2},x\right)-\alpha^{2},\epsilon_{2}^{2}\leq-\pi_{2}^{2}\left(z_{2}^{2},x\right)-\alpha^{2})\\ &+Pr\left(-\pi_{1}^{2}\left(z_{1}^{2},x\right)\leq\epsilon_{1}^{2}\leq-\pi_{1}^{2}\left(z_{1}^{2},x\right)-\alpha^{2},\epsilon_{2}^{2}\leq\epsilon_{1}^{2}+\pi_{1}^{2}\left(z_{1}^{2},x\right)-\pi_{2}^{2}\left(z_{2}^{2},x\right)\right),\end{aligned}$$

where $\pi_{1}^{i}\left(z_{1}^{i},x\right)=z_{1}^{i}+\beta_{1}^{i}x+\delta_{1}^{i}\left(P_{11}^{-i}\left(x\right)+P_{10}^{-i}\left(x\right)\right)$ and $\pi_{2}^{i}\left(z_{2}^{i},x\right)=z_{2}^{i}+\beta_{2}^{i}x+\delta_{2}^{i}\left(P_{11}^{-i}\left(x\right)+P_{01}^{-i}\left(x\right)\right)$.

The equilibrium of the game is defined as the fixed point

$$[P_{11}^{1*}(x),P_{01}^{1*}(x),P_{10}^{1*}(x),P_{11}^{2*}(x),P_{01}^{2*}(x),P_{10}^{2*}(x)]$$

of the following system:[9]

$$P_{11}^{1}\left(x\right) = \mathbb{E}[Pr\left(\epsilon_{1}^{1}\geq-\pi_{1}^{1}\left(Z_{1}^{1},x\right)-\alpha^{1},\epsilon_{2}^{1}\geq-\pi_{2}^{1}\left(Z_{2}^{1},x\right)-\alpha^{1}\right)|x],$$

$$\begin{aligned}P_{01}^{1}\left(x\right) &= \mathbb{E}\left[Pr\left(\epsilon_{1}^{1}\leq-\pi_{1}^{1}\left(Z_{1}^{1},x\right),\epsilon_{2}^{1}\geq-\pi_{2}^{1}\left(Z_{2}^{1},x\right)\right)|x\right]\\ &+\mathbb{E}\left[Pr\left(-\pi_{1}^{1}\left(Z_{1}^{1},x\right)\leq\epsilon_{1}^{1}\leq-\pi_{1}^{1}\left(Z_{1}^{1},x\right)-\alpha^{1},\epsilon_{2}^{1}\geq\epsilon_{1}^{1}+\pi_{1}^{1}\left(Z_{1}^{1},x\right)-\pi_{2}^{1}\left(Z_{2}^{1},x\right)\right)|x\right],\end{aligned}$$

[9] Note that there are 3 free probabilities for each player, and they uniquely characterize the pure strategy for each player.

$$\begin{aligned} P_{10}^{1}(x) &= \mathbb{E}\left[Pr\left(\epsilon_1^1 \geq -\pi_1^1\left(Z_1^1,x\right)-\alpha^1, \epsilon_2^1 \leq -\pi_2^1\left(Z_2^1,x\right)-\alpha^1\right)|x\right] \\ &+\mathbb{E}\left[Pr\left(-\pi_1^1\left(Z_1^1,x\right) \leq \epsilon_1^1 \leq -\pi_1^1\left(Z_1^1,x\right)-\alpha^i, \epsilon_2^1 \leq \epsilon_1^1+\pi_1^1\left(Z_1^1,x\right)-\pi_2^1\left(Z_2^1,x\right)\right)|x\right], \end{aligned}$$

$$P_{11}^{2}(x) = \mathbb{E}\left[Pr\left(\epsilon_1^2 \geq -\pi_1^2\left(Z_1^2,x\right)-\alpha^2, \epsilon_2^2 \geq -\pi_2^2\left(Z_2^2,x\right)-\alpha^2\right)|x\right],$$

$$\begin{aligned} P_{01}^{2}(x) &= \mathbb{E}\left[Pr\left(\epsilon_1^2 \leq -\pi_1^2\left(Z_1^2,x\right), \epsilon_2^2 \geq -\pi_2^2\left(Z_2^2,x\right)\right)|x\right] \\ &+\mathbb{E}\left[Pr\left(-\pi_1^2\left(Z_1^2,x\right) \leq \epsilon_1^2 \leq -\pi_1^2\left(Z_1^2,x\right)-\alpha^2, \epsilon_2^2 \geq \epsilon_1^2+\pi_1^2\left(Z_1^2,x\right)-\pi_2^2\left(Z_2^2,x\right)\right)|x\right], \end{aligned}$$

$$\begin{aligned} P_{10}^{2}(x) &= \mathbb{E}\left[Pr(\epsilon_1^2 \geq -\pi_1^2\left(Z_1^2,x\right)-\alpha^2, \epsilon_2^2 \leq -\pi_2^2\left(Z_2^2,x\right)-\alpha^2)|x\right] \\ &+\mathbb{E}\left[Pr\left(-\pi_1^2\left(Z_1^2,x\right) \leq \epsilon_1^2 \leq -\pi_1^2\left(Z_1^2,x\right)-\alpha^2, \epsilon_2^2 \leq \epsilon_1^2+\pi_1^2\left(Z_1^2,x\right)-\pi_2^2\left(Z_2^2,x\right)\right)|x\right], \end{aligned}$$

where $\pi_1^i\left(Z_1^i,x\right) = Z_1^i + \beta_1^i x + \delta_1^i\left(P_{11}^{-i}(x) + P_{10}^{-i}(x)\right)$ and
$\pi_2^i\left(Z_2^i,x\right) = Z_2^i + \beta_2^i x + \delta_2^i\left(P_{11}^{-i}(x) + P_{01}^{-i}(x)\right)$.

Similarly, if $\alpha^1 > 0$ and $\alpha^2 > 0$, the conditional choice probabilities are

$$\begin{aligned} P_{11}^{1}\left(z_1^1,z_2^1,x\right) &= Pr\left(-\pi_1^1\left(z_1^1,x\right)-\alpha^1 \leq \epsilon_1^1 \leq -\pi_1^1\left(z_1^1,x\right), \epsilon_2^1 \geq -\epsilon_1^1-\pi_1^1\left(z_1^1,x\right)-\pi_2^1\left(z_2^1,x\right)-\alpha^1\right) \\ &+Pr\left(\epsilon_1^1 \geq -\pi_1^1\left(z_1^1,x\right), \epsilon_2^1 \geq -\pi_2^1\left(z_2^1,x\right)-\alpha^1\right), \end{aligned}$$

$$P_{01}^{1}\left(z_1^1,z_2^1,x\right) = Pr\left(\epsilon_1^1 \leq -\pi_1^1\left(z_1^1,x\right)-\alpha^1, \epsilon_2^1 \geq -\pi_2^1\left(z_2^1,x\right)\right),$$

$$P_{10}^{1}\left(z_1^1,z_2^1,x\right) = Pr\left(\epsilon_1^1 \geq -\pi_1^1\left(z_1^1,x\right), \epsilon_2^1 \leq -\pi_2^1\left(z_2^1,x\right)-\alpha^1\right),$$

$$\begin{aligned} P_{00}^{1}\left(z_1^1,z_2^1,x\right) &= Pr\left(-\pi_1^1\left(z_1^1,x\right)-\alpha^1 \leq \epsilon_1^1 \leq -\pi_1^1\left(z_1^1,x\right), \epsilon_2^1 \leq -\epsilon_1^1-\pi_1^1\left(z_1^1,x\right)-\pi_2^1\left(z_2^1,x\right)-\alpha^1\right) \\ &+Pr\left(\epsilon_1^1 \leq -\pi_1^1\left(z_1^1,x\right)-\alpha^1, \epsilon_2^1 \leq -\pi_2^1\left(z_2^1,x\right)\right), \end{aligned}$$

$$
\begin{aligned}
P_{11}^{2}\left(z_1^2, z_2^2, x\right) &= Pr\left(-\pi_1^2\left(z_1^2, x\right) - \alpha^2 \leq \epsilon_1^2 \leq -\pi_1^2\left(z_1^2, x\right), \epsilon_2^2 \geq -\epsilon_1^2 - \pi_1^2\left(z_1^2, x\right) - \pi_2^2\left(z_2^2, x\right) - \alpha^2\right) \\
&\quad + Pr\left(\epsilon_1^2 \geq -\pi_1^2\left(z_1^2, x\right), \epsilon_2^2 \geq -\pi_2^2\left(z_2^2, x\right) - \alpha^2\right),
\end{aligned}
$$

$$
P_{01}^{2}\left(z_1^2, z_2^2, x\right) = Pr\left(\epsilon_1^2 \leq -\pi_1^2\left(z_1^2, x\right) - \alpha^2, \epsilon_2^2 \geq -\pi_2^2\left(z_2^2, x\right)\right),
$$

$$
P_{10}^{2}\left(z_1^2, z_2^2, x\right) = Pr\left(\epsilon_1^2 \geq -\pi_1^2\left(z_1^2, x\right), \epsilon_2^2 \leq -\pi_2^2\left(z_2^2, x\right) - \alpha^2\right),
$$

$$
\begin{aligned}
P_{00}^{2}\left(z_1^2, z_2^2, x\right) &= Pr\left(-\pi_1^2\left(z_1^2, x\right) - \alpha^2 \leq \epsilon_1^2 \leq -\pi_1^2\left(z_1^2, x\right), \epsilon_2^2 \leq -\epsilon_1^2 - \pi_1^2\left(z_1^2, x\right) - \pi_2^2\left(z_2^2, x\right) - \alpha^2\right) \\
&\quad + Pr\left(\epsilon_1^2 \leq -\pi_1^2\left(z_1^2, x\right) - \alpha^2, \epsilon_2^2 \leq -\pi_2^2\left(z_2^2, x\right)\right),
\end{aligned}
$$

where $\pi_1^i\left(z_1^i, x\right) = z_1^i + \beta_1^i x + \delta_1^i\left(P_{11}^{-i}(x) + P_{10}^{-i}(x)\right)$ and $\pi_2^i\left(z_2^i, x\right) = z_2^i + \beta_2^i x + \delta_2^i\left(P_{11}^{-i}(x) + P_{01}^{-i}(x)\right)$.

The equilibrium of the game is defined as the fixed point

$$
[P_{11}^{1*}(x), P_{01}^{1*}(x), P_{10}^{1*}(x), P_{11}^{2*}(x), P_{01}^{2*}(x), P_{10}^{2*}(x)]
$$

of the following system:

$$
\begin{aligned}
P_{11}^{1}(x) &= \mathbb{E}\left[Pr\left(-\pi_1^1\left(Z_1^1, x\right) - \alpha^1 \leq \epsilon_1^1 \leq -\pi_1^1\left(Z_1^1, x\right), \epsilon_2^1 \geq -\epsilon_1^1 - \pi_1^1\left(Z_1^1, x\right) - \pi_2^1\left(Z_2^1, x\right) - \alpha^1\right)|x\right] \\
&\quad + \mathbb{E}\left[Pr\left(\epsilon_1^1 \geq -\pi_1^1\left(Z_1^1, x\right), \epsilon_2^1 \geq -\pi_2^1\left(Z_2^1, x\right) - \alpha^1\right)|x\right],
\end{aligned}
$$

$$
P_{01}^{1}(x) = \mathbb{E}\left[Pr\left(\epsilon_1^1 \leq -\pi_1^1\left(Z_1^1, x\right) - \alpha^1, \epsilon_2^1 \geq -\pi_2^1\left(Z_2^1, x\right)\right)|x\right],
$$

$$
P_{10}^{1}(x) = \mathbb{E}\left[Pr\left(\epsilon_1^1 \geq -\pi_1^1\left(Z_1^1, x\right), \epsilon_2^1 \leq -\pi_2^1\left(Z_2^1, x\right) - \alpha^1\right)|x\right],
$$

$$
\begin{aligned}
P_{11}^{2}(x) &= \mathbb{E}\left[Pr\left(-\pi_1^2\left(Z_1^2, x\right) - \alpha^2 \leq \epsilon_1^2 \leq -\pi_1^2\left(Z_1^2, x\right), \epsilon_2^2 \geq -\epsilon_1^2 - \pi_1^2\left(Z_1^2, x\right) - \pi_2^2\left(Z_2^2, x\right) - \alpha^2\right)|x\right] \\
&\quad + \mathbb{E}\left[Pr\left(\epsilon_1^2 \geq -\pi_1^2\left(Z_1^2, x\right), \epsilon_2^2 \geq -\pi_2^2\left(Z_2^2, x\right) - \alpha^2\right)|x\right],
\end{aligned}
$$

$$
P_{01}^{2}(x) = \mathbb{E}\left[Pr\left(\epsilon_1^2 \leq -\pi_1^2\left(Z_1^2, x\right) - \alpha^2, \epsilon_2^2 \geq -\pi_2^2\left(Z_2^2, x\right)\right)|x\right],
$$

$$P_{10}^2(x) = \mathbb{E}\left[Pr\left(\epsilon_1^2 \geq -\pi_1^2\left(Z_1^2, x\right), \epsilon_2^2 \leq -\pi_2^2\left(Z_2^2, x\right) - \alpha^2\right)|x\right],$$

where $\pi_1^i(z_1^i, x) = z_1^i + \beta_1^i x + \delta_1^i\left(P_{11}^{-i}(x) + P_{10}^{-i}(x)\right)$, and $\pi_2^i(z_2^i, x) = z_2^i + \beta_2^i x + \delta_2^i\left(P_{11}^{-i}(x) + P_{01}^{-i}(x)\right)$. For the cases of $\alpha^1 > 0$, $\alpha^2 < 0$ or $\alpha^1 < 0$, $\alpha^2 > 0$, the exact forms for the equilibrium conditions are different, but an equilibrium always exist by Brouwer's fixed point theorem under Assumption 2.2.3.

Example 2.2.1 (Equilibrium Beliefs)**.** *As an illustration, consider a game with* $x \in \{-1, -0.5, 0.5, 1\}$, $\epsilon_k^i \overset{i.i.d}{\sim} \mathrm{U}[-4, 4]$, *and* $Z_k^i \overset{i.i.d.}{\sim} \mathrm{U}[-10, 10]$. *Parameter values for the game are given below:*

Table 2.1: Parameter Values

β_1^1	β_2^1	δ_1^1	δ_2^1	α^1	β_1^2	β_2^2	δ_1^2	δ_2^2	α^2
1.5	1	−1	−0.5	−1	1	1.5	−1	−1.5	−2

Equilibrium beliefs for different x *are calculated and presented in the following table:*

Table 2.2: Calculated Equilibrium Beliefs

	$P_{11}^1(x)$	$P_{01}^1(x)$	$P_{10}^1(x)$	$P_{11}^2(x)$	$P_{01}^2(x)$	$P_{10}^2(x)$
$x = -1$	0.1389	0.2832	0.2455	0.0970	0.2582	0.2994
$x = -0.5$	0.1627	0.2818	0.2570	0.1165	0.2722	0.2996
$x = 0.5$	0.2155	0.2737	0.2748	0.1605	0.2952	0.2949
$x = 1$	0.2445	0.2670	0.2811	0.1850	0.3043	0.2900

2.3 Sequential Identification

In this section we study the identification of the game. The parameters of interest are α^i and $[\beta_k^i, \delta_k^i]^\top$ for $i = 1, 2$ and $k = 1, 2$. We use a sequential identification strategy that is widely used in the econometrics of games literature, see Bajari et al. (2013), Aguirregabiria and Mira (2007) and references therein. Following Bajari et al. (2013), Aguirregabiria and Mira (2007) we assume there is a unique equilibrium in the data.

Assumption 2.3.1 (Unique Equilibrium)**.** *For each given $X = x$, data is rationalized by a single equilibrium.*

Under Assumption 2.2.4 and 2.3.1, the beliefs are identified. From now on we treat the beliefs as known.

2.3.1 Step1: Identification of the Sign of the Complementarity Parameter

Lemma 2.3.1. *Under Assumption 2.2.1-2.3.1, the sign of α^i is identified for $i = 1, 2$.*

Proof. See Appendix B.2. □

2.3.2 Step2: Identification of Structural Parameters

Given the identified sign of the complementarity parameter, in this section we establish the identification of structural parameters. For $i = 1, 2$, define $S_1^i = \epsilon_1^i + \beta_1^i X + \delta_1^i P_1^{-i}(X)$, $S_2^i = \epsilon_2^i + \beta_2^i X + \delta_2^i P_2^{-i}(X)$, $\widetilde{S}_1^i = -\epsilon_1^i - \beta_1^i X - \delta_1^i P_1^{-i}(X) - \alpha^i$, $\widetilde{S}_2^i = -\epsilon_2^i - \beta_2^i X - \delta_2^i P_2^{-i}(X) - \alpha^i$. Define $S_1'^i = \epsilon_1^i + \beta_1^i X + \delta_1^i P_1^{-i}(X) + \alpha^i$, $S_2'^i = -\epsilon_2^i - \beta_2^i X - \delta_2^i P_2^{-i}(X)$, $\widetilde{S}_1'^i = -\epsilon_1^i - \beta_1^i X - \delta_1^i P_1^{-i}(X)$, and $\widetilde{S}_2'^i = \epsilon_2^i + \beta_2^i X + \delta_2^i P_2^{-i}(X) + \alpha^i$. Define $v_k^i(X) = [X, P_k^{-i}(X)]^\top$.

Assumption 2.3.2 (Support)**.** *Given each $X = x$, the conditional support of $\mathbf{S}^i$ is a subset of the conditional support of $(-z_1^i, -z_2^i)$, the conditional support of $\widetilde{\mathbf{S}}^i$ is a subset of the conditional support of (z_1^i, z_2^i), the conditional support of $\mathbf{S}'^i$ is a subset of the conditional support of $(-z_1^i, z_2^i)$, the conditional support of $\widetilde{\mathbf{S}}'^i$ is a subset of the conditional support of $(z_1^i, -z_2^i)$.*

Example 2.3.1 (Illustration of Support Condition)**.** *The support condition is easily satisfied if the support of z_k^i is relatively large while the support of ϵ_k^i is relatively small given any x. To see this, consider the game we studied in Example 2.2.1. The support condition is satisfied in this game. The following table illustrates the support condition for $x = 1$ (illustration for other x are given in Appendix B.1):*

Table 2.3: Illustration of Support Condition ($x = 1$)

Support of S_1^1 and S_2^1:	[-2.9751,5.0249]	[-3.2446,4.7554]
Support of S_1^2 and S_2^2:	[-3.5256,4.4744]	[-3.2672,4.7328]
Support of $-z_1^i$ and $-z_2^i$:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^1$ and $\widetilde{S}_2^1$:	[-4.0249,3.9751]	[-3.7554,4.2446]
Support of $\widetilde{S}_1^2$ and $\widetilde{S}_2^2$:	[-2.4744,5.5256]	[-2.7328,5.2672]
Support of z_1^i and z_2^i:	[-10,10]	[-10,10]
Support of $S_1^{'1}$ and $S_2^{'1}$:	[-3.9751,4.0249]	[-4.7554,3.2446]
Support of $S_1^{'2}$ and $S_2^{'2}$:	[-5.5256,2.4744]	[-4.7328,3.2672]
Support of $-z_1^i$ and z_2^i:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^{'1}$ and $\widetilde{S}_2^{'1}$:	[-5.0249,2.9751]	[-4.2446,3.7554]
Support of $\widetilde{S}_1^{'2}$ and $\widetilde{S}_2^{'2}$:	[-4.4744,3.5256]	[-5.2672,2.7328]
Support of z_1^i and $-z_2^i$:	[-10,10]	[-10,10]

Theorem 2.3.2. *Suppose Assumptions 2.2.1-2.3.2 hold, and suppose $\mathbb{E}[v_k^i(X)\, v_k^i(X)^\top]$ has full rank for $i = 1, 2$ and $k = 1, 2$. Given the identified sign of α^i, $[\beta_k^i, \delta_k^i]^\top$ and the magnitude of α^i are identified as follows:*[10]

(i). If $\alpha^i \leq 0$,

$$\begin{bmatrix} \beta_1^i \\ \delta_1^i \end{bmatrix} = -\left(\mathbb{E}\left[v_1^i(X)\, v_1^i(X)^\top\right]\right)^{-1} \mathbb{E}\left[v_1^i(X) \int_L^U z_1^i \int_L^U \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right],$$

$$\begin{bmatrix} \beta_2^i \\ \delta_2^i \end{bmatrix} = -\left(\mathbb{E}\left[v_2^i(X)\, v_2^i(X)^\top\right]\right)^{-1} \mathbb{E}\left[v_2^i(X) \int_L^U z_2^i \int_L^U \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_1^i dz_2^i\right],$$

$$\alpha^i = -\mathbb{E}\left[\int_L^U z_1^i \int_L^U \frac{\partial^2 P_{11}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right] + \mathbb{E}\left[\int_L^U z_1^i \int_L^U \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right].$$

(ii). If $\alpha^i \geq 0$,

[10] Note that I could replace the integration in the following formulas with expectation, but since we are not going to use this estimator, we will leave the original form here.

$$\begin{bmatrix} \beta_1^i \\ \delta_1^i \end{bmatrix} = \left(\mathbb{E}\left[v_1^i(X)\, v_1^i(X)^\top\right]\right)^{-1} \mathbb{E}\left[v_1^i(X) \int_L^U z_1^i \int_L^U \frac{\partial^2 P_{10}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right],$$

$$\begin{bmatrix} \beta_2^i \\ \delta_2^i \end{bmatrix} = \left(\mathbb{E}\left[v_2^i(X)\, v_2^i(X)^\top\right]\right)^{-1} \mathbb{E}\left[v_2^i(X) \int_L^U z_2^i \int_L^U \frac{\partial^2 P_{01}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_1^i dz_2^i\right],$$

$$\alpha^i = \mathbb{E}\left[\int_L^U z_1^i \int_L^U \frac{\partial^2 P_{01}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right] - \mathbb{E}\left[\int_L^U z_1^i \int_L^U \frac{\partial^2 P_{10}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right].$$

Proof. See Appendix B.3. □

Remark 2.3.1 (More than two actions)**.** *Notice that just like in Fox and Lazzati (2017), this identification strategy doesn't trivially extend to the case when we have more than two actions. More restrictions are needed. One important restriction is the existence of a vector of actions that is known by the econometrician such that if the vector is a "local maximizer", then it is also a "global maximizer".*[11]

2.3.3 Simple Alternative Forms for the Identified Structural Parameters

An estimator directly built on Theorem 2.3.2 (or its equivalent form with integration replaced with expectation) requires estimation of cross derivatives of conditional choice probability, which is cumbersome to implement. Thus, based on the insights from Lewbel (2000) and Lewbel et al. (2012), we propose simpler alternative forms for the identified parameters (under slightly different rank conditions). The resulting estimators for the structural parameters that circumvent the need to estimate cross derivatives. These estimators make use of knowledge of the supports of the latent component of utilities, which are identified as stated in the following lemma.

[11] See Appendix D of Fox and Lazzati (2017) for details.

Lemma 2.3.2. *Under Assumption 2.2.1-2.3.2, for $i = 1, 2$ and $k = 1, 2$, the support of S_k^i, $\widetilde{S}_k^i$, $S_k'^i$ and $\widetilde{S}_k'^i$ are identified.*

Proof. See Appendix B.4. □

Given the result in Lemma 2.3.2, we also identify regions in the support of special regressors that go beyond the support of these latent part of utilities. We call these intervals *near-tail intervals* from now on, as they are intervals close to the tail of distribution of Z_1^i or Z_2^i. Specifically let $\mathcal{I}_2^i(L)$ be the identified interval near L in the support of Z_2^i, let $\mathcal{I}_1^i(L)$ be an identified interval near L in the support of Z_1^i, let $\mathcal{I}_2^i(U)$ be an identified interval near U in the support of Z_2^i, let $\widetilde{\mathcal{I}}_2'^i(L)$ be an identified interval near L in the support of Z_2^i, let $\mathcal{I}_1'^i(L)$ be an identified interval near L in the support of Z_2^i, and let $\mathcal{I}_2'^i(U)$ be an identified interval near U in the support of Z_2^i. The exact forms of $\mathcal{I}_2^i(L)$, $\mathcal{I}_1^i(L)$, $\widetilde{\mathcal{I}}_2^i(U)$, $\widetilde{\mathcal{I}}_2'^i(L)$, $\mathcal{I}_1'^i(L)$, and $\mathcal{I}_2'^i(U)$ are given as follows: $\mathcal{I}_2^i(L) = [L, -U_{s2}^i]$, $\mathcal{I}_1^i(L) = [L, -U_{s1}^i]$, $\widetilde{\mathcal{I}}_2^i(U) = [\widetilde{U}_{s2}^i, U]$, $\mathcal{I}_1'^i(L) = [L, -U_{s1}'^i]$, and $\widetilde{\mathcal{I}}_2'^i(L) = [L, -\widetilde{U}_{s2}'^i]$.

Let $W_{2L}^i = \mathbb{1}(Z_2^i \in \mathcal{I}_2^i(L))$, let $W_{1L}^i = \mathbb{1}(Z_1^i \in \mathcal{I}_1^i(L))$, let $\widetilde{W}_{2L}^i = \mathbb{1}\left(Z_2^i \in \widetilde{\mathcal{I}}_2'^i(L)\right)$, and let $W_{1L}'^i = \mathbb{1}\left(Z_1^i \in \mathcal{I}_1'^i(L)\right)$.[12] Given the identified sign of complementarity parameter and the identified supports, the identification of structural parameters are given as follows:

Theorem 2.3.3. *Suppose Assumptions 2.2.1-2.3.2 hold, and suppose matrices $\mathbb{E}\left[v_1^i(X)\, v_1^i(X)^\top W_{2L}^i\right]$, $\mathbb{E}\left[v_2^i(X)\, v_2^i(X)^\top W_{1L}^i\right]$, $\mathbb{E}\left[v_1^i(X)\, v_1^i(X)^\top \widetilde{W}_{2L}^i\right]$, $\mathbb{E}\left[v_2^i(X)\, v_2^i(X)^\top W_{1L}'^i\right]$ are of full rank. Without loss of generality, for some constant c known to be in the support of $-S_k^i$, $\widetilde{S}_k^i$, $-S_1'^i$, $S_2'^i$ and $\widetilde{S}_1'^i$ and $-\widetilde{S}_2'^i$, it holds that*[13]

(i). If $\alpha^i \leq 0$, $[\beta_1^i, \delta_1^i]^\top$ is identified as

$$\left(\mathbb{E}\left[v_1^i(X)\, v_1^i(X)^\top W_{2L}^i\right]\right)^{-1} \mathbb{E}\left[v_1^i(X)\left(-\frac{D_{00}^i + \mathbb{1}(Z_1^i \geq c) - 1}{f(Z_1^i | Z_2^i, X)} - c\right) W_{2L}^i\right],$$

[12] In general, these *near-tail intervals* also changes with x, but we suppress x in the notations for brevity.

[13] Note that this constant could be different across S_k^i, $\widetilde{S}_k^i$, $S_k'^i$ and $\widetilde{S}_k'^i$ for $i = 1, 2$ and $k = 1, 2$. We do not give additional superscript or subscript to c as it doesn't change the identifying equations but complicates notation.

$[\beta_2^i, \delta_2^i]^\top$ *is identified as*

$$\left(\mathbb{E}\left[v_2^i\left(X\right)v_2^i\left(X\right)^\top W_{1L}^i\right]\right)^{-1}\mathbb{E}\left[v_2^i(X)\left(-\frac{D_{00}^i+\mathbb{1}\left(Z_2^i\geq c\right)-1}{f\left(Z_2^i|Z_1^i,X\right)}-c\right)W_{1L}^i\right],$$

α^i *is identified as*

$$\mathbb{E}\left[\frac{D_{11}^i-\mathbb{1}\left(Z_1^i\geq c\right)}{f\left(Z_1^i|Z_2^i,X\right)}\Big|Z_2^i\in\widetilde{\mathcal{I}}_2^i\left(U\right)\right]+\mathbb{E}\left[\frac{D_{00}^i+\mathbb{1}\left(Z_1^i\geq c\right)-1}{f(Z_1^i|Z_2^i,X)}\Big|Z_2^i\in\mathcal{I}_2^i\left(L\right)\right],$$

(ii). If $\alpha^i\geq 0$, $[\beta_1^i, \delta_1^i]^\top$ is identified as

$$\left(\mathbb{E}\left[v_1^i\left(X\right)v_1^i\left(X\right)^\top \widetilde{W}_{2L}^i\right]\right)^{-1}\mathbb{E}\left[v_1^i\left(X\right)\left(\frac{D_{10}^i-\mathbb{1}\left(Z_1^i\geq c\right)}{f\left(Z_1^i|Z_2^i,X\right)}-c\right)\widetilde{W}_{2L}^i\right],$$

$[\beta_2^i, \delta_2^i]^\top$ *is identified as*

$$\left(\mathbb{E}\left[v_2^i\left(X\right)v_2^i\left(X\right)^\top W_{1L}'^i\right]\right)^{-1}\mathbb{E}\left[v_2^i\left(X\right)\left(-\frac{D_{01}^i-\mathbb{1}\left(Z_2^i\geq c\right)}{f\left(Z_2^i|Z_1^i,X\right)}+c\right)W_{1L}'^i\right],$$

α^i *is identified as*

$$-\mathbb{E}\left[\frac{D_{01}^i-\mathbb{1}\left(Z_1^i\leq c\right)}{f\left(Z_1^i|Z_2^i,X\right)}\Big|Z_2^i\in\mathcal{I}_2'^i\left(U\right)\right]-\mathbb{E}\left[\frac{D_{10}^i-\mathbb{1}\left(Z_1^i\geq c\right)}{f\left(Z_1^i|Z_2^i,X\right)}\Big|Z_2^i\in\widetilde{\mathcal{I}}_2'^i\left(L\right)\right],$$

Remark 2.3.2. *The forms of these identifying expressions could be further simplified if $Z_1^i\perp Z_2^i|X$ and/or $\mathbf{Z}^i\perp X$. For example, suppose $\alpha^i\leq 0$, and both $Z_1^i\perp Z_2^i|X$ and $\mathbf{Z}^i\perp X$ hold. Then the identifying formula for $[\beta_1^i,\delta_1^i]^\top$ can be simplified to:*

$$\left(\mathbb{E}\left[v_1^i\left(X\right)v_1^i\left(X\right)^\top\right]\right)^{-1}\mathbb{E}\left[v_1^i(X)\left(-\frac{D_{00}^i+\mathbb{1}\left(Z_1^i\geq c\right)-1}{f\left(Z_1^i\right)}-c\right)\Big|Z_2^i\in\mathcal{I}_2^i\left(L\right)\right].$$

2.4 Estimation

Since our identification is constructive, a semiparametric estimator is directly available.

2.4.1 Preparation: Estimation of Near-Tail Intervals

The *near-tail intervals* could be consistently estimated in a number of ways. Here we introduce a very simple estimator that exploits the rectangular supports of both unobservables and exclusive observable variables. As an illustration, consider the estimation of $\mathcal{I}_2^i(L) = [L, -U_{s2}^i]$, which could be obtained by recovering (U_{s1}^i, U_{s2}^i). Based on the observation that for any given x, it holds that $P_{00}^i(z_1^i, z_2^i, x) = 1$ if $-z_1^i \geq U_{s1}^i$ and $-z_2^i \geq U_{s2}^i$ (or equivalently, $L \leq z_1^i \leq -U_{s1}^i$ and $L \leq z_2^i \leq -U_{s2}^i$), which implies that D_{00}^i always equal to 1 as long as $(z_1^i, z_2^i) \in \mathcal{I}_1^i(L) \times \mathcal{I}_2^i(L)$. Given the rectangular shape of supports, the region $\mathcal{I}_1^i(L) \times \mathcal{I}_2^i(L)$ is uniquely determined by (U_{s1}^i, U_{s2}^i). Thus an estimator for $U_s^i \equiv (U_{s1}^i, U_{s2}^i)$ could be constructed as the solution to the constrained maximization problem:

$$-\widehat{U}_s^i \equiv (-\widehat{U}_{s1}^i, -\widehat{U}_{s2}^i) = \underset{(z_1^i, z_2^i) \in \{(Z_{1m}^i, Z_{2m}^i)\}_{m=1}^M}{\operatorname{argmax}} (z_1^i - L)(z_2^i - L),$$

subject to:

all observations with $Z_{1m}^i \leq z_1^i$ and $Z_{2m}^i \leq z_2^i$ have their corresponding $D_{00m}^i = 1$.

The following lemma establishes consistency of this estimator.

Lemma 2.4.1. *Under Assumption 2.2.1-2.3.2, it holds that* $\widehat{U}_s^i \xrightarrow{p} U_s^i$.

Proof. See Appendix B.6. □

Consistency of this estimator implies consistency of $[L, -\widehat{U}_{s2}^i]$ and $[L, -\widehat{U}_{s1}^i]$ for $\mathcal{I}_2^i(L)$ and $\mathcal{I}_1^i(L)$ respectively. Similarly, we also obtain estimates for *near-tail intervals* $\mathcal{I}_2^i(U)$, $\widetilde{\mathcal{I}}_2^{\prime i}(L)$, $\mathcal{I}_1^{\prime i}(L)$, and $\mathcal{I}_2^{\prime i}(U)$, all of which are consistent by the same logic as in Lemma 2.4.1. Thus, from now on we proceed as if these quantities are known.[14]

2.4.2 A Three-Stage Least Squares Estimator

Step 1. Estimate the sign of complementarity parameter and near-tail intervals.

$$\widehat{\operatorname{sgn}(\alpha^i)} = \operatorname{sgn}\left(\frac{1}{M}\sum_{m=1}^{M}\frac{\partial \widehat{P}_1^i(Z_{1m}^i, Z_{2m}^i, X_m)}{\partial z_2^i}\right), \tag{2.2}$$

[14]Ignoring estimation error from some initial stage estimation of subset of the support is common in the econometrics of games literature. See Xu (2014) and Lewbel and Tang (2015), both involve consistent estimation of a subset of the support in an initial stage.

where for generic (z_1^i, z_2^i, x)

$$\frac{\partial \widehat{P}_1^i\left(z_1^i, z_2^i, x\right)}{\partial z_2^i} = \frac{\partial\left(\frac{\sum_{m=1}^M D_{1m}^i K_h\left(\left(Z_{1m}^i, Z_{2m}^i\right)-\left(z_1^i, z_2^i\right)\right)\mathbb{1}(X_m=x)}{\sum_{m=1}^M K_h\left(\left(Z_{1m}^i, Z_{2m}^i\right)-\left(z_1^i, z_2^i\right)\right)\mathbb{1}(X_m=x)}\right)}{\partial z_2^i}, \tag{2.3}$$

where $K_h\left(\left(Z_{1m}^i, Z_{2m}^i\right)-\left(z_1^i, z_2^i\right)\right) = \frac{1}{h^2}K\left(\frac{\left(Z_{1m}^i, Z_{2m}^i\right)-\left(z_1^i, z_2^i\right)}{h}\right)$, and $K()$ is a bivariate (product) kernel. As will be discussed in the next section, the estimated sign of complementarity parameter will have no effect on the asymptotic distribution of the remaining structural parameters.

Step 2. Estimate the equilibrium beliefs and construct the regressors.

For some generic x, the corresponding beliefs could be estimated as:

$$\widehat{P}_1^{-i}(x) = \frac{\sum_{m=1}^M\left(\mathbb{1}\left(D_m^{-i}=(1,0), X_m=x\right)+\mathbb{1}\left(D_m^{-i}=(1,1), X_m=x\right)\right)}{\sum_{m=1}^M \mathbb{1}\left(X_m=x\right)}. \tag{2.4}$$

$$\widehat{P}_2^{-i}(x) = \frac{\sum_{m=1}^M\left(\mathbb{1}\left(D_m^{-i}=(0,1), X_m=x\right)+\mathbb{1}\left(D_m^{-i}=(1,1), X_m=x\right)\right)}{\sum_{m=1}^M \mathbb{1}\left(X_m=x\right)}, \tag{2.5}$$

The regressors corresponding to action 1 and 2 for observation m are constructed as

$$\widehat{v}_1^i(X_m) = [X_m, \widehat{P}_1^{-i}\left(X_m\right)]^\top, \text{ and } \widehat{v}_2^i(X_m) = [X_m, \widehat{P}_2^{-i}\left(X_m\right)]^\top.$$

Step 3. Estimate the structural parameters based on the estimated sign of complementarity parameter and estimated near-tail intervals.

For example, if $\alpha^i \leq 0$, and suppose that we set $c = 0$, then $[\beta_1^i, \delta_1^i]^\top$ could be estimated as

$$\begin{bmatrix}\widehat{\beta}_1^i \\ \widehat{\delta}_1^i\end{bmatrix} = -\left(\frac{\sum_{m=1}^M \widehat{v}_1^i\left(X_m\right)\widehat{v}_1^i\left(X_m\right)^\top W_{2Lm}^i}{M}\right)^{-1}\frac{\sum_{m=1}^M \widehat{v}_1^i(X_m)\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i, X_m\right)}W_{2Lm}^i}{M},$$

and $[\beta_2^i, \delta_2^i]^\top$ could be estimated as

$$\begin{bmatrix} \widehat{\beta}_2^i \\ \widehat{\delta}_2^i \end{bmatrix} = -\left(\frac{\sum_{m=1}^{M} \widehat{v}_2^i\left(X_m\right) \widehat{v}_2^i\left(X_m\right)^\top W_{1Lm}^i}{M} \right)^{-1} \frac{\sum_{m=1}^{M} \widehat{v}_2^i(X_m) \frac{D_{00m}^i + \mathbb{1}\left(Z_{2m}^i \geq 0\right) - 1}{\widehat{f}\left(Z_{2m}^i | Z_{1m}^i, X_m\right)} W_{1Lm}^i}{M}.$$

α^i will be estimated as

$$\frac{\sum_{m=1}^{M} \frac{D_{11m}^i - \mathbb{1}\left(Z_{1m}^i \geq 0\right)}{\widehat{f}\left(Z_{1m}^i | Z_{2m}^i, X_m\right)} W_{2Um}^i}{\sum_{m=1}^{M} W_{2Um}^i} + \frac{\sum_{m=1}^{M} \frac{D_{00m}^i + \mathbb{1}\left(Z_{1m}^i \geq 0\right) - 1}{\widehat{f}(Z_{1m}^i | Z_{2m}^i, X_m)} W_{2Lm}^i}{\sum_{m=1}^{M} W_{2Lm}^i}. \tag{2.6}$$

2.4.3 Trimming

Given definition of our estimator, for the asymptotic theory to be developed in the next section to work, trimming for Z_{1m}^i and Z_{2m}^i is required. We use a fixed trimming strategy that excludes a small but fixed boundary range from the support of Z_{1m}^i and Z_{2m}^i. Following Lewbel and Tang (2015), we do not explicitly write out the trimming function in our estimator. In the proof, this is explicitly accounted for by only evaluating estimated unknown functions on compact subsets of the corresponding supports that exclude the boundary ranges. Note that focusing on the subsets do not affect identification (or consistency) because as we have shown in the identification part, every value for Z_{1m}^i and Z_{2m}^i in the *near-tail-intervals* can deliver identification, and we do not need to use all of them.

2.5 Asymptotic Theory

In this section we establish the asymptotic properties of our proposed estimator. In Section 2.5.1 we will show that the estimated sign of complementarity parameter converges at rate that is faster than root-n. These facts imply that the first step estimation will not affect the asymptotic distributions of structural parameters estimated in later steps. Thus in Section 2.5.2 we will establish root-n asymptotic normality for estimated structural parameters as if *near-tail intervals* and sign of complementarity parameters are known.

Define $Y_{\alpha m}^{i}=(d_{1m}^{i},Z_{1m}^{i},Z_{2m}^{i},X_{m})$, $Y_{\theta m}^{i}=(d_{00m}^{i},Z_{1m}^{i},Z_{2m}^{i},X_{m})$, $\mathbf{X}_{m}^{i}=(Z_{1m}^{i},Z_{2m}^{i},X_{m})$, $\gamma_{0}^{i}=f(z_{1}^{i},z_{2}^{i}|x)f(x)$, $\gamma_{1}^{i}=\mathbb{E}\left(D_{1}^{i}|z_{1}^{i},z_{2}^{i},x\right)f\left(z_{1}^{i},z_{2}^{i}|x\right)f\left(x\right)$, and $\gamma_{2}^{i}=\mathbb{E}\left(D_{00}^{i}|z_{1}^{i},z_{2}^{i},x\right)f\left(z_{1}^{i},z_{2}^{i}|x\right)f\left(x\right)$. Define $\boldsymbol{\gamma}_{\alpha}^{i}=[\gamma_{0}^{i},\gamma_{1}^{i}]^{\top}$ and $\boldsymbol{\gamma}_{\theta}^{i}=\gamma_{0}^{i}$. Let

$$\widehat{\gamma}_{0}^{i}=\frac{1}{M}\sum_{m=1}^{M}K_{h}\left(\left(Z_{1m}^{i},Z_{2m}^{i}\right)-\left(z_{1}^{i},z_{2}^{i}\right)\right)\mathbb{1}\left(X_{m}=x\right),$$

and

$$\widehat{\gamma}_{1}^{i}=\frac{1}{M}\sum_{m=1}^{M}D_{1m}^{i}K_{h}\left(\left(Z_{1m}^{i},Z_{2m}^{i}\right)-\left(z_{1}^{i},z_{2}^{i}\right)\right)\mathbb{1}\left(X_{m}=x\right),$$

Define $\widehat{\gamma}_{\alpha}^{i}=[\widehat{\gamma}_{0}^{i},\widehat{\gamma}_{1}^{i}]^{\top}$, and define $\boldsymbol{\gamma}_{\alpha}^{i*}=[\gamma_{0}^{i*},\gamma_{1}^{i*}]^{\top}$ to be the true functions. Let $\overline{d}^{i}=[1,d_{1}^{i},d_{00}^{i}]$, $d_{\alpha}^{i}=[d_{1}^{i},1]^{\top}$ and $d_{\theta}^{i}=1$. We use $\nabla_{z_{2}^{i}}$ to denote the operator of partial derivative with respect to z_{2}^{i}.

2.5.1 Consistency and Rate of Convergence for Estimated Sign of Complementarity

Define $q_{\alpha}^{i}(\mathbf{x},\gamma_{\alpha}^{i})=\nabla_{z_{2}^{i}}\frac{\gamma_{1}^{i}}{\gamma_{0}^{i}}$ and

$$Q_{\alpha}^{i}\left(\mathbf{x},\gamma_{\alpha}^{i}\right)=\nabla_{z_{2}^{i}}\left([-\frac{\gamma_{1}^{i*}}{\gamma_{0}^{i*2}},\frac{1}{\gamma_{0}^{i*}}]\gamma_{\alpha}^{i}\right)=\nabla_{z_{2}^{i}}\left([-\frac{\gamma_{1}^{i*}}{\gamma_{0}^{i*2}},\frac{1}{\gamma_{0}^{i*}}]\right)\gamma_{\alpha}^{i}+[-\frac{\gamma_{1}^{i*}}{\gamma_{0}^{i*2}},\frac{1}{\gamma_{0}^{i*}}]\nabla_{z_{2}^{i}}\gamma_{\alpha}^{i}.$$

Let $F_{\mathbf{x}}^{*}$ be the true distribution of $\mathbf{X}$, and let $\int dF_{\mathbf{x}}^{*}$ denote expectation with respect to $\mathbf{X}$. For $\int Q_{\alpha}^{i}(\mathbf{x},\gamma_{\alpha}^{i})dF_{\mathbf{x}}^{*}$, applying integration by part to its second term gives

$$\int Q_{\alpha}^{i}(\mathbf{x},\gamma_{\alpha}^{i})dF_{\mathbf{x}}^{*}=\int\sum_{s=0,1}\psi_{\alpha s}^{i}\left(\mathbf{x}\right)\gamma_{s}^{i}d\mathbf{x},\tag{2.7}$$

where the exact form of $\psi_{\alpha s}^{i}\left(\mathbf{x}\right)$ are given in appendix. Define $\mathbf{d}_{\alpha}^{i}=[1,d_{1}^{i}]$ and $\mathbf{D}_{\alpha}^{i}=[1,D_{1}^{i}]$. Define $\psi_{\alpha}^{i}\left(\mathbf{x}\right)=[\psi_{\alpha0}^{i}\left(\mathbf{x}\right),\psi_{\alpha1}^{i}\left(\mathbf{x}\right)]^{\top}$, and $\nu_{\alpha}^{i}\left(y_{\alpha}^{i}\right)=\psi_{\alpha}^{i}(\mathbf{x})\mathbf{d}_{\alpha}^{i}-\mathbb{E}\left[\psi_{\alpha}^{i}\left(\mathbf{X}\right)\mathbf{D}_{\alpha}^{i}\right]$. In this section, we show that $\frac{1}{M}\sum_{m=1}^{M}\frac{\partial\widehat{P}_{1}^{i}\left(Z_{1m}^{i},Z_{2m}^{i},X_{m}\right)}{\partial z_{2}^{i}}$ converges at $\sqrt{n}$ rate, which implies that its sign converges at faster than $\sqrt{n}$ rate. Similar to Lewbel and Tang (2015), the proof builds on chapter 8 in Newey and McFadden (1994).

Assumption 2.5.1 (Unknown functions)**.** *(i).* γ_0^{i*}, γ_1^{i*} *and* γ_2^{i*} *are continuously differentiable in* $\mathbf{z}^i$ *up to an order* $\bar{q} \geq 2$ *given any* x*, and the derivatives are continuous and uniformly bounded over the support of* $\mathbf{z}^i$*; (ii).* γ_0^{i*} *is bounded away from zero over the support of* $\mathbf{z}^i$ *given any* x*; (iii). There exists* $p \geq 4$ *such that* $\mathbb{E}\left[\left\|\overline{D}^i\right\|^p\right] < \infty$ *and for any given* x*,* $\mathbb{E}\left[\left\|\overline{D}^i\right\|^p |\mathbf{z}^i, x\right] f(\mathbf{z}^i|x) < \infty$ *uniformly over the support of* $\mathbf{z}^i$*.*

Note that under Assumption 2.5.1 the derivatives and second derivatives of $q_\alpha^i(\mathbf{x}, \gamma_\alpha^i)$ with respect to γ_α^i are all bounded over $\mathbf{x}$. The next assumption imposes restrictions on an additional derivative term created via integration by parts.

Assumption 2.5.2 (Derivative-α)**.** $\nabla_{z_2^i}(J_\alpha^i \gamma_0^{i*})$ *is continuous in* $\mathbf{z}^i$ *and bounded over* $\mathbf{x}$*.*

Assumption 2.5.3 (Kernel)**.** *(i).* $K(\cdot)$ *is bounded and differentiable of order* $\overline{q}$ *and partial derivatives are all bounded; (ii).* $\int K(t)dt = 1$*,* $K(\cdot)$ *has 0 moments up to order* $q \leq \overline{q}$*,* $\int \|t\|^{\bar{q}} |K(t)| \, dt < \infty$*; (iii).* $K(\cdot) = 0$ *outside a bounded set.*

Assumption 2.5.4 (Bandwidth)**.** $\sqrt{M}h^q \to 0$*,* $\left(\frac{\sqrt{M}}{logM}\right) h^4 \to \infty$ *as* $M \to \infty$*.*

Assumption 2.5.5. *(i). There exists a constant* $c_\alpha^i > 0$ *such that* $\mathbb{E}\left[\sup_{\|\eta\| \leq c_\alpha^i} \left\|\psi_{\alpha k}^i(\mathbf{X} + \eta)\right\|^4\right] < \infty$ *for* $k = 1, 2$*; (ii).* $\mathbb{E}\left[\left\|q_\alpha^i(\mathbf{X}, \gamma_\alpha^{i*}) + \nu_\alpha^i(Y_\alpha^i)\right\|^2\right] < \infty$*.*

Lemma 2.5.1. *Under Assumption 2.2.1-2.5.5,* $\widehat{\operatorname{sgn}(\alpha^i)} = \operatorname{sgn}(\alpha^i) + o_p\left(M^{-1/2}\right)$*.*

Proof. See Appendix B.7. □

2.5.2 Root-n Consistency and Asymptotic Normality for Estimated Structural Parameters

Without loss of generality, we establish the root-n consistency and asymptotic normality for estimated structural parameters when $\alpha^i \leq 0$. In this case, the forms of the estimators are given in Section 2.4. Let $\theta_1^i = [\beta_1^i, \delta_1^i]$ and $\widehat{\theta}_1^i = [\widehat{\beta}_1^i, \widehat{\delta}_1^i]$. Define

$$\begin{aligned} & q_{1\theta}^i \left(\mathbf{x}, \gamma_\theta^i\right) \\ = \;& v_1^i(x) w_{2L}^i \frac{D_{00}^i + \mathbb{1}\left(z_1^i \geq 0\right) - 1}{f\left(z_1^i | z_2^i, x\right)} \\ = \;& v_1^i(x) w_{2L}^i \frac{\left(D_{00}^i + \mathbb{1}\left(z_1^i \geq 0\right) - 1\right) \int \gamma_0^i\left(\mathbf{t}\right) dt_1}{\gamma_0^i}. \end{aligned}$$

Define $Q_{1\theta}^i\left(\mathbf{x}, \gamma_\theta^i\right) = J_{1\theta}^i \gamma_\theta^i$. A major step in the proof shows that $Q_{1\theta}^i\left(\mathbf{x}, \gamma_\theta^i\right)$ has the following representation:

$$\int Q_{1\theta}^i(\mathbf{x}, \gamma_\theta^i) dF_{\mathbf{x}}^* = \int \psi_{1\theta 0}^i\left(\mathbf{x}\right) \gamma_0^i d\mathbf{x}, \tag{2.8}$$

where the exact form of $\psi_{1\theta 0}^i\left(\mathbf{x}\right)$ are given in appendix. Define $\nu_{1\theta}^i\left(y_\theta^i\right) = \psi_{1\theta}^i(\mathbf{x}) - \mu_{1\theta}^i$, where $\mu_{1\theta}^i \equiv \mathbb{E}\left[\psi_{1\theta}^i\left(\mathbf{X}\right)\right]$.

Assumption 2.5.6. *(i). There exists a constant* $c_\theta^i > 0$ *such that* $\mathbb{E}\left[\sup_{\|\eta\| \leq c_\theta^i} \left\|\psi_{1\theta 0}^i\left(\mathbf{X} + \eta\right)\right\|^4\right] < \infty$*; (ii).* $\mathbb{E}\left[\left\|q_{1\theta}^i\left(\mathbf{X}, \gamma_\theta^{i*}\right) + \nu_{1\theta}^i\left(Y_\theta^i\right)\right\|^2\right] < \infty$.

The next theorem gives asymptotic distributions for our estimated parameters:

Theorem 2.5.1. *Suppose Assumptions 2.2.1-2.5.4, 2.5.6 hold, and the rank conditions in Theorem 2.3.3 are satisfied. Then it holds that for* $k = 1, 2$

$$\sqrt{M}\left(\widehat{\theta}_k^i - \theta_k^i\right) \xrightarrow{d} N\left(0, \Omega_{\theta_k^i}\right), \tag{2.9}$$

$$\text{and } \sqrt{M}\left(\widehat{\alpha}^i - \alpha^i\right) \xrightarrow{d} N\left(0, \Omega_{\alpha^i}\right). \tag{2.10}$$

Proof. See Appendix B.8 □

2.6 Monte Carlo Simulations

In this section we conduct a small Monte Carlo study of the performance of our estimator in finite samples given knowledge of *near-tail intervals*. We construct two DGPs: a game with substitutable actions and a game with complementary actions. Like in Lewbel and Tang (2015), we use the tri-weight kernel function $K(t) = \frac{35}{32}\left(1-t^2\right)^3 \mathbb{1}(|t| \leq 1)$. We set bandwidth to be $h = 4M^{-1/9}$.

DGP1.(Game with Substitutable Actions) Parameters for both players are set according to Table 2.1. X follows discrete uniform distribution over $\{-1, -0.5, 0.5, 1\}$, $\epsilon_k^i \overset{\text{i.i.d}}{\sim} \text{U}[-4, 4]$, $z_k^i \overset{\text{i.i.d.}}{\sim} \text{U}[-10, 10]$.

Table 2.4 and Table 2.5 shows the performance for our estimator for $[\beta_1^1, \delta_1^1]$ and $[\beta_2^1, \delta_2^1]$. In these tables "LQ" denotes the 25% quantile, "UQ" denotes the 75% quantile and "median no. of obs" denotes the median number of observations contained in the *near-tail intervals*, which gives an estimate of the number of observations actually used in estimation. Given that the mean is close to the true value for both sample sizes, comparison of the "MSE" column between these two tables show clear evidence of convergence to the true value as sample size increases.

Table 2.4: Estimator for $[\beta_1^1, \delta_1^1]^\top$ and $[\beta_2^1, \delta_2^1]^\top$, $M = 1500$, 200 repetitions

	true value	mean	median	LQ	UQ	MSE	median no. of obs
$\widehat{\beta_1^1}$	1.5	1.341	1.358	1.096	1.590	0.173	300
$\widehat{\delta_1^1}$	-1	-0.909	-0.915	-1.387	-0.427	0.505	300
$\widehat{\beta_2^1}$	1	0.871	0.847	0.641	1.114	0.173	299
$\widehat{\delta_2^1}$	-0.5	-0.394	-0.441	-0.817	-0.009	0.400	299

Table 2.5: Estimator for $[\beta_1^1, \delta_1^1]^\top$, and $[\beta_2^1, \delta_2^1]^\top$, $M = 3000$, 200 repetitions

	true value	mean	median	LQ	UQ	MSE	median no. of obs
$\widehat{\beta_1^1}$	1.5	1.542	1.538	1.320	1.762	0.122	599
$\widehat{\delta_1^1}$	-1	-1.073	-1.094	-1.343	-0.732	0.215	599
$\widehat{\beta_2^1}$	1	0.947	0.960	0.758	1.157	0.095	599
$\widehat{\delta_2^1}$	-0.5	-0.452	-0.432	-0.780	-0.083	0.293	599

Table 2.6: Estimator for α^1, $M = 1500$ and 3000, 200 repetitions

	true value	mean	median	LQ	UQ	MSE
$M = 1500$	-1	-0.853	-0.866	-1.199	-0.541	0.252
$M = 3000$	-1	-0.962	-0.968	-1.216	-0.722	0.111

Table 2.6 represent performance for estimation of α^1, the parameter that quantifies the degree of substitutability between two actions. We can see that the estimator is well centered and its MSE decreases with sample size.

DGP2.(Game with Complementary Actions) Everything else is the same as in **DGP1** except that now $\alpha^1 = 1$ and $\alpha^2 = 2$.

Table 2.7 and Table 2.8 shows the performance for our estimator for $[\beta_1^1, \delta_1^1]$ and $[\beta_2^1, \delta_2^1]$. Given that the mean is close to the true value for both sample sizes, comparison of the "MSE" column between these two tables show clear evidence of convergence to the true value as sample size increases.

Table 2.7: Estimator for $[\beta_1^1, \delta_1^1]^\top$ and $[\beta_2^1, \delta_2^1]^\top$, $M = 1500$, 200 repetitions

	true value	mean	median	LQ	UQ	MSE	median no. of obs
$\widehat{\beta_1^1}$	1.5	1.343	1.378	1.070	1.615	0.181	300
$\widehat{\delta_1^1}$	-1	-0.894	-0.939	-1.281	-0.534	0.345	300
$\widehat{\beta_2^1}$	1	0.800	0.802	0.488	1.111	0.235	299
$\widehat{\delta_2^1}$	-0.5	-0.376	-0.408	-0.759	-0.009	0.338	299

Table 2.8: Estimator for $[\beta_1^1, \delta_1^1]^\top$, and $[\beta_2^1, \delta_2^1]^\top$, $M = 3000$, 200 repetitions

	true value	mean	median	LQ	UQ	MSE	median no. of obs
$\widehat{\beta_1^1}$	1.5	1.457	1.471	1.229	1.675	0.113	599
$\widehat{\delta_1^1}$	-1	-1.003	-1.040	-1.327	-0.590	0.274	599
$\widehat{\beta_2^1}$	1	0.979	0.971	0.774	1.183	0.074	599
$\widehat{\delta_2^1}$	-0.5	-0.456	-0.483	-0.802	-0.006	0.314	599

Table 2.9: Estimator for α^1, $M = 1500$ and 3000, 200 repetitions

	true value	mean	median	LQ	UQ	MSE
$M = 1500$	1	0.954	0.972	0.585	1.295	0.211
$M = 3000$	1	0.999	0.957	0.787	1.217	0.102

Table 2.9 represent performance for estimation of α^1, the parameter that quantifies the degree of complementarity between two actions. We can see that the estimator is well centered and its MSE decreases with sample size.

2.7 Concluding Remarks

In this paper we study the identification and estimation of static games of incomplete information in which each player faces two complementary (substitutable) actions, and could take one, or both, or neither of them. Each player has two pieces of unobservable private information, one for each action and drawn from a bivariate distribution that is unknown to the econometrician. Identification relies on player-action specific state variables that are private information among players when the game is played, but available to the econometrician afterwards. We show that the structural parameters in this model are identified when such player-action specific state variables are additively separable and have large supports. A simple estimator for the structural parameters is proposed based on this identification strategy. The estimator could be implemented easily by running a three-stage least squares. We establish the root-n consistency and asymptotic normality of this estimator. A small Monte Carlo simulation shows the efficacy of our methods in finite samples with moderate sample sizes. There are at least two directions for future research. One direction is to study the asymptotic distribution of our estimator when the estimation error involved in the estimation of *near-tail* intervals are also accounted for. Another direction is to explore ways that could make use of more data for estimation instead of concentrating on a small amount of data near the tails. Additional restrictions could be helpful. For example, if the error term is known to be conditionally symmetric, by the mathematical equivalence of a demand model for bundles and a potential game of complete information (see Fox and Lazzati (2017)), the identification and estimation procedure developed by Zhou (2019) for games of complete information under symmetry restriction could be directly applied here (the added complication in our setting will be a generated regressor).

Chapter 3

IDENTIFICATION AND ESTIMATION OF BINARY GAMES OF INCOMPLETE INFORMATION UNDER SYMMETRY OF THE UNOBSERVED PRIVATE INFORMATION

3.1 Introduction

In the recent two decades, there is a burgeoning literature on the econometric analysis of game theoretical models, see Bajari et al. (2013) for a detailed review. In the study of incomplete information games, a class of games in which some information is only privately observed when the game is played, most applications assume a known parametric distribution of the unobservable private information (or the error term for short), see Seim (2006), Sweeting (2009) and references therein.[1]

Recently, efforts have been made to relax such distributional assumption. Aradillas-Lopez (2010) studies identification and estimation of a discrete game of incomplete information without a known error distribution. He assumes statistical independence between the error term and all observed states, thus ruling out heteroskedastic games. Tang (2010) studies identification and estimation of an incomplete information game under the median independence assumption between error term and state variables. He establishes consistency for his estimators without establishing rates of convergence. Wan and Xu (2014) studies identification and estimation of a two-player binary game with correlated private information. They propose a modified maximum score type estimator which is shown to converge at $N^{-1/3}$. Lewbel and Tang (2015) studies identification and estimation of binary games of incomplete information using the idea of special regressors developed by Lewbel (2000). Lewbel and Tang (2015) uses a generated special regressor that is a function of equilibrium conditional

[1] Throughout this paper we use the term unobservable private information and error term interchangeably.

choice probabilities (CCPs) for identification, and their model could adapt heteroskedasticity. However, their special regressor based identification strategy requires the special regressor to have a support that is at least as large as the support for the error term, which equals the whole real line if the error follows the commonly assumed normal or logistic distribution. Magnac and Maurin (2007) points out that identification based on large support assumption can only be used when the conditional choice probability increases from 0 to 1 over the support of the special regressor, which is admittedly restrictive, and "*represents a potential obstacle to empirical applications.*" For root-N normality, Lewbel and Tang (2015) further requires an additional assumption for relative tail thickness between the error term and the special regressor: when the error term has bounded support, it needs to be strictly contained in the support of the special regressor; when the error term has unbounded support (important special cases such as normal or logistic error belong to this category), the special regressor needs to have a support that is at least as large with infinite variance. Without such restrictions, the rate of convergence for their estimator will in general vary between $N^{-1/4}$ and $N^{-1/2}$ (Khan and Tamer (2010)). The fundamental reason for such stringent requirement is that identification of the structural parameters requires identification of the whole error distribution using variation of the special regressor, and point identification will be lost, for example, if an arbitrarily small part of the error support in the tail region is not covered by the special regressor.

As ignoring heteroskedasticity and misspecification of the error distribution both lead to invalid estimation and inference, an estimator that could adapt heteroskedasticity, is robust to misspecification of error distribution, does not require stringent support or tail conditions, and converges at root-N rate is highly called for. This motivates our paper.

Our paper provides an alternative method for identification and estimation of binary games of incomplete information compared to the existing methods: we exploit the restriction implied by symmetry of the error distribution. To the best of our knowledge, this is the first paper that studies identification and estimation of a game of incomplete information under symmetry restriction. Our method is inspired by Lewbel et al. (2020), which uses the central symmetry restriction of the error distribution for identification and estimation of multinomial discrete choice models. We show that equilibrium condition and this shape restriction suffices to identify a binary game of incomplete information without requirement of large support as in Lewbel and Tang (2015). This means that we could allow for dataset in which the observed range of conditional choice probability is any non-degenerate subset of $[0, 1]$ instead of exactly $[0, 1]$. For root-N asymptotic normality, we do not need restriction on relative tail thickness as in Lewbel and Tang (2015). Support or tail requirement could be relaxed here, because our identification strategy is more robust to possible nonidentification of the whole error distribution. In our case, an excluded regressor could have bounded support that doesn't cover the error support, which could be bounded or unbounded. Excluded regressor with bounded support is the case in most empirical applications. Although inspired by Lewbel et al. (2020), we differ from their work in that they use the implied restriction for pdf while we use the implied restriction for cdf when we construct identification condition as well as the resulting estimator. The way we exploit symmetry is closer to the one used in Chen (2000). Compared to Lewbel et al. (2020)'s approach, our identifying restriction allows for a larger class of error distribution. In particular, uniform distribution or distribution that is flat around the symmetric point could not be handled by a density-based identifying restriction, but is allowed by our identifying restriction. An additional difference from Lewbel et al. (2020) is that in our problem there is a generated regressor due to the first-step estimation of equilibrium belief, which is handled using linear expansion with respect to the generated regressor.

Our paper is related to a number of papers using central symmetry for semiparametric identification and estimation. In a binary choice model setting, Chen (2000) studies estimation of binary choice models under symmetry, and in particular, examines the efficiency gain from imposing the symmetry restriction. Chen et al. (2016) finds that in a binary choice model, the combination of central symmetry of error term and conditional independence between error distribution and one regressor suffices to identify index coefficient and results in a root-N consistent estimator of index coefficient. Zhou (2019) uses the central symmetry restriction for identification and estimation of a two-player binary static game of complete information. Lewbel et al. (2020) use the central symmetry restriction for identification and estimation of a multinomial discrete choice model.

The rest of the paper is organized as follows. Section 3.2 introduces the baseline game and defines equilibrium. Section 3.3 discusses identification. Section 3.4 introduces our two-step minimum distance estimator. Section 3.5 proves consistency and root-N asymptotic normality. Section 3.6 presents results from a Monte Carlo simulation. Section 3.7 discusses extensions. Section 3.8 concludes.

Notation For a generic random variable Q, let Ω_Q denote its support. For all random variables except ϵ, we use uppercase letters to denote the random variables and lowercase to denote their realized values. For a $L-$dimensional vector v, let $v[l]$ with $1 \leq l \leq L$ be its l-th element and $v[-l]$ be the elements other than l. Let $\mathcal{P}_{m,n}$ denote the sset of all 2-element permutations in $\{l : i \in \mathbb{Z}, 1 \leq l \leq N\}$, where $\mathbb{Z}$ denotes the set of integers. For a positive integer L, let $\mathbf{0}_L$ denote a $L-$dimensional vector of zeros. For any finite set A, let $\text{card}(A)$ denote its cardinality. Throughout this paper, superscript i denotes player i.

3.2 The Baseline Game and Equilibrium

Without loss of generality, consider a binary game of incomplete information with 2 players.[2] For $i = 1, 2$, player i chooses an action D^i from the action space $\mathcal{D} = \{0, 1\}$. Let $-i \equiv \{1, 2\} \setminus i$, i.e., the player index for player i's opponent. The payoff function of choosing action 0 is normalized to 0, and the payoff function of choosing action 1 is specified as $\dot{U}^i\left(Z^i, X\right) - \dot{\epsilon}^i = \dot{\alpha}^i Z^i + X\dot{\beta}_1^i + \dot{\beta}_2^i D^{-i} - \dot{\epsilon}^i$, where $Z^i \in \Omega_{Z^i} \subset \mathbb{R}$ is a scalar with nonzero coefficient, $X \in \Omega_X \subset \mathbb{R}^L$ is a (row) vector of state variables, and $\dot{\epsilon}^i \in \Omega_{\epsilon^i} \subset \mathbb{R}$ is player i's private information. $\dot{\alpha}^i$, $\dot{\beta}_2^i$ are scalar coefficients, and $\dot{\beta}_1^i$ is a (column) vector coefficient.[3]

[2]All discussions in this paper trivially generalize to a game with I players.

[3]The requirement for player-specific covariates to have nonzero coefficients is used in Lewbel and Tang (2015). A similar assumption is also used in papers on single agent discrete choice models such as Lewbel (2000), Chen et al. (2016) and Lewbel et al. (2020).

Assumption 3.2.1 (Information on State Variables)**.** *(i). The realization of X is common knowledge among the two players; (ii). The realization of Z^i is player i's private information when the game is played, and observable to the econometrician (after the game is played).*

From the econometrician's perspective, this assumption introduces two pieces of private information for player i: the observable private informaiton Z^i and the unobservable private information $\dot{\epsilon}^i$. This assumption could be justified by the institutional background under which players interact, by knowledge of players' main focus in decision making, or by the data collecting process. When lacking such information, it could be justified by examining whether the probability of $D^{-i} = 1$ changes significantly with Z^i after controlling for other variables that affect player $-i$'s payoff. In section 3.7, we study the situation in which Assumption 3.2.1 (ii) is changed to the situation in which (Z^1, Z^2) are also common knowledge.

Assumption 3.2.2 (Continuity and Conditional Independence)**.** *For $i = 1, 2$, suppose the conditional distributions of ϵ^i, Z^i and $\epsilon^i + \alpha^i Z^i$ given $X = x$ are absolutely continuous with respect to the Lebesgue measure and have bounded and positive Radon-Nikodym densities everywhere on their conditional supports. We assume the following independence conditions hold: (i). $(Z^1, \epsilon^1) \perp (Z^2, \epsilon^2)|X$; (ii). For $i = 1, 2$, $Z^i \perp \epsilon^i|X$.*

The assumption that private information is independent across players conditional on common information is a widely maintained assumption in the econometrics of games literature (See De Paula and Tang (2012) and Bajari et al. (2013)). Also notice that this assumption implies that $F_{\dot{\epsilon}^i|(Z^i,X)}(\cdot) = F_{\dot{\epsilon}^i|X}(\cdot)$, and similar to Lewbel (2000) and Lewbel et al. (2020), it allows for arbitrary forms of heteroskedasticity with respect to X. We also focus on pure strategies defined as follows:

Definition 3.2.1 (Pure Strategy)**.** *A pure strategy for player i in this game is a mapping $g^i : \Omega_{Z^i} \times \Omega_X \times \Omega_{\dot{\epsilon}^i} \to \mathcal{D}$.*

By definition $D^i = g^i(Z^i, X, \dot{\epsilon}^i)$. Based on his available information, player i forms a belief about player $-i$'s action. In a binary game, this belief is completely characterized by the probability of event $D^{-i}=1$ given player i's available information. Formally, we have

$$\Pr\left(g^{-i}\left(Z^{-i},X,\dot{\epsilon}^{-i}\right)=1|Z^i,X,\dot{\epsilon}^i\right)=\Pr\left(g^{-i}\left(Z^{-i},X,\dot{\epsilon}^{-i}\right)=1|X\right)=\Pr(D^{-i}=1|X),$$

where the first equality holds by Assumption 3.2.1 and Assumption 3.2.2 (i).[4] Note that the belief only depends on commonly observed state variable X. Intuitively, once X is known, Z^i and $\dot{\epsilon}^i$ doesn't contain any additional information about player $-i$'s action. Given $Z^i, X, \dot{\epsilon}^i$ and belief $\Pr\left(D^{-i}=1|X\right)$, player i's optimal strategy follows a threshold crossing rule:

$$D^i=\mathbb{1}\left(\dot{\alpha}^i Z^i+X\dot{\beta}_1^i+\dot{\beta}_2^i\Pr(D^{-i}=1|X)-\dot{\epsilon}^i\geq 0\right).$$

A pure strategy Bayesian Nash equilibrium (BNE) of the game is defined as follows

Definition 3.2.2 (Equilibrium)**.** *Given X, a BNE of the game is defined as the fixed point $(P^{1*}(x),P^{2*}(x))$ of the following system in $(P^1(x),P^2(x))$:*

$$P^1(X)=\mathbb{E}\left(\mathbb{1}\left(\dot{\alpha}^1Z^1+X\dot{\beta}_1^1+\dot{\beta}_2^1P^2(X)-\dot{\epsilon}^1\geq 0\right)|X\right),$$
$$P^2(X)=\mathbb{E}\left(\mathbb{1}\left(\dot{\alpha}^2Z^2+X\dot{\beta}_1^2+\dot{\beta}_2^2P^1(X)-\dot{\epsilon}^2\geq 0\right)|X\right).$$

From now on, we use $P^i(x)$ to denote equilibrium belief given common observed state $X=x$ and use $P^i(z^i,x)$ to denote the equilibrium conditional choice probability for $D^i=1$ given $(Z^i,X)=(z^i,x)$, for $i=1,2$. As an illustration, we introduce a very simple game in the following example, and compute its equilibrium beliefs:

Example 3.2.1 (Equilibrium Beliefs)**.** *Suppose for $i=1,2$, $Z^i\sim$ Unif$[-10,10]$, $\dot{\epsilon}^i\sim$ Normal$(0,2)$. $\Omega_X=\{-1,1\}$ and X takes each value with probability 0.5, and the true constant term is zero. The parameters of the game are set as follows:*

---

[4]Note that if private information is correlated across players conditional on X, belief will depend on (Z^i,X,ϵ^i).

Table 3.1: Value of Parameters

$\dot{\alpha}^1$	$\dot{\beta}_1^1$	$\dot{\beta}_2^1$	$\dot{\alpha}^2$	$\dot{\beta}_1^2$	$\dot{\beta}_2^2$
-1	0.8	-0.5	-1	0.7	-0.6

The calculated equilibrium beliefs of this game are given in the following table:

Table 3.2: Calculated Equilibrium Beliefs

	Equilibrium belief for player 1	Equilibrium belief for player 2
$X = -1$	0.449	0.452
$X = 1$	0.527	0.519

3.3 Two-Step Identification

Suppose the density of X with respect to the product of Lebesgue and counting measure is given by $f(x) \equiv f(x_c|x_d)P_{xd}$, where P_{xd} is the probability mass function for discrete components of x and $x \in \Omega_X$. Suppose the econometrician has access to cross-sectional data on players' actions and state variables:

Assumption 3.3.1 (Data)**.** *The econometrician observes a random sample on players' actions and state variables* $\{(Z_n^1, Z_n^2, X_n, D_n^1, D_n^2)\}_{n=1}^N$.

We use a two-step identification strategy that is widely used in the econometrics of games literature, see Aguirregabiria and Mira (2007), Bajari et al. (2013) and references therein. In two-step identification, equilibrium beliefs are identified from data first, and structural parameters are identified in the second step after plugging in identified beliefs.

3.3.1 Identification of Equilibrium Beliefs

Similar to Aguirregabiria and Mira (2007), Bajari et al. (2013) we assume that data is rationalized by a single equilibrium:

Assumption 3.3.2 (Equilibrium)**.** *Given each* X*, the observed actions are rationalized by a single equilibrium.*

Following Aguirregabiria and Mira (2007), Bajari et al. (2013) and Lewbel and Tang (2015), under Assumption 3.3.1 and Assumption 3.3.2, $P^1(X)$ and $P^2(X)$ are identified for any X. Thus we treat $P^1(X)$ and $P^2(X)$ as known from now on.

3.3.2 Identification of Structural Parameters

As in standard binary choice models, structural parameters in binary games are only identified up to scale and normalization is required. Without loss of generality, we assume $\dot{\alpha}^i < 0$ and normalize $\alpha^i \equiv \frac{\dot{\alpha}^i}{|\dot{\alpha}^i|} = -1$.[5] After normalization, we have that for $i = 1, 2$ the equilibrium response is given by:

$$D^i = \mathbb{1}\left(-Z^i + X\beta_1^i + \beta_2^i P^{-i}(X) - \epsilon^i \geq 0\right),$$

where $\beta_1^i \equiv \frac{\dot{\beta}_1^i}{|\dot{\alpha}^i|}$, $\beta_2^i \equiv \frac{\dot{\beta}_2^i}{|\dot{\alpha}^i|}$ and $\epsilon^i \equiv \frac{\dot{\epsilon}^i}{|\dot{\alpha}^i|}$. We normalize the coefficient of Z^i to -1 (instead of 1) to keep the sign of private information variables the same.

After plugging in the identified equilibrium beliefs, identification of structural parameters combines the idea in Chen (2000) and Lewbel et al. (2020).

Assumption 3.3.3 (Conditional Central Symmetry)**.** *For $i = 1, 2$, given any $x \in \Omega_X$, $F_{\epsilon^i}(e^i|x) = 1 - F_{\epsilon^i}(-e^i|x)$ for any $e^i \in \Omega_{\epsilon^i|x}$.*

Just like in Lewbel et al. (2020), the designation of the true center at zero is without loss of generality, as any nonzero center could be "absorbed" by a constant term in the payoff function (see the remark in the next section). Conditional central symmetry is satisfied by a wide class of distributions including normal distribution, t-distribution, logistic distribution, uniform distribution and Laplace distribution.

Remark 3.3.1. *Assumption 3.3.3 implies that $\mathbb{E}(\epsilon^i|X) = 0$ as long as the conditional expectation exists. In such cases, ϵ^i and X are mean independent, which implies uncorrelatedness. This means that ϵ^i and X will be statistically independent if they are jointly*

[5] If the sign of α^i is unknown, it could be identified as $\mathbb{E}(\frac{\partial P^i(Z^i,X)}{\partial z^i})$ and be estimated at faster than root-N rate, see Lewbel (2000).

normal. Except for this special case, X *are* ϵ^i *are not independent in general, and heteroskedasticity, the situation in which* X *affects the conditional variance of* ϵ^i*, could be allowed. For example, suppose* $X \sim Bernoulli(0.5)$ *on* $\{1, 2\}$*, and the conditional distribution of* ϵ^i *given* $X = x$ *is* $N(0, x^2)$*.* X *and* ϵ^i *are not independent here, as it holds that* $Pr(\epsilon^i \leq e) = 0.5\Phi(e) + 0.5\Phi(\frac{e}{2})$ *and* $Pr(\epsilon^i \leq e|X = 1) = \Phi(e)$*, which implies* $Pr(\epsilon^i \leq e) - Pr(\epsilon^i \leq e|X = 1) = 0.5[\Phi(\frac{e}{2}) - \Phi(e)] \neq 0$ *for* $e \neq 0$*.*

For $i = 1, 2$, let $\mathbf{X}_p^i \equiv (X, P^{-i}(X)) \in \Omega_X \times [0, 1] \subset \mathbb{R}^{L+1}$ be the (row) vector of all covariates excluding Z^i for player i, let $\mathbf{X}^i \equiv (Z^i, X) \in \Omega_{(Z^i,X)} \subset \mathbb{R}^{L+1}$ denote all the state variables that are observable to the econometrician, and let $X_c \in \Omega_{X_c} \subset \mathbb{R}^{L_c}$ and $X_d \in \Omega_{X_d} \subset \mathbb{R}^{L_d}$ denote continuous and discrete state variables in X respectively, where $L_c + L_d = L$. Similarly, let $\mathbf{X}_c^i$ and $\mathbf{X}_d^i$ denote continuous and discrete state variables in $\mathbf{X}^i$ respectively. By construction $\mathbf{X}_c^i = (Z^i, X_c)$, $\mathbf{X}_d^i = X_d$, and dimension of $\mathbf{X}_c^i$ is $L_c + 1$.

Intuition for Identification

For notational convenience, let $\boldsymbol{\beta}^i = (\beta_1^{i\top}, \beta_2^i)^\top$ and let $\boldsymbol{\beta}_0^i$ be its true value. Notice that under Assumption 3.3.2, the value of $\mathbf{X}_p^i$ is completely determined by the value of X. The conditional choice probability of $D^i = 1$ for player i given $Z^i = z^i$ and $X = x$ is

$$P^i(z^i, x) \equiv \mathbb{E}(D^i|z^i, x) = \Pr\left(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i \geq \epsilon^i\right) = F_{\epsilon^i|z^i,x}(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i) = F_{\epsilon^i|x}(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i),$$

where the last equality hold by Assumption 3.2.2 (ii). The conditional choice probability $\mathbb{E}(D^i|z^i, x)$ as a function of (z^i, x) is nonparametrically identified on $\Omega_{(Z^i,X)}$. The fact that $\mathbb{E}(D^i|z^i, x) = F_{\epsilon^i|x}(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i)$ implies that for a generic point $(z^{i*}, x^*) \in \Omega_{(Z^i,X)}$, $\mathbb{E}(D^i|z^{i*}, x^*)$ identifies $F_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i)$, which is the value of $F_{\epsilon^i|x^*}(\cdot)$ evaluated at $-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$.

On the other hand, for some candidate value $\boldsymbol{\beta}^i$ that we choose from the parameter space, if we evaluate the function $\mathbb{E}(D^i|z^i, x)$ at $(-z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i, x^*) \in \Omega_{(Z^i,X)}$, we get

$$\mathbb{E}(D^i| - z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i, x^*) = F_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i).$$

If the candidate value $\boldsymbol{\beta}^i$ happens to be the true value $\boldsymbol{\beta}_0^i$, then we have

$$\mathbb{E}(D^i| - z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i, x^*) = F_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = F_{\epsilon^i|x^*}(z^{i*} - \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i).$$

As under Assumption 3.3.3, it holds that

$$F_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = 1 - F_{\epsilon^i|x^*}(z^{i*} - \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i),$$

we have

$$\mathbb{E}(D^i|z^{i*}, x^*) = 1 - \mathbb{E}(D^i| - z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i, x^*),$$

for any $(z^{i*}, x^{i*}) \in \Omega_{(Z^i,X)}$. As (z^{i*}, x^{i*}) varies, this creates a set of equations that will provide enough restrictions for identifying $\boldsymbol{\beta}_0^i$, as for $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$, this set of equations do not hold simultaneously in general. Based on this fact, define the following distance function

$$\begin{aligned} d^i(\boldsymbol{\beta}^i; z^{i*}, x^*) &= \mathbb{E}(D^i|z^{i*}, x^*) - (1 - \mathbb{E}(D^i| - z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i, x^*)) \\ &= P^i(z^{i*}, x^*) - (1 - P^i(-z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i, x^*)). \end{aligned}$$

The identification restriction could be rewritten using the distance function as:

$$d^i(\boldsymbol{\beta}^i; z^{i*}, x^*) = 0. \tag{3.1}$$

By construction, it holds that $d^i(\boldsymbol{\beta}_0^i; z^{i*}, x^*) = 0$ for any (z^{i*}, x^*); for $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$, $d^i(\boldsymbol{\beta}^i; z^{i*}, x^*) = 0$ does not hold in general (at least for some (z^{i*}, x^*)). We will use this distance function to distinguish $\boldsymbol{\beta}_0^i$ from $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$.

Remark 3.3.2. *Another way to look at this restriction is that for a distribution $F_{\epsilon^i|x^*}(\cdot)$ that is known to be symmetric and continuous, the restriction $F_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = 1 - F_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i)$ implies that the center of the distribution is at $\mathbf{x}_p^{i*}(-\boldsymbol{\beta}^i + \boldsymbol{\beta}_0^i)$, which coincides with zero (for any $\mathbf{x}_p^{i*}$) when $\boldsymbol{\beta}^i$ equals $\boldsymbol{\beta}_0^i$.*

Example 3.3.1 (Illustration of Identification)**.** *Continue with the game introduced in Example 3.1, we illustrate our idea of identification in the following table:*

Table 3.3: Illustration of Identification

Support of Z^1	$(-10, 10)$
Support of ϵ^1	$\mathbb{R}$
Identified region in the support of ϵ^1 when $X = -1$	$(-11.026, 8.974)$
Identified region in the support of ϵ^1 when $X = 1$	$(-9.459, 10.541)$
Restriction implied by Equation (3.1) when $X = -1$	$-1(\beta_1^1 - 0.8) + 0.452(\beta_2^1 + 0.5) = 0$
Restriction implied by Equation (3.1) when $X = 1$	$(\beta_1^1 - 0.8) + 0.519(\beta_2^1 + 0.5) = 0$

The two implied restrictions on the parameters uniquely determine $\beta_1^1 = 0.8$ and $\beta_2^1 = -0.5$.

Remark 3.3.3. *Suppose the true center of ϵ^i given x is some (unknown) λ_0^i instead of zero, and suppose the first argument in $\mathbf{x}_p^i$ is one so that there is a constant term in the payoff function. Let $\mathbf{x}_p^i[-1]$, $\boldsymbol{\beta}^i[-1]$ and $\boldsymbol{\beta}_0^i[-1]$ denote the corresponding vectors without their first elements. We have*

$$\mathbb{E}(D^i|z^{i*}, x^*) = F_{\epsilon^i|z^{i*},x^*}(-z^{i*} + \boldsymbol{\beta}_0^i[1] + \mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}_0^i[-1]) = F_{\epsilon^i|x^*}(-z^{i*} + \boldsymbol{\beta}_0^i[1] + \mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}_0^i[-1]).$$

If we evaluate the function $\mathbb{E}(D^i|z^i, x)$ at $(-z^{i*} + 2\boldsymbol{\beta}^i[1] + 2\mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}^i[-1], x^*) \in \Omega_{(Z^i,X)}$, we get

$$\begin{aligned} & \mathbb{E}(D^i| - z^{i*} + 2\boldsymbol{\beta}^i[1] + 2\mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}^i[-1], x^*) \\ = \; & F_{\epsilon^i|x^*}(z^{i*} - 2\boldsymbol{\beta}^i[1] - 2\mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}^i[-1] + \boldsymbol{\beta}_0^i[1] + \mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}_0^i[-1]). \end{aligned}$$

Then imposing

$$F_{\epsilon^i|x^*}(-z^{i*}+\boldsymbol{\beta}_0^i[1]+\mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}_0^i[-1]) = 1-F_{\epsilon^i|x^*}(z^{i*}-2\boldsymbol{\beta}^i[1]-2\mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}^i[-1]+\boldsymbol{\beta}_0^i[1]+\mathbf{x}_p^{i*}[-1]\boldsymbol{\beta}_0^i[-1])$$

implies that the center is at $(-\boldsymbol{\beta}^i[1]+\boldsymbol{\beta}_0^i[1])+\mathbf{x}_p^{i*}[-1](-\boldsymbol{\beta}^i[-1]+\boldsymbol{\beta}_0^i[-1])$. We have $(-\boldsymbol{\beta}^i[1]+\boldsymbol{\beta}_0^i[1])+\mathbf{x}_p^{i*}[-1](-\boldsymbol{\beta}^i[-1]+\boldsymbol{\beta}_0^i[-1]) = \lambda_0^i$. This equation should hold for any $\mathbf{x}_p^{i*}$, and the only way this could hold is when $\boldsymbol{\beta}^i[1] = \boldsymbol{\beta}_0^i[1] - \lambda_0^i$ and $\boldsymbol{\beta}^i[-1] = \boldsymbol{\beta}_0^i[-1]$. From this we can see that our equality based on conditional central symmetry only identifies the difference between *the true constant* $\boldsymbol{\beta}_0^i[1]$ *and the true center* λ_0^i. *Thus we could designate the true center* λ_0^i *to be zero as a (location) normalization. With the true center sitting at zero, the identifying restriction implies* $\boldsymbol{\beta}^i[1] = \boldsymbol{\beta}_0^i[1]$. *Given this fact, we focus on the baseline specification of the game with error term satisfying Assumption 3.3.3 with the knowledge in mind that if the center of the error term becomes some unknown constant, simply adding a constant in our payoff function will restore Assumption 3.3.3.*

Formal Definition of Identification and Regularity Conditions

In this section we provide the formal definition of identification as well as a set of regularity conditions that are sufficient to deliver identification based on the distance function we defined. We first introduce the parameter space as follows:

Assumption 3.3.4 (Parameter Space)**.** *The true parameter* $\boldsymbol{\beta}_0^i$ *is in the interior of parameter space* $\mathcal{B}^i$, *where* $\mathcal{B}^i$ *is a compact subset of* $\mathbb{R}^{L+1}$.

Following Torgovitsky (2017), Zhou (2019), and Lewbel et al. (2020), we introduce the following definition of point identification:

Definition 3.3.1 (Point Identification)**.** *For every* $\boldsymbol{\beta}^i \in \mathcal{B}^i$, *define the following set of covariates for player* i*:*

$$\mathcal{D}^i(\boldsymbol{\beta}^i) \equiv \{(z^i, x) \in \Omega_{(Z^i,X)} : (-z^i + 2\mathbf{x}_p^i\boldsymbol{\beta}^i, x) \in \Omega_{(Z^i,X)}, \text{ and } d^i(\boldsymbol{\beta}^i; z^i, x) \neq 0\}.$$

The true parameter $\boldsymbol{\beta}_0^i \in \mathcal{B}^i$ *is point identified when* $Pr((Z^i, X) \in \mathcal{D}^i(\boldsymbol{\beta}^i)) = 0$ *if and only if* $\boldsymbol{\beta}^i = \boldsymbol{\beta}_0^i$, *where* $\boldsymbol{\beta}^i \in \mathcal{B}^i$.

Next we introduce regularity conditions on the data generating process such that $\boldsymbol{\beta}_0^i$ is identified in the sense of Definition 3.3.1. First, notice that for any given x, the distance function is only defined when evaluation points z^i and $-z^i + 2\mathbf{x}_p^i\boldsymbol{\beta}^i$ both belong to $\Omega_{Z^i|x}$. Without further restrictions, it is possible that for some $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$, $Pr((Z^i, X) \in \mathcal{D}^i(\boldsymbol{\beta}^i)) = 0$ simply because we cannot find any (z^i, x) such that $(-z^i + 2\mathbf{x}_p^i\boldsymbol{\beta}^i, x) \in \Omega_{(Z^i,X)}$. Thus we first introduce the following set of regularity conditions, which includes the condition that for any given $\boldsymbol{\beta}^i$ we can always find "sufficient amount" of covariates with well-defined distance function.

Assumption 3.3.5 (Regularity Conditions on Covariates)**.** *(i). For any constant vector $c \in \mathbb{R}^{L+1}$, $Pr(\mathbf{X}_p^i c = 0) = 1$ if and only if $c = \mathbf{0}_{L+1}$. (ii). For every $x \in \Omega_x$, the conditional distribution of of Z^i given $X = x$ is absolutely continuous with respect to the Lebesgue measure and has bounded positive Radon-Nikodym density everywhere over $\Omega_{z^i|x}$. (iii). Given any $X = x$, there exists a subset $\mathcal{S}_{Z^i|x} \subset \Omega_{Z^i|x}$ with strictly positive measure, such that for any $z^i \in \mathcal{S}_{Z^i|x}$ and any $\boldsymbol{\beta}^i \in \mathcal{B}^i$, $-z^i + 2\mathbf{x}_p^i\boldsymbol{\beta}^i \in \Omega_{Z^i|x}$.*

In Assumption 3.3.5, (i) is a standard identification condition ruling out perfect colinearity among variables in $\mathbf{X}_p^i$ and it is also used in Chen et al. (2016) and Lewbel et al. (2020). (ii) guarantees continuity of Z^i, and the identification of $\mathbb{E}(D^i|z^i, x)$ as a function of z^i. (iii) is a sufficient condition that makes sure that given any $X = x$, there is enough variation in z^i to guarantee that the distance function is well-defined over the parameter space.

There is one situation in which the identifying restriction $\mathbb{E}(D^i|z^{i*}, x^*) = 1 - \mathbb{E}(D^i| - z^{i*} + 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i, x^*)$ (or in terms of the conditional cdf of ϵ^i: $F_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = 1 - F_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i))$ loses its identifying power, that is when $-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ is below the lower bound of $\Omega_{\epsilon^i|x}$ and $z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ is above the upper bound of $\Omega_{\epsilon^i|x}$ (or $-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ is above the upper bound of $\Omega_{\epsilon^i|x}$ and $z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ is below the lower bound of $\Omega_{\epsilon^i|x}$). For example, when $\Omega_{\epsilon^i|x}$ is bounded, for sufficiently large z^{i*}, $-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ could go below the lower bound of $\Omega_{\epsilon^i|x}$, and at the same time $z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ could go above the upper bound of $\Omega_{\epsilon^i|x}$. The identifying restriction could not distinguish all $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$ from $\boldsymbol{\beta}_0^i$ as long as $z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ remains above the upper bound of $\Omega_{\epsilon^i|x}$. Thus we impose the following restriction to make sure that the equality restriction is truly effective:

Assumption 3.3.6 (Regularity Conditions on the Error Terms)**.** *Given any $X = x$, $\mathcal{S}_{\epsilon^i|x} \equiv \{-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i : z^i \in \Omega_{Z^i|x}\} \cup \{z^i - 2\mathbf{x}_p^i\boldsymbol{\beta}^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i : z^i \in \Omega_{Z^i|x}$ and $\boldsymbol{\beta}^i \in \mathcal{B}^i\}$ is a subset of $int(\Omega_{\epsilon^i|x})$.*

Essentially, given any $X = x$, $\mathcal{S}_{\epsilon^i|x}$ is the set of (paired) points at which we evaluate $F_{\epsilon^i|x}(\cdot)$ as z^i and $\boldsymbol{\beta}^i$ vary. We also define $\mathcal{S}_{\epsilon^i|x}(\boldsymbol{\beta}^i) \equiv \{-z^i + \mathbf{x}_p^i \boldsymbol{\beta}_0^i : z^i \in \Omega_{Z^i|x}\} \cup \{z^i - 2\mathbf{x}_p^i \boldsymbol{\beta}^i + \mathbf{x}_p^i \boldsymbol{\beta}_0^i : z^i \in \Omega_{Z^i|x}\}$, which is the set of (paired) points at which we evaluate $F_{\epsilon^i|x}(\cdot)$ as z^i varies for some fixed $\boldsymbol{\beta}^i$. In applications with normal or logistic error distribution, we have $\Omega_{\epsilon^i|x} = \mathbb{R}$, and Ω_X, $\Omega_{Z^i|x}$ and $\mathcal{B}^i$ are bounded, then the requirement in Assumption 3.3.6 is always satisfied. The following theorem presents result on point identification:

Theorem 3.3.2 (Point Identification)**.** *Under Assumptions 3.2.1-3.2.2 and 3.3.1-3.3.6,* $\boldsymbol{\beta}_0^i \in \mathcal{B}^i$ *is point identified in the sense of Definition 3.3.1.*

Proof. See Appendix C.1. □

3.3.3 Discussion: the Choice of Distance Function

While our identification strategy is similar to Lewbel et al. (2020) in that we all rely on equality constraint on the true parameter implied by central symmetry, we differ in the choice of the distance function (or the exact form of equality constraint). Lewbel et al. (2020) constructs the distance function using pdf, while we constructs the distance function using cdf. In Lewbel et al. (2020), because of the multinomial nature of his problem, a pdf based distance function has simpler mathematical form compared to a cdf based distance function.[6] A pdf based distance function like Lewbel et al. (2020) could also be a choice in our problem, but like in their problem we will need an additional assumption of unique local center (Assumption 5. (b) in their paper).This is because when Z^i only has a relatively small support compared to the error term, the equality restriction is only created at a subset of evaluation points in the support of the error term, and this assumption guarantees that the equality constraint based on pdf is truly effective in restricting the parameter on such subset. The implication is that such subset cannot correspond to a flat part of the density, i.e., it cannot be a density that looks like a uniform density. For example, suppose the true error density is the uniform distribution on $[-1, 1]$, and suppose that $\mathcal{S}_{\epsilon^i|x} = [-0.5, 0.5]$, then the

[6]In the multinomial case, we need to work with the joint pdf or cdf of at least two random variables.

pdf based restriction which requires that $f_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = f_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i)$ puts no restriction on $\boldsymbol{\beta}^i$ at all, as $f_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = f_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = \frac{1}{2}$ for any $\boldsymbol{\beta}^i$ in the parameter space.[7] Such an identifying restriction is supposed to have low identifying power if the density is close to flat. In contrast, our identification strategy will still have identifying power even for flat densities, as in restriction $F_{\epsilon^i|x^*}(-z^{i*} + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i) = 1 - F_{\epsilon^i|x^*}(z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i)$, by strict monotonicity of cdf for the continuously distributed ϵ^i conditional on x, there is always effective restriction placed on $\boldsymbol{\beta}^i$. The only case in which our restriction places no real restriction is when $z^{i*} - 2\mathbf{x}_p^{i*}\boldsymbol{\beta}^i + \mathbf{x}_p^{i*}\boldsymbol{\beta}_0^i$ goes out of the conditional support of ϵ^i, but this is ruled out by the fact that the conditional distribution of ϵ^i is continuous, and that Assumption 3.3.6 holds.

3.4 Two-Step Estimation

In this section, we introduce a two-step estimation strategy following our identification strategy, in which the equilibrium beliefs and conditional choice probabilities are nonparametrically estimated in the first step (using the simple Nadaraya-Watson estimator), and the structural parameters are estimated in the second step. Trimming is necessary not only to guarantee the validity of evaluation points for the distance function, but also to guarantee fast enough convergence rate for all nonparametrically estimated functions.

3.4.1 The Trimming Function

Consistency and asymptotic normality of structural parameters require uniform consistency of the first-step estimators over a compact subset of the relevant supports. In addition, as we have discussed, for any generic $\boldsymbol{\beta}^i$, as $d^i(\boldsymbol{\beta}^i; z^i, x)$ is only defined when both (z^i, x) and $(-z^i + 2\boldsymbol{x}_p^i\boldsymbol{\beta}^i, x)$ belong to $\Omega_{(Zi,X)}$, and thus we only want to evaluate $d^i(\boldsymbol{\beta}^i; z^i, x)$ at such points. For this purpose, we use a "fixed trimming" strategy so that the distance function and the belief function are only evaluated on a compact subset of the corresponding support,

[7] Similarly, the restriction in Zhou (2019) which requires that integration on two reflected intervals of equal length being equal puts no restriction on the parameter.

and a small and fixed proportion of observations near the boundary is trimmed out. Formally, we introduce $\mathcal{X}$ as

$$\{(x_c, x_d) : x_c \in \text{a compact subset of } \Omega_{X_c|x_d} \text{ excluding the boundary range for any } x_d \in \Omega_{X_d}\},$$

and define $\mathcal{Z}^i_x$ as a compact subset of $\Omega_{Z^i|x}$ that excludes the boundary range. Let $\underline{z}_x$ and $\overline{z}_x$ be the lower and upper bound for $\Omega_{Z^i|x}$, then $\mathcal{Z}^i_x$ takes the form of $[\underline{z}_x + \delta, \overline{z}_x - \delta]$ for a small constant $\delta > 0$. Similar to Lewbel et al. (2020), we introduce a trimming function:

$$\tau^i_p(z^i, x) = \tau^i_{o,Z}(z^i, x)\tau^i_{p,cs,Z}(z^i, x)\tau^i_X(z^i, x),$$

where $\tau^i_{o,Z}(Z^i, X) \equiv \mathbb{1}(\underline{z}_X + \delta \leq Z^i \leq \overline{z}_X - \delta)$, $\tau^i_{p,cs,Z}(Z^i, X) \equiv \tau^i_{p,cs,Z,min}(Z^i, X)\tau^i_{p,cs,Z,max}(Z^i, X)$ and $\tau^i_X(z^i, x) \equiv \mathbb{1}(X \in \mathcal{X})$. We define $\tau^i_{p,cs,Z,min}(Z^i, X) \equiv \mathbb{1}(\underline{z}_X + \delta \leq -Z^i + 2\mathbf{X}^i_p\boldsymbol{\beta}^i_{min}(X))$ and $\tau^i_{p,cs,Z,min}(Z^i, X) \equiv \mathbb{1}(-Z^i + 2\mathbf{X}^i_p\boldsymbol{\beta}^i_{max}(X) \leq \overline{z}_X - \delta)$, where $\boldsymbol{\beta}^i_{min}(X)$ and $\boldsymbol{\beta}^i_{max}(X)$ are the values of $\boldsymbol{\beta}^i$ at which minimum and maximum for $\mathbf{X}^i_p\boldsymbol{\beta}^i$ are achieved, respectively. Notice that $\tau^i_{o,Z}(z^i, x)$ makes sure that the z^i- argument of the first term in the distance function is always evaluated on $\mathcal{Z}^i_x$, $\tau^i_{cs,Z}(z^i, x)$ makes sure that the z^i- argument of the second term in the distance function is always evaluated on $\mathcal{Z}^i_x$, and $\tau^i_X(z^i, x)$ makes sure that the first stage estimated belief and the $x-$ arguments of the distance function are only evaluated on $\mathcal{X}$. Like in Lewbel et al. (2020), the trimming function does not depend on any particular value of $\boldsymbol{\beta}^i$. In our problem, estimated belief from the first step will become part of the regressors. But as beliefs (true or estimated) always belong to $[0, 1]$, the belief times its coefficient achieves extreme value only when beliefs equals 0 or 1. For example, when β^i_2 is always negative, we can replace the beliefs inside the trimming function with 1 to get rid of estimated belief inside the trimming function:

$$\tau^i(z^i, x) = \tau^i_{o,Z}(z^i, x)\tau^i_{cs,Z}(z^i, x)\tau^i_X(z^i, x),$$

where $\tau^i_{o,Z}(Z^i, X) \equiv \mathbb{1}(\underline{z}_X + \delta \leq Z^i \leq \overline{z}_X - \delta)$, $\tau^i_{cs,Z}(Z^i, X) \equiv \tau^i_{cs,Z,min}(Z^i, X)\tau^i_{cs,Z,max}(Z^i, X)$ and $\tau^i_X(z^i, x) \equiv \mathbb{1}(X \in \mathcal{X})$. We define $\tau^i_{cs,Z,min}(Z^i, X) \equiv \mathbb{1}(\underline{z}_X + \delta \leq -Z^i + 2X^i\beta^i_{1,min}(X) +$

$\beta^i_{2,min}$) and $\tau^i_{cs,Z,max}(Z^i, X) \equiv \mathbb{1}(-Z^i + 2X^i\beta^i_{1,max}(X) + \beta^i_{2,max} \leq \overline{z}_X - \delta)$.

Let $\mathcal{S}^{i,NT}_{(Z^i,X)}$ be the set of covariates that survives the trimming, and by construction $\mathcal{S}^{i,NT}_{(Z^i,X)} = \{(z^i, x) : z^i \in \mathcal{Z}^i_x, -z^i + 2x^i_p\boldsymbol{\beta}^i \in \mathcal{Z}^i_x$ for any $\boldsymbol{\beta}^i$ and any belief, and $x \in \mathcal{X}\}$. For any $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}^i_0$, let $\mathscr{X}(\boldsymbol{\beta}^i)$ be the set of x such that $\mathbf{x}^i_p(\boldsymbol{\beta}^i - \boldsymbol{\beta}^i_0) \neq \mathbf{0}$. By Assumption 3.3.5 (i), $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}^i_0$ implies $\Pr(X \in \mathscr{X}(\boldsymbol{\beta}^i)) > 0$.[8] Thus, as long as the parameter deviates from the true value, it's always possible to find x such that $\mathbf{x}^i_p(\boldsymbol{\beta}^i - \boldsymbol{\beta}^i_0) \neq \mathbf{0}$. Define $\mathscr{S}(\boldsymbol{\beta}^i) \equiv \{(z^i, x) : z^i \in \mathcal{S}_{Z^i|x},\ x \in \mathscr{X}(\boldsymbol{\beta}^i)\}$, and the trimming function must make sure that such a set is not entirely trimmed out. Thus we have the following assumption on the trimming function:

Assumption 3.4.1 (Trimming Function)**.** *For any $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}^i_0$, $\mathscr{S}(\boldsymbol{\beta}^i) \bigcap \mathcal{S}^{i,NT}_{(Z^i,X)}$ has strictly positive measure, and $\tau^i(z^i, x) > 0$ on $\mathcal{S}^{i,NT}_{(Z^i,X)}$ and equals zero elsewhere.*

Example 3.4.1 (Trimming Function)**.** *Continue with Example 3.1, suppose the parameter space is $\boldsymbol{\mathcal{B}} = [\underline{\beta}^i_1, \overline{\beta}^i_1] \times [\underline{\beta}^i_2, \overline{\beta}^i_2]$ with $\underline{\beta}^i_1 < 0, \underline{\beta}^i_2 < 0, \overline{\beta}^i_1 > 0$, and $\overline{\beta}^i_2 > 0$. Let $c = 10 - \delta$ for some small $\delta > 0$. One possible trimming function for this example could be*

$$\tau^i(z^i, x) = \tau^i_o(z^i, x)\tau^i_{cs,min}(-z^i + 2\boldsymbol{x}^i_p\boldsymbol{\beta}^i_{min}, x)\tau^i_{cs,max}(-z^i + 2\boldsymbol{x}^i_p\boldsymbol{\beta}^i_{max}, x), \tag{3.2}$$

where

$$\tau^i_o(z^i, x) = \mathbb{1}(|z^i| \leq c),$$

$$\begin{aligned} & \tau^i_{cs,max}(-z^i + 2\boldsymbol{x}^i_p\boldsymbol{\beta}^i_{max}, x) \\ = \ & \mathbb{1}(|-z^i - 2\underline{\beta}^i_1 + 2\overline{\beta}^i_2| \leq c)\mathbb{1}(x = -1) + \mathbb{1}(|-z^i + 2\overline{\beta}^i_1 + 2\overline{\beta}^i_2| \leq c)\mathbb{1}(x = 1), \end{aligned}$$

$$\begin{aligned} & \tau^i_{cs,min}(-z^i + 2\boldsymbol{x}^i_p\boldsymbol{\beta}^i_{min}, x) \\ = \ & \mathbb{1}(|-z^i - 2\overline{\beta}^i_1 + 2\underline{\beta}^i_2| \leq c)\mathbb{1}(x = -1) + \mathbb{1}(|-z^i + 2\underline{\beta}^i_1 + 2\underline{\beta}^i_2| \leq c)\mathbb{1}(x = 1). \end{aligned}$$

[8] Note that this doesn't rule out the possibility that for some x, the index might be zero.

By construction, as long as $(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i,X)}$, it is guaranteed that $-z^i + 2[x, p]\boldsymbol{\beta}^i \in \mathcal{Z}^i_x$ uniformly for any $\boldsymbol{\beta}^i \in \mathcal{B}^i$ and any $p \in [0, 1]$.

3.4.2 The Population Objective Function

Combining the distance function and the trimming function, we define the population objective function as

$$Q^i(\boldsymbol{\beta}^i) = \frac{1}{2}\,\mathbb{E}[(\tau(Z^i, X)d^i(\boldsymbol{\beta}^i; Z^i, X))]^2.$$

The following theorem presents identification result using this objective function, i.e., the true parameter is the unique minimizer of the population objective function.

Theorem 3.4.2. *Under Assumptions 3.2.1-3.2.2, 3.3.1-3.3.6 and 3.4.1,* $Q^i(\boldsymbol{\beta}^i) \geq 0$ *for any* $\boldsymbol{\beta}^i \in \mathcal{B}^i$*, and* $Q^i(\boldsymbol{\beta}^i) = 0$ *if and only if* $\boldsymbol{\beta}^i = \boldsymbol{\beta}^i_0$.

Proof. See Appendix C.2. □

The result in Theorem 3.4.2 is used later in the proof of consistency.

3.4.3 A Two-Step Minimum Distance Estimator

Step-1. Estimate equilibrium beliefs and construct generated regressor.

When all variables in X are discrete, the belief could be estimated using a frequency estimator:

$$\widehat{P}^{-i}(x) = \frac{\sum_{n=1}^N \mathbb{1}(D_n^{-i} = 1, X_n = x)}{\sum_{n=1}^N \mathbb{1}(X_n = x)}.$$

In general, belief could be estimated with

$$\widehat{P}^{-i}(x) = \frac{\sum_{n=1}^N D_n^{-i}\boldsymbol{K}_h(X_{n,c} - x_c)\mathbb{1}(X_{n,d} = x_d)}{\sum_{n=1}^N \boldsymbol{K}_h(X_{n,c} - x_c)\mathbb{1}(X_{n,d} = x_d)},$$

where $\boldsymbol{K}_h(X_{n,c} - x_c)$ is the product kernel $\prod_{l=1}^{L_c-1} K_h(X_{n,c}[l] - x_c[l])$, $K_h(X_{n,c}[l] - x_c[l]) = \frac{1}{h}k(\frac{X_{n,c}[l]-x_c[l]}{h})$, and $k(\cdot)$ is the kernel function. Plug in the estimated belief we get the generated regressor:

$$\widehat{\mathbf{X}}^i_p = [x\ \widehat{P}^{-i}(x)].$$

Step-2. Estimate structural parameters.

First, we estimate the distance function using a leave-one-out estimator:

$$\begin{aligned}
\widehat{d}^i_{-n}(\boldsymbol{\beta}^i; Z^i_n, X_n) &= \widehat{P}^i_{-n}(Z^i_n, X_n) - (1 - \widehat{P}^i_{-n}(-Z^i_n + 2\widehat{\mathbf{X}}^i_{pn}\boldsymbol{\beta}^i, X_n)) \\
&\equiv \widehat{\varphi}^i_{o,-n}(Z_n, X_n) - \widehat{\varphi}^i_{cs,-n}(Z_n, X_n, \boldsymbol{\beta}^i)
\end{aligned}$$

where for generic (z^i, x),

$$\widehat{P}^i_{-n}(z^i, x) = \frac{\sum_{m=1, m\neq n}^N D^i_m K_h(Z^i_m - z^i)\boldsymbol{K}_h(X_{m,c} - x_c)\mathbb{1}(X_{m,d} = x_d)}{\sum_{m=1, m\neq n}^N K_h(Z^i_m - z^i)\boldsymbol{K}_h(X_{m,c} - x_c)\mathbb{1}(X_{m,d} = x_d)},$$

and $K_h(Z^i_m - z^i) = \frac{1}{h}k(\frac{Z^i_m - z^i}{h})$ with $k(\cdot)$ being the kernel function. Plug in the estimated distance function and apply the trimming function, the sample objective function is constructed as

$$Q^i_N(\boldsymbol{\beta}^i) = \frac{1}{2N}\sum_{n=1}^N [\tau^i(Z^i_n, X_n)\widehat{d}^i_{-n}(\boldsymbol{\beta}^i; Z^i_n, X_n)]^2.$$

Thus our semiparametric two-step minimum distance (MD) estimator is defined as

$$\widehat{\boldsymbol{\beta}}^i = \underset{\boldsymbol{\beta}\in\mathcal{B}^i}{\operatorname{argmin}} Q^i_N(\boldsymbol{\beta}^i). \tag{3.3}$$

Note that this estimator is closely related to the one introduced in Zhou (2019) and Lewbel et al. (2020) although we use a different distance function and we also need to estimate equilibrium beliefs in the first step. Similar semiparametric minimum distance estimator is also introduced in other problems that rely on equality restriction to identify and estimate parameters. For example, Torgovitsky (2017) uses a semiparametric minimum distance estimator to estimate nonseparable instrumental variables models.

3.5 Asymptotic Theory

In this section, we provide results on consistency and asymptotic normality of our proposed estimator. Define $g^i(z^i, x) = P^i(Z^i_n, X_n)f^i(z^i, x)$, and $f^i(z^i, x) = f^i(z^i, x_c|x_d)P_{x_d}$.

3.5.1 The Roadmap of the Proof

Define

$$q_N^i(\boldsymbol{\beta}^i) = \nabla_{\boldsymbol{\beta}^i} Q_N^i(\boldsymbol{\beta}^i) \text{ and } H_N^i(\boldsymbol{\beta}^i) = \nabla_{\boldsymbol{\beta}^i} \nabla_{\boldsymbol{\beta}^{i\top}} Q_N^i(\boldsymbol{\beta}^i).$$

Smoothness of the sample objective function implies that first-order condition is satisfied at $\widehat{\boldsymbol{\beta}}^i$: $q_N^i(\widehat{\boldsymbol{\beta}}^i) = \mathbf{0}_{L+1}$. The proof of consistency relies on showing that our sample objective function $Q_N^i(\boldsymbol{\beta}^i)$ satisfies the conditions in Theorem 2.1 of Newey and McFadden (1994). Given consistency, the proof of asymptotic normality follows the following steps.

Firstly, application of mean value theorem at for $q_N^i(\widehat{\boldsymbol{\beta}}^i)$ at $\boldsymbol{\beta}_0^i$ gives

$$\mathbf{0}_{L+1} = q_N^i(\widehat{\boldsymbol{\beta}}^i) = q_N^i(\boldsymbol{\beta}_0^i) + H_N^i(\widetilde{\boldsymbol{\beta}}^i)(\widehat{\boldsymbol{\beta}}^i - \boldsymbol{\beta}_0^i),$$

where $\widetilde{\boldsymbol{\beta}}^i$ is a point between $\boldsymbol{\beta}_0^i$ and $\widehat{\boldsymbol{\beta}}^i$ in parameter space $\mathcal{B}^i$. Then solve for $\widehat{\boldsymbol{\beta}}^i - \boldsymbol{\beta}_0^i$:

$$\widehat{\boldsymbol{\beta}}^i - \boldsymbol{\beta}_0^i = -[H_N^i(\widetilde{\boldsymbol{\beta}}^i)]^{-1} q_N^i(\boldsymbol{\beta}_0^i).$$

The proof of asymptotic normality relies on showing $H_N^i(\widetilde{\boldsymbol{\beta}}^i) \xrightarrow{p} H^i$, where

$$H^i \equiv \mathbb{E}[\tau^i(Z_n^i, X_n)^2] \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}^i; Z^i, X)) \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}^i; Z^i, X))^\top],$$

and showing $\sqrt{N} q_N^i(\boldsymbol{\beta}_0^i) \xrightarrow{d} N(\mathbf{0}_{L+1}, \Omega^i)$.

3.5.2 Consistency

In this section, we lay out assumptions needed for consistency of our estimator, and we present the theorem on consistency. First, we need to impose smoothness assumptions on the unknown functions involved:

Assumption 3.5.1. *For $i = 1, 2$, $f(z^i|x)$ and $\mathbb{E}(D^i|z^i, x)$ are s ($s \geq 2$) times continuously differentiable in (z^i, x_c) for all $(z^i, x_c) \in \Omega_{(Z^i, X_c)|x_d}$ and have bounded derivatives;* $\mathrm{card}(\Omega_{X_d}) < \infty$; *$P^i(x)$ is s times continuously differentiable in x_c for all $x_c \in \Omega_{X_c|x_d}$ and has bounded derivatives; $f^i(z^i, x)$ is bounded away from zero on $\mathcal{S}^{i,NT}_{(Z^i,X)}$, and $\frac{\partial \mathbb{E}(D^i|-z^i+2\mathbf{x}_p^i\boldsymbol{\beta}^i, x)}{\partial z^i}$ is uniformly bounded on $\mathcal{S}^{i,NT}_{(Z^i,X)}$.*

Note that Assumption 3.5.1 implies that $g^i(z^i, x)$ is also s times continuously differentiable in (z^i, x_c) for all $(z^i, x_c) \in \Omega_{(Z^i,X_c)|x_d}$ and has bounded derivatives. Assumptions on the kernel function are listed as follows:

Assumption 3.5.2. *The kernel function $k(\cdot)$ is a $s-th$ order kernel satisfying the following conditions:*
(i). $k(u) = k(-u)$ for any u in its support and $\int k(u)du = 1$;
(ii). $\int |u|^j k(u)du < \infty$ for $0 \leq j \leq s$;
(iii). $\int u^j k(u)du = 0$ for $0 < j < s$ and $\int u^j k(u)du \neq 0$ if $j = s$;
(iv). $k(u)$ has bounded support;
(v). $\sup_u |k(u)|^2 < \infty$.

Assumptions on the bandwidth are listed as follows:

Assumption 3.5.3 (Bandwidth)**.** *The bandwidth h satisfies:*
(i). $h \to 0$,
(ii). $\sqrt{N} \times \frac{logN}{Nh^{L_c+3}} \to 0$, and $\sqrt{N} \times h^s \to 0$,
as $N \to \infty$.

Remark 3.5.1. *The listed bandwidth conditions imply the following conditions:*
(i). $\sqrt{N} \times \frac{logN}{Nh^{L_c}} \to 0$, which is required for uniform consistency of estimated belief at faster than $N^{-1/4}$ rate;
(ii). $\sqrt{N} \times \frac{logN}{Nh^{L_c+1}} \to 0$, which is required for uniform consistency of estimated distance function at faster than $N^{-1/4}$ rate;
(iii). $\sqrt{N} \times h^{2s} \to 0$, which is a small bias condition required for uniform consistency of estimated distance function and estimated belief at faster than $N^{-1/4}$ rate.
(iv). $N \times h^{L_c+1} \to \infty$ and $N \times h^{L_c} \to \infty$, which are part of the conditions required for Hoeffding decomposition in the proof of asymptotic normality;

The following theorem states the consistency result:

Theorem 3.5.1 (Consistency)**.** *Under Assumptions 3.2.1-3.2.2, 3.3.1-3.3.6, 3.4.1 and 3.5.1-3.5.3, the semiparametric two-step MD estimator* $\widehat{\boldsymbol{\beta}}^i$ *defined in equation (3.3) converges in probability to* $\boldsymbol{\beta}_0^i$.

Proof. See Appendix C.2. □

3.5.3 Asymptotic Normality

In this section, we lay out additional assumptions required for asymptotic normality and present the theorem on asymptotic normality. For asymptotic normality to hold, an additional assumption is imposed on the population hessian matrix:

Assumption 3.5.4 (Hessian)**.** *For* $i = 1, 2$, *the population hessian matrix* H^i *is nonsingular.*

Theorem 3.5.2. *Assumptions 3.2.1-3.2.2, 3.3.1-3.3.6, 3.4.1 and 3.5.1-3.5.4, for* $i = 1, 2$, *it holds that:*
(i).(Asymptotic Linearity)
The two-step MD estimator has the following asymptotic linear representation:

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^i - \boldsymbol{\beta}_0^i) = (H^i)^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} t_n^i + o_p(1),$$

with

$$\begin{aligned} t_n^i &= \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)(D_n^i - P^i(Z_n^i, X_n)) \\ &+ \frac{\xi^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n, Z_n^i)f^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)}(D_n^i - P^i(Z_n^i, X_n)) \\ &+ 2\beta_{20}^i C^i(X_n)(D_n^{-i} - P^{-i}(X_n)), \end{aligned}$$

where $\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)$ *and* $C^i(X_n)$ *are defined in Appendix C.4.*

(ii).(Asymptotic Normality)

The two-step MD estimator is asymptotically normal:

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^i - \boldsymbol{\beta}_0^i) \xrightarrow{d} N\left(\mathbf{0}_{L+1}, \left(H^i\right)^{-1} \Omega^i \left(H^i\right)^{-1}\right),$$

where $\Omega^i = \mathbb{E}(t_n^i t_n^{i\top})$.

Proof. See Appendix C.4. □

Notice that each of the three components in t_n^i has some intuitive meaning. The first term corresponds to contribution of the first term in the estimated distance function, the second term corresponds to contribution of the second term in the estimated distance function and the third term corresponds to the additional contribution of estimated belief. In addition, we have further interpretations for the second and third term in the linear representation:
(i). The contribution to asymptotic variance of the second term will be large if the ratio $\frac{f^i(-Z_n^i+2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)}$ is large. Intuitively, the numerator in this ratio is measuring the value of density at which we are evaluating the second term in the distance function, and the denominator in this ratio is measuring the value of density at which we are estimating the second term in the distance function. A large ratio means we are frequently evaluating the second term in the distance function at points around which there aren't enough observations for estimation (and thus is poorly estimated), which results in large asymptotic variance.
(ii). If the true strategic interaction coefficient β_{20}^i is zero, then estimated belief doesn't contribute to asymptotic variance.

Notice that the influence function takes a rather complicated form, which might make direct plug-in implementation for inference cumbersome, and bootstrap inference might be preferred. We do not attempt to introduce and develop the asymptotic theory for a bootstrap procedure here, and leave it for future research.

3.6 Monte Carlo Simulations

In this section, we conduct a small Monte Carlo study of the performance of our estimator compared to that of two-step pseudo maximum likelihood (PML) estimation of a probit model under three designs: a homoskedastic probit game and a heteroskedastic probit game and a uniform game. The results of the simulations are presented in Table 3.4, Table 3.5 and Table 3.6. The estimator in all tables are obtained using grid search. We use the triweight kernel as in Lewbel and Tang (2015) and set the bandwidth according to $h = 1.7\sigma_{Z^i} N^{-1/7}$, where σ_{Z^i} is the standard deviation for Z^i.[9] Bias and MSE are both calculated using 100 Monte Carlo repetitions. We focus on estimating the structural parameters for player 1.

DGP1. The DGP for the homoskedastic game is the one introdued in Example 3.1.

Table 3.4: Homoskedastic Probit Game: (Semiparametric) Two-step MD vs Two-step PML

		Two-step MD		Two-step PML	
		Bias	MSE	Bias	MSE
N=2000	β_1^i	-0.005	0.014	-0.023	0.013
	β_2^i	-0.005	0.068	0.022	0.045
N=3000	β_1^i	0.01	0.009	0.002	0.009
	β_2^i	0.036	0.035	-0.005	0.037

DGP2. Everthing else in the heteroskedastic probit game is the same as **DGP1** except that now X follows a discrete uniform distribution with support $\{-1, -0.5, 0.5, 1\}$, and that the conditional variance for ϵ^i equals 0.5 when $X = -1$, equals 1 when $X = -0.5$, equals 24 when $X = 0.5$, and equals 25 when $X = 1$.

[9]We also run the simulation with fourth order Epanechnikov kernel and get very similar results. Second order kernel is also used in the simulation of Lewbel and Tang (2015) and Lewbel et al. (2020), see Powell et al. (1989) for a discussion on the advantage of second order kernel over higher order kernel in finite samples.

Table 3.5: Heteroskedastic Probit Game: (Semiparametric) Two-step MD vs Two-step PML

		Two-step MD		Two-step PML	
		Bias	MSE	Bias	MSE
N=2000	β_1^i	0.02	0.160	-0.032	0.086
	β_2^i	0.018	0.467	**-0.447**	0.426
N=3000	β_1^i	-0.01	0.108	-0.015	0.091
	β_2^i	-0.012	0.346	**-0.495**	0.415

DGP3. In the uniform game, the distribution of ϵ^i is $Unif[-20, 20]$, X follows a discrete uniform distribution with support $\{-1, -0.5, 0.5, 1\}$, and everything else is the same as in **DGP1**.

Table 3.6: Uniform Game: (Semiparametric) Two-step MD vs Two-step PML

		Two-step MD		Two-step PML	
		Bias	MSE	Bias	MSE
N=2000	β_1^i	-0.047	0.286	**0.235**	0.552
	β_2^i	-0.077	0.480	**-0.32**	1.645
N=3000	β_1^i	-0.02	0.212	**0.285**	0.487
	β_2^i	0.01	0.285	**-0.225**	0.818

Here in both **DGP1** and **DGP2**, Z^i has a bounded support, while ϵ^i has support $\mathbb{R}$; in **DGP3**, the support of Z^i is strictly smaller than the support of ϵ^i. All DGPs are cases in which parameters are not identified using Lewbel and Tang (2015).

Table 3.4 and Table 3.5 illustrate the robustness of our estimator with respect to the presence of heteroskedasticity. Comparison of MSE column for the semiparametric two-step MD estimator in Table 3.4 and Table 3.5 across sample sizes shows evidence of convergence to true value as measured by the small bias and shrink in MSE when sample size increases. Comparison of the MSE and bias of the semiparametric two-step MD estimator and two-step PML of probit in Table 3.4 shows that when the homoskedastic probit game is correctly specified, PML is comparable to the semiparametric two-step MD estimator and is slightly more precise as measured by MSE. Comparison of the MSE and bias of the Semiparametric two-step MD estimator and the two-step PML of probit in Table 3.5 shows that when the probit model is incorrectly specified (when there is conditional heteroskedasticity in the error), two-step PML of probit is obviously worse than the semiparametric two-step MD estimator. While the magnitude of MSE for two-step PML is more or less comparable to that of the semiparametric two-step MD, the magnitude of the bias for the strategic coefficient is more than 40 times larger than that of a two-step MD for both sample sizes. This is evidence that the two-step PML in this case is not converging to the true value, and suggests that estimates not properly taking care of heteroskedasticity are not reliable.

Comparing Table 3.4 and Table 3.6 shows the robustness of our estimator with respect to misspecification of the distribution family of ϵ^i. In particular, when the true distribution of the error term is uniform, but we fit a probit model, the absolute value of the bias of the two-step PML is much larger than that of the two-step MD for both parameters at different sample sizes. The MSE of the two-step PML is also at least 2 times larger for both parameters at different sample sizes.

3.7 Extensions

In this section, we discuss three extensions of our method. In the first extension, we study a game in which (Z^1, Z^2) are also common knowledge. In the second extension, we allow for multiple equilibrium beliefs rationalizing our data. In the third extension, we allow for symmetrically distributed random coefficients.

3.7.1 (Z^1, Z^2) Are Also Common Knowledge

In the baseline game, we study the case in which the econometrician observes (Z^1, Z^2, X) but the two players only commonly observe X, a situation that is mentioned in Aradillas-Lopez (2010). In this section, we extend our method to the case in which (Z^1, Z^2) are also common knowledge. Similar to Lewbel and Tang (2015), we also allow the strategic coefficient to depend on X. In this case, each player in the game has access to exactly the same information about its opponent as the econometrician, which is the information setting in Lewbel and Tang (2015). In this case, the equilibrium of the game is defined as follows:

Definition 3.7.1 (Equilibrium: Common Knowledge on (Z^1, Z^2, X))**.** *Given* $(Z^1, Z^2, X) = (z^1, z^2, x)$*, a BNE of the game is defined as the fixed point* $(P^{1*}(z^1, z^2, x), P^{2*}(z^1, z^2, x))$ *of the following system in* $(P^1(z^1, z^2, x), P^2(z^1, z^2, x))$*:* [10]

$$P^1(z^1, z^2, x) = F_{\epsilon^1|(z^1,z^2,x)}\left(-z^1 + x\beta_1^1 + \beta_2^1(x)P^2(z^1, z^2, x)\right),$$

$$P^2(z^1, z^2, x) = F_{\epsilon^2|(z^1,z^2,x)}\left(-z^2 + x\beta_1^2 + \beta_2^2(x)P^1(z^1, z^2, x)\right).$$

To extend our identification procedure to this situation, a slightly different independence assumption is required:

Assumption 3.7.1 (Independence Relationships)**.** *The following independence conditions hold: (i).* $\epsilon^1 \perp \epsilon^2|(Z^1, Z^2, X)$*; (ii). For* $i = 1, 2$*,* $(Z^1, Z^2) \perp \epsilon^i|X$*.*

Assumption 3.7.1 is a combination of Assumption A1 and Assumption A3 (ii) in Lewbel and Tang (2015). For $i = 1, 2$, let β_{10}^i and $\beta_{20}^i(x)$ be the true values of the parameters. We focus on player 1 for illustrative convenience. Under Assumption 3.7.1, it holds that

$$P^1(z^1, z^2, x) = F_{\epsilon^1|x}\left(-z^1 + x\beta_{10}^1 + \beta_{20}^1(x)P^2(z^1, z^2, x)\right). \tag{3.4}$$

Thus $P^1(z^1, z^2, x)$ identifies $F_{\epsilon^1|x}(\cdot)$ evaluated at $-z^1 + x\beta_{10}^1 + \beta_{20}^1(x)P^2(z^1, z^2, x)$. However, our previous identification argument which relies on index linearity no longer works here, as for some generic β_1^1 and $\beta_2^1(x)$, if we evaluate the first argument of $P^1(z^1, z^2, x)$ at $-z^1 + 2x\beta_1^1 + 2\beta_2^1(x)P^2(z^1, z^2, x)$, we get

$$\begin{aligned} & P^1(-z^1 + 2x\beta_1^1 + 2\beta_2^1(x)P^2(z^1, z^2, x), z^2, x) \\ = \; & F_{\epsilon^1|x}(z^1 - 2x\beta_1^1 - 2\beta_2^1(x)P^2(z^1, z^2, x) + x\beta_{10}^1 \\ & +\beta_{20}^1(x)P^2(-z^1 + 2x\beta_1^1 + 2\beta_2^1(x)P^2(z^1, z^2, x), z^2, x)), \end{aligned}$$

[10] As optimal strategy is invariant to scale normalization, here we define the equilibrium using the normalized parameters and error term.

where inside $F_{\epsilon^1|x}(\cdot)$ we no longer have a linear index because z^1 also appears in the equilibrium belief. Thus here we use an extended identification strategy that apply the trick in Lewbel and Tang (2015) and identify $\beta_2^1(x)$ in an initial step. Partial derivative with respect to z^1 on both sides of the Equation (3.4) gives

$$\frac{\partial P^1(z^1,z^2,x)}{\partial z^1} = f_{\epsilon^1|x}\left(-z^1 + x\beta_1^1 + \beta_2^1(x)P^2(z^1,z^2,x)\right)(-1+\beta_2^1(x)\frac{\partial P^2(z^1,z^2,x)}{\partial z^1}).$$

Similarly, partial derivative with respect to z^2 gives

$$\frac{\partial P^1(z^1,z^2,x)}{\partial z^2} = f_{\epsilon^1|x}\left(-z^1 + x\beta_1^1 + \beta_2^1(x)P^2(z^1,z^2,x)\right)\beta_2^1(x)\frac{\partial P^2(z^1,z^2,x)}{\partial z^2}.$$

Eliminate $f_{\epsilon^1|x}(-z^1 + x\beta_1^1 + \beta_2^1 P^2(z^1,z^2,x))$ and solve for $\beta_2^1(x)$ gives

$$\beta_2^1(x) = \frac{-\frac{\partial P^1(z^1,z^2,x)}{\partial z^2}}{\frac{\partial P^1(z^1,z^2,x)}{\partial z^1}\frac{\partial P^2(z^1,z^2,x)}{\partial z^2} - \frac{\partial P^1(z^1,z^2,x)}{\partial z^2}\frac{\partial P^2(z^1,z^2,x)}{\partial z^1}}. \tag{3.5}$$

Here overidentification for many different (z^1,z^2) implies that we could do an additional average over (Z^1,Z^2) given $X=x$ and obtain:

$$\beta_2^1(x) = \mathbb{E}[\frac{-\frac{\partial P^1(Z^1,Z^2,x)}{\partial z^2}}{\frac{\partial P^1(Z^1,Z^2,x)}{\partial z^1}\frac{\partial P^2(Z^1,Z^2,x)}{\partial z^2} - \frac{\partial P^1(Z^1,Z^2,x)}{\partial z^2}\frac{\partial P^2(Z^1,Z^2,x)}{\partial z^1}}|x]. \tag{3.6}$$

Identification of other structural parameters relies on the following assumption:

Assumption 3.7.2 (Degenerate Subset for Strategic Effect)**.** *There exists $\mathcal{E} \subset \Omega_X$ with strictly positive measure such that $\beta_2^1(x) = 0$ for $x \in \mathcal{E}$.*[11]

Using Equation (3.6), $\mathcal{E}$ is identified from data. The identification of the other structural parameters will be done conditional on $\mathcal{E}$. On this set, we have that

$$P^1(z^1,z^2,x) = F_{\epsilon^1|x}\left(-z^1 + x\beta_{10}^1\right),$$

[11]Note that for pure identification purpose X could be continuous, but if we want $\beta_2^1(x)$ to still have root-N rate for any x using an average of the corresponding kernel-estimated CCPs on this x, then X has to be discrete.

and that

$$P^1(-z^1 + 2x\beta_1^1, z^2, x) = F_{\epsilon^1|x}\left(z^1 - 2x\beta_1^1 + x\beta_{10}^1\right).$$

Thus by Assumption 3.3.3, it holds that

$$P^1(z^1, z^2, x) = 1 - P^1(-z^1 + 2x\beta_{10}^1, z^2, x).$$

As this equality in general do not hold for $\beta_1^1 \neq \beta_{10}^1$, we can define an extended distance function for $x \in \mathcal{E}$:

$$d^1(\boldsymbol{\beta}^i; z^1, z^2, x) = P^1(z^1, z^2, x) - 1 + P^1(-z^1 + 2x\beta_1^1, z^2, x), \tag{3.7}$$

which could be used for identification and used as the basis for estimation similar to what we did for the baseline model.[12] When X is discrete with a finite number of points in the support, $\mathcal{E}$ could be estimated using a consistent testing procedure as follows: we test the hypothesis $H_0 : \beta_2^1(x) = 0$ point by point for $x \in \Omega_X$ using test statistic $T_n = \sqrt{N}|\frac{1}{N}\sum_{n=1}^N \frac{\partial \widehat{P}^1(Z_n^1, Z_n^2, x)}{\partial z^2}|$ (or using the estimated $\beta_2^1(x)$), where $\widehat{P}^1(Z_n^1, Z_n^2, x)$ is a kernel estimator for $P^1(Z_n^1, Z_n^2, x)$. We use critical value CV_N that goes to infinity at slower than root$-N$ rate, then our test will reject with probability 1 if H_0 is not true and not reject with probability 1 if H_0 is true. Let $\widehat{\mathcal{E}}$ be the collection of points such that the test does not reject. Then it holds that $\Pr(\widehat{\mathcal{E}} = \mathcal{E}) \to 1$ as $N \to \infty$. Using only data with $X \in \widehat{\mathcal{E}}$, we have could estimate β_1^1 using the distance function in Equation (3.7) and the resulting estimator is root-N consistent. We do not repeat the regularity conditions, the estimation and asymptotic theory here as they are trivial extensions of the baseline case given knowledge of $\widehat{\mathcal{E}}$.

Remark 3.7.1. *Although Assumption 3.7.2 could also be used in Lewbel and Tang (2015) to identify the rest of the structural parameters, they still rely on additional large support assumption for identification and thick tail condition for root-N normality.*

[12] As here we do not have continuous regressors, identification requires that card($\mathcal{E}$) should be larger than or equal to the number of parameters.

3.7.2 Identification with Multiple Equilibria

Multiple equilibria in discrete games of incomplete information has been studied in Xiao (2018), Aguirregabiria and Mira (2018), and Fan et al. (2020), all of which assume a known error distribution. In this section, we make a straightforward extension that allows us to identify structural parameters even when data is rationalized by multiple equilibria without knowing the error distribution. Let $\omega \in \Omega_{\omega|x}$ be the index for equilibria given x. Let $\mathbf{x}_{p,\omega}$ be the regressor vector with belief from equilibrium ω plugged in. Let I be the number of players. Divide the I players into three groups, such that the third group has exactly one player for odd I and two players for even I, and each of the first two groups has $\widetilde{I}$ players. Thus, $I = 2\widetilde{I}+1$ when I is odd; and $I = 2\widetilde{I}+2$ when I is even. Player group i is denoted as g_i for $i = 1, 2, 3$. By definition, $\bigcup_{i=1}^{3} g_i = \{1, ..., I\}$. For each group, we create a group action variable, denoted by D^{g_1}, D^{g_2}, and D^{g_3}. We have $D^{g_1}, D^{g_2} \in \left\{0, ..., 2^{\widetilde{I}} - 1\right\}$, and $D^{g_3} \in \mathcal{D}$ if there is one player in group 3 and $D^{g_3} \in \{0, ..., 2^2 - 1\}$ if there are two players in group 3. We introduce the following notation to denote the matrix composed of CCPs for group action D^{g_i} for $i = 1, 2$ on each latent state:

$$\mathbf{P}_{g_i \mathbf{z}} = \left(\Pr\left(D^{g_i} = j \mid x, \omega\right)\right)_{j=0,\omega=1}^{2^{\widetilde{I}}-1, \left|\Omega_{\omega|x}\right|}.$$

Assumption 3.7.3 (Identification of Equilibrium Beliefs Under Multiple Equilibria)**.** *(i) X is a discrete random vector with a finite number of points in its support; (ii) $I \geq 3$ and $(J+1)^{\widetilde{I}} \geq \left|\Omega_{\omega|x}\right|$; (iii) For each x, there exists a partition $(D^{g_1}, D^{g_2}, D^{g_3})$ of joint actions $\left(D^1, ..., D^I\right)$ such that $\mathbf{P}_x^{g_1}$ and $\mathbf{P}_x^{g_2}$ both have full column rank; (iii) $2^{\widetilde{I}} > \left|\Omega_{\omega|x}\right|$ for any x.*

Assumption 3.7.3 guarantees the identification of the number of equilibria, and the equilibrium beliefs (up to a label swapping). Given the identification of equilibrium beliefs, we layout a set of sufficient conditions such that structural parameters are identified using identified equilibrium beliefs from all equilibria.

Definition 3.7.2 (Point Identification)**.** *For every $\boldsymbol{\beta}^i \in \mathcal{B}^i$, define the following set of co-*

variates for player i:

$$\mathcal{D}^i(\boldsymbol{\beta}^i) \equiv \{(z^i, x) \in \Omega_{(Z^i,X)} : (-z^i + 2\boldsymbol{x}_{p,\omega}^{i\top}\boldsymbol{\beta}^i, x) \in \Omega_{(Z^i,X)}, \forall \omega \in \Omega_{\omega|x} \text{ and } d^i(\boldsymbol{\beta}^i; z^i, x) \neq 0\}.$$

The true parameter $\boldsymbol{\beta}_0^i \in \mathcal{B}^i$ *is point identified when* $Pr((Z^i, X) \in \mathcal{D}^i(\boldsymbol{\beta}^i)) = 0$ *if and only if* $\boldsymbol{\beta}^i = \boldsymbol{\beta}_0^i$*, where* $\boldsymbol{\beta}^i \in \mathcal{B}^i$*.*

Assumption 3.7.4 (Regularity Conditions on Covariates)**.** *(i). For any constant vector* $c \in \mathbb{R}^{L+1}$*,* $Pr(\mathbf{X}_{p,\omega}^i c = 0) = 1$ *if and only if* $c = \mathbf{0}_{L+1}$*. (ii). For every* $x \in \Omega_x$*, the conditional distribution of of* ϵ^i*,* Z^i *and* $\epsilon^i + \alpha^i Z^i$ *given* $X = x$ *is absolutely continuous with respect to the Lebesgue measure and has bounded positive Radon-Nikodym densities everywhere over their conditional supports. (iii). Given any* $X = x$*, there exists a subset* $\mathcal{S}_{Z^i|x} \subset \Omega_{Z^i|x}$ *with strictly positive measure, such that for any* $z^i \in \mathcal{S}_{Z^i|x}$ *and any* $\boldsymbol{\beta}^i \in \mathcal{B}^i$*,* $-z^i + 2\mathbf{x}_{p,\omega}^i\boldsymbol{\beta}^i \in \Omega_{Z^i|x}$ *for any* $\omega \in \Omega_{\omega|x}$*.*

Assumption 3.7.5 (Regularity Conditions on the Error Terms)**.** *Given any* $X = x$*,* $\mathcal{S}_{\epsilon^i|x} \equiv \{-z^i + \mathbf{x}_{p,\omega}^i\boldsymbol{\beta}_0^i : z^i \in \Omega_{Z^i|x}, \omega \in \Omega_{\omega|x}\} \cup \{z^i - 2\mathbf{x}_{p,\omega}^i\boldsymbol{\beta}^i + \mathbf{x}_{p,\omega}^i\boldsymbol{\beta}_0^i : z^i \in \Omega_{Z^i|x}, \omega \in \Omega_{\omega|x} \text{ and } \boldsymbol{\beta}^i \in \mathcal{B}^i\}$ *is a subset of* $int(\Omega_{\epsilon^i|x})$*.*

Proposition 3.7.1. *Under Assumptions 3.2.1-3.2.2, 3.3.1, 3.3.3-3.3.4, 3.7.3, 3.7.4-3.7.5,* $\boldsymbol{\beta}_0^i \in \mathcal{B}^i$ *is point identified in the sense of Definition 3.7.2.*

Proof. See Appendix C.5. □

3.7.3 Random Coefficients

Consider (normalized) payoff function given by $U^i(Z^i, X) - \epsilon^i = -Z^i + X\beta_{1R}^i + \beta_{2R}^i D^{-i} - \epsilon^i$, where β_{1R}^i is a random coefficients vector with continuous and symmetric distributions centered at β_1^i, and β_{2R}^i is a continuous and symmetric scalar random variable with center β_2^i. More specifically, for $k = 1, 2$, we can write them as $\beta_{kR}^i = \beta_k^i - r_k^i$, where r_k^i is the disturbance with continuous and symmetric distributions centered at the origin. Suppose the realizations of the disturbances are private information. Let $r^i = [r_1^{i\top}, r_2^i]^\top$. In addition, suppose the following independence conditions hold: (i). $(Z^1, \epsilon^1, r^1) \perp (Z^2, \epsilon^2, r^2)|X$; (ii). For $i = 1, 2$, $Z^i \perp (\epsilon^i, r^i)|X$. Under these two assumptions, take expectation with respect to the action of the opponent player gives $-Z^i + X\beta_1^i + \beta_2^i Pr(D^{-i} = 1|X) - Xr_1^i - Pr(D^{-i} = 1|X)r_2^i - \epsilon^i$. Then the equilibrium of this game could be defined similarly as the fixed point of the following system:

$$
\begin{aligned}
P^1(X) &= \mathbb{E}\left(\mathbb{1}\left(-Z^1 + X\beta_1^1 + \beta_2^1 P^2(X) - Xr_1^1 - P^2(X)r_2^1 - \epsilon^1 \geq 0\right)|X\right), \\
P^2(X) &= \mathbb{E}\left(\mathbb{1}\left(-Z^2 + X\beta_1^2 + \beta_2^2 P^1(X) - Xr_1^2 - P^1(X)r_2^2 - \epsilon^2 \geq 0\right)|X\right).
\end{aligned}
$$

Our method is applicable to the identification and estimation of the center of the distribution of random coefficients, as now the composite error will satisfy all the aforementioned assumptions. Note that this composite error is by construction conditionally heteroskedastic. Our method provides a complementary tool to the study of games of incomplete information with random coefficients compared to existing methods such as Yoon (2019). Yoon (2019) assumes that random coefficients are independent of all covariates and the realization of random coefficients are common knowledge and that the researcher knows which coefficients are random, and he aims to recover the whole distribution of random coefficients. Our method on the other hand, allows random coefficients to be dependent on the common knowledge variables X and we assume the realization of random coefficients are private information, and in addition, we allow the researcher to be agnostic about which coefficients are random. We only aim to recover the center of the distribution of random coefficients.

3.8 Conclusions

In this paper, we study the semiparametric identification and estimation of a binary game of incomplete information under the restriction of error symmetry. We use a two-step identification strategy that is inspired by and combines Chen (2000) and Lewbel et al. (2020). We propose a two-step minimum distance estimator, and prove its root-N consistency. Compared to existing semiparametric method in the literature, our estimator could adapt heteroskedasticity, does not require stringent support and tail conditions, and could adapt random coefficients without knowing which coefficients are random. A small Monte Carlo study demonstrates the efficacy and robustness of our estimator compared to the popular two-step quasi-likelihood method. One direction for future research is to extend our method to multinomial games and dynamic games. Another direction for future research is a systematic examination of gains in efficiency when the symmetry restriction is imposed on existing method such as Aradillas-Lopez (2010) and Lewbel and Tang (2015).

BIBLIOGRAPHY

AGUIRREGABIRIA, V. AND P. MIRA (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75, 1–53.

——— (2018): "Identification of Games of Incomplete Information with Multiple Equilibria and Unobserved Heterogeneity," Working Paper.

——— (2019): "Identification of Games of Incomplete Information with Multiple Equilibria and Common Unobserved Heterogeneity," *Quantitative Economics*, 10, 1659–1701.

ALLMAN, E. S., C. MATIAS, AND J. A. RHODES (2009): "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *The Annals of Statistics*, 37, 3099–3132.

ANDREWS, D. W. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–563.

ANDREWS, D. W. AND X. CHENG (2012): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

ANDREWS, D. W., X. CHENG, AND P. GUGGENBERGER (2020): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Journal of Econometrics*.

ARADILLAS-LOPEZ, A. (2010): "Semiparametric estimation of a simultaneous game with incomplete information," *Journal of Econometrics*, 157, 409–431.

ATHEY, S. AND S. STERN (1998): "An empirical framework for testing theories about complimentarity in organizational design," Tech. rep., National Bureau of Economic Research.

AUGEREAU, A., S. GREENSTEIN, AND M. RYSMAN (2006): "Coordination versus differentiation in a standards war: 56K modems," *The RAND Journal of Economics*, 37, 887–909.

Bajari, P., H. Hong, and D. Nekipelov (2013): "Game Theory and Econometrics: A Survey of Some Recent Research," in *Advances in Economics and Econometrics, 10th World Congress*, vol. 3, 3–52.

Berry, S., A. Khwaja, V. Kumar, A. Musalem, K. C. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. M. Hanemann, P. Jeziorski, et al. (2014): "Structural models of complementary choices," *Marketing Letters*, 25, 245–256.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016): "Estimating Multivariate Latent-Structure Models," *The Annals of Statistics*, 44, 540–563.

Chen, S. (2000): "Efficient estimation of binary choice models under symmetry," *Journal of econometrics*, 96, 183–199.

Chen, S., S. Khan, and X. Tang (2016): "Informational content of special regressors in heteroskedastic binary response models," *Journal of econometrics*, 193, 162–182.

Cheng, X. (2015): "Robust Inference in Nonlinear Models with Mixed Identification Strength," *Journal of Econometrics*, 189, 207–228.

De Paula, A. (2013): "Econometric Analysis of Games with Multiple Equilibria," *Annual Review of Economics*, 5, 107–131.

De Paula, A. and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80, 143–172.

Deng, Y. and G. Picone (2018): "An empirical analysis of entry and location decisions by bars and liquor stores," *Empirical Economics*, 1–32.

Dong, Y. and A. Lewbel (2015): "A simple estimator for binary choice models with endogenous regressors," *Econometric Reviews*, 34, 82–105.

Duan, J.-C. (2019): "Variable Selection with Big Data based on Zero Norm and via Sequential Monte Carlo," Working Paper.

Fan, Y., S. Jiang, and X. Shi (2020): "Estimation and Inference in Games of Incomplete Information with Non-separable Unobserved Heterogeneity," Working Paper.

Fox, J. T. and N. Lazzati (2017): "A note on identification of discrete choice models for bundles and binary games," *Quantitative Economics*, 8, 1021–1036.

Gentzkow, M. (2007): "Valuing new goods in a model with complementarity: Online newspapers," *American Economic Review*, 97, 713–744.

Grieco, P. L. (2014): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," *The RAND Journal of Economics*, 45, 303–340.

Hansen, B. (2020): *Econometrics*, Unpublished Manuscript.

Hotz, V. J. and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497–529.

Hu, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144, 27–61.

Jia, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76, 1263–1316.

Kasahara, H. and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77, 135–175.

Khan, S. and E. Tamer (2010): "Irregular identification, support conditions, and inverse weight estimation," *Econometrica*, 78, 2021–2042.

Leeb, H. and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21–59.

——— (2008): "Can One Estimate the Unconditional Distribution of Post-Model-Selection Estimators?" *Econometric Theory*, 24, 338–376.

Lenzo, J. (2006): "Market structure and profit complementarity: the case of spect and pet," *manuscript. Department of Economics. Boston University.*

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

Lewbel, A. and X. Tang (2015): "Identification and estimation of games with incomplete information using excluded regressors," *Journal of Econometrics*, 189, 229–244.

Lewbel, A., J. Yan, and Z. Yu (2020): "Semiparametric Identification and Estimation of Multinomial Discrete Choice Models using Error Symmetry," Working Paper.

Lewbel, A. et al. (2012): *An overview of the special regressor method*, Boston College, Department of Economics.

Luo, Y. (2020): "Unobserved Heterogeneity in Auctions under Restricted Stochastic Dominance," *Journal of Econometrics*, 216, 354–374.

Luo, Y., P. Xiao, and R. Xiao (2018): "Identification of Dynamic Games With Unobserved Heterogeneity and Multiple Equilibria: Global Fast Food Chains in China," Working Paper.

Magnac, T. and E. Maurin (2007): "Identification and information in monotone binary models," *Journal of Econometrics*, 139, 76–104.

Newey, K. and D. McFadden (1994): "Large sample estimation and hypothesis," *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, 2112–2245.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric estimation of index coefficients," *Econometrica: Journal of the Econometric Society*, 1403–1430.

Ren, J. (2017): "Essays on Empirical Industrial Organization," Ph.D. thesis, The Ohio State University.

Robin, J.-M. and R. J. Smith (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175.

Seim, K. (2006): "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 37, 619–640.

Sweeting, A. (2009): "The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis Using Multiple Equilibria," *The RAND Journal of Economics*, 40, 710–742.

Tang, X. (2010): "Estimating simultaneous games with incomplete information under median restrictions," *Economics Letters*, 108, 273–276.

Torgovitsky, A. (2017): "Minimum distance from independence estimation of nonseparable instrumental variables models," *Journal of Econometrics*, 199, 35–48.

Vives, X. (2005): "Complementarities and games: New developments," *Journal of Economic Literature*, 43, 437–479.

Wan, Y. and H. Xu (2014): "Semiparametric identification and estimation of binary decision games of incomplete information with correlated private signals," Tech. Rep. 1.

Xiao, R. (2018): "Identification and Estimation of Incomplete Information Games with Multiple Equilibria," *Journal of Econometrics*, 203, 328–343.

Xu, H. (2014): "Estimation of discrete games with correlated types," *The Econometrics Journal*, 17, 241–270.

Yoon, J. (2019): "Inference in Game Theoretic Models of Incomplete Information with Random Coefficients," Working Paper.

Zhou, Y. (2019): "Identification and estimation of entry games under symmetry of unobservables," *Available at SSRN 3328270.*

Zhu, T. and V. Singh (2009): "Spatial competition with endogenous location choices: An application to discount retailing," *QME*, 7, 1–35.

Appendix A

GAMES OF INCOMPLETE INFORMATION WITH NON-SEPARABLE UNOBSERVED HETEROGENEITY: ESTIMATION, INFERENCE, AND COMPUTATION

A.1 Identification of CCPs and an Algorithm for SMC

A.1.1 Identification of CCPs via the Approach in Xiao (2018)

The Simple Game We first recall some notations from Section 2 of the paper. For $i = 1,2,3$, $p_i(\mathbf{z}, A) \equiv \Pr(d_i = 1 \mid \mathbf{z}, A)$, $p_i(\mathbf{z}, B) \equiv \Pr(d_i = 1 \mid \mathbf{z}, B)$, $p^A(\mathbf{z}) \equiv \Pr(k = A \mid \mathbf{z})$, and $p^B(\mathbf{z}) \equiv \Pr(k = B \mid \mathbf{z})$.

For $k, k' \in \{A, B\}$ and $k \neq k'$, let

$$\mathbf{P}_{i\mathbf{z}} = \begin{bmatrix} p_i(\mathbf{z}, k) & p_i(\mathbf{z}, k') \\ 1 - p_i(\mathbf{z}, k) & 1 - p_i(\mathbf{z}, k') \end{bmatrix} \text{ for } i = 1,2,3.$$

Define the vector of mixing weights as $W_{k|\mathbf{z}} = \left(p^k(\mathbf{z}), p^{k'}(\mathbf{z})\right)^\top$. Its diagonal form is written as $D_{k|\mathbf{z}} = \operatorname{diag}\left(W_{k|\mathbf{z}}^\top\right)$. Let the diagonal matrix containing player i's CCPs of choosing action $d_i = 0,1$ on two latent states be

$$D^i_{1\mathbf{z}} \equiv \operatorname{diag}(p_i(\mathbf{z}, k), p_i(\mathbf{z}, k')) \text{ and } D^i_{0\mathbf{z}} \equiv \operatorname{diag}(1 - p_i(\mathbf{z}, k), 1 - p_i(\mathbf{z}, k')).$$

To illustrate the identification strategy, consider identifying CCPs for player 1 on observed state $\mathbf{z}$. Define the following population contingency tables:

$$\begin{aligned}
A^{12}_{1\mathbf{z}} &\equiv \begin{bmatrix} \Pr\left((d_1,d_2,d_3)=(1,1,1)\mid\mathbf{z}\right), & \Pr\left((d_1,d_2,d_3)\equiv(1,0,1)\mid\mathbf{z}\right) \\ \Pr\left((d_1,d_2,d_3)=(0,1,1)\mid\mathbf{z}\right), & \Pr\left((d_1,d_2,d_3)=(0,0,1)\mid\mathbf{z}\right) \end{bmatrix}, \\
A^{12}_{\mathbf{z}} &\equiv \begin{bmatrix} \Pr\left((d_1,d_2)=(1,1)\mid\mathbf{z}\right), & \Pr\left((d_1,d_2)=(1,0)\mid\mathbf{z}\right) \\ \Pr\left((d_1,d_2)=(0,1)\mid\mathbf{z}\right), & \Pr\left((d_1,d_2)=(0,0)\mid\mathbf{z}\right) \end{bmatrix}, \\
A^{13}_{\mathbf{z}} &\equiv \begin{bmatrix} \Pr\left((d_1,d_3)=(1,1)\mid\mathbf{z}\right), & \Pr\left((d_1,d_3)=(1,0)\mid\mathbf{z}\right) \\ \Pr\left((d_1,d_3)=(0,1)\mid\mathbf{z}\right), & \Pr\left((d_1,d_3)=(0,0)\mid\mathbf{z}\right) \end{bmatrix}, \text{ and} \\
A^{1}_{\mathbf{z}} &\equiv \begin{bmatrix} \Pr\left(d_1=1\mid\mathbf{z}\right) \\ \Pr\left(d_1=0\mid\mathbf{z}\right) \end{bmatrix}.
\end{aligned}$$

The matrices $A^{12}_{1\mathbf{z}}$, $A^{12}_{\mathbf{z}}$, $A^{13}_{\mathbf{z}}$, and $A^{1}_{\mathbf{z}}$ are identified from data. By construction, the population contingency tables can be written as products of CCPs and mixing weights:

$$A^{12}_{1\mathbf{z}} = \mathbf{P}_{1\mathbf{z}} D^3_{1\mathbf{z}} D_{k|\mathbf{z}} \mathbf{P}^{\top}_{2\mathbf{z}},\ A^{12}_{\mathbf{z}} = \mathbf{P}_{1\mathbf{z}} D_{k|\mathbf{z}} \mathbf{P}^{\top}_{2\mathbf{z}},\ A^{13}_{\mathbf{z}} = \mathbf{P}_{1\mathbf{z}} D_{k|\mathbf{z}} \mathbf{P}^{\top}_{3\mathbf{z}},\ \text{and } A^{1}_{\mathbf{z}} = \mathbf{P}_{1\mathbf{z}} W_{k|\mathbf{z}}.$$

The CCPs are identified using the eigendecomposition method from Xiao (2018) as follows. First, CCPs for player 1 are identified as the eigenvectors (of the left hand side observable matrix) with column sum being 1: $A^{12}_{1\mathbf{z}} \left(A^{12}_{\mathbf{z}}\right)^{-1} = \mathbf{P}_{1\mathbf{z}} D^3_{1\mathbf{z}} \mathbf{P}^{-1}_{1\mathbf{z}}$ and the vector of mixing weights is identified as $W_{k|\mathbf{z}} = \left(\mathbf{P}_{1\mathbf{z}}\right)^{-1} A^{1}_{\mathbf{z}}$. Second, given the recovered $\mathbf{P}_{1\mathbf{z}}$, $W_{k|\mathbf{z}}$ (and $D_{k|\mathbf{z}}$), CCPs for players 2 and 3 are identified as $\mathbf{P}_{2\mathbf{z}} = \left(A^{12}_{\mathbf{z}}\right)^{\top} \left(D^{\top}_{k|\mathbf{z}} \mathbf{P}^{\top}_{1\mathbf{z}}\right)^{-1}$ and $\mathbf{P}_{3\mathbf{z}} = \left(A^{13}_{\mathbf{z}}\right)^{\top} \left(D^{\top}_{k|\mathbf{z}} \mathbf{P}^{\top}_{1\mathbf{z}}\right)^{-1}$.

Note that if we change the order of the two columns of the eigenvector matrix $\mathbf{P}_{1\mathbf{z}}$ and eigenvalue matrix $D^3_{1\mathbf{z}}$ at the same time, equation $A^{12}_{1\mathbf{z}} \left(A^{12}_{\mathbf{z}}\right)^{-1} = \mathbf{P}_{1\mathbf{z}} D^3_{1\mathbf{z}} \mathbf{P}^{-1}_{1\mathbf{z}}$ still holds and equations $W_{k|\mathbf{z}} = \left(\mathbf{P}_{1\mathbf{z}}\right)^{-1} A^{1}_{\mathbf{z}}$, $\mathbf{P}_{2\mathbf{z}} = \left(A^{12}_{\mathbf{z}}\right)^{\top} \left(D^{\top}_{k|\mathbf{z}} \mathbf{P}^{\top}_{1\mathbf{z}}\right)^{-1}$, and $\mathbf{P}_{3\mathbf{z}} = \left(A^{13}_{\mathbf{z}}\right)^{\top} \left(D^{\top}_{k|\mathbf{z}} \mathbf{P}^{\top}_{1\mathbf{z}}\right)^{-1}$ inherit the order of the unobserved states adopted in equation $A^{12}_{1\mathbf{z}} \left(A^{12}_{\mathbf{z}}\right)^{-1} = \mathbf{P}_{1\mathbf{z}} D^3_{1\mathbf{z}} \mathbf{P}^{-1}_{1\mathbf{z}}$. Thus the CCPs for three players are identified up to a common label swapping.

The General Case Consider the general game we introduced in Section 5. Let $A_{\mathbf{z}}^{g_1 g_2}$ be the joint contingency table for player group g_1 and player group g_2 on observed state $\mathbf{z}$. With unknown $|S_k|$ and possibly multiple equilibria, we could first identify the number of mixtures on each $\mathbf{z}$ under Assumption 5.4: $\left|S_{\omega|\mathbf{z}}\right| = \operatorname{rank}\left(A_{\mathbf{z}}^{g_1 g_2}\right)$. Part (c) of Assumption 5.4 (iii) implies that $|S_k|$ is identified as $|S_k| = \min_{\mathbf{z}} \left\{\operatorname{rank}\left(A_{\mathbf{z}}^{g_1 g_2}\right)\right\}$.

Define $A_{d\mathbf{z}}^{g_1 g_2}$ as the joint contingency table for player groups g_1 and g_2 fixing player group g_3's action at $d_{g_3} = d$. Let $A_{\mathbf{z}}^{gi}$ and $A_{\mathbf{z}}^{ig}$ be the joint contingency tables of some generic player group g and individual player i. For each $\mathbf{z}$, CCPs are identified following Xiao (2018). First, by summing up rows and columns of $A_{\mathbf{z}}^{g_1 g_2}$ (thus collapsing the actions of player group 1 and player group 2), we can create $A_{\mathbf{z}}^{\widetilde{g}_1 \widetilde{g}_2}$ with $\operatorname{rank}\left(A_{\mathbf{z}}^{\widetilde{g}_1 \widetilde{g}_2}\right) = \left|S_{\omega|\mathbf{z}}\right|$, where $\widetilde{g}_1$ and $\widetilde{g}_2$ denote player group 1 and player group 2 with collapsed actions. Let $\mathbf{P}_{\widetilde{g}_i}$ denote the matrix storing CCPs for player group g_i with collapsed actions (each column correspond to a different ω). Second, we use eigendecomposition to identify $\mathbf{P}_{\widetilde{g}_1\mathbf{z}}$ as $A_{d\mathbf{z}}^{\widetilde{g}_1 \widetilde{g}_2}\left(A_{\mathbf{z}}^{\widetilde{g}_1 \widetilde{g}_2}\right)^{-1} = \mathbf{P}_{\widetilde{g}_1\mathbf{z}} D_{d\mathbf{z}}^{g_3} \left(\mathbf{P}_{\widetilde{g}_1\mathbf{z}}\right)^{-1}$, where $D_{d\mathbf{z}}^{g_3}$ is the diagonal matrix storing the conditional choice probabilities of $d_{g_3} = d$ for all ω on the diagonal. The vector of mixing weights is identified as $W_{\omega|\mathbf{z}} = \left(\mathbf{P}_{\widetilde{g}_1\mathbf{z}}\right)^{-1} A_{\mathbf{z}}^{\widetilde{g}_1}$. Define $D_{\omega|\mathbf{z}} = \operatorname{diag}\left(W_{\omega|\mathbf{z}}^{\top}\right)$, the equilibrium CCPs for player group 2 with collapsed actions are then identified as $\mathbf{P}_{\widetilde{g}_2\mathbf{z}} = \left(A_{0\mathbf{z}}^{\widetilde{g}_1 \widetilde{g}_2}\right)^{\top} \left(D_{\omega|\mathbf{z}}^{\top} \mathbf{P}_{\widetilde{g}_1\mathbf{z}}^{\top}\right)^{-1}$. For equilibrium CCPs of individual player $i \in g_2 \bigcup g_3$, we obtain $\mathbf{P}_{i\mathbf{z}}^{\top} = \left(\mathbf{P}_{\widetilde{g}_1\mathbf{z}} D_{\omega|\mathbf{z}}\right)^{-1} A_{\mathbf{z}}^{\widetilde{g}_1 i}$. For equilibrium CCPs of individual player $i \in g_1$, we have $\mathbf{P}_{i\mathbf{z}}^{\top} = A_{\mathbf{z}}^{i\widetilde{g}_2}\left(D_{\omega|\mathbf{z}} \mathbf{P}_{\widetilde{g}_2\mathbf{z}}^{\top}\right)^{-1}$.

A.1.2 Numerical Computation via Sequential Monte Carlo

The following discussion follows from Section 4.3. We define an intermediate target distribution function as

$$f_{n\gamma}(u) \propto \left[\frac{\exp\left(-J_n(u)\right)}{I(u)}\right]^{\gamma} I(u),$$

where $\propto$ denotes proportionality, γ is a parameter that advances from 0 to 1 in a self-adapted way described below, and $I(u)$ is the initial distribution. Our algorithm adapts the procedure in Duan (2019) to account for the special structure of our problem, which further speeds up the searching.

Step 1. Initial sampling. At $\gamma_0 = 0$, draw a random sample $(U_1, ..., U_M)$ of size M from the initial distribution $I(u)$. In each sampled permutation (or particle), the first element is fixed. And for each subsequent element, we sample either the first row or the second row (but not both) corresponding to each observed state.

Step 2. Finding the next γ. Define the incremental importance weight for sample point U_i between the current γ and some generic γ' as $W_{\gamma,\gamma'}(U_i) \equiv f_{n\gamma'}(U_i)/f_{n\gamma}(U_i)$ and the effective sample size (ESS) as

$$ESS_{\gamma,\gamma'} \equiv \frac{\left(\sum_{i=1}^{M} W_{\gamma,\gamma'}(U_i)\right)^2}{\sum_{i=1}^{M} W_{\gamma,\gamma'}^2(U_i)}.$$

For each point in the grid $(\gamma_0, 1]$, we obtain $(W_{\gamma,\gamma_0}(U_1), .., W_{\gamma,\gamma_0}(U_M))$ as the incremental weights. Find γ^* such that ESS_{γ^*,γ_0} is no less than (or close to) ηM for $\eta \in (0,1)$. We set $\gamma_1 = \gamma^*$.

Step 3. Reweighting, resampling, and support boosting. Define $S_{\gamma_1,\gamma_0} \equiv \sum_{i=1}^{M} W_{\gamma_1,\gamma_0}(U_i)$. Now we have the weighted sample:

$$\left\{(U_1, ..., U_M), \left(\frac{W_{\gamma_1,\gamma_0}(U_1)}{S_{\gamma_1,\gamma_0}}, .., \frac{W_{\gamma_1,\gamma_0}(U_M)}{S_{\gamma_1,\gamma_0}}\right)\right\},$$

which represents our first intermediate target $f_{n\gamma_1}(u)$. However, the problem with this sample is that the incremental weights are too volatile, resulting in low ESS. To improve ESS, we resample from $\left\{(U_1, ..., U_M), \left(\frac{W_{\gamma_1,\gamma_0}(U_1)}{S_{\gamma_1,\gamma_0}}, .., \frac{W_{\gamma_1,\gamma_0}(U_M)}{S_{\gamma_1,\gamma_0}}\right)\right\}$ according to the normalized incremental weights, and get a sample with equal weight: $\left\{(U_1^*, ..., U_M^*), \left(\frac{1}{M}, .., \frac{1}{M}\right)\right\}$. This sample has good ESS, but its support has shrunk substantially. We next boost its support using a sufficient number of Metropolis Hastings (MH) moves. When implementing the MH moves, the structure of our problem is used again, i.e., for each new particle proposed, we only replace an element in the original particle with the other row of the same observed state. Having finished MH support boosting, we obtain the improved sample $\left\{\left(\widetilde{U}_1^*, ..., \widetilde{U}_M^*\right), \left(\frac{1}{M}, .., \frac{1}{M}\right)\right\}$, which is a good enough sample representing $f_{n\gamma_1}(u)$.

Step 4. Updating γ and iterating Step 2-3 until reaching $\gamma^* = 1$. For example, for each point in the grid $(\gamma_1, 1]$, we obtain the incremental weights $(W_{\gamma,\gamma_1}(U_1), .., W_{\gamma,\gamma_1}(U_M))$, find γ^* such that ESS is close to ηM, and set $\gamma_2 = \gamma^*$. Then we repeat the reweighting, resampling, and support boosting steps. The resulting sample will be a good sample representing $f_{n\gamma_2}(u)$. The whole process is repeated until we reach $\gamma^* = 1$.

Step 5. Computing the mode. We collapse the sampled permutations into combinations and find the highest frequency combination.

This algorithm makes estimation under large state space feasible because the running time is of order $O(M)$, which is not growing exponentially with the number of states. High precision could be achieved because exploration is focused on the more promising subsets of the selection vector space.

A.2 Technical Proofs

Proof of Lemma 2.1: Under Assumption 2.1-2.5, CCPs are identified up to a label swapping, which enables us to identify expected payoffs (via $F^{-1}(\cdot)$) and set up system (2.7). Under Assumption 2.6, the systems corresponding to $c \neq c_0$ has no solution. Note that system corresponding to c_0 always has a solution, as π_0 generates the system corresponding to c_0. Under the condition that Γ_{c_0} has full column rank, this π_0 is uniquely determined by $\overline{\pi}_{c_0}$ and Γ_{c_0}. □

Proof of Theorem G1: The proof follows from the similar argument in the proof of Theorem 1 of Andrews (1999). For any $c \neq c_0$, Assumption G2 (iii) implies that

$$\min_{\pi\in\Pi} \|G_{nc}(\pi)\|^2_{W_n(c)} \xrightarrow{p} \min_{\pi\in\Pi} \|G_c(\pi)\|^2_{W(c)} > 0, \tag{A.1}$$

where the inequality follows from Assumption G1 (iv) and G2 (ii). By Assumption G1 (iv), we have that $G_{c_0}(\pi_0) = \mathbf{0}$. Then by Assumption G2 (i), it holds that $G_{nc_0}(\pi_0) = o_p(1)$.

Therefore, we obtain that

$$\min_{\pi\in\Pi}\|G_{nc_0}(\pi)\|^2_{W_n(c)} \leq \|G_{nc_0}(\pi_0)\|^2_{W_n(c)} = o_p(1). \tag{A.2}$$

Combining Equations (A.1) and (A.2), we conclude that $\widehat{c} = c_0$ with probability approaching one.

It remains to show that $\widehat{\pi} \xrightarrow{p} \pi_0$. Define $\widetilde{\pi} \equiv \arg\min_{\pi\in\Pi}\|G_{nc_0}(\pi)\|^2_{W_n(c_0)}$. We have $\widetilde{\pi}$ converges in probability to its population counterpart π_0 by the standard argument for consistency of a GMM estimator provided that Assumption G1 (v) holds. Thus for any $\epsilon > 0$ and $\delta > 0$, we can find N_1 such that $\Pr\left(\|\widetilde{\pi} - \pi_0\| > \frac{\epsilon}{2}\right) < \frac{\delta}{2}$ for $n \geq N_1$. Because $\widehat{c} = c_0$ implies $\widetilde{\pi} = \widehat{\pi}$, we have $\Pr(\widetilde{\pi} = \widehat{\pi}) \geq \Pr(\widehat{c} = c_0) \to 1$ by the fact that $\widehat{c} = c_0$ with probability approaching one. Thus for the given $\epsilon > 0$ and $\delta > 0$, we can find N_2 such that $\Pr\left(\|\widetilde{\pi} - \widehat{\pi}\| > \frac{\epsilon}{2}\right) < \frac{\delta}{2}$ for $n \geq N_2$. Combing the above two results, we have that for any given $\epsilon > 0$ and $\delta > 0$, there exists $N \equiv \max\{N_1, N_2\}$ such that for $n > N$,

$$\begin{aligned} & \Pr(\|\widehat{\pi} - \pi_0\| > \epsilon) = \Pr(\|\widehat{\pi} - \widetilde{\pi} + \widetilde{\pi} - \pi_0\| > \epsilon) \\ \leq\ & \Pr\left(\|\widehat{\pi} - \widetilde{\pi}\| > \frac{\epsilon}{2}\right) + \Pr\left(\|\widetilde{\pi} - \pi_0\| > \frac{\epsilon}{2}\right) \leq \delta. \end{aligned}$$

Therefore, the result $\widehat{\pi} \xrightarrow{p} \pi_0$ holds. □

Proof of Theorem G2: To show part (i), we prove the result in two cases: $l_R < l_\pi$ and $l_R = l_\pi$. When $l_R < l_\pi$, the null space of R has dimension $l_\pi - l_R$. Let Ψ be a $l_\pi \times (l_\pi - l_R)$ matrix storing a basis of the null space. Then there exists $\pi_f \in R^{l_\pi - l_R}$ and μ such that any π_0 satisfying H_0 can be written as $\pi_0 = \Psi\pi_f + \mu$. By imposing H_0 on the sample "moment" functions, we obtain that

$$G_{nc}(\pi_0) = G_{nc}(\Psi\pi_f + \mu) = \overline{\pi}_{nc} - \Gamma_{nc}\mu - \Gamma_{nc}\Psi\pi_f.$$

Define $\widehat{\pi}_f(c) \equiv \arg\min_{\pi\in\Pi}\|G_{nc}(\Psi\pi_f + \mu)\|^2_{W_n(c)}$, where $W_n(c) = W(c) + o_p(1)$ by Assumption G3 (ii). Define $\pi^*_f(c) \equiv \operatorname{plim}\widehat{\pi}_f(c)$. For each given c, we have the solution $\widehat{\pi}_f(c)$ to the minimum distance problem with a corresponding "pseudo-true" value $\pi_f^*(c)$ defined as its probability limit. The "pseudo-true" value at c_0 delivers the true value as $\pi_0 = \Psi\pi_f^*(c_0) + \mu$. Let $\Sigma(c)$ be the asymptotic variance of $\sqrt{n}\left(G_{nc}\left(\Psi\pi_f^*(c)+\mu\right) - G_c\left(\Psi\pi_f^*(c)+\mu\right)\right)$. We know that $\Sigma(c_0) = \Omega_0$ by definition. Our test statistic is equivalent to

$$T_n = \min_{c\in\mathscr{C}} nG_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right)^\top W_n(c)\, G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right).$$

Intuitively, we reject the null for large values of T_n. To finalize the test, we need to find a critical value for this test statistic such that the asymptotic size of the test is controlled over the null parameter space. For this purpose we derive its asymptotic distribution under drifting sequences. This is achieved by the following steps.

Step 1. For any given c, applying mean value expansion of $G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right)$ at $\pi_f^*(c)$, we get

$$G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right) = G_{nc}\left(\Psi\pi_f^*(c)+\mu\right) - \Gamma_{nc}\Psi\left(\widehat{\pi}_f(c) - \pi_f^*(c)\right). \tag{A.3}$$

By construction, $\widehat{\pi}_f(c)$ satisfies the following first order condition:

$$\Gamma_{nc}\Psi^\top W_n(c)\, G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right) = 0.$$

Step 2. Multiply both sides of Equation (A.3) by $a_{nc} \equiv \Gamma_{nc}\Psi^\top W_n(c)$. We have that

$$a_{nc}G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right) = a_{nc}G_{nc}\left(\Psi\pi_f^*(c)+\mu\right) - a_{nc}\Gamma_{nc}\Psi\left(\widehat{\pi}_f(c) - \pi_f^*(c)\right).$$

After rearrangement, we have that

$$\begin{aligned}\left(\widehat{\pi}_f(c) - \pi_f^*(c)\right) &= \left(a_{nc}\Gamma_{nc}\Psi\right)^{-1} a_{nc}G_{nc}\left(\Psi\pi_f^*(c)+\mu\right)\\ &\quad - \left(a_{nc}\Gamma_{nc}\Psi\right)^{-1} a_{nc}G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right).\end{aligned} \tag{A.4}$$

Multiply both sides of (A.4) with $\Gamma_{nc}\Psi$ we get that

$$\begin{aligned}\Gamma_{nc}\Psi\left(\widehat{\pi}_f(c) - \pi_f^*(c)\right) &= \Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1} a_{nc}G_{nc}\left(\Psi\pi_f^*(c)+\mu\right)\\ &\quad - \Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1} a_{nc}G_{nc}\left(\Psi\widehat{\pi}_f(c)+\mu\right).\end{aligned} \tag{A.5}$$

Step 3. Combining Equation (A.3) with Equation (A.5), we have that

$$\begin{aligned}
&\sqrt{n}\left(G_{nc}\left(\Psi\widehat{\pi}_f\left(c\right)+\mu\right)-G_{nc}\left(\Psi\pi_f^*\left(c\right)+\mu\right)\right)\\
&=-\Gamma_{nc}\Psi\sqrt{n}\left(\widehat{\pi}_f\left(c\right)-\pi_f^*\left(c\right)\right)\\
&=\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\sqrt{n}G_{nc}\left(\Psi\widehat{\pi}_f\left(c\right)+\mu\right)\\
&-\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\sqrt{n}G_{nc}\left(\Psi\pi_f^*\left(c\right)+\mu\right).
\end{aligned}$$

Thus it holds that

$$\begin{aligned}
&\sqrt{n}G_{nc}\left(\Psi\widehat{\pi}_f\left(c\right)+\mu\right)\\
=\;&\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\sqrt{n}G_{nc}\left(\Psi\widehat{\pi}_f\left(c\right)+\mu\right)\\
&+\left(I_{\|c\|_0}-\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\right)\sqrt{n}G_{nc}\left(\Psi\pi_f^*\left(c\right)+\mu\right)\\
=\;&\left(I_{\|c\|_0}-\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\right)\sqrt{n}\left(G_{nc}\left(\Psi\pi_f^*\left(c\right)+\mu\right)-G_c\left(\Psi\pi_f^*\left(c\right)+\mu\right)\right)\\
&+\left(I_{\|c\|_0}-\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\right)\sqrt{n}G_c\left(\Psi\pi_f^*\left(c\right)+\mu\right),
\end{aligned}\tag{A.6}$$

where the second equality follows from $a_{nc}\sqrt{n}G_{nc}\left(\Psi\widehat{\pi}_f\left(c\right)+\mu\right)=0$. After plugging in a_{nc}, we have that

$$\left(I_{\|c\|_0}-\Gamma_{nc}\Psi\left(a_{nc}\Gamma_{nc}\Psi\right)^{-1}a_{nc}\right)=\left(I_{\|c\|_0}-\Gamma_c\Psi\left(\Gamma_c\Psi^\top W\left(c\right)\Gamma_c\Psi\right)^{-1}\Gamma_c\Psi^\top W\left(c\right)\right)+o_p\left(1\right).$$

For $c=c_0$, by condition (iii) in Assumption G3, $W\left(c_0\right)=\Omega_0^{-1}$ is symmetric and positive definite, and as a result it admits Cholesky decomposition $W\left(c_0\right)=A\left(c_0\right)^\top A\left(c_0\right)$. Thus it holds that

$$\begin{aligned}
&nG_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)^\top W_n\left(c_0\right)G_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)\\
=\;&nG_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)^\top W\left(c_0\right)G_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)+o_p\left(1\right)\\
=\;&\left(A\left(c_0\right)\sqrt{n}G_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)\right)^\top\left(A\left(c_0\right)\sqrt{n}G_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)\right)+o_p\left(1\right).
\end{aligned}$$

As $W\left(c_0\right)=\Sigma\left(c_0\right)^{-1}$, $\Sigma\left(c_0\right)=A\left(c_0\right)^{-1}\left(A\left(c_0\right)^\top\right)^{-1}$, under Assumption G3 (i) and by equation (A.6), we have that

$$
\begin{aligned}
& A\left(c_0\right) \sqrt{n} G_{n c_0}\left(\Psi \widehat{\pi}_f\left(c_0\right)+\mu\right) \\
\xrightarrow{d}\ & \left(I_{\left\|c_0\right\|_0}-A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top}\right) Z \\
& +\left(A\left(c_0\right)-A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1}\right) \sqrt{n} G_{c_0}\left(\Psi \pi_f^*\left(c_0\right)+\mu\right) \\
=\ & \left(I_{\left\|c_0\right\|_0}-A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top}\right) Z,
\end{aligned}
$$

where $Z \sim N\left(0, I_{\left\|c_0\right\|_0}\right)$.

It is easy to verify that the matrix

$$
I_{\left\|c_0\right\|_0}-A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top}
$$

is idempotent. It has the rank $\left\|c_0\right\|_0-\left(l_\pi-l_R\right)$ verified by the following derivation:

$$
\begin{aligned}
& \operatorname{rank}\left[\left(I_{\left\|c_0\right\|_0}-A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top}\right)\right] \\
=\ & \operatorname{trace}\left[\left(I_{\left\|c_0\right\|_0}-A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top}\right)\right] \\
=\ & \operatorname{trace}\left(I_{\left\|c_0\right\|_0}\right)-\operatorname{trace}\left(A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1} \Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top}\right) \\
=\ & \operatorname{trace}\left(I_{\left\|c_0\right\|_0}\right)-\operatorname{trace}\left(\Gamma_{c_0} \Psi^{\top} A\left(c_0\right)^{\top} A\left(c_0\right) \Gamma_{c_0} \Psi\left(\Gamma_{c_0} \Psi^{\top} \Sigma\left(c_0\right)^{-1} \Gamma_{c_0} \Psi\right)^{-1}\right) \\
=\ & \left\|c_0\right\|_0-\left(l_\pi-l_R\right),
\end{aligned}
$$

where the first equality follows from the property of idempotent matrix and third equality follows from the invariance property of trace under cyclic permutations. For $c \neq c_0$, it holds that

$$
\begin{aligned}
& n G_{n c}\left(\Psi \widehat{\pi}_f(c)+\mu\right)^{\top} W_n(c) G_{n c}\left(\Psi \widehat{\pi}_f(c)+\mu\right) \\
=\ & n G_{n c}\left(\Psi \widehat{\pi}_f(c)+\mu\right)^{\top} W(c) G_{n c}\left(\Psi \widehat{\pi}_f(c)+\mu\right)+o_p(1) \\
\xrightarrow{d}\ & \left(\left(I_{\|c\|_0}-\Gamma_c \Psi\left(a_c \Gamma_c \Psi\right)^{-1} a_c\right) \Sigma(c)^{\frac{1}{2}} Z\right)^{\top} W(c)\left(\left(I_{\|c\|_0}-\Gamma_c \Psi\left(a_c \Gamma_c \Psi\right)^{-1} a_c\right) \Sigma(c)^{\frac{1}{2}} Z\right) \\
& +\left(\left(I_{\|c\|_0}-\Gamma_c \Psi\left(a_c \Gamma_c \Psi\right)^{-1} a_c\right) \lambda_c\right)^{\top} W(c)\left(\left(I_{\|c\|_0}-\Gamma_c \Psi\left(a_c \Gamma_c \Psi\right)^{-1} a_c\right) \lambda_c\right)+o_p(1),
\end{aligned}
$$

where $Z \sim N\left(0, I_{\|c\|_0}\right)$, $a_c=\Gamma_c \Psi^{\top} W(c)$ and $\lambda_c \equiv \lim_{n \rightarrow \infty} \sqrt{n} G_c\left(\Psi \pi_f^*(c)+\mu\right)$. Elements in λ_c are allowed to be infinite. For any $c \in \mathscr{C}$, we call the selected system incorrect if $\|\lambda_c\| = +\infty$, and nearly correct if $\|\lambda_c\| \neq 0$ and $\|\lambda_c\| < +\infty$. Thus, under drifting sequence of model parameters, when there is no nearly correct systems, i.e., $\|\lambda_c\| = \infty$ for any $c \neq c_0$, our test statistic $T_n \xrightarrow{d} \chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right]}$. When there are nearly correct systems, i.e., exists at least one $\widetilde{c} \neq c_0$ such that $0 < \|\lambda_{\widetilde{c}}\| < \infty$, T_n is asymptotically stochastically dominated by $\chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right]}$. In either case, using the $1-\alpha$ quantile of $\chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right]}$ denoted as $\chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right],1-\alpha}$, achieves asymptotic size control:

$$\limsup_{n\to\infty} \sup_{\xi\in\Xi_R} \Pr_\xi \left(T_n > \chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right],1-\alpha}\right) = \alpha.$$

The proof for the case where $l_R = l_\pi$ is the same with $\pi_0 = R^{-1}r$. Therefore, part (i) of the theorem holds.

To show part (ii), notice that when $\xi \notin \Xi_R$, i.e., under H_1, T_n diverges to infinity under Assumption G2. Since the CV is finite, it holds that

$$\lim_{n\to\infty} \Pr_\xi \left(T_n > \chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right],1-\alpha}\right) = 1.$$

Thus, the second part of the theorem holds. $\square$

Proof of Proposition 4.1: Without loss of generality, we focus on proving the consistency of estimated payoff for player 1 in his observed state $z_1 = L$. We proceed by verifying that Assumption G1 and Assumption G2 for the Generic Model hold under primitive conditions of the Simple Game. Firstly notice that Assumption G1 is satisfied under Assumptions 2.1-2.6. Let $\mathbf{p}$ be the vector that stores the CCPs of all players on all observed and latent states ($\mathbf{z} \in \{\mathbf{z}^1, \ldots, \mathbf{z}^9\}, k \in \{A, B\}$) and $\widehat{\mathbf{p}}$ be its consistent estimator obtained via the eigendecomposition method. Because the estimated coefficient matrix in the Simple Game is a smooth function of $\widehat{\mathbf{p}}$, we write it in the form of $\Gamma_n \equiv \mathbf{e}\left(\widehat{\mathbf{p}}\right)$ for some smooth function $\mathbf{e}\left(\cdot\right)$. Similarly, the estimated expected payoff vector can be written in the form of $\overline{\pi}_n \equiv \mathbf{F}^{-1}\left(\widehat{\mathbf{p}}\right)$, where $\mathbf{F}^{-1}\left(\widehat{\mathbf{p}}\right)$ stacks $F^{-1}\left(p_{1\mathbf{z}k}\right)$ for any $\mathbf{z} \in \{\mathbf{z}^1, \cdots, \mathbf{z}^9\}, k \in \{A, B\}$. Thus the sample "moment" functions can be written as $G_n\left(\pi\right) = \mathbf{F}^{-1}\left(\widehat{\mathbf{p}}\right) - \mathbf{e}\left(\widehat{\mathbf{p}}\right)\pi$. By differentiability of

$\mathbf{F}^{-1}(\cdot)$ and $\mathbf{e}(\cdot)$, given any π, the mean value expansion of $G_n(\pi)$ with respect to $\mathbf{p}$ gives

$$\sqrt{n}\left(G_n(\pi) - G(\pi)\right) = \sqrt{n} D_{\mathbf{p}^*}\left(\widehat{\mathbf{p}} - \mathbf{p}\right),$$

where $D_{\mathbf{p}^*} = \frac{\partial \mathbf{F}^{-1}(\mathbf{p}^*)}{\partial \mathbf{p}'} - \frac{\partial(\mathbf{e}(\mathbf{p}^*)\pi)}{\partial \mathbf{p}'}$ and $\mathbf{p}^*$ is a point between $\widehat{\mathbf{p}}$ and $\mathbf{p}$. Note that $D_{\mathbf{p}^*}$ is bounded. Under Assumption 2.1-2.6 and 4.2, Lemma C.2 in Xiao (2018) implies that that $\widehat{\mathbf{p}}$ is consistent and $\sqrt{n}\left(\widehat{\mathbf{p}} - \mathbf{p}\right)$ converges to a normal distribution. Under Assumption 4.1 (ii), we have that $D_{\mathbf{p}^*} \xrightarrow{p} D_{\mathbf{p}} \equiv \frac{\partial \mathbf{F}^{-1}(\mathbf{p})}{\partial \mathbf{p}'} - \frac{\partial(\mathbf{e}(\mathbf{p})\pi)}{\partial \mathbf{p}'}$ and

$$\sqrt{n}\left(G_n(\pi) - G(\pi)\right) = O_p(1).$$

Thus, condition (i) in Assumption G2 is verified. Condition (ii) in Assumption G2 is satisfied under Assumption 4.1. For condition (iii) in Assumption G2, it suffices to prove that for any c

$$\max_{\pi \in \Pi} \left| G_{nc}(\pi)^\top W_n(c) G_{nc}(\pi) - G_c(\pi)^\top W(c) G_c(\pi) \right| = o_p(1),$$

where max is used rather than sup because Π is compact by Assumption 4.3 and both $G_{nc}(\cdot)$ and $G_c(\cdot)$ are continuous in π. The triangular inequality provides that

$$\begin{aligned} & \max_{\pi \in \Pi} \left| G_{nc}(\pi)^\top W_n(c) G_{nc}(\pi) - G_c(\pi)^\top W(c) G_c(\pi) \right| \\ \leq\ & \max_{\pi \in \Pi} \left| G_{nc}(\pi)^\top W_n(c) G_{nc}(\pi) - G_{nc}(\pi)^\top W(c) G_{nc}(\pi) \right| \\ & + \max_{\pi \in \Pi} \left| G_{nc}(\pi)^\top W(c) G_{nc}(\pi) - G_c(\pi)^\top W(c) G_c(\pi) \right|. \end{aligned}$$

By the property of the matrix infinity norm,[1] the first term on the right hand side of the inequality is bounded above by $\max_{\pi \in \Pi} \|c\|_0 \|G_{nc}(\pi)\|^2 \|W_n(c) - W(c)\|_\infty$. Π is bounded by Assumption 4.3. Moreover, Γ_n and $\overline{\pi}_n$ are both $O_p(1)$ by the continuity of $\mathbf{e}(\cdot)$ and $\mathbf{F}^{-1}(\cdot)$ and the consistency of $\widehat{\mathbf{p}}$. Therefore, we have $\max_{\pi \in \Pi} \|G_{nc}(\pi)\|^2 = O_p(1)$. By Assumption 4.1 (i), it holds that $\|W_n(c) - W(c)\|_\infty = o_p(1)$. Thus, the first term is $o_p(1)$. For the second

[1] For a generic matrix E, $\|E\|_\infty$ denotes the matrix infinity norm that equals the maximum absolute row sum of matrix E.

term, it holds that

$$\begin{aligned}
& \max_{\pi\in\Pi}\left|G_{nc}(\pi)^{\top}W(c)\,G_{nc}(\pi)-G_{c}(\pi)^{\top}W(c)\,G_{c}(\pi)\right| \\
= \; & \max_{\pi\in\Pi}\left|G_{nc}(\pi)^{\top}W(c)\left(G_{nc}(\pi)-G_{c}(\pi)+G_{c}(\pi)\right)-G_{c}(\pi)^{\top}W(c)\,G_{c}(\pi)\right| \\
\leq \; & \max_{\pi\in\Pi}\left|G_{nc}(\pi)^{\top}W(c)\left(G_{nc}(\pi)-G_{c}(\pi)\right)\right| \\
& +\max_{\pi\in\Pi}\left|G_{nc}(\pi)^{\top}W(c)\,G_{c}(\pi)-G_{c}(\pi)^{\top}W(c)\,G_{c}(\pi)\right| \\
\leq \; & \max_{\pi\in\Pi}\left\|G_{nc}(\pi)^{\top}W(c)\right\|_{\infty}\max_{\pi\in\Pi}\left\|G_{nc}(\pi)-G_{c}(\pi)\right\| \\
& +\max_{\pi\in\Pi}\left\|G_{nc}(\pi)-G_{c}(\pi)\right\|\max_{\pi\in\Pi}\left\|W(c)\,G_{c}(\pi)\right\|_{\infty} \\
= \; & o_{p}(1),
\end{aligned}$$

where the first and second inequality follow from the triangular inequality and applying matrix norm, and last equality holds by Γ_n and $\overline{\pi}_n$ being $O_p(1)$ and the compactness of Π. □

Proof of Proposition 4.2: The proof follows from the similar argument as in the proof for Proposition 4.1. For the games in Sections 3.1 and 3.2, $\mathbf{p}$ includes CCPs for all players on all observed and latent states, and for the game in Section 3.3, $\mathbf{p}$ includes CCPs for the representative player on all observed and latent states. In all three games, the coefficient matrices are smooth in $\mathbf{p}$, and they only differ in the exact form of function $\mathbf{e}(\cdot)$. □

Proof of Proposition 4.3: We prove Proposition 4.3 by verifying the conditions in Theorem G2. Firstly we show that the "moment" functions for the Simple Game admits an asymptotic linear representation under drifting sequence. Without loss of generality, we prove it for player 1. The proof of asymptotic linear representation builds on two lemmas provided after this proof. We first focus on finding the asymptotic linear representation for the first observed state $\mathbf{z}=\mathbf{z}^1$, denoted as $G_{n\mathbf{z}^1}(\pi)$, and then stack the asymptotic linear representations of $G_{n\mathbf{z}}(\pi)$ for $\mathbf{z}=\mathbf{z}^1,\ldots,\mathbf{z}^9$ in the end to obtain the asymptotic linear representation for $G_n(\pi)$. Because we have fixed the player and the observed state variable, from now on we will suppress the subscript i, z_i, and $\mathbf{z}$ when there is no confusion. With slight abuse of notation, define the 6×1 vector of equilibrium CCPs and its estimator as:

$$\begin{aligned} p_{\mathbf{z}^1} &\equiv \left[p_1\left(\mathbf{z}^1, k\right), p_1\left(\mathbf{z}^1, k'\right), \ldots, p_3\left(\mathbf{z}^1, k\right), p_3\left(\mathbf{z}^1, k'\right)\right]^\top \equiv \left[p_1, \ldots, p_6\right]^\top \text{ and} \\ \widehat{p}_{\mathbf{z}^1} &\equiv \left[\widehat{p}_1, \ldots, \widehat{p}_6\right]^\top . \end{aligned} \tag{A.7}$$

Let $\overline{\pi} \equiv \left[F^{-1}\left(p_1\left(\mathbf{z}^1, k\right)\right), F^{-1}\left(p_1\left(\mathbf{z}^1, k'\right)\right)\right]^\top$ be the 2×1 equilibrium expected payoff vector, and define $\overline{\pi}_n$ as the estimated expected payoff vector. The true matrix storing the joint probabilities of opponents' actions is $\Gamma = \left[\mathbf{p}_{-1}\left(\mathbf{z}^1, k\right)^\top, \ \ \mathbf{p}_{-1}\left(\mathbf{z}^1, k'\right)^\top\right]^\top$. The estimated matrix for opponents actions is denoted as Γ_n. Let the 8×1 vector $a_{\mathbf{z}^1}$ denote the free joint probabilities in the contingency table (conditional on $\mathbf{z} = \mathbf{z}^1$), and let the 9×1 vector q denote the free unconditional joint probabilities that generate the contingency table on $\mathbf{z}^1$:[2]

$$a_{\mathbf{z}^1} \equiv \left[a_1, \ldots, a_8\right]^\top \text{ and } q_{\mathbf{z}^1} \equiv \left[q_1, \ldots, q_9\right]^\top \text{, where} \tag{A.8}$$

$$\begin{aligned} &a_1 \equiv \Pr\left(\left[d_1, d_2, d_3\right] = [1,1,1] \mid \mathbf{z}^1\right), \ \ a_2 \equiv \Pr\left(\left[d_1, d_2, d_3\right] = [0,1,1] \mid \mathbf{z}^1\right), \\ &a_3 \equiv \Pr\left(\left[d_1, d_2, d_3\right] = [1,0,1] \mid \mathbf{z}^1\right), \ \ a_4 \equiv \Pr\left(\left[d_1, d_2\right] = [1,1] \mid \mathbf{z}^1\right), \\ &a_5 \equiv \Pr\left(\left[d_1, d_2\right] = [0,1] \mid \mathbf{z}^1\right), \ \ a_6 \equiv \Pr\left(\left[d_1, d_3\right] = [1,1] \mid \mathbf{z}^1\right), \\ &a_7 \equiv \Pr\left(\left[d_1, d_3\right] = [0,1] \mid \mathbf{z}^1\right), \ \ a_8 \equiv \Pr\left(d_1 = 1 \mid \mathbf{z}^1\right) \text{ and} \end{aligned} \tag{A.9}$$

$$\begin{aligned} &q_1 \equiv \Pr\left(\left[d_1, d_2, d_3\right] = [1,1,1], \mathbf{z} = \mathbf{z}^1\right), \ldots, \\ &q_8 \equiv \Pr\left(d_1 = 1, \mathbf{z} = \mathbf{z}^1\right), \ \ q_9 \equiv \Pr\left(\mathbf{z} = \mathbf{z}^1\right). \end{aligned} \tag{A.10}$$

For $i = 1, ..., 8$, it holds that $a_i = q_i / q_9$. Note that $q_{\mathbf{z}^1}$ includes three probabilities for the joint actions of three players on $\mathbf{z}^1$, two probabilities for the joint actions of player 1 and player 2 on $\mathbf{z}^1$, and two probabilities for the joint actions of player 1 and player 3 on $\mathbf{z}^1$, one

[2] Note that in Appendix A.1.1 although there are in total 14 probabilities in the contingency tables that generate CCPs via eigendecomposition, only 8 of them are free (not linear combinations of other probabilities): $a_1, \ldots, a_8$.

probability for the action of player 1 on $\mathbf{z}^1$ and the probability of $\mathbf{z} = \mathbf{z}^1$. The estimators $\widehat{q}_{\mathbf{z}^1}$ and $\widehat{a}_{\mathbf{z}^1}$ are calculated as

$$\widehat{q}_{\mathbf{z}^1} = \frac{1}{n}\sum_{l=1}^{n} \eta_l \equiv [\widehat{q}_1, \ldots, \widehat{q}_9]^\top \text{ and } \widehat{a}_{\mathbf{z}^1} = \left[\frac{\widehat{q}_1}{\widehat{q}_9}, \ldots, \frac{\widehat{q}_8}{\widehat{q}_9}\right]^\top \equiv [\widehat{a}_1, \ldots, \widehat{a}_8]^\top, \tag{A.11}$$

where $\eta_{l\mathbf{z}^1}$ is defined as

$$\eta_{l\mathbf{z}^1} \equiv \left[\mathbb{1}\left([d_{1l}, d_{2l}, d_{3l}] = [1,1,1], \mathbf{z}_l = \mathbf{z}^1\right), \ldots, \mathbb{1}\left(d_{1l} = 1, \mathbf{z}_l = \mathbf{z}^1\right), \mathbb{1}\left(\mathbf{z}_l = \mathbf{z}^1\right)\right]^\top. \tag{A.12}$$

By Lemmas A.2.1 and A.2.2 below, the second order Taylor expansion of $G_{n\mathbf{z}^1}(\pi)$ at $p_{\mathbf{z}^1}$ and with respect to $p_{\mathbf{z}^1}$ gives:

$$\sqrt{n}\left(G_{n\mathbf{z}^1}(\pi) - G_{\mathbf{z}^1}(\pi)\right) = \sqrt{n} D_{p\mathbf{z}^1}\left(\widehat{p}_{\mathbf{z}^1} - p_{\mathbf{z}^1}\right) + \sum_j \sum_k \frac{H_{p^*-j,k}}{2}\sqrt{n}\left(\widehat{p}_j - p_j\right)\left(\widehat{p}_k - p_k\right),$$

where $D_{p\mathbf{z}^1} \equiv \begin{bmatrix} dp_{11}, \ldots, dp_{16} \\ dp_{21}, \ldots, dp_{26} \end{bmatrix}_{2\times 6}$, in which

$$\begin{aligned}
dp_{11} &= \frac{1}{F'\left(F^{-1}\left(p_1\left(\mathbf{z}^1, k\right)\right)\right)}, \;\; dp_{22} = \frac{1}{F'\left(F^{-1}\left(p_1\left(\mathbf{z}^1, k'\right)\right)\right)}, \\
dp_{12} &= dp_{14} = dp_{16} = dp_{21} = dp_{23} = dp_{25} = 0, \\
dp_{13} &= \left(p_3\left(\mathbf{z}^1, k\right), \left(1 - p_3\left(\mathbf{z}^1, k\right)\right), -p_3\left(\mathbf{z}^1, k\right), -\left(1 - p_3\left(\mathbf{z}^1, k\right)\right)\right)\pi, \\
dp_{15} &= \left(p_2\left(\mathbf{z}^1, k\right), -p_2\left(\mathbf{z}^1, k\right), 1 - p_2\left(\mathbf{z}^1, k\right), -\left(1 - p_2\left(\mathbf{z}^1, k\right)\right)\right)\pi, \\
dp_{24} &= \left(p_3\left(\mathbf{z}^1, k'\right), \left(1 - p_3\left(\mathbf{z}^1, k'\right)\right), -p_3\left(\mathbf{z}^1, k'\right), -\left(1 - p_3\left(\mathbf{z}^1, k'\right)\right)\right)\pi, \\
dp_{26} &= \left(p_2\left(\mathbf{z}^1, k'\right), -p_2\left(\mathbf{z}^1, k'\right), 1 - p_2\left(\mathbf{z}^1, k'\right), -\left(1 - p_2\left(\mathbf{z}^1, k'\right)\right)\right)\pi.
\end{aligned}$$

First, $D_{p\mathbf{z}^1}$ is of full row rank. Second, by Assumption 4.3, the denominator in dp_{11} and dp_{22} is bounded away from 0, and all other elements in $D_{p\mathbf{z}^1}$ are bounded for any $\pi \in \Pi$. Thus, $D_{p\mathbf{z}^1}$ is bounded. There are only two types of non-zero elements in all Hessian vectors: payoff parameters and $-\frac{F''\left(F^{-1}\left(p_1\left(\mathbf{z}^1,k\right)\right)\right)}{\left[F'\left(F^{-1}\left(p_1\left(\mathbf{z}^1,k\right)\right)\right)\right]^3}$ or $-\frac{F''\left(F^{-1}\left(p_1\left(\mathbf{z}^1,k'\right)\right)\right)}{\left[F'\left(F^{-1}\left(p_1\left(\mathbf{z}^1,k'\right)\right)\right)\right]^3}$. Thus by Assumption 4.3 and 4.4 (i), all elements in the Hessian vectors are bounded, i.e., there exists M^* such that for any

j and k, it holds that $\sum_j \sum_k \left\| \frac{H_{p^*-j,k}}{2} \right\| < M^*$. For large enough n, the following inequality holds:

$$\left\| \sum_j \sum_k \frac{H_{p^*-j,k}}{2} \sqrt{n} \left(\widehat{p}_j - p_j\right) \left(\widehat{p}_k - p_k\right) \right\| \leq \sqrt{n} \sum_j \sum_k \left\| \frac{H_{p^*-j,k}}{2} \right\| \left\| \widehat{p}_{\mathbf{z}^1} - p_{\mathbf{z}^1} \right\|^2$$
$$\leq M^* \sqrt{n} \left\| \widehat{p}_{\mathbf{z}^1} - p_{\mathbf{z}^1} \right\|^2 = o_p(1).$$

This implies that

$$\begin{aligned}
&\sqrt{n} \left(G_{n\mathbf{z}^1}(\pi) - G_{\mathbf{z}^1}(\pi)\right) \\
&= D_{p\mathbf{z}^1} \sqrt{n} \left(\widehat{p}_{q\mathbf{z}^1} - p_{q\mathbf{z}^1}\right) + o_p(1) \\
&= D_{p\mathbf{z}^1} D_{a\mathbf{z}^1} \left(\sqrt{n} \left(\widehat{a}_{q\mathbf{z}^1} - a_{q\mathbf{z}^1}\right) + o_p(1)\right) + o_p(1) \\
&= D_{p\mathbf{z}^1} D_{a\mathbf{z}^1} \left(D_{q\mathbf{z}^1} \sqrt{n} \left(\widehat{q}_{q\mathbf{z}^1} - q_{q\mathbf{z}^1}\right) + o_p(1)\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{l=1}^{n} D_{p\mathbf{z}^1} D_{a\mathbf{z}^1} D_{q\mathbf{z}^1} \left(\eta_{l\mathbf{z}^1} - q_{\mathbf{z}^1}\right) + o_p(1) \text{ and}
\end{aligned}$$

$$G_{n\mathbf{z}^1}(\pi) = \frac{1}{n} \sum_{l=1}^{n} \left[D_{p\mathbf{z}^1} D_{a\mathbf{z}^1} D_{q\mathbf{z}^1} \left(\eta_{l\mathbf{z}^1} - q_{\mathbf{z}^1}\right) + G_{\mathbf{z}^1}(\pi)\right] + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where definitions of $D_{a\mathbf{z}^1}$ and $D_{q\mathbf{z}^1}$ can be found in Lemma A.2.1 and A.2.2. Without loss of generality, Lemma A.2.1 and A.2.2 prove the result for $\mathbf{z} = \mathbf{z}^1$. The same result holds for $\mathbf{z} = \mathbf{z}^2, \ldots, \mathbf{z}^9$, and such asymptotic linear representation of $G_{n\mathbf{z}}(\pi)$ is available for every observed state $\mathbf{z}$. Stacking the asymptotic linear representation for $\mathbf{z} = \mathbf{z}^1, \ldots, \mathbf{z}^9$, we obtain the asymptotic linear representation for player 1's "moment" function when holding his exclusive observed state at $z_1 = L$. The asymptotic linear representation is expressed as

$$G_n(\pi) = \frac{1}{n} \sum_{l=1}^{n} \left[D_p D_a D_q \left(\eta_l - q\right) + G(\pi)\right] + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where D_p, D_a, D_q, η_l, and q are obtained by stacking $D_{p\mathbf{z}}$, $D_{a\mathbf{z}}$, $D_{q\mathbf{z}}$, $\eta_{l\mathbf{z}}$, and $q_{\mathbf{z}}$ for $\mathbf{z} = \mathbf{z}^1, \ldots, \mathbf{z}^9$.

The CCPs, contingency tables, and indicator functions are different across $\mathbf{z}$. This implies that the coefficient matrix in front of $\left[[\eta_{l\mathbf{z}^1} - q_{\mathbf{z}^1}]^\top, \ldots, [\eta_{l\mathbf{z}^9} - q_{\mathbf{z}^9}]^\top\right]^\top$ is block diagonal, and each block has full row rank. Lastly, note that on $\mathbf{z}^9$, as $\Pr(\mathbf{z} = \mathbf{z}^9) = 1 - \sum_{s=1}^{8} \Pr(\mathbf{z} = \mathbf{z}^s)$ and $\mathbb{1}(\mathbf{z} = \mathbf{z}^9) = 1 - \left(\sum_{s=1}^{8} \mathbb{1}(\mathbf{z} = \mathbf{z}^s)\right)$, we drop these two elements from $q_{\mathbf{z}^9}$ and $\eta_{l\mathbf{z}^9}$, and denote the resulting vectors as $\widetilde{q}_{\mathbf{z}^9}$ and $\widetilde{\eta}_{l\mathbf{z}^s 9}$. The coefficient matrix in front of $\left[[\eta_{l\mathbf{z}^1} - q_{\mathbf{z}^1}]^\top, \ldots, [\widetilde{\eta}_{l\mathbf{z}^9} - \widetilde{q}_{\mathbf{z}^9}]^\top\right]^\top$ could be reduced to a block diagonal matrix (with each block having full row rank) via elementary transformations. Thus the scaled and demeaned "moment" function converges in distribution to a normal distribution with positive definite variance covariance matrix, and condition (i) in Assumption G3 is verified. Under Assumption 4.4, conditions (ii) and (iii) in Assumption G3 are satisfied. In the Simple Game, the degrees of freedom for the chi-squared distribution is calculated as: $\|c_0\|_0 - (l_\pi - l_R) = 5 + l_R$. By Theorem G2, the desired results follow. □

Lemma A.2.1. *Under Assumptions 2.3 and 4.4, it holds that for any $\xi \in \Xi_R$ and the parameter sequence $\{\xi_n\} \in \Xi_R(\xi)$,*

$$\sqrt{n}\left(\widehat{q}_{\mathbf{z}^1} - q_{\mathbf{z}^1}\right) = O_p(1) \text{ and} \tag{A.13}$$

$$\sqrt{n}\left(\widehat{a}_{\mathbf{z}^1} - a_{\mathbf{z}^1}\right) = \sqrt{n} D_{q\mathbf{z}^1}\left(\widehat{q}_{\mathbf{z}_1} - q_{\mathbf{z}^1}\right) + o_p(1), \text{ where} \tag{A.14}$$

$$D_{q\mathbf{z}^1} \equiv \begin{pmatrix} \frac{1}{q_9} & 0 & \ldots & 0 & -\frac{q_1}{q_9^2} \\ 0 & \frac{1}{q_9} & \ldots & 0 & -\frac{q_2}{q_9^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{1}{q_9} & -\frac{q_8}{q_9^2} \end{pmatrix}_{8\times 9}$$

in which $\widehat{q}_{\mathbf{z}^1}$, $q_{\mathbf{z}^1}$, $\widehat{a}_{\mathbf{z}^1}$, and $a_{\mathbf{z}^1}$ are defined in (A.8)-(A.11).

Proof of Lemma A.2.1: Since each element in $\eta_{l\mathbf{z}^1}$ (defined in (A.12)) is less than or equal to 1, the Lindeberg Condition is satisfied. Under Assumption 2.3, by Lyapunov CLT, Equation (A.13) holds. To obtain Equation (A.14), apply the Taylor Expansion of $\sqrt{n}\left(\widehat{a}_{\mathbf{z}^1} - a_{\mathbf{z}^1}\right)$

around q:

$$\sqrt{n}\left(\widehat{a}_{\mathbf{z}^1}-a_{\mathbf{z}^1}\right)=\sqrt{n}D_{q\mathbf{z}^1}\left(\widehat{q}_{\mathbf{z}^1}-q_{\mathbf{z}^1}\right)+\sqrt{n}\sum_j\sum_k\frac{H_{q^*-j,k}}{2}\left(\widehat{q}_j-q_j\right)\left(\widehat{q}_k-q_k\right),$$

where $D_{q\mathbf{z}^1}$ is defined in the lemma, and $H_{q^*-j,k}$ is the Hessian vector that stores the second order derivatives with respect to q_j and q_k evaluated at $q^*_{\mathbf{z}^1}$, which is a point that lies between $q_{\mathbf{z}^1}$ and $\widehat{q}_{\mathbf{z}^1}$. Each element in $H_{q^*-j,k}$ is either $-\frac{1}{q_9^{*2}}$, $\frac{2q_j^*}{q_9^{*3}}$, or 0. Notice that

$$\left\|\sqrt{n}\sum_j\sum_k\frac{H_{q^*-j,k}}{2}\left(\widehat{q}_j-q_j\right)\left(\widehat{q}_k-q_k\right)\right\|\leq\sqrt{n}\sum_j\sum_k\left\|\frac{H_{q^*-j,k}}{2}\right\|\left\|\widehat{q}_{\mathbf{z}^1}-q_{\mathbf{z}^1}\right\|^2.$$

For any $\epsilon>0$, there exists $N_{\epsilon,q}>0$ such that $\Pr\left(\left\|\widehat{q}_{\mathbf{z}^1}-q_{\mathbf{z}^1}\right\|<\frac{\delta_1}{2}\right)\geq1-\epsilon$ for $n\geq N_{\epsilon,q}$ by triangular array WLLN; and since $q_9=\Pr\left(\mathbf{z}=\mathbf{z}^1\right)\geq\delta_1>0$, we obtain that $\Pr\left(q_9^*\geq\frac{\delta_1}{2}\right)\geq 1-\epsilon$ for $n\geq N_{\epsilon,q}$. For $q_9^*\geq\frac{\delta_1}{2}$, there exists M_1 such that $\sum_j\sum_k\left\|\frac{H_{q^*-j,k}}{2}\right\|\leq M_1<\infty$. Therefore, by Equation (A.13) and triangular array WLLN, we have

$$\sqrt{n}\sum_j\sum_k\left\|\frac{H_{q^*-j,k}}{2}\right\|\left\|\widehat{q}_{\mathbf{z}^1}-q_{\mathbf{z}^1}\right\|^2\leq M_1\sqrt{n}\left\|\widehat{q}_{\mathbf{z}^1}-q_{\mathbf{z}^1}\right\|^2$$

with probability at least $1-\epsilon$ for n sufficiently large, where $\epsilon>0$ is arbitrary. Because $\left\|\widehat{q}_{\mathbf{z}^1}-q_{\mathbf{z}^1}\right\|=O_p\left(n^{-1/2}\right)$, Equation (A.14) holds. $\square$

Lemma A.2.2. *Under Assumption 2.1-2.5, Assumption 4.2, and Assumption 4.4, it holds that*

$$\sqrt{n}\left(\widehat{p}_{\mathbf{z}^1}-p_{\mathbf{z}^1}\right)=\sqrt{n}D_{a\mathbf{z}^1}\left(\widehat{a}_{\mathbf{z}^1}-a_{\mathbf{z}^1}\right)+o_p(1)\qquad\text{(A.15)}$$

for any $\xi\in\Xi_R$ *and the parameter sequence* $\{\xi_n\}\in\Xi_R(\xi)$, *where* $D_{a\mathbf{z}^1}\equiv\frac{\partial p_{\mathbf{z}^1}}{\partial a^\top_{\mathbf{z}^1}}$, *and* $\widehat{p}_{\mathbf{z}^1}$ *and* $p_{\mathbf{z}^1}$ *are defined in equation (A.7).*

Proof of Lemma A.2.2: Second order Taylor expansion of $\widehat{p}_{\mathbf{z}^1}$ at a provides that

$$\sqrt{n}\left(\widehat{p}_{\mathbf{z}^1}-p_{\mathbf{z}^1}\right)=\sqrt{n}D_{a\mathbf{z}^1}\left(\widehat{a}_{\mathbf{z}^1}-a_{\mathbf{z}^1}\right)+\sqrt{n}\sum_j\sum_k\frac{H_{a^*-j,k}}{2}\left(\widehat{a}_j-a_j\right)\left(\widehat{a}_k-a_k\right),$$

where $H_{a^*-j,k}$ is the Hessian vector that stores second order derivatives with respect to a_j and a_k evaluated at $a^*_{\mathbf{z}^1}$, which is a point between $a_{\mathbf{z}^1}$ and $\widehat{a}_{\mathbf{z}^1}$. Under Assumption 4.2, there exists a neighborhood around the true value of a such that in this neighborhood the eigenvector function of $A^{12}_{1\mathbf{z}^1}\left(A^{12}_{\mathbf{z}^1}\right)^{-1}$ (a matrix whose elements are continuous functions of a) is analytic. CCPs for player 1 are delivered by the eigenvector function, and CCPs for other players are continuously differentiable transformation of player 1's CCPs. Because $\widehat{a}_{\mathbf{z}^1} \xrightarrow{p} a_{\mathbf{z}^1}$, for large enough n, $H^*_{a-j,k}$ is bounded with probability close to 1. By a similar reasoning as in the previous lemma, the claim in the lemma holds. □

Proof of Proposition 4.4: For (i), after adapting Assumptions 2.1-2.6, 4.1-4.4 to the games in Sections 3.1-3.3, the proof of size control is the same as for the Simple Game except that now the degrees of freedom for the chi-squared distribution is different: it equals $13 + l_R$ for the game in Section 3.1, $6 + l_R$ for the game in Section 3.2, and $2 + l_R$ for the game in Section 3.3. Thus the same test statistic together with $\chi^2_{[13+l_R],1-\alpha}$ for the game in Section 3.1, with $\chi^2_{[6+l_R],1-\alpha}$ for the game in Section 3.2, and with $\chi^2_{[2+l_R],1-\alpha}$ for the game in Section 3.3 achieve asymptotic size control when $\xi \in \Xi_R$ and achieve consistency when $\xi \notin \Xi_R$. □

Proof of Proposition 4.5: It suffices to show that $W^b_n(c) \xrightarrow{p} W^b(c)$, where $W^b(c)$ is positive definite for any c and $W^b(c_0) = \Omega_0^{-1}$. Under Assumption 2.5, Γ_{c_0} is of full column rank and $\operatorname{rank}(\Gamma_{c_0}\Psi) = l_\pi - l_R$. For n sufficiently large, we have $\operatorname{rank}(\Gamma_{nc_0}\Psi) = l_\pi - l_R$ and $\arg\min_{\pi\in\Pi} \|G_{nc_0}(\Psi\pi_f + \mu)\|^2$ is unique. Therefore, $W^b_n(c_0) = \left(\Sigma^b_n(c_0, \widehat{\pi}_f(c_0))\right)^{-1}$ for large enough n.

Latent states are matched across bootstrap draws with probability approaching one based on the known function $\psi(\cdot)$ described in Assumption 4.5. Under Assumptions 2.1-2.6, 4.2, 4.3 and Assumption 4.4 (i) and (ii) the "moment" functions for the Simple Game admits an asymptotic linear representation by Proposition 4.3. Thus it holds that

$$G_{nc_0}^{(b)}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)=\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l^{(b)},\theta,\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)+o_p\left(1\right).$$

Since $\widehat{\pi}_f\left(c_0\right)\xrightarrow{p}\pi_f^*\left(c_0\right)$ and $\pi_0=\Psi\pi_f^*\left(c_0\right)+\mu$ under the null hypothesis, where $\pi_f^*\left(c_0\right)=\arg\min_{\pi\in\Pi}\left\|G_{c_0}\left(\Psi\pi_f+\mu\right)\right\|^2$, we have $\Psi\widehat{\pi}_f\left(c_0\right)+\mu\xrightarrow{p}\pi_0$. Under Assumptions 2.1-2.5, 4.2, 4.3, and Assumption 4.4 (i) and (ii), it holds that

$$\sqrt{n}\left(G_{nc_0}\left(\Psi\pi_f^*\left(c_0\right)+\mu\right)-G_{c_0}\left(\Psi\pi_f^*\left(c_0\right)+\mu\right)\right)\xrightarrow{d}N\left(0,\Omega_0\right)$$

and

$$\begin{aligned}
&\sqrt{n}\left(G_{nc_0}^{(b)}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)-G_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)\right)\\
=&\sqrt{n}\left(\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l^{(b)},\theta,\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)-\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l,\theta,\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)\right)+o_p^*\left(1\right)\\
=&\sqrt{n}\left(\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l^{(b)},\theta,\pi_0\right)-\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l,\theta,\pi_0\right)\right)+o_p^*\left(1\right)\xrightarrow{d^*}N\left(0,\Omega_0\right),
\end{aligned}$$

where the definitions of $o_p^*\left(1\right)$ and $\xrightarrow{d^*}$ can be found in Hansen (2020). The first equality holds because conditional on data, the difference between $G_{nc_0}^{(b)}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)$ and its linear representation is $o_p\left(1\right)$, and the difference between $G_{nc_0}\left(\Psi\widehat{\pi}_f\left(c_0\right)+\mu\right)$ and its linear representation is $o\left(1\right)$. Conditional on data, convergence in distribution holds by Theorem 10.8 in Hansen (2020) as $\left\|\phi_{c_0}\left(Q_l,\theta,\pi_0\right)\right\|^2$ is uniformly square integrable by the proof of Proposition 4.3. By Theorem 10.13 in Hansen (2020), $\Sigma_n^b\left(c_0,\widehat{\pi}_f\left(c_0\right)\right)=\Omega_0+o_p\left(1\right)$ holds if $y_n^{(b)}$ is uniformly square integrable, where

$$y_n^{(b)}\equiv\sqrt{n}\left(\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l^{(b)},\theta,\pi_0\right)-\frac{1}{n}\sum_{l=1}^{n}\phi_{c_0}\left(Q_l,\theta,\pi_0\right)\right).$$

Because the uniform square integrability of a vector is implied by each element of the vector being uniformly square integrable, let $y_n^{(b)}\left(i\right)$ be its ith element. It holds that $E^*\left|y_n^{(b)}\left(i\right)\right|^4=$

$\frac{\widehat{\mu}_4(i)-3\widehat{\sigma}^4(i)}{n}+3\widehat{\sigma}^4(i)$, where

$$\begin{aligned}\widehat{\mu}_4(i) &= \frac{1}{n}\sum_{l=1}^{n}\left|\phi_{c0}^{(i)}\left(Q_l^{(b)},\theta,\pi_0\right)-\frac{1}{n}\sum_{l=1}^{n}\phi_{c0}^{(i)}\left(Q_l,\theta,\pi_0\right)\right|^4 \text{ and}\\ \widehat{\sigma}^2(i) &= \frac{1}{n}\sum_{l=1}^{n}\left|\phi_{c0}^{(i)}\left(Q_l^{(b)},\theta,\pi_0\right)-\frac{1}{n}\sum_{l=1}^{n}\phi_{c0}^{(i)}\left(Q_l,\theta,\pi_0\right)\right|^2,\end{aligned}$$

with $\phi_{c0}^{(i)}\left(Q_l^{(b)},\theta,\pi_0\right)$ and $\phi_{c0}^{(i)}\left(Q_l,\theta,\pi_0\right)$ being the ith element in $\phi_{c_0}\left(Q_l^{(b)},\theta,\pi_0\right)$ and $\phi_{c_0}\left(Q_l,\theta,\pi_0\right)$ respectively. Uniform square integrability of $\left|\phi_{c0}^{(i)}\left(Q_l,\theta,\pi_0\right)\right|^2$ implies that $\frac{\widehat{\mu}_4(i)}{n}=o_p(1)$ and $\widehat{\sigma}^2(i)=O_p(1)$, which imply that $E^*\left|y_n^{(b)}(i)\right|^4=O_p(1)$. Therefore $y_n^{(b)}$ is uniformly square integrable and $\Sigma_n^b\left(c_0,\widehat{\pi}_f(c_0)\right)=\Omega_0+o_p(1)$ holds. Ω_0 is positive definite, because the asymptotic variance matrix of

$$\sqrt{n}\left(G_n\left(\Psi\pi_f^*(c_0)+\mu\right)-G\left(\Psi\pi_f^*(c_0)+\mu\right)\right)$$

is positive definite and Ω_0 is its submatrix with the corresponding rows and columns selected by c_0.

It remains to prove that for $c\neq c_0$, $W_n^b(c)$ converges in probability to some positive definite matrix. If $\operatorname{rank}(\Gamma_{nc}\Psi)<l_\pi-l_R$, then $W_n^b(c)=W_P(c)$, which is positive definite. If $\operatorname{rank}(\Gamma_c\Psi)=l_\pi-l_R$, then $\operatorname{rank}(\Gamma_{nc}\Psi)=l_\pi-l_R$ and $\arg\min_{\pi\in\Pi}\left\|G_{nc}\left(\Psi\pi_f+\mu\right)\right\|^2$ is unique for sufficiently large n. By a similar argument as the one for $W_n^b(c_0)$, it holds that $W_n^b(c)=\left(\Sigma_n^b\left(c,\widehat{\pi}_f(c)\right)\right)^{-1}=\Omega_c^{-1}+o_p(1)$, where Ω_c is the asymptotic variance of $\sqrt{n}\left(G_{nc}\left(\Psi\pi_f^*(c)+\mu\right)-G_c\left(\Psi\pi_f^*(c)+\mu\right)\right)$ with $\pi_f^*(c)=\operatorname{plim}\widehat{\pi}_f(c)$. Ω_c is positive definite because the asymptotic variance matrix of

$$\sqrt{n}\left(G_n\left(\Psi\pi_f^*(c)+\mu\right)-G\left(\Psi\pi_f^*(c)+\mu\right)\right)$$

is positive definite, and Ω_c is its submatrix with the corresponding rows and columns selected by c. $\square$

Proof of Lemma 5.1: First note that Assumption 5.1-Assumption 5.4 are sufficient for identification of CCPs up to a label swapping. As $\text{rank}\left([\overline{\pi}_c, \Gamma_c]\right) > \text{rank}\left(\Gamma_c\right)$ for any $c \notin \mathscr{C}\mathscr{Z}$, $\mathscr{C}\mathscr{Z}$ is identified in $\mathscr{C}$. Furthermore, c_0 is defined as the vector in $\mathscr{C}\mathscr{Z}$ that selects the most "moments", which happens only when all equilibria corresponding to a latent state is selected. Thus c_0 is unique. Γ_{c_0} has full column rank implies that the system selected by c_0 uniquely determines π_0. $\square$

Proof of Theorem EG1: Under Assumption EG1 and Assumption EG2, the proof for $\widehat{c} = c_0$ wp $\to 1$ follows exactly the proof of Theorem 1 (a) in Andrews (1999). Given $\widehat{c} = c_0$, the proof for consistency of $\widehat{\pi}$ is exactly the same as the proof given in Theorem G1. $\square$

Proof of Theorem EG2: Without loss of generality, suppose $\mathscr{C}\mathscr{Z} = \{c_0, c_1, \ldots, c_q\}$. Under Assumption EG1, EG3 and when H_0 is true, it holds that

$$T_n \leq \min\left\{\chi^2_{\left[\|c_0\|_0-(l_\pi-l_R)\right],1-\alpha}, \ldots, \chi^2_{\left[\|c_q\|_0-(l_\pi-l_R)\right],1-\alpha}\right\} \leq \chi^2_{[cl-(l_\pi-l_R)],1-\alpha}.$$

Thus, using $\chi^2_{[cl-(l_\pi-l_R)],1-\alpha}$ achieves asymptotic size control. The consistency of the test follows from the similar argument in the proof of Theorem G2. $\square$

Proof of Proposition 5.1: Under Assumptions 5.1-5.7, following a similar argument as in the proof of Proposition 4.3, Assumption EG1 and EG2 are verified. Thus we apply Theorem EG1 and the results hold. $\square$

Proof of Proposition 5.2: Under Assumptions 5.1-5.5, 5.7, and 5.8, Assumption EG1 and EG3 are verified. Applying Theorem EG2, the results hold. $\square$

A.3 Three Variants of the Extended Game

Similar to the three variants of the Simple Game introduced in Section 3, we introduce three variants of the Extended Game below by specifying their corresponding $\overline{\pi}$ and Γ in the generic "moments" function $G\left(\pi\right) = \overline{\pi} - \Gamma\pi$.

A.3.1 The Extended Game with Common Observed State

Suppose that the common observed state variable x takes values in $\left\{x^1, ..., x^V\right\}$. We normalize the payoff functions for the first action $d_i = 0$ to 0. Player 1's payoff function when choosing $d_1 = j$ is given by

$$\pi_1\left(j, \mathbf{d}_{-1}, z_1, x, k\right) = \beta_{1kj} x + \pi_1\left(1, \mathbf{d}_{-1}, z_1, k\right).$$

Player 1's expected payoff vector and coefficient matrix when his exclusive observed state z_1 is fixed at the smallest value in S_z are as follows:

$$\begin{aligned}
\overline{\pi} &= \left[\overline{\pi}_1\left(\mathbf{z}^1, x^1\right), \ldots, \overline{\pi}_1\left(\mathbf{z}^T, x^1\right), \ldots, \overline{\pi}_1\left(\mathbf{z}^1, x^V\right), \ldots, \overline{\pi}_1\left(\mathbf{z}^T, x^V\right)\right]^\top \text{ and} \\
\Gamma &= \left[\Gamma_1\left(\mathbf{z}^1, x^1\right)^\top, \ldots, \Gamma_1\left(\mathbf{z}^T, x^1\right)^\top, \ldots, \Gamma_1\left(\mathbf{z}^1, x^V\right)^\top, \ldots, \Gamma_1\left(\mathbf{z}^T, x^V\right)^\top\right]^\top,
\end{aligned}$$

where the submatrices on each observed state are defined as

$$\overline{\pi}_1\left(\mathbf{z}, x\right) = \begin{bmatrix} \overline{\pi}_1\left(1, \mathbf{z}, x, \omega\left(1, \mathbf{z}, x\right)\right) \\ \vdots \\ \overline{\pi}_1\left(J, \mathbf{z}, x, \omega\left(1, \mathbf{z}, x\right)\right) \\ \vdots \\ \overline{\pi}_1\left(1, \mathbf{z}, x, \omega\left(\left|S_{\omega|(\mathbf{z},x)}\right|, \mathbf{z}, x\right)\right) \\ \vdots \\ \overline{\pi}_1\left(J, \mathbf{z}, x, \omega\left(\left|S_{\omega|(\mathbf{z},x)}\right|, \mathbf{z}, x\right)\right) \end{bmatrix} \text{ and } \Gamma_1\left(\mathbf{z}, x\right) = \begin{bmatrix} \iota_1\left(1, \mathbf{z}, x\right) \\ \vdots \\ \iota_1\left(\left|S_{\omega|(\mathbf{z},x)}\right|, \mathbf{z}, x\right) \end{bmatrix}.$$

Note that matrix $\iota\left(s, \mathbf{z}, x\right)$ with $s \in \left\{1, \ldots, \left|S_{\omega|(\mathbf{z},x)}\right|\right\}$ is a block diagonal matrix with J identical blocks defined as:

$$\iota_1\left(s, \mathbf{z}, x\right) = \begin{bmatrix} \left(x,\ \mathbf{p}_{-1}\left(\mathbf{z}, x, \omega\left(\left(s, \mathbf{z}, x\right)\right)\right)\right. & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \left(x,\ \mathbf{p}_{-1}\left(\mathbf{z}, x, \omega(s, \mathbf{z}, x)\right)\right) \end{bmatrix},$$

with $\mathbf{p}_{-1}\left(\mathbf{z}, x, \omega\left(s, \mathbf{z}, x\right)\right)$ being a 1 by $\left(J+1\right)^{N-1}$ vector consisting of the probabilities of joint actions for all other players. Vector $\overline{\pi}$ is of dimension $\sum_{t=1}^{T} \sum_{v=1}^{V} J \left|S_{\omega|(\mathbf{z}^t, x^v)}\right|$, and Γ is of dimension $\sum_{t=1}^{T}\sum_{v=1}^{V} J\left|S_{\omega|(\mathbf{z}^t,x^v)}\right|$ by $J+J\left(J+1\right)^{N-1}$.

For this game, there are $|S_k|$ selection vector spaces defined as follows: for $s=1,...,|S_k|$,

$$\mathscr{C}^s \equiv \left\{\begin{array}{c} \left(c_1,...,c_{TV}\right)^\top : c_1 = \left[c_{1,1},\ldots,c_{1,\left|S_{\omega|(\mathbf{z}^1,x^1)}\right|}\right] \text{ with } c_{1,s}=\mathbf{e}_1 \text{ and } c_{1,j}=\mathbf{e}_0 \text{ for } j\neq s; \\ \text{for } t=2,\ldots,TV,\ c_t = \left[c_{t,1},\ldots,c_{t,\left|S_{\omega|(\mathbf{z}^p,x^q)}\right|}\right] \text{ with } c_{t,w}\in\left\{\mathbf{e}_1,\mathbf{e}_0\right\}, \\ \text{where } w\in\left\{1,\ldots,\left|S_{\omega|(\mathbf{z}^p,x^q)}\right|\right\} \text{ and } p\neq 1,\ \text{or } q\neq 1 \end{array}\right\}.$$

For any $\mathscr{C}^s$, $cl = JTV$. Given some selection vector space with the first selected latent state being k, the unique solution to the system $G_{c_0}(\pi)=\mathbf{0}$ gives us the true payoff vector written as $\left[\beta_{1k1},\pi_1\left(1,z_1,k\right),\ldots,\beta_{1kJ},\pi_1\left(J,z_1,k\right)\right]^\top$, where for $j=1,\ldots,J$,

$$\pi_1\left(j,z_1,k\right)=\left[\pi_1\left(j,\left(0,\ldots,0\right),z_1,k\right),\ldots,\pi_1\left(j,\left(J,\ldots,J\right),z_1,k\right)\right].$$

The true payoff vector is of dimension $J+J\left(J+1\right)^{N-1}$ by 1.

A.3.2 An Extended Parametric Game with Game-Level Random Coefficients

Let $\left\{\mathbf{z}^1,\ldots,\mathbf{z}^{|S_z|T}\right\}\equiv S_{\mathbf{z}}$ be the collection of observed state vectors. Player i's payoff function when choosing $d_i=j$ is

$$\pi_i\left(j,\mathbf{d}_{-i},z_i,k\right)=z_i\theta_{ikj}+\delta_{ikj}\left(\sum_{q\neq i}\mathbb{1}\left(d_q=j\right)\right). \tag{A.16}$$

Player i's expected payoff vector and coefficient matrix are written as:

$$\overline{\pi}=\left[\overline{\pi}_i\left(\mathbf{z}^1\right),\ldots,\overline{\pi}_i\left(\mathbf{z}^{|S_z|T}\right)\right]^\top \text{ and } \Gamma=\left[\Gamma_i\left(\mathbf{z}^1\right)^\top,\ldots,\Gamma_i\left(\mathbf{z}^{|S_z|T}\right)^\top\right]^\top.$$

The submatrices for each observed state are defined as

$$\overline{\pi}_i(\mathbf{z}) = \begin{bmatrix} \overline{\pi}_i(1, \mathbf{z}, \omega(1, \mathbf{z})) \\ \vdots \\ \overline{\pi}_i(J, \mathbf{z}, \omega(1, \mathbf{z})) \\ \vdots \\ \overline{\pi}_i\left(1, \mathbf{z}, \omega\left(\left|S_{\omega|\mathbf{z}}\right|, \mathbf{z}\right)\right) \\ \vdots \\ \overline{\pi}_i\left(J, \mathbf{z}, \omega\left(\left|S_{\omega|\mathbf{z}}\right|, \mathbf{z}\right)\right) \end{bmatrix} \text{ and } \Gamma_i(\mathbf{z}) = \begin{bmatrix} \iota_i(1, \mathbf{z}) \\ \vdots \\ \iota_i\left(\left|S_{\omega|\mathbf{z}}\right|, \mathbf{z}\right) \end{bmatrix}.$$

Note that matrix $\iota(s, \mathbf{z})$ with $s \in \{1, \ldots, \left|S_{\omega|\mathbf{z}}\right|\}$ is a block diagonal matrix with J blocks defined as:

$$\iota_i(s, \mathbf{z}) = \begin{bmatrix} \left(z_i, \sum_{q \neq i} p_{q1}(\mathbf{z}, \omega(s, \mathbf{z}))\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \left(z_i, \sum_{q \neq i} p_{qJ}(\mathbf{z}, \omega(s, \mathbf{z}))\right) \end{bmatrix},$$

where $p_{qj}(\mathbf{z}, \omega(s, \mathbf{z}))$ is player q's equilibrium CCP of choosing action j for given $\mathbf{z}$ and $\omega(s, \mathbf{z})$. Vector $\overline{\pi}$ is of dimension $\sum_{t=1}^{|S_z|T} J\left|S_{\omega|\mathbf{z}^t}\right|$, and matrix Γ is of dimension $\sum_{t=1}^{|S_z|T} J\left|S_{\omega|\mathbf{z}^t}\right|$ by $2J$.

For this game, there are $|S_k|$ selection vector spaces to work with. They are defined as follows: for $s = 1, ..., |S_k|$,

$$\mathscr{C}^s \equiv \left\{ \begin{array}{c} \left(c_1, ..., c_{|S_z|T}\right)^\top : c_1 = \left[c_{1,1}, \ldots, c_{1,\left|S_{\omega|\mathbf{z}^1}\right|}\right] \text{ with } c_{1,s} = \mathbf{e}_1 \text{ and } c_{1,j} = \mathbf{e}_0 \text{ for } j \neq s; \\ \text{for } t = 2, \ldots, |S_z|\, T,\ c_t = \left[c_{t,1}, \ldots, c_{t,\left|S_{\omega|\mathbf{z}^t}\right|}\right] \text{ with } c_{t,w} \in \{\mathbf{e}_1, \mathbf{e}_0\}, \\ \text{where } w \in \left\{1, \ldots, \left|S_{\omega|\mathbf{z}^t}\right|\right\} \end{array} \right\}.$$

Given any selection vector space $\mathscr{C}^s$ for $s = 1, ..., |S_k|$, $cl = J\,|S_z|\,T$. Given some selection vector space with the first selected latent state being k, the unique solution to the system $G_{c_0}(\pi) = \mathbf{0}$ gives us the true payoff vector written as $(\theta_{ik1}, \delta_{ik1}, ..., \theta_{ikJ}, \delta_{ikJ})^\top$.[3] The true

[3]By convention, θ_{ik0} and δ_{ik0} are both normalized to 0.

payoff vector is of dimension $2J$.

A.3.3 An Extended Symmetric Game with Random Coefficients

Equation (3.6) provides the specification of the payoff function. Suppose the common observed state variable is x taking values in $\left\{x^1, ..., x^V\right\}$. The payoff function for $d_i = j$ is parametrized as

$$\pi_i\left(j, \mathbf{d}_{-i}, x, k\right) = x\beta_{kj} + \frac{1}{(N-1)}\delta_{kj}\sum_{q\neq i}\mathbb{1}\left(d_q = j\right) \equiv \pi\left(j, \mathbf{d}_{-i}, x, k\right). \tag{A.17}$$

The expected payoff vector of the representative player and coefficient matrix can be written as:

$$\overline{\pi} = \left[\overline{\pi}\left(x^1\right), \ldots, \overline{\pi}\left(x^V\right)\right]^\top \text{ and } \Gamma = \left[\Gamma\left(x^1\right)^\top, \ldots, \Gamma\left(x^V\right)^\top\right]^\top.$$

The submatrices for each observed state are defined as

$$\overline{\pi}\left(x\right) = \begin{bmatrix} \overline{\pi}\left(1, x, \omega\left(1, x\right)\right) \\ \vdots \\ \overline{\pi}\left(J, x, \omega\left(1, x\right)\right) \\ \vdots \\ \overline{\pi}\left(1, x, \omega\left(\left|S_{\omega|x}\right|, x\right)\right) \\ \vdots \\ \overline{\pi}\left(J, x, \omega\left(\left|S_{\omega|x}\right|, x\right)\right) \end{bmatrix} \text{ and } \Gamma\left(x\right) = \begin{bmatrix} \iota\left(1, x\right) \\ \vdots \\ \iota\left(\left|S_{\omega|x}\right|, x\right) \end{bmatrix},$$

Note that matrix $\iota\left(s, x\right)$ with $s \in \left\{1, \ldots, \left|S_{\omega|x}\right|\right\}$ is a block diagonal matrix with J blocks defined as:[4]

$$\iota\left(s, x\right) = \begin{bmatrix} \left(x, p_1\left(x, \omega\left(s, x\right)\right)\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \left(x, p_J\left(x, \omega\left(s, x\right)\right)\right) \end{bmatrix},$$

[4] Note that the block diagonal structure is a consequence of $J > 1$, as now each row in $\overline{\pi}_{1z_1}$ corresponds to a single action for player 1 while the payoff vector stacks other actions as well.

where $p_j(x, \omega(s, x))$ is the representative player's equilibrium CCP of choosing action j for given x and $\omega(s, x)$. Vector $\overline{\pi}$ is of dimension $\sum_{v=1}^{V} J\left|S_{\omega|x^v}\right|$, and matrix Γ is of dimension $\sum_{v=1}^{V} J\left|S_{\omega|x^v}\right|$ by $2J$.

For this game, there are $|S_k|$ selection vector spaces to work with. They are defined as follows: for $s = 1, ..., |S_k|$,

$$\mathscr{C}^s \equiv \left\{ \begin{array}{c} (c_1, ..., c_V)^\top : c_1 = \left[c_{1,1}, \ldots, c_{1,\left|S_{\omega|x^1}\right|}\right] \text{ with } c_{1,s} = \mathbf{e}_1 \text{ and } c_{1,j} = \mathbf{e}_0 \text{ for } j \neq s; \\ \text{for } \; t = 2, \ldots, V, \; c_t = \left[c_{t,1}, \ldots, c_{t,\left|S_{\omega|x^t}\right|}\right] \text{ with } c_{t,w} \in \{\mathbf{e}_1, \mathbf{e}_0\}, \\ \text{where } w \in \left\{1, \ldots, \left|S_{\omega|x^t}\right|\right\} \end{array} \right\}.$$

We have $cl = JV$ for any $\mathscr{C}^s$. Given some selection vector space with the first selected latent state being k, the unique solution to the system $G_{c_0}(\pi) = \mathbf{0}$ gives us the true payoff vector written as $(\theta_{k1}, \delta_{k1}, ..., \theta_{kJ}, \delta_{kJ})^\top$. The true payoff vector is of dimension $2J$.

A.3.4 Estimation and Inference

Consistency of the estimated payoff vector and selected matching in the three variants of the Extended Game holds by following similar arguments to those in the Extended Game and adapting Assumptions 5.1-5.7 to the variants of the Extended Game with $\pi_0 = [\beta_{1k1}, \pi_1(1, z_1, k), ..., \beta_{1kJ}, \pi_1(J, z_1, k)]^\top$ for the extended game with common observed state; $\pi_0 = (\theta_{ik1}, \delta_{ik1}, ..., \theta_{ikJ}, \delta_{ikJ})^\top$ for the extended parametric game with game-level random coefficient; and $\pi_0 = (\theta_{k1}, \delta_{k1}, ..., \theta_{kJ}, \delta_{kJ})^\top$ for the extended symmetric game with random coefficients. That is, $\widehat{c} = c_0$ wp $\to 1$ and $\widehat{\pi} \xrightarrow{p} \pi_0$ hold for each game.

Adapting Assumptions 5.1-5.5, 5.7, and 5.8 to the three variants of the Extended Game, we conclude that tests based on the test statistics defined in (4.3) and critical values $\chi^2_{[JTV-J(J+1)^{N-1}-J+l_R],1-\alpha}$, $\chi^2_{[JT|S_z|-2J+l_R],1-\alpha}$, and $\chi^2_{[JV-2J+l_R],1-\alpha}$ for the three variants of the Extended Game achieve asymptotic size control in the sense of (5.2). Moreover, adapting assumptions imposed on the Extended Game for the validity of the bootstrap to all variants of the Extended Game, we have that the test implemented using bootstrap variance achieves asymptotic size control and is consistent for each game.

Appendix B

A SIMPLE SEMIPARAMETRIC ESTIMATOR FOR STATIC GAMES OF INCOMPLETE INFORMATION WITH ACTION COMPLEMENTARITY

B.1 Support Condition Illustration for other x in Example 1

Table B.1: Illustration of Support Condition ($x = -1$)

Support of S_1^1 and S_2^1:	[-5.8964,2.1036]	[-5.1776,2.8224]
Support of S_1^2 and S_2^2:	[-5.3844,2.6156]	[-6.1332,1.8668]
Support of $-z_1^i$ and $-z_2^i$:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^1$ and $\widetilde{S}_2^1$:	[-1.1036,6.8964]	[-1.8224,6.1776]
Support of $\widetilde{S}_1^2$ and $\widetilde{S}_2^2$:	[-0.61563,7.3844]	[0.13316,8.1332]
Support of z_1^i and z_2^i:	[-10,10]	[-10,10]
Support of $S_1^{'1}$ and $S_2^{'1}$:	[-6.8964,1.1036]	[-2.8224,5.1776]
Support of $S_1^{'2}$ and $S_2^{'2}$:	[-7.3844,0.61563]	[-1.8668,6.1332]
Support of $-z_1^i$ and z_2^i:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^{'1}$ and $\widetilde{S}_2^{'1}$:	[-2.1036,5.8964]	[-6.1776,1.8224]
Support of $\widetilde{S}_1^{'2}$ and $\widetilde{S}_2^{'2}$:	[-2.6156,5.3844]	[-8.1332,-0.13316]
Support of z_1^i and $-z_2^i$:	[-10,10]	[-10,10]

Table B.2: Illustration of Suppor Condition ($x = -0.5$)

Support of S_1^1 and S_2^1:	[-5.1661,2.8339]	[-4.6944,3.3056]
Support of S_1^2 and S_2^2:	[-4.9197,3.0803]	[-5.4167,2.5833]
Support of $-z_1^i$ and $-z_2^i$:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^1$ and $\widetilde{S}_2^1$:	[-1.8339,6.1661]	[-2.3056,5.6944]
Support of $\widetilde{S}_1^2$ and $\widetilde{S}_2^2$:	[-1.0803,6.9197]	[-0.58333,7.4167]
Support of z_1^i and z_2^i:	[-10,10]	[-10,10]
Support of $S_1^{'1}$ and $S_2^{'1}$:	[-6.1661,1.8339]	[-3.3056,4.6944]
Support of $S_1^{'2}$ and $S_2^{'2}$:	[-6.9197,1.0803]	[-2.5833,5.4167]
Support of $-z_1^i$ and z_2^i:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^{'1}$ and $\widetilde{S}_2^{'1}$:	[-2.8339,5.1661]	[-5.6944,2.3056]
Support of $\widetilde{S}_1^{'2}$ and $\widetilde{S}_2^{'2}$:	[-3.0803,4.9197]	[-7.4167,0.58333]
Support of z_1^i and $-z_2^i$:	[-10,10]	[-10,10]

Table B.3: Illustration of Support Condition ($x = 0.5$)

Support of S_1^1 and S_2^1:	[-3.7054,4.2946]	[-3.7279,4.2721]
Support of S_1^2 and S_2^2:	[-3.9903,4.0097]	[-3.9837,4.0163]
Support of $-z_1^i$ and $-z_2^i$:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^1$ and $\widetilde{S}_2^1$:	[-3.2946,4.7054]	[-3.2721,4.7279]
Support of $\widetilde{S}_1^2$ and $\widetilde{S}_2^2$:	[-2.0097,5.9903]	[-2.0163,5.9837]
Support of z_1^i and z_2^i:	[-10,10]	[-10,10]
Support of $S_1^{'1}$ and $S_2^{'1}$:	[-4.7054,3.2946]	[-4.2721,3.7279]
Support of $S_1^{'2}$ and $S_2^{'2}$:	[-5.9903,2.0097]	[-4.0163,3.9837]
Support of $-z_1^i$ and z_2^i:	[-10,10]	[-10,10]
Support of $\widetilde{S}_1^{'1}$ and $\widetilde{S}_2^{'1}$:	[-4.2946,3.7054]	[-4.7279,3.2721]
Support of $\widetilde{S}_1^{'2}$ and $\widetilde{S}_2^{'2}$:	[-4.0097,3.9903]	[-5.9837,2.0163]
Support of z_1^i and $-z_2^i$:	[-10,10]	[-10,10]

B.2 *Proof of Lemma 2.3.1*

We layout the steps in details, but essentially it is the argument in Fox and Lazzati (2017) with the utility function replaced by the expected payoff function for each player. In particular, the presence of equilibrium belief doesn't affect this argument as the equilibrium belief is a function of x only.

We first show that if $\alpha_i \geq 0$, the equilibrium expected payoff $\pi^{\dagger i}(\mathbf{d}^i, \mathbf{z}^i, x, \boldsymbol{\epsilon}^i)$ is supermodular in (d_1, d_2). This is established as follows:

$$\begin{aligned} & \pi^{\dagger i}\left([(1,1) \vee (0,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left([(1,1) \wedge (0,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) \\ = \; & \pi^{\dagger i}\left((1,1), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left((0,0), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right), \end{aligned}$$

$$\begin{aligned} & \pi^{\dagger i}\left([(1,0) \vee (0,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left([(1,0) \wedge (0,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) \\ = \; & \pi^{\dagger i}\left((1,0), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left((0,0), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right), \end{aligned}$$

$$\begin{aligned} & \pi^{\dagger i}\left([(0,1) \vee (0,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left([(0,1) \wedge (0,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) \\ = \; & \pi^{\dagger i}\left((0,1), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left((0,0), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right), \end{aligned}$$

$$\begin{aligned} & \pi^{\dagger i}\left([(1,1) \vee (1,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left([(1,1) \wedge (1,0)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) \\ = \; & \pi^{\dagger i}\left((1,1), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left((1,0), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right), \end{aligned}$$

$$\begin{aligned} & \pi^{\dagger i}\left([(1,1) \vee (0,1)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left([(1,1) \wedge (0,1)], \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) \\ = \; & \pi^{\dagger i}\left((1,1), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right) + \pi^{\dagger i}\left((0,1), \mathbf{z}^i, x, \boldsymbol{\epsilon}^i\right), \end{aligned}$$

$$\begin{aligned}
& \pi^{\dagger i}\left([(1,0)\vee(0,1)],\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right)+\pi^{\dagger i}\left([(1,0)\wedge(0,1)],\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right) \\
= \; & \pi^{\dagger i}\left((1,1),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right)+\pi^{\dagger i}\left((0,0),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right) \\
= \; & \pi_{1}^{i}\left(z_{1}^{i},x\right)-\epsilon_{1}^{i}+\pi_{2}^{i}\left(z_{2}^{i},x\right)-\epsilon_{2}^{i}+\alpha^{i} \\
\geq \; & \pi_{1}^{i}\left(z_{1}^{i},x\right)-\epsilon_{1}^{i}+\pi_{2}^{i}\left(z_{2}^{i},x\right)-\epsilon_{2}^{i} \\
= \; & \pi^{\dagger i}\left((1,0),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right)+\pi^{\dagger i}\left((0,1),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right).
\end{aligned}$$

Note that the first five cases hold by definition, and the last case makes use of the fact that $\alpha^i \geq 0$. It remains to show that $\pi^{\dagger i}(d^i, \mathbf{z}^i, x, \boldsymbol{\epsilon}^i)$ has increasing differences in both (d_1^i, z_1^i) and (d_2^i, z_1^i) at least for some values of the state variables. Firstly we show that $\pi^{\dagger i}(d^i, \mathbf{z}^i, x, \boldsymbol{\epsilon}^i)$ has increasing differences in (d_1^i, z_1^i). Notice that

$$\begin{aligned}
\Delta_{d_1^i}\pi^{\dagger i}(d^i,\mathbf{z}^i,x,\boldsymbol{\epsilon}^i) &= \pi^{\dagger i}\left(\left(1,d_2^i\right),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right)-\pi^{\dagger i}\left(\left(0,d_2^i\right),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right) \\
&= z_1^i+\beta_1^i x+\delta_1^i P_1^{-i}(x)-\epsilon_1^i+\alpha^i d_2^i,
\end{aligned}$$

$$\text{and}\quad \frac{\partial \Delta_{d_1^i}\pi^{\dagger i}(d^i,\mathbf{z}^i,x,\boldsymbol{\epsilon}_i)}{\partial z_1^i}=1>0.$$

This shows $\pi^{\dagger i}(d^i, \mathbf{z}^i, x, \boldsymbol{\epsilon}^i)$ has increasing difference in (d_1^i, z_1^i).

Next we show that $\pi^{\dagger i}(d^i, x, \mathbf{z}^i, \boldsymbol{\epsilon}^i)$ has increasing differences in (d_2^i, z_1^i). Notice that:

$$\begin{aligned}
\Delta_{d_2^i}\pi^{*i}\left(d^i,x,\mathbf{z}^i,\boldsymbol{\epsilon}^i\right) &= \pi^{\dagger i}\left(\left(d_1^i,1\right),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right)-\pi^{\dagger i}\left((d_{i1},0),\mathbf{z}^{i},x,\boldsymbol{\epsilon}^{i}\right) \\
&= z_2^i+\beta_{i2}x_2+\delta_{i2}P_2^{-i}(x)-\epsilon_2^i+\alpha^i d_1^i,
\end{aligned}$$

$$\frac{\partial \Delta_{d_2^i}\pi^{\dagger i}(d^i,\mathbf{z}^i,x,\boldsymbol{\epsilon}^i)}{\partial z_1^i}=0\geq 0.$$

This shows $\pi^{\dagger i}(d^i, \mathbf{z}^i, x, \boldsymbol{\epsilon}^i)$ has increasing difference in (d_2^i, z_1^i). Define

$$g^i\left(\mathbf{Z}^i, X, \boldsymbol{\epsilon}^i\right) = \text{argmax}\{\pi^{\dagger i}(d^i, \mathbf{Z}^i, x, \boldsymbol{\epsilon}^i) : d^i \in \{0,1\}^2\}$$

Under Assumption 2.2.3, this maximizer is unique with probability 1. It follows by Topkis' theorem that $g^i\left(\mathbf{Z}^i, X, \boldsymbol{\epsilon}^i\right)$ increases (in the coordinatewise order) in z_1^i with probability 1. Thus, for all $z_1'^i > z_1^i$ and every upper set Ω in $\{0,1\}^2$, by Assumption 2.2.2 (ii), we have that

$$Pr\left(g^i\left(\mathbf{Z}^i, X, \boldsymbol{\epsilon}^i\right) \in \Omega | z_1'^i, z_2^i, x\right) \geq Pr\left(g^i\left(\mathbf{Z}^i, X, \boldsymbol{\epsilon}^i\right) \in \Omega | z_1^i, z_2^i, x\right).$$

That is, given (z_1^i, z_2^i, x), the random vector $g^i\left((\mathbf{Z}^i, X, \boldsymbol{\epsilon}_i\right)$ increases (in the sense of first order stochastic dominance) in z_1^i. Because stochastic dominance is preserved under marginalization, in the data, $Pr(d_2^i = 1|z_1^i, z_2^i, x)$ increases in z_1^i. Similarly, we can show that $Pr(d_1^i = 1|z_1^i, z_2^i, x)$ increases in z_2^i.

Similarly, when $\alpha_i \leq 0$, the equilibrium expected payoff function is supermodular in $(d_1^i, -d_2^i)$. In addition, the equilibrium expected payoff function has increasing differences in both (d_1^i, z_1^i) and (d_2^i, z_1^i). Following similar arguments, we get that $Pr(d_2^i = 1|z_1^i, z_2^i, x)$ decreases in z_1^i. Similarly, $Pr(d_1^i = 1|z_1^i, z_2^i, x)$ decreases in z_2^i.

Lastly, notice that cross-product effect being exactly zero could be detected, because as long as $\alpha_i \neq 0$, then $Pr(d_2^i = 1|z_1^i, z_2^i, x)$ and $Pr(d_1^i = 1|z_1^i, z_2^i, x)$ are not constant as functions of z_1^i and z_2^i.

B.3 Proof of Theorem 2.3.2

Without loss of generality, for $k = 1, 2$, let the support of S_k^i, $\widetilde{S}_k^i$, $S_k'^i$ and $\widetilde{S}_k'^i$ be $[L_{sk}^i, U_{sk}^i]$, $[\widetilde{L}_{sk}^i, \widetilde{U}_{sk}^i]$, $[L_{sk}'^i, U_{sk}'^i]$ and $[\widetilde{L}_{sk}'^i, \widetilde{U}_{sk}'^i]$, respectively. Assumption 2.3.2 implies the following set inclusion relationships:

- $[L_{sk}^i, U_{sk}^i] \subset [-U, -L]$ for $i = 1, 2$ and $k = 1, 2$;

- $[\widetilde{L}_{sk}^i, \widetilde{U}_{sk}^i] \subset [L, U]$;

- $[L'^{i}_{s1}, U'^{i}_{s1}] \subset [-U, -L]$ and $[L'^{i}_{s2}, U'^{i}_{s2}] \subset [L, U]$;
- $[\widetilde{L}'^{i}_{s1}, \widetilde{U}'^{i}_{s1}] \subset [L, U]$ and $[\widetilde{L}'^{i}_{s2}, \widetilde{U}'^{i}_{s2}] \subset [-U, -L]$.

These relationships and their implications are used in the proof of this theorem as well as in the proof of Lemma 2.3.2 and Theorem 2.3.2.

B.3.1 $\alpha^i \leq 0$

When $\alpha^i \leq 0$, we make use of $P^i_{00}(z^i_1, z^i_2, x)$ and $P^i_{11}(z^i_1, z^i_2, x)$ for identification of structural parameters. As

$$\begin{aligned} P^i_{00}\left(z^i_1, z^i_2, x\right) &= Pr\left(\epsilon^i_1 \leq -\pi^i_1\left(z^i_1, x\right), \epsilon^i_2 \leq -\pi^i_2\left(z^i_2, x\right)\right) \\ &= Pr\left(\epsilon^i_1 + \beta^i_1 x + \delta^i_1 P^{-i}_1(x) \leq -z^i_1, \epsilon^i_2 + \beta^i_2 x + \delta^i_2 P^{-i}_2(x) \leq -z^i_2\right) \\ &= Pr\left(S^i_1 \leq -z^i_1, S^i_2 \leq -z^i_2\right), \end{aligned}$$

we have

$$\begin{aligned} F_{\mathbf{S}^i|x}(-z^i_1, -z^i_2) &= \int_{L^i_{s2}}^{-z^i_2} \int_{L^i_{s1}}^{-z^i_1} f_{\mathbf{S}^i|x}(s^i_1, s^i_2) ds^i_1 ds^i_2 \\ &= P^i_{00}(z^i_1, z^i_2, x). \end{aligned}$$

Thus

$$f_{\mathbf{S}^i|x}(-z^i_1, -z^i_2) = \frac{\partial^2 P^i_{00}(z^i_1, z^i_2, x)}{\partial z^i_1 \partial z^i_2}.$$

The marginal density for S_1^i given x is calculated as:

$$
\begin{aligned}
f_{S_1^i|x}(-z_1^i) &= \int_{L_{s2}^i}^{U_{s2}^i} f_{\mathbf{S}^i|x}(-z_1^i, s_2^i) ds_2^i \\
&= \int_{-L_{s2}^i}^{-U_{s2}^i} f_{\mathbf{S}^i|x}(-z_1^i, -z_2^i)(-dz_2^i) \\
&= -\int_{-L_{s2}^i}^{-U_{s2}^i} f_{\mathbf{S}^i|x}(-z_1^i, -z_2^i) dz_2^i \\
&= \int_{-U_{s2}^i}^{-L_{s2}^i} f_{\mathbf{S}^i|x}(-z_1^i, -z_2^i) dz_2^i \\
&= \int_{L}^{U} f_{\mathbf{S}^i|x}(-z_1^i, -z_2^i) dz_2^i \\
&= \int_{L}^{U} \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, x)}{\partial z_1^i \partial z_2^i} dz_2^i
\end{aligned}
$$

Under Assumption 2.2.2 (iii),

$$
\begin{aligned}
\mathbb{E}(S_1^i|x) &= \beta_1^i x + \delta_1^i P_1^{-i}(x) \\
&= v_1^i(x)^\top \begin{bmatrix} \beta_1^i \\ \delta_1^i \end{bmatrix} \\
&= \int_{L_{s1}^i}^{U_{s1}^i} s_1^i f_{S_1^i|x}(s_1^i) ds_1^i \\
&= \int_{-L_{s1}^i}^{-U_{s1}^i} (-z_1^i) f_{S_1^i|x}(-z_1^i)(-dz_1^i) \\
&= -\int_{-U_{s1}^i}^{-L_{s1}^i} z_1^i f_{S_1^i|x}(-z_1^i) dz_1^i \\
&= -\int_{L}^{U} z_1^i f_{S_1^i|x}(-z_1^i) dz_1^i \\
&= -\int_{L}^{U} z_1^i \int_{L}^{U} \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, x)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i \\
&= -\int_{L}^{U} z_1^i \int_{L}^{U} \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, x)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i.
\end{aligned}
$$

Thus

$$\begin{bmatrix} \beta_1^i \\ \delta_1^i \end{bmatrix} = -\left(\mathbb{E}\left[v_1^i(X)\, v_1^i(X)^\top\right]\right)^{-1} \mathbb{E}\left[v_1^i(X) \int_L^U z_1^i \int_L^U \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right],$$

where the last equality follows from law of iterated expectations. The expression for $[\beta_2^i\ \delta_2^i]^\top$ could be obtained similarly. Also, because

$$\begin{aligned} & P_{11}^i(z_1^i, z_2^i, x) \\ = \;& Pr(\epsilon_1^i \geq -\pi_1^i(z_1^i, x) - \alpha^i, \epsilon_2^i \geq -\pi_2^i(z_2^i, x) - \alpha^i) \\ = \;& Pr(\epsilon_1^i \geq -z_1^i - \beta_1^i x - \delta_1^i P_1^{-i}(x) - \alpha^i, \epsilon_2^i \geq -z_2^i - \beta_2^i x - \delta_2^i P_2^{-i}(x) - \alpha^i) \\ = \;& Pr(z_1^i \geq -\epsilon_1^i - \beta_1^i x - \delta_1^i P_1^{-i}(x) - \alpha^i, z_2^i \geq -\epsilon_2^i - \beta_2^i x - \delta_2^i P_2^{-i}(x) - \alpha^i) \\ = \;& Pr(z_1^i \geq \widetilde{S}_1^i, z_2^i \geq \widetilde{S}_2^i), \end{aligned}$$

the conditional cdf of $\widetilde{\mathbf{S}}^i$ is identified

$$F_{\widetilde{\mathbf{S}}^i|x}(z_1^i, z_2^i) = \int_{\widetilde{L}_{s2}^i}^{z_2^i} \int_{\widetilde{L}_{s1}^i}^{z_1^i} f_{\widetilde{\mathbf{S}}^i|x}\left(\widetilde{s}_1^i, \widetilde{s}_s^i\right) d\widetilde{s}_1^i d\widetilde{s}_2^i = P_{11}^i(z_1^i, z_2^i, x). \tag{B.1}$$

Thus

$$f_{\widetilde{\mathbf{S}}^i|x}(z_1^i, z_2^i) = \frac{\partial^2 P_{11}^i(z_1^i, z_2^i, x)}{\partial z_1^i \partial z_2^i}. \tag{B.2}$$

The marginal pdf of $\widetilde{S}_1^i$ is then identified as:

$$f_{\widetilde{S}_1^i|x}(z_1^i) = \int_{\widetilde{L}_{s2}^i}^{\widetilde{U}_{s2}^i} f_{\widetilde{\mathbf{S}}^i|x}(z_1^i, \widetilde{s}_2^i) d\widetilde{s}_2^i = \int_L^U f_{\widetilde{\mathbf{S}}^i|x}(z_1^i, \widetilde{s}_2^i) d\widetilde{s}_2^i = \int_L^U \frac{\partial^2 P_{11}^i(z_1^i, z_2^i, x)}{\partial z_1^i \partial z_2^i} dz_2^i. \tag{B.3}$$

As

$$\begin{aligned} E(\widetilde{S}_1^i|x) &= -\beta_1^i x - \delta_1^i P_1^{-i}(x) - \alpha^i \\ &= \int_L^U z_1^i \int_L^U \frac{\partial^2 P_{11}^i(z_1^i, z_2^i, x)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i. \end{aligned}$$

Thus

$$\begin{aligned}\alpha^i &= -\mathbb{E}\left(\widetilde{S}_1^i\right) - \mathbb{E}\left(S_1^i\right) \\ &= -\mathbb{E}\left[\int_L^U z_1^i \int_L^U \frac{\partial^2 P_{11}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right] + \mathbb{E}\left[\int_L^U z_1^i \int_L^U \frac{\partial^2 P_{00}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right].\end{aligned}$$

B.3.2 $\alpha^i \geq 0$

We use $P_{01}^i(z_1^i, z_2^i, x)$ to identify $\mathbb{E}\left(S_1'^i|x\right)$ and $\mathbb{E}\left(S_2'^i|x\right)$. Then $\mathbb{E}\left(S_2'^i|x\right)$ is used in the identification of $[\beta_2^i\ \delta_2^i]^\top$; We use $P_{10}^i(z_1^i, z_2^i, x)$ to identify $\mathbb{E}\left(\widetilde{S}_1'^i|x\right)$, which is used in the identification of $[\beta_1^i\ \delta_1^i]^\top$; We use the difference between $\mathbb{E}\left(S_1'^i|x\right)$ and $-\mathbb{E}\left(\widetilde{S}_1'^i|x\right)$ to identify α^i.

By definition, it holds that

$$\begin{aligned}P_{01}^i\left(z_1^i, z_2^i, x\right) &= Pr\left(\epsilon_1^i \leq -\pi_1^i - \alpha^i, \epsilon_2^i \geq -\pi_2^i\right) \\ &= Pr\left(\epsilon_1^i \leq -z_1^i - \beta_1^i x - \delta_1^i P_1^{-i}(x) - \alpha^i, \epsilon_2^i \geq -z_2^i - \beta_2^i x - \delta_2^i P_2^{-i}(x)\right) \\ &= Pr\left(\epsilon_1^i + \beta_1^i x + \delta_1^i P_1^{-i}(x) + \alpha^i \leq -z_1^i, z_2^i \geq -\epsilon_2^i - \beta_2^i x - \delta_2^i P_2^{-i}(x)\right) \\ &= Pr\left(\epsilon_1^i + \beta_1^i x + \delta_1^i P_1^{-i}(x) + \alpha^i \leq -z_1^i, -\epsilon_2^i - \beta_2^i x - \delta_2^i P_2^{-i}(x) \leq z_2^i\right) \\ &= Pr\left(S_1'^i \leq -z_1^i, S_2'^i \leq z_2^i\right) \\ &= F_{\mathbf{S}'^i|x}\left(-z_1^i, z_2^i\right) \\ &= \int_{L_{s1}'^i}^{-z_1^i} \int_{L_{s2}'^i}^{z_2^i} f_{\mathbf{S}'^i|x}\left(s_1'^i, s_2'^i\right) ds_2'^i ds_1'^i.\end{aligned}$$

This means that

$$\frac{\partial^2 P_{01}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} = -f_{\mathbf{S}'^i|x}\left(-z_1^i, z_2^i\right). \tag{B.4}$$

The marginal density is then identified as

$$\begin{aligned}
f_{S_2'^i|x}\left(z_2^i\right) &= \int_{L_{s1}'^i}^{U_{s1}'^i} f_{\mathbf{S}'^i|x}\left(s_1'^i, z_2^i\right) ds_1'^i \\
&= \int_{-L_{s1}'^i}^{-U_{s1}'^i} f_{\mathbf{S}'^i|x}\left(-z_1^i, z_2^i\right)\left(-dz_1^i\right) \\
&= \int_{-L_{s1}'^i}^{-U_{s1}'^i} -f_{\mathbf{S}'^i|x}\left(-z_1^i, z_2^i\right) dz_1^i \\
&= \int_{-U_{s1}'^i}^{-L_{s1}'^i} f_{\mathbf{S}'^i|x}\left(-z_1^i, z_2^i\right) dz_1^i \\
&= -\int_L^U \frac{\partial^2 P_{01}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} dz_1^i.
\end{aligned}$$

As the expectation is identified as

$$\begin{aligned}
\mathbb{E}\left(S_2'^i|x\right) &= -[x \;\; P_2^{-i}\left(x\right)] \begin{bmatrix} \beta_2^i \\ \delta_2^i \end{bmatrix} \\
&= -v_2^i\left(x\right)^\top \begin{bmatrix} \beta_2^i \\ \delta_2^i \end{bmatrix} \\
&= \int_{L_{s2}'^i}^{U_{s2}'^i} z_2^i f_{S_2'^i|x}\left(z_2^i\right) dz_2^i \\
&= \int_L^U z_2^i f_{S_2'^i|x}\left(z_2^i\right) dz_2^i \\
&= -\int_L^U z_2^i \int_L^U \frac{\partial^2 P_{01}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} dz_1^i dz_2^i.
\end{aligned}$$

This implies that

$$\begin{bmatrix} \beta_2^i \\ \delta_2^i \end{bmatrix} = \mathbb{E}[v_2^i\left(X\right) v_2^i\left(X\right)^\top]^{-1}\, \mathbb{E}[v_2^i\left(X\right) \int_L^U z_2^i \int_L^U \frac{\partial^2 P_{01}^i\left(z_1^i, z_2^i, X\right)}{\partial z_1^i \partial z_2^i} dz_1^i dz_2^i]. \tag{B.5}$$

Also note that

$$
\begin{aligned}
f_{S_1'^i|x}\left(-z_1^i\right) &= \int_{L_{s2}'^i}^{U_{s2}'^i} f_{\mathbf{S}'^i|x}\left(-z_1^i, s_2'^i\right) ds_2'^i \\
&= \int_L^U f_{\mathbf{S}'^i|x}\left(-z_1^i, s_2'^i\right) ds_2'^i \\
&= \int_L^U f_{\mathbf{S}'^i|x}\left(-z_1^i, z_2^i\right) dz_2^i \\
&= -\int_L^U \frac{\partial^2 P_{01}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} dz_2^i.
\end{aligned}
$$

This implies

$$
\begin{aligned}
\mathbb{E}\left(S_1'^i|x\right) &= \int_{L_{s1}'^i}^{U_{s1}'^i} s_1'^i f_{S_1'^i|x}\left(s_1'^i\right) ds_1'^i \\
&= \int_{-L_{s1}'^i}^{-U_{s1}'^i} \left(-z_1^i\right) f_{S_1'^i|x}\left(-z_1^i\right)\left(-dz_1^i\right) \\
&= \int_{-L_{s1}'^i}^{-U_{s1}'^i} z_1^i f_{S_1'^i|x}\left(-z_1^i\right) dz_1^i \\
&= -\int_{-U_{s1}'^i}^{-L_{s1}'^i} z_1^i f_{S_1'^i|x}\left(-z_1^i\right) dz_1^i \\
&= -\int_L^U z_1^i f_{S_1'^i|x}\left(-z_1^i\right) dz_1^i \\
&= \int_L^U z_1^i \int_L^U \frac{\partial^2 P_{01}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
P_{10}^{i}\left(z_1^i, z_2^i, x\right) &= Pr\left(\epsilon_1^i \geq -\pi_1^i, \epsilon_2^i \leq -\pi_2^i - \alpha^i\right) \\
&= Pr\left(\epsilon_1^i \geq -z_1^i - \beta_1^i x - \delta_1^i P_1^{-i}\left(x\right), \epsilon_2^i \leq -z_2^i - \beta_2^i x - \delta_2^i P_2^{-i}\left(x\right) - \alpha^i\right) \\
&= Pr\left(-\epsilon_1^i - \beta_1^i x - \delta_1^i P_1^{-i}\left(x\right) \leq z_1^i, \epsilon_2^i + \beta_2^i x + \delta_2^i P_2^{-i}\left(x\right) + \alpha^i \leq -z_2^i\right) \\
&= Pr\left(\widetilde{S}_1'^i \leq z_1^i, \widetilde{S}_2'^i \leq -z_2^i\right) \\
&= F_{\widetilde{\mathbf{S}}'^i|x}\left(z_1^i, -z_2^i\right) \\
&= \int_{\widetilde{L}_{s2}'^i}^{-z_2^i} \int_{\widetilde{L}_{s1}'^i}^{z_1^i} f_{\widetilde{\mathbf{S}}'^i|x}\left(\widetilde{s}_1'^i, \widetilde{s}_2'^i\right) d\widetilde{s}_1'^i d\widetilde{s}_2'^i.
\end{aligned}
$$

This implies

$$
\frac{\partial^2 P_{10}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} = -f_{\widetilde{\mathbf{S}}'^i|x}\left(z_1^i, -z_2^i\right).
$$

So

$$
\begin{aligned}
f_{\widetilde{S}_1'^i|x}\left(z_1^i\right) &= \int_{\widetilde{L}_{s2}'^i}^{\widetilde{U}_{s2}'} f_{\widetilde{\mathbf{S}}_1'^i|x}\left(z_1^i, \widetilde{s}_2'^i\right) d\widetilde{s}_2'^i \\
&= \int_{-\widetilde{L}_{s2}'}^{-\widetilde{U}_{s2}'} f_{\widetilde{\mathbf{S}}_1'^i|x}\left(z_1^i, -z_2^i\right)\left(-dz_2^i\right) \\
&= \int_{-\widetilde{U}_{s2}'}^{-\widetilde{L}_{s2}'} f_{\widetilde{\mathbf{S}}_1'^i|x}\left(z_1^i, -z_2^i\right) dz_2^i \\
&= \int_{L}^{U} f_{\widetilde{\mathbf{S}}_1'^i|x}\left(z_1^i, -z_2^i\right) dz_2^i \\
&= -\int_{L}^{U} \frac{\partial^2 P_{10}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} dz_2^i
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(\widetilde{S}_1'^i|x\right) &= \int_{\widetilde{L}_{s1}'^i}^{\widetilde{U}_{s1}'^i} \widetilde{s}_1'^i f_{\widetilde{S}_1'^i|x}\left(\widetilde{s}_1'^i\right) d\widetilde{s}_1'^i \\
&= \int_{L}^{U} z_1^i f_{\widetilde{S}_1'^i|x}\left(z_1^i\right) dz_1^i \\
&= -\int_{L}^{U} z_1^i \int_{L}^{U} \frac{\partial^2 P_{10}^i\left(z_1^i, z_2^i, x\right)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i
\end{aligned}
$$

Noting that

$$\begin{aligned}\mathbb{E}(\widetilde{S}'^{i}|x) &= -[x\ P_1^{-i}(x)]\begin{bmatrix}\beta_1^i\\ \delta_1^i\end{bmatrix}\\ &= -v_1^i(x)\begin{bmatrix}\beta_1^i\\ \delta_1^i\end{bmatrix}\end{aligned}$$

This implies that

$$\begin{bmatrix}\beta_1^i\\ \delta_1^i\end{bmatrix} = \mathbb{E}[v_1^i(X)\, v_1^i(X)^\top]\,\mathbb{E}[v_1^i(X)\int_L^U z_1^i\int_L^U \frac{\partial^2 P_{10}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i] \tag{B.6}$$

The complementarity parameter is identified as:

$$\begin{aligned}\alpha^i &= \mathbb{E}(S_1'^{i}) + \mathbb{E}(\widetilde{S}_1'^{i})\\ &= \mathbb{E}\left[\int_L^U z_1^i\int_L^U \frac{\partial^2 P_{01}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right] - \mathbb{E}\left[\int_L^U z_1^i\int_L^U \frac{\partial^2 P_{10}^i(z_1^i, z_2^i, X)}{\partial z_1^i \partial z_2^i} dz_2^i dz_1^i\right].\end{aligned}$$

B.4 Proof of Lemma 2.3.2

B.4.1 $\alpha^i \leq 0$

To identify the support of S_1^i and S_2^i, notice that under Assumption 2.3.2 we have

$$\begin{aligned}P_{00}^i\left(z_1^i, L, x\right) &= \int_{L_{s1}^i}^{-z_1^i}\int_{L_{s2}^i}^{-L} f_{\mathbf{S}^i|x}\left(s_1^i, s_2^i\right) ds_2^i ds_1^i\\ &= \int_{L_{s1}^i}^{-z_1^i} f_{S_1^i|x}\left(s_1^i\right) ds_1^i\\ &= F_{S_1^i|x}\left(-z_1^i\right).\end{aligned}$$

Thus it follows that $[L_{s1}^i, U_{s1}^i]$ is identified as the closure of the set $\{-z_1^i : 0 < P_{00}^i(z_1^i, L, x) < 1\}$. Similarly, $[L_{s2}^i, U_{s2}^i]$ is identified as the closure of the set $\{-z_2^i : 0 < P_{00}^i(L, z_2^i, x) < 1\}$.

To identify the support of $\widetilde{S}_1^i$ and $\widetilde{S}_2^i$, notice that under Assumption 2.3.2 we have

$$
\begin{aligned}
P_{11}^i\left(z_1^i, U, x\right) &= \int_{\widetilde{L}_{s1}^i}^{z_1^i} \int_{\widetilde{L}_{s2}^i}^{U} f_{\widetilde{\mathbf{S}}^i|x}\left(\widetilde{s}_1^i, \widetilde{s}_2^i\right) d\widetilde{s}_2^i d\widetilde{s}_1^i \\
&= \int_{\widetilde{L}_{s1}^i}^{z_1^i} f_{\widetilde{S}_1^i|x}\left(\widetilde{s}_1^i\right) d\widetilde{s}_1^i \\
&= F_{\widetilde{S}_1^i|x}\left(z_1^i\right).
\end{aligned}
$$

It follows that $[\widetilde{L}_{s1}^i, \widetilde{U}_{s1}^i]$ is identified as the closure of the set $\{z_1^i : 0 < P_{11}^i(z_1^i, U, x) < 1\}$. Similarly, $[\widetilde{L}_{s2}^i, \widetilde{U}_{s2}^i]$ is identified as the closure of the set $\{z_2^i : 0 < P_{11}^i(U, z_2^i, x) < 1\}$.

B.4.2 $\alpha^i \geq 0$

To identify the support of $S_1'^i$ and $S_2'^i$, notice that under Assumption 2.3.2 we have

$$
\begin{aligned}
P_{01}^i\left(z_1^i, U, x\right) &= \int_{L_{s1}'^i}^{-z_1^i} \int_{L_{s2}'^i}^{U} f_{\mathbf{S}'^i|x}\left(s_1'^i, s_2'^i\right) ds_1'^i ds_2'^i \\
&= \int_{L_{s1}'^i}^{-z_1^i} f_{S_1'^i|x}\left(s_1'^i\right) ds_1'^i \\
&= F_{S_1'^i|x}\left(-z_1^i\right).
\end{aligned}
$$

It then follows that $[L_{s1}'^i, U_{s1}'^i]$ is identified as the closure of the set $\{-z_1^i : 0 < P_{01}^i\left(z_1^i, U, x\right) < 1\}$. Similarly, $[L_{s2}'^i, U_{s2}'^i]$ is identified as the closue of the set $\{z_2^i : 0 < P_{01}^i\left(L, z_2^i, x\right) < 1\}$.

To identify the support of $\widetilde{S}_1'^i$ and $\widetilde{S}_2'^i$, notice that under Assumption 2.3.2 we have

$$
\begin{aligned}
P_{10}^i\left(z_1^i, L, x\right) &= \int_{\widetilde{L}_{s1}'^i}^{z_1^i} \int_{\widetilde{L}_{s2}'^i}^{-L} f_{\widetilde{\mathbf{S}}'|x}\left(\widetilde{s}_1'^i, \widetilde{s}_2'^i\right) d\widetilde{s}_1'^i d\widetilde{s}_2'^i \\
&= \int_{\widetilde{L}_{s1}'^i}^{z_1^i} f_{\widetilde{S}_1'^i|x}\left(\widetilde{s}_1'^i\right) d\widetilde{s}_1'^i \\
&= F_{\widetilde{S}_1'^i|x}\left(z_1^i\right).
\end{aligned}
$$

It then follows that $[\widetilde{L}_{s1}'^i, \widetilde{U}_{s1}'^i]$ is identified as the closure of the set $\{z_1^i : 0 < P_{10}^i\left(z_1^i, L, x\right) < 1\}$. Similarly, $[\widetilde{L}_{s2}'^i, \widetilde{U}_{s2}'^i]$ is identified as the closure of the set $\{-z_2^i : 0 < P_{10}^i\left(U, z_2^i, x\right) < 1\}$.

B.5 Proof of Theorem 2.3.3

B.5.1 $\alpha^i \leq 0$

To show the simple form for $[\beta_1^i \ \delta_1^i]^\top$, we first show that

$$\mathbb{E}(S_1^i|x) = -\int_{L_{s1}^i}^{U_{s1}^i} \left(P_{00}^i\left(-s_1^i, -U_{s2}^i, x\right) + \mathbb{1}\left(-s_1^i \geq c\right) - 1\right) ds_1^i - c \qquad \text{(B.7)}$$

Note that $F_{\mathbf{S}^i|x}(s_1^i, s_2^i) = P_{00}^i(-s_1^i, -s_2^i, x)$ implies $F_{S_1^i|x}(s_1^i) = F_{\mathbf{S}^i|x}(s_1^i, U_{s2}^i) = P_{00}^i(-s_1^i, -U_{s2}^i, x)$. So

$$\begin{aligned}
RHS &= -\int_{L_{s1}^i}^{U_{s1}^i} \Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(-s_1^i \geq c\right) - 1\Big) ds_1^i - c \\
&= -\int_{L_{s1}^i}^{U_{s1}^i} \Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right) - 1\Big) \frac{\partial s_1^i}{\partial s_1^i} ds_1^i - c \\
&= -\Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right) - 1\Big) s_1^i\Big|_{L_{s1}^i}^{U_{s1}^i} + \int_{L_s}^{U_s} \frac{\partial \Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right) - 1\Big)}{\partial s_1^i} s_1^i ds_1^i - c \\
&= -\Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right) - 1\Big) s_1^i\Big|_{L_{s1}^i}^{-c} - \Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right) - 1\Big) s_1^i\Big|_{-c}^{U_{s1}^i} \\
&\quad + \int_{L_{s1}^i}^{-c} \frac{\partial \Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right)\Big)}{\partial s_1^i} s_1^i ds_1^i + \int_{-c}^{U_{s1}^i} \frac{\partial \Big(F_{S_1^i|x}(s_1^i) + \mathbb{1}\left(s_1^i \leq -c\right)\Big)}{\partial s_1^i} s_1^i ds_1^i - c \\
&= -\Big(F_{S_1^i|x}\left(-c\right) + 1 - 1\Big)\left(-c\right) + \left(0 + 1 - 1\right) L_{s1}^i - \left(1 + 0 - 1\right) U_{s1}^i \\
&\quad + \Big(F_{S_1^i|x}\left(-c\right) + 0 - 1\Big)\left(-c\right) + \int_{L_{s1}^i}^{-c} f_{S_1^i|x}\left(s_1^i\right) s_1^i ds_1^i + \int_{-c}^{U_{s1}^i} f_{S_1^i|x}\left(s_1^i\right) ds_1^i - c \\
&= c + \mathbb{E}\left(S_1^i|x\right) - c \\
&= \mathbb{E}\left(S_1^i|x\right).
\end{aligned}$$

Next do change of variable $z_1^i = -s_1^i$ (so that $ds_1^i = -dz_1^i$), and for any $z_2^i \in \mathcal{I}_2^i(L) = [L, -U_{s2}^i]$, we get

$$\begin{aligned}
&\int_{-L_{s1}^i}^{-U_{s1}^i} \left(P_{00}^i\left(z_1^i, -U_{s2}^i, x\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1\right) dz_1^i - c \\
= &\int_{-L_{s1}^i}^{-U_{s1}^i} \frac{\left(P_{00}^i\left(z_1^i, z_2^i, x\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1\right) f(z_1^i|z_2^i, x)}{f(z_1^i|z_2^i, x)} dz_1^i - c \\
= &-\int_{-U_{s1}^i}^{-L_{s1}^i} \frac{\left(P_{00}^i\left(z_1^i, z_2^i, x\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1\right) f(z_1^i|z_2^i, x)}{f(z_1^i|z_2^i, x)} dz_1^i - c \\
= &-\int_{L}^{U} \frac{\left(P_{00}^i\left(z_1^i, z_2^i, x\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1\right) f(z_1^i|z_2^i, x)}{f(z_1^i|z_2^i, x)} dz_1^i - c, \\
= &-\int_{L}^{U} \frac{\left(\mathbb{E}\left(D_{00}^i|z_1^i, z_2^i, x\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1\right) f(z_1^i|z_2^i, x)}{f(z_1^i|z_2^i, x)} dz_1^i - c, \\
= &-\int_{L}^{U} \mathbb{E}\left(\frac{\left(D_{00}^i + \mathbb{1}\left(z_1^i \geq c\right) - 1\right)}{f(z_1^i|z_2^i, x)} | z_1^i, z_2^i, x\right) f(z_1^i|z_2^i, x) dz_1^i - c, \\
= &-\mathbb{E}\left(\frac{\left(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1\right)}{f(Z_1^i|z_2^i, x)} | z_2^i, x\right) - c
\end{aligned}$$

where the third equality holds because $[-U_{s1}^i, -L_{s1}^i] \subset [L, U]$[1] and the integrand is zero outside $[-U_{s1}^i, -L_{s1}^i]$. To see this, notice that $z_1^i > -L_{s1}^i$ implies $-z_1^i < L_{s1}^i$, $S_1^i \leq -z_1^i$ is impossible because $-z_1^i$ is below the lower bound of the support. $z_1^i > -L_{s1}^i$ also implies that $z_1^i > c$ as c is in the support of $-S_1^i$. Thus $\mathbb{1}\left(\mathbf{d}_1^i = (0,0)\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1 = 0 + 1 - 1 = 0$, where $\mathbb{1}\left(\mathbf{d}_1^i = (0,0)\right) = 0$ because S_1^i is below the lower bound of its support; Also when $z_1^i \leq -U_s$, $S_1^i \leq -z_1^i$ with $-z_1^i \geq U_s$, $\mathbb{1}\left(\mathbf{d}_1^i = (0,0)\right) + \mathbb{1}\left(z_1^i \geq c\right) - 1 = 1 + 0 - 1 = 0$, where $\mathbb{1}\left(\mathbf{d}_1^i = (0,0)\right) = 1$ because S_1^i is above the upper bound of its support and $\mathbb{1}\left(z_1^i \geq c\right) = 0$ as z_1^i is below the lower bound of the support of $-S_1^i$. The fifth equality holds by linearity of the expectation operator because z_1^i, z_2^i and x are given. Thus it holds that

$$\mathbb{E}\left(S_1^i|x\right) = -\mathbb{E}\left(\frac{\left(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1\right)}{f(Z_1^i|z_2^i, x)} | z_2^i, x\right) - c \tag{B.8}$$

[1] This is implied by the fact that the support of S_1^i is a subset of the support of $-z_1^i$.

Take an additional average with respect to Z_2^i over $\mathcal{I}_2^i(L)$ gives

$$\mathbb{E}\left(S_1^i|x\right) = -\mathbb{E}\left(\frac{(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1)}{f(Z_1^i|Z_2^i, x)}|Z_2^i \in \mathcal{I}_2^i(L), x\right) - c. \tag{B.9}$$

As

$$\begin{bmatrix}\beta_1^i & \delta_1^i\end{bmatrix}^\top = \left(\mathbb{E}\left[v_1^i\left(X\right) v_1^i\left(X\right)^\top |Z_2^i \in \mathcal{I}_2^i(L)\right]\right)^{-1} \mathbb{E}\left[v_1^i(X)\,\mathbb{E}(S_1^i|X)|Z_2^i \in \mathcal{I}_2^i(L)\right], \tag{B.10}$$

$[\beta_1^i\ \delta_1^i]^\top$ is identified as:

$$\left(\mathbb{E}\left[v_1^i\left(X\right) v_1^i\left(X\right)^\top |Z_2^i \in \mathcal{I}_2^i(L)\right]\right)^{-1} \mathbb{E}\left[v_1^i(X)\left(-\frac{(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1)}{f(Z_1^i|Z_2^i, X)} - c\right)|Z_2^i \in \mathcal{I}_2^i(L)\right].$$

As

$$\mathbb{E}\left[v_1^i\left(X\right) v_1^i\left(X\right)^\top |Z_2^i \in \mathcal{I}_2^i(L)\right] = \frac{\mathbb{E}\left[v_1^i\left(X\right) v_1^i\left(X\right)^\top \mathbb{1}\left(Z_2^i \in \mathcal{I}_2^i(L)\right)\right]}{Pr\left(Z_2^i \in \mathcal{I}_2^i(L)\right)},$$

and

$$\begin{aligned}&\mathbb{E}\left[v_1^i(X)\left(-\frac{(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1)}{f(Z_1^i|Z_2^i, X)} - c\right)|Z_2^i \in \mathcal{I}_2^i(L)\right]\\ =\ &\frac{\mathbb{E}\left[v_1^i(X)\left(\frac{-\left(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1\right)}{f(Z_1^i|Z_2^i, X)} - c\right)\mathbb{1}\left(Z_2^i \in \mathcal{I}_2^i\left(L\right)\right)\right]}{Pr\left(Z_2^i \in \mathcal{I}_2^i\left(L\right)\right)},\end{aligned}$$

where the second equality follows from law of iterated expectation. We could rewrite the identifying equation for $[\beta_1^i, \delta_1^i]^\top$ as

$$\left(\mathbb{E}\left[v_1^i\left(X\right) v_1^i\left(X\right)^\top W_{2L}^i\right]\right)^{-1} \mathbb{E}\left[v_1^i(X)\left(-\frac{(D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1)}{f(Z_1^i|Z_2^i, X)} - c\right) W_{2L}^i\right].$$

The simple form for $[\beta_2^i\ \delta_2^i]^\top$ could be derived similarly by symmetry.

To obtain the simple form for α^i, we first derive the simple form for $\mathbb{E}\left(\widetilde{S}_1^i\right)$.

First, we show that

$$\mathbb{E}(\widetilde{S}_1^i|x) = -\int_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} \left(P_{11}^i\left(\widetilde{s}_1^i, \widetilde{U}_{s2}^i, x\right) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) d\widetilde{s}_1^i + c \tag{B.11}$$

$$\begin{aligned}
RHS &= -\int_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} \left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) d\widetilde{s}_1^i + c \\
&= -\int_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} \left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) \frac{\partial \widetilde{s}_1^i}{\partial \widetilde{s}_1^i} d\widetilde{s}_1^i + c \\
&= -\left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) \widetilde{s}_1^i \Big|_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} + \int_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} \frac{\partial \left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right)}{\partial \widetilde{s}_1^i} \widetilde{s}_1^i d\widetilde{s}_1^i + c \\
&= -\left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) \widetilde{s}_1^i \Big|_{\widetilde{L}_{s1}^i}^{c} - \left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) \widetilde{s}_1^i \Big|_{c}^{\widetilde{U}_{s1}^i} + c \\
&\quad + \int_{\widetilde{L}_{s1}^i}^{c} \frac{\partial \left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right)}{\partial \widetilde{s}_1^i} \widetilde{s}_1^i d\widetilde{s}_1^i + \int_{c}^{\widetilde{U}_{s1}^i} \frac{\partial \left(F_{\widetilde{S}_1^i|x}(\widetilde{s}_1^i) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right)}{\partial \widetilde{s}_1^i} \widetilde{s}_1^i d\widetilde{s}_1^i + c \\
&= -\left(F_{\widetilde{S}_1^i|x}(c) - 0\right) c + (0 - 0)\, \widetilde{L}_{s1}^i - (1 - 1)\, \widetilde{U}_{s1}^i + \left(F_{\widetilde{S}_1^i|x}(c) - 1\right) c \\
&\quad + \int_{\widetilde{L}_{s1}^i}^{c} f_{\widetilde{S}_1^i|x}\left(\widetilde{s}_1^i\right) \widetilde{s}_1^i d\widetilde{s}_1^i + \int_{c}^{\widetilde{U}_{s1}^i} f_{\widetilde{S}_1^i|x}\left(\widetilde{s}_1^i\right) \widetilde{s}_1^i d\widetilde{s}_1^i + c \\
&= \int_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} f_{\widetilde{S}_1^i|x}\left(\widetilde{s}_1^i\right) \widetilde{s}_1^i d\widetilde{s}_1^i \\
&= \mathbb{E}\left(\widetilde{S}_1^i|x\right).
\end{aligned}$$

Next, notice that for any $z_2^i \in \mathcal{I}_2^i(U) = [\widetilde{U}_{s2}^i, U]$, it holds that

$$
\begin{aligned}
& -\int_{\widetilde{L}_{s1}^i}^{\widetilde{U}_{s1}^i} \left(P_{11}^i\left(\widetilde{s}_1^i, \widetilde{U}_{s2}^i, x\right) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right)\right) d\widetilde{s}_1^i + c \\
= & -\int_L^U \left(P_{11}^i\left(z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \geq c\right)\right) dz_1^i + c \\
= & -\int_L^U \frac{\left(\mathbb{E}\left(D_{11}^i | z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \geq c\right)\right) f\left(z_1^i | z_2^i, x\right)}{f\left(z_1^i | z_2^i, x\right)} dz_1^i + c \\
= & -\int_L^U \mathbb{E}\left(\frac{\left(D_{11}^i - \mathbb{1}\left(z_1^i \geq c\right)\right)}{f\left(z_1^i | z_2^i, x\right)} | z_1^i, z_2^i, x\right) f\left(z_1^i | z_2^i, x\right) dz_1^i + c \\
= & -\int_L^U \frac{\left(\mathbb{E}\left(D_{11}^i | z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \geq c\right)\right) f\left(z_1^i | z_2^i, x\right)}{f\left(z_1^i | z_2^i, x\right)} dz_1^i + c \\
= & -\mathbb{E}\left(\frac{\mathbb{E}\left(D_{11}^i | z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \geq c\right)}{f\left(Z_1^i | z_2^i, x\right)} | z_2^i, x\right) + c,
\end{aligned}
$$

where the first equality follows because when $\widetilde{s}_1^i \geq \widetilde{U}_s$, the integrand $P_{11}^i\left(\widetilde{s}_1^i, z_2^i, x\right) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right) = 1 - 1 = 0$; when $\widetilde{s}_1^i \leq \widetilde{L}_s$, the integrand $P_{11}^i\left(\widetilde{s}_1^i, z_2^i, x\right) - \mathbb{1}\left(\widetilde{s}_1^i \geq c\right) = 0 - 0 = 0$. Note that as c is in the support of $\widetilde{S}_1^i$, the integration over z_1^i doesn't need to exactly touch the boundary of the support of z_1^i, for example, it could be written as integration over $[L+\epsilon, U-\epsilon]$, which still gives expectation wrt Z_1^i as the integrand beyond this range is zero (This is also true for other cases). Thus it holds that

$$
\mathbb{E}\left(\widetilde{S}_1^i | x\right) = -\mathbb{E}\left(\frac{P_{11}^i\left(Z_1^i, z_2^i, x\right) - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i | z_2^i, x\right)} | z_2^i, x\right) + c. \tag{B.12}
$$

Law of iterated expectation gives

$$
\begin{aligned}
\mathbb{E}\left(\widetilde{S}_1^i\right) &= -\mathbb{E}\left(\frac{P_{11}^i\left(Z_1^i, Z_2^i, x\right) - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i | Z_2^i, x\right)} | Z_2^i \in \mathcal{I}_2^i(U)\right) + c. \\
&= -\mathbb{E}\left(\frac{\mathbb{E}\left(D_{11}^i | Z_1^i, Z_2^i, X\right) - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i | Z_2^i, x\right)} | Z_2^i \in \mathcal{I}_2^i(U)\right) + c. \\
&= -\mathbb{E}\left(\frac{D_{11}^i - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i | Z_2^i, X\right)} | Z_2^i \in \mathcal{I}_2^i(U)\right) + c.
\end{aligned}
$$

Thus the simple expression for α^i is

$$\mathbb{E}\left[\frac{D_{11}^i - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i | Z_2^i, X\right)} \Big| Z_2^i \in \widetilde{\mathcal{I}}_2^i\left(U\right)\right] + \mathbb{E}\left[\frac{D_{00}^i + \mathbb{1}\left(Z_1^i \geq c\right) - 1}{f(Z_1^i | Z_2^i, X)} \Big| Z_2^i \in \mathcal{I}_2^i\left(L\right)\right].$$

B.5.2 $\alpha^i \geq 0$

First we show the following two equalities:

$$\mathbb{E}\left(\widetilde{S}_1^{\prime i} | x\right) = -\,\mathbb{E}\left[\frac{D_{10}^i - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i | Z_2^i, x\right)} \Big| Z_2^i \in \widetilde{\mathcal{I}}_2^{\prime}\left(L\right), x\right] + c, \tag{B.13}$$

$$\mathbb{E}\left(S_2^{\prime i} | x\right) = -\,\mathbb{E}\left[\frac{D_{01}^i - \mathbb{1}\left(Z_2^i \geq c\right)}{f\left(Z_2^i | Z_1^i, x\right)} \Big| Z_1^i \in \mathcal{I}_1^{\prime}\left(L\right), x\right] + c. \tag{B.14}$$

As these Z_1^i and Z_2^i appears in symmetric positions in these two equations, we only need to prove one of them and the other one hold by the same logic. Here we prove the second equation. First, we show that

$$\mathbb{E}\left(S_2^{\prime i} | x\right) = -\int_{L_{s2}^{\prime i}}^{U_{s2}^{\prime i}} \left(P_{01}^i\left(-U_{s1}^{\prime i}, s_2^{\prime i}, x\right) - \mathbb{1}\left(s_2^{\prime i} \geq c\right)\right) ds_2^{\prime i} + c \tag{B.15}$$

Recall that $P_{01}^i(z_1^i, z_2^i, x)$ is related to the underlying distribution of $(S_1^{\prime i}, S_2^{\prime i})$ by the following equation:

$$\begin{aligned} P_{01}^i\left(z_1^i, z_2^i, x\right) &= Pr\left(S_1^{\prime i} \leq -z_1^i, S_2^{\prime i} \leq z_2^i\right) \\ &= F_{\mathbf{S}^{\prime i}|x}\left(-z_1^i, z_2^i\right) \\ &= \int_{L_{s1}^{\prime i}}^{-z_1^i} \int_{L_{s2}^{\prime i}}^{z_2^i} f_{\mathbf{S}^{\prime i}|x}\left(s_1^{\prime i}, s_2^{\prime i}\right) ds_2^{\prime i} ds_1^{\prime i} \end{aligned}$$

So $P_{01}^i\left(-U_{s1}^{\prime i}, s_2^{\prime i}, x\right) = F_{\mathbf{S}^{\prime i}|x}\left(U_{s1}^{\prime i}, z_2^i\right) = F_{S_2^{\prime i}|x}\left(z_2^i\right)$.

$$
\begin{aligned}
RHS &= -\int_{L'^i_{s2}}^{U'^i_{s2}} \left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right) \frac{\partial s'^i_2}{\partial s'^i_2} ds'^i_2 + c \\
&= -\left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right) s'^i_2 \Big|_{L'^i_{s2}}^{U'^i_{s2}} + \int_{L'^i_{s2}}^{U'^i_{s2}} \frac{\partial\left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right)}{\partial s'^i_2} s'^i_2 ds'^i_2 + c \\
&= -\left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right) s'^i_2 \Big|_{L'^i_{s2}}^{c} - \left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right) s'^i_2 \Big|_{c}^{U'^i_{s2}} \\
&\quad + \int_{L'^i_{s2}}^{c} \frac{\partial\left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right)}{\partial s'^i_2} s'^i_2 ds'^i_2 + \int_{c}^{U'^i_{s2}} \frac{\partial\left(F_{S'^i_2|x}\left(s'^i_2\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right)}{\partial s'^i_2} s'^i_2 ds'^i_2 + c \\
&= -\left(F_{S'^i_2|x}(c) - 0\right) c + \left(F_{S'^i_2|x}\left(L'^i_{s2}\right) - 0\right) L'^i_{s2} - \left(F_{S'^i_2|x}\left(U'^i_{s2}\right) - 1\right) U'^i_{s2} \\
&\quad + \left(F_{S'^i_2|x}(c) - 1\right) c + \int_{L'^i_{s2}}^{U'^i_{s2}} \frac{\partial F_{S'^i_2|x}\left(s'^i_2\right)}{\partial s'^i_2} s'^i_2 ds'^i_2 + c \\
&= \mathbb{E}\left(S'^i_2|x\right)
\end{aligned}
$$

Also, for any $z^i_1 \in \mathcal{I}'_1(L) = [L, -U'^i_{s1}]$

$$
\begin{aligned}
& -\int_{L'^i_{s2}}^{U'^i_{s2}} \left(P^i_{01}\left(-U'^i_{s2}, s'^i_2, x\right) - \mathbb{1}\left(s'^i_2 \geq c\right)\right) ds'^i_2 + c \\
&= -\int_{L}^{U} \left(P^i_{01}\left(z^i_1, z^i_2, x\right) - \mathbb{1}\left(z^i_2 \geq c\right)\right) dz^i_2 + c \\
&= -\int_{L}^{U} \frac{\left(P^i_{01}\left(z^i_1, z^i_2, x\right) - \mathbb{1}\left(z^i_2 \geq c\right)\right) f\left(z^i_2|z^i_1, x\right)}{f\left(z^i_2|z^i_1, x\right)} dz^i_2 + c \\
&= -\mathbb{E}\left(\frac{\left(P^i_{01}\left(z^i_1, Z^i_2, x\right) - \mathbb{1}\left(Z^i_2 \geq c\right)\right)}{f\left(Z^i_2|z^i_1, x\right)} \Big| z^i_1, x\right) + c \\
&= -\mathbb{E}\left(\frac{\left(P^i_{01}\left(z^i_1, Z^i_2, x\right) - \mathbb{1}\left(Z^i_2 \geq c\right)\right)}{f\left(Z^i_2|z^i_1, x\right)} \Big| z^i_1, x\right) + c
\end{aligned}
$$

where the first equality holds because $[L'^i_{s2}, U'^i_{s2}] \subset [L, U]$ and the integrand is 0 outside $[L'^i_{s2}, U'^i_{s2}]$. To see this notice that when $z^i_2 > U'^i_{s2}$, $P^i_{01}(L, z^i_2, x) - \mathbb{1}(z^i_2 \geq c) = 1 - 1 = 0$.

When $z_2^i < L_{s2}^{\prime i}$, $P_{01}^i (L, z_2^i, x) - \mathbb{1}(z_2^i \geq c) = 0 - 0 = 0$. Thus it holds that

$$\mathbb{E}\left(S_2^{\prime i}|x\right) = -\mathbb{E}\left(\frac{\left(P_{01}^i\left(z_1^i, Z_2^i, x\right) - \mathbb{1}\left(Z_2^i \geq c\right)\right)}{f\left(Z_2^i|z_1^i, x\right)}\Big|z_1^i, x\right) + c \tag{B.16}$$

Law of iterated expectation gives

$$\mathbb{E}\left(S_2^{\prime i}|x\right) = -\mathbb{E}\left(\frac{\left(D_{01}^i - \mathbb{1}\left(Z_2^i \geq c\right)\right)}{f\left(Z_2^i|Z_1^i, x\right)}\Big|Z_1^i \in \mathcal{I}_1^{\prime}\left(L\right), x\right) + c. \tag{B.17}$$

By symmetry, for $\widetilde{\mathcal{I}}_2^{\prime}(L) = [L, -\widetilde{U}_{s2}^{\prime i}]$, it holds that

$$\mathbb{E}\left(\widetilde{S}_1^{\prime i}|x\right) = -\mathbb{E}\left[\frac{D_{10}^i - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i|Z_2^i, x\right)}\Big|Z_2^i \in \widetilde{\mathcal{I}}_2^{\prime}\left(L\right), x\right] + c. \tag{B.18}$$

Given that $\mathbb{E}(\widetilde{S}_1^{\prime i}|x) = -v_1^i(x)\begin{bmatrix}\beta_1^i \\ \delta_1^i\end{bmatrix}$, it holds that

$$\begin{bmatrix}\beta_1^i \\ \delta_1^i\end{bmatrix} = \left(\mathbb{E}\left[v_1^i\left(X\right)v_1^i\left(X\right)^{\top}\widetilde{W}_{2L}^i\right]\right)^{-1}\mathbb{E}\left[v_1^i\left(X\right)\left(\frac{D_{10}^i - \mathbb{1}\left(Z_1^i \geq c\right)}{f\left(Z_1^i|Z_2^i, X\right)} - c\right)\widetilde{W}_{2L}^i\right],$$

The simple expression for $[\beta_2^i, \delta_2^i]^{\top}$ could be obtained by symmetry.

To find the expression for α^i, notice that

$$P_{01}^i\left(-s_1^{\prime i}, U_{s2}^{\prime}, x\right) = \int_{L_{s1}^{\prime i}}^{s_1^{\prime i}}\int_{L_{s2}^{\prime}}^{U_{s2}^{\prime}} f_{\mathbf{S}^{\prime i}|x}\left(s_1^{\prime i}, s_2^{\prime i}\right) ds_2^{\prime i} ds_1^{\prime i} = F_{S_1^{\prime i}|x}\left(s_1^{\prime i}\right). \tag{B.19}$$

First we show that

$$\mathbb{E}\left(S_1^{\prime i}|x\right) = -\int_{L_{s1}^{\prime}}^{U_{s1}^{\prime}}\left(P_{01}^i\left(-s_1^{\prime i}, U_{s2}^{\prime i}, x\right) - \mathbb{1}\left(-s_1^{\prime i} \leq c\right)\right) ds_1^{\prime i} - c \tag{B.20}$$

Note that

$$
\begin{aligned}
RHS &= -\int_{L_{s1}'^i}^{U_{s1}'^i} \left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq -c\right)\right) ds_1'^i - c \\
&= -\int_{L_{s1}'^i}^{U_{s1}'^i} \left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq -c\right)\right) \frac{\partial s_1'^i}{\partial s_1'^i} ds_1'^i - c \\
&= -\left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq c\right)\right) s_1'^i \Big|_{L_{s1}'^i}^{U_{s1}'^i} + \int_{L_{s1}'^i}^{U_{s1}'^i} \frac{\partial\left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq -c\right)\right)}{\partial s_1'^i} s_1'^i ds_1'^i - c \\
&= -\left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq -c\right)\right) \Big|_{L_{s1}'^i}^{-c} - \left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq -c\right)\right) \Big|_{-c}^{U_{s1}'^i} \\
&+ \int_{L_{s1}'^i}^{-c} \frac{\partial\left(F_{S_1'^i|x}\left(s'^i\right) - 1\right)}{\partial s_1'^i} s_1'^i ds_1'^i + \int_{-c}^{U_{s1}'^i} \frac{\partial F_{S'^i|x}\left(s_1'^i\right)}{\partial s_1'^i} s_1'^i ds_1'^i - c \\
&= -\left(F_{S_1'^i|x}(-c) - 0\right)(-c) - (0-0)\, L_{s1}'^i - (1-1)\, U_{s1}'^i + \left(F_{S_1'^i|x}(-c) - 1\right)(-c) \\
&\quad + \int_{L_{s1}'^i}^{U_{s1}'^i} f_{S_1'^i|x}\left(s_1'^i\right) ds_1'^i - c \\
&= \mathbb{E}\left(S_1'^i|x\right)
\end{aligned}
$$

At the same time, for any $z_2^i \in \mathcal{I}_2'(U) = [U_{s2}', U]$, it holds that

$$
\begin{aligned}
& -\int_{L_{s1}'^i}^{U_{s1}'^i} \left(F_{S_1'^i|x}\left(s_1'^i\right) - \mathbb{1}\left(s_1'^i \geq -c\right)\right) ds_1'^i - c \\
&= -\int_{L_{s1}'^i}^{U_{s1}'} \left(P_{01}^i\left(-s_1'^i, z_2^i, x\right) - \mathbb{1}\left(-s_1'^i \leq c\right)\right) ds_1'^i - c \\
&= \int_{-L_{s1}'^i}^{-U_{s1}'^i} \left(P_{01}^i\left(z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \leq c\right)\right) dz_1^i - c \\
&= -\int_{-U_s'^i}^{-L_s'^i} \left(P_{01}^i\left(z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \leq c\right)\right) dz_1^i - c \\
&= -\int_{L}^{U} \frac{\left(P_{01}^i\left(z_1^i, z_2^i, x\right) - \mathbb{1}\left(z_1^i \leq c\right)\right) f\left(z_1^i|z_2^i, x\right)}{f\left(z_1^i|z_2^i, x\right)} dz_1^i - c \\
&= -\mathbb{E}\left[\frac{P_{01}^i\left(Z_1^i, z_2^i, x\right) - \mathbb{1}\left(Z_1^i \leq c\right)}{f\left(Z_1^i|z_2^i, x\right)} \Big| z_2^i, x\right] - c,
\end{aligned}
$$

where the last equality holds because $[-U_{s1}'^i, -L_{s1}'^i] \subset [L, U]$ and the integrand is zero outside

$[-U'^{i}_{s1}, -L'^{i}_{s1}]$. To see this notice that $c \in [-U'^{i}_{s1}, -L'^{i}_{s1}]$, when $z_1^i > -L'^{i}_{s1}$, $P^i_{01}(z_1^i, U, x) - \mathbb{1}(z_1^i \leq c) = 0 - 0 = 0$, and when $z_1^i < -U'^{i}_{s1}$, $P^i_{01}(z_1^i, U, x) - \mathbb{1}(z_1^i \leq c) = 1 - 1 = 0$. Thus it holds that

$$\mathbb{E}\left(S'^{i}_1|x\right) = -\mathbb{E}\left[\frac{P^i_{01}(Z_1^i, z_2^i, x) - \mathbb{1}(Z_1^i \leq c)}{f(Z_1^i|z_2^i, x)}\big|z_2^i, x\right] - c. \tag{B.21}$$

By law of iterated expectation it holds that

$$\mathbb{E}\left(S'^{i}_1|x\right) = -\mathbb{E}\left[\frac{D^i_{01} - \mathbb{1}(Z_1^i \leq c)}{f(Z_1^i|Z_2^i, x)}\big|Z_2^i \in \mathcal{I}'_2(U), x\right] - c. \tag{B.22}$$

Thus the simple expression for α^i is

$$-\mathbb{E}\left[\frac{D^i_{01} - \mathbb{1}(Z_1^i \leq c)}{f(Z_1^i|Z_2^i, X)}\Big|Z_2^i \in \mathcal{I}'^{i}_2(U)\right] - \mathbb{E}\left[\frac{D^i_{10} - \mathbb{1}(Z_1^i \geq c)}{f(Z_1^i|Z_2^i, X)}\Big|Z_2^i \in \widetilde{\mathcal{I}}'^{i}_2(L)\right].$$

B.6 Proof of Lemma 2.4.1

Proof. Let $Q \equiv \mathcal{I}_1^i(L) \times \mathcal{I}_2^i(L)$ denote the rectangle on the bottom-left corner on which D^i_{00} always equal 1 (including the boundaries). Let $B \equiv (-\widehat{U}^i_{s1}, -\widehat{U}^i_{s2})$ and let $A \equiv (-U^i_{s1}, -U^i_{s2})$. Let $d(B, A)$ be the Euclidean distance between point A and point B. For any $\epsilon > 0$, we analyze the probability that $Pr(d(B, A) > \epsilon) = Pr(B \notin Ball(A, \epsilon))$.

Case 1.$B \notin Ball(A, \epsilon)$ and $B \notin Q$. We will show that asymptotically B is not feasible. As if B is outside Q, the bottom-left rectangular region of point B will overlap with the complement of Q, where the probability of observing $D^i_{00} = 0$ is strictly positive. As M goes to infinity, with probability 1, there will be a point $\widetilde{B}$ in this region with the corresponding $D^i_{00m} = 0$, which implies that B is not a feasible point.

Case 2.$B \notin Ball(A, \epsilon)$ and $B \in Q$. In this case, we can show that, in the limit, B will not be optimal. As in this case, look at the rectengular region between B and A. As M goes to infinity, with probability 1, there will be a point $B^\dagger$ falling in this rectangle. This point is feasible, and this point gives a strictly larger area of rectangle to its bottom-left than point

B, which implies that point B is optimal as defined. Thus it holds that

$$\begin{aligned}
& \lim_{M\to\infty} Pr(d(A,B) > \epsilon) \\
= & \lim_{M\to\infty} Pr(B \notin Ball(A,\epsilon)) \\
= & \lim_{M\to\infty} Pr(B \notin Ball(A,\epsilon) \text{ and } B \notin Q) + \lim_{M\to\infty} Pr(B \notin Ball(A,\epsilon) \text{ and } B \in Q) \\
= & \ 0
\end{aligned}$$

□

B.7 Proof of Lemma 2.5.1

It suffices to show the root-n normality of $\sqrt{M}\left(\frac{1}{M}\sum_{m=1}^{M}\frac{\partial \widehat{P}_1^i\left(Z_{1m}^i, Z_{2m}^i, X_m\right)}{\partial z_2^i} - \mathbb{E}\left(\frac{\partial P_1^i\left(Z_{1m}^i, Z_{2m}^i, X_m\right)}{\partial z_2^i}\right)\right)$.

Step 1. We show that $\frac{1}{M}\sum_{m=1}^{M}\left(q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i\right) - q_\alpha^i\left(\mathbf{x}_m, \gamma_\alpha^{i*}\right)\right) = \int Q_\alpha^i\left(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) dF_{\mathbf{X}}^* + o_p\left(M^{-1/2}\right)$. This could be established by showing

$$\frac{1}{M}\sum_{m=1}^{M}\left(q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i\right) - q_\alpha^i\left(\mathbf{x}_m, \gamma_\alpha^{i*}\right)\right) = \frac{1}{M}\sum_{m=1}^{M} Q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) + o_p\left(M^{-1/2}\right) \quad \text{(B.23)}$$

and

$$\frac{1}{M}\sum_{m=1}^{M} Q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) = \int Q_\alpha^i\left(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) dF_{\mathbf{X}}^* + o_p\left(M^{-1/2}\right). \quad \text{(B.24)}$$

To show equation (B.23), note that

$$\left\|q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i\right) - q_{1\alpha}^i\left(\mathbf{x}_m, \gamma_\alpha^{i*}\right) - Q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right)\right\| \leq b_\alpha^i\left(\gamma_\alpha^{i*}, \mathbf{x}_m\right) \sup_{\mathbf{x}\in\mathcal{H}}\left\|\widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right\|^2,$$

where $\mathcal{H}$ is a compact subset of the support of $\mathbf{x}$ (note that expectation with respect to $\mathbf{Z}^i$ can focus on a compact subset of the support). Under Assumption 2.5.1, 2.5.3 and 2.5.4, it holds that $\sup_{\mathbf{x}\in\mathcal{H}}\left\|\widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right\|^2 = o_p\left(M^{-1/2}\right)$. Equation (B.23) then holds by triangle inequality and weak law of large numbers.

Define $\overline{\gamma}_\alpha^i = \mathbb{E}(\widehat{\gamma}_\alpha^i)$. To show equation (B.24), notice that

$$\begin{aligned}
& \frac{1}{M}\sum_{m=1}^{M} Q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) - \int Q_\alpha^i\left(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) dF_{\mathbf{X}}^* \\
= \; & \frac{1}{M}\sum_{m=1}^{M} Q_\alpha^i\left(\mathbf{x}_m, \widehat{\gamma}_\alpha^i - \overline{\gamma}_\alpha^{i}\right) - \int Q_\alpha^i\left(\mathbf{x}, \widehat{\gamma}_\alpha^i - \overline{\gamma}_\alpha^{i}\right) dF_{\mathbf{X}}^* \\
+ \; & \frac{1}{M}\sum_{m=1}^{M} Q_\alpha^i\left(\mathbf{x}_m, \overline{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) - \int Q_\alpha^i\left(\mathbf{x}, \overline{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) dF_{\mathbf{X}}^*.
\end{aligned}$$

Under Assumption 2.5.3 and 2.5.4 the first difference on the RHS is $o_p(M^{-1/2})$ by Lemma 8.4 in Newey and McFadden (1994). For the second difference, notice that by linearity of $Q_\alpha^i(\mathbf{x}, \cdot)$ and the fact that γ_0^{i*} is uniformly bounded away from 0 (Assumption 2.5.1), it holds that there exists constant $C < \infty$ such that $\mathbb{E}\,||Q_\alpha^i\left(\mathbf{x}_m, \overline{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right)||^2 \leq C \sup_{\mathbf{x}\in\mathcal{H}} ||\overline{\gamma}_\alpha^i - \gamma_\alpha^{i*}||^2 \to 0$ under Assumption 2.5.4. Also note that $\mathbb{E}[Q_\alpha^i\left(\mathbf{x}_m, \overline{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right)] = o(M^{-1/2})$ under Assumption 2.5.4. The second difference is $o_p(M^{-1/2})$ by Chebyshev's inequality.

Step 2. We show that $\int Q_\alpha^i\left(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) dF_{\mathbf{X}}^* = \frac{1}{M}\sum_{m=1}^{M} \nu_\alpha^i\left(\mathbf{x}_m\right) + o_p\left(M^{-1/2}\right)$ for some function $\nu_{1\alpha}^i\left(\mathbf{x}_m\right)$. This could be established by showing

$$\int Q_\alpha^i\left(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}\right) dF_{\mathbf{X}}^* = \int \left(\psi_\alpha^i\left(\mathbf{x}\right)\mathbf{d}_\alpha^i - \mu_\alpha^i\right) d\widehat{F}_{Y_\alpha^i}, \tag{B.25}$$

and

$$\begin{aligned}
& \int \left(\psi_\alpha^i\left(\mathbf{x}\right)\mathbf{d}_\alpha^i - \mu_\alpha^i\right) d\widehat{F}_{Y_\alpha^i} \\
= \; & \frac{1}{M}\sum_{m=1}^{M}\left(\psi_\alpha^i\left(\mathbf{x}_m\right)\mathbf{d}_{\alpha m}^i - \mu_\alpha^i\right) + o_p\left(M^{-1/2}\right),
\end{aligned} \tag{B.26}$$

where $\widehat{F}_{Y_\alpha^i}$ is a kernel-smoothed version of the empirical distribution of Y_α^i. Then set $\psi_\alpha^i\left(\mathbf{x}_m\right)\mathbf{d}_{\alpha m}^i - \mu_\alpha^i = \nu_\alpha^i\left(\mathbf{x}_m\right)$ gives the desired result. To show equation (B.25), first we show that for generic γ_α^i, it holds that

$$\int Q_\alpha^i(\mathbf{x}, \gamma_\alpha^i) dF_{\mathbf{x}}^* = \int \psi_\alpha^i(\mathbf{x}) \gamma_\alpha^i d\mathbf{x}, \tag{B.27}$$

for some function $\psi_\alpha^i(\mathbf{x})$. Note that

$$Q_\alpha^i(\mathbf{x}, \gamma_\alpha^i) = \nabla_{z_2^i}(J_\alpha^i \gamma_\alpha^i) = \nabla_{z_2^i}(J_\alpha^i)\gamma_\alpha^i + J_\alpha^i \nabla_{z_2^i}\gamma_\alpha^i,$$

where $J_\alpha^i = [-\frac{\gamma_1^{i*}}{\gamma_0^{i*2}}, \frac{1}{\gamma_0^{i*}}]$. So

$$\begin{aligned}
\int Q_\alpha^i(\mathbf{x}, \gamma_\alpha^i) dF_{\mathbf{x}}^* &= \int \left(\nabla_{z_2^i}(J_\alpha^i)\gamma_\alpha^i + J_\alpha^i \nabla_{z_2^i}\gamma_\alpha^i\right) dF_{\mathbf{x}}^* \\
&= \int \gamma_0^{i*}\nabla_{z_2^i}(J_\alpha^i)\gamma_\alpha^i d\mathbf{x} + \int J_\alpha^i \gamma_0^{i*}\nabla_{z_2^i}\gamma_\alpha^i d\mathbf{x} \\
&= \int \psi_\alpha^i(\mathbf{x})\gamma_\alpha^i d\mathbf{x},
\end{aligned}$$

where $\psi_\alpha^i(\mathbf{x}) = \gamma_0^{i*}\nabla_{z_2^i}(J_\alpha^i) + J_\alpha^i\gamma_0^{i*}\mathbb{1}(z_2^i = U) - J_\alpha^i\gamma_0^{i*}\mathbb{1}(z_2^i = L) + \nabla_{z_2^i}(J_\alpha^i\gamma_0^{i*})$ and the last equality follows by integration by part. Equation (B.27) implies that

$$\int Q_\alpha^i(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}) dF_{\mathbf{x}}^* = \int \psi_\alpha^i(\mathbf{x})(\widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}) d\mathbf{x}. \tag{B.28}$$

Plugging the expression for $\widehat{\gamma}_\alpha^i$ and use the definition of μ_α^i gives

$$\int Q_\alpha^i(\mathbf{x}, \widehat{\gamma}_\alpha^i - \gamma_\alpha^{i*}) dF_{\mathbf{X}}^* = \int \left(\psi_\alpha^i(\mathbf{x})\mathbf{d}_\alpha^i - \mu_\alpha^i\right) d\widehat{F}_{Y_\theta^i}, \tag{B.29}$$

where we define

$$\int a(y_\alpha^i) d\widehat{F}_{Y_\alpha^i} = \frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{K}\int a(d_{1m}^i, \mathbf{z}_m^i, x_m) K_h(\mathbf{z}_m^i - \mathbf{z}^i) d\mathbf{z}^i \mathbb{1}(x_m = x_k). \tag{B.30}$$

To show equation (B.26), notice that

$$\int \left(\psi^i_\alpha\left(\mathbf{x}\right)\mathbf{d}^i_\alpha-\mu^i_\alpha\right)d\widehat{F}_{Y^i_\alpha}-\frac{1}{M}\sum_{m=1}^{M}\left(\psi^i_\alpha\left(\mathbf{x}_m\right)\mathbf{d}^i_{\alpha m}-\mu^i_\alpha\right)$$

$$\int \left(\psi^i_{\alpha 0}\left(\mathbf{x}\right)+\psi^i_{\alpha 1}\left(\mathbf{x}\right)d^i_1-\mu^i_\alpha\right)d\widehat{F}_{Y^i_\alpha}-\frac{1}{M}\sum_{m=1}^{M}\left(\psi^i_{\alpha 0}\left(\mathbf{x}_m\right)+\psi^i_{\alpha 1}\left(\mathbf{x}_m\right)d^i_{1m}-\mu^i_\alpha\right)$$

$$= \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{s=0,1}d^i_{\alpha[s]m}\left(\sum_{k=1}^{K}\int \psi^i_{\alpha s}(\mathbf{z}^i,x_k)K_h\left(\mathbf{z}^i_m-\mathbf{z}^i\right)d\mathbf{z}^i\mathbb{1}\left(x_m=x_k\right)-\psi^i_{\alpha s}\left(\mathbf{x}_m\right)\right)\right),$$

where $d^i_{\alpha[0]m}=1$ and $d^i_{\alpha[1]m}=d^i_{1m}$. So it suffices to show that

$$\left\|\mathbb{E}\left[\sum_{s=0,1}D^i_{\alpha[s]m}\left(\sum_{k=1}^{K}\int \psi^i_{\alpha s}(\mathbf{z}^i,x_k)K_h\left(\mathbf{Z}^i_m-\mathbf{z}^i\right)d\mathbf{z}^i\mathbb{1}\left(X_m=x_k\right)-\psi^i_{\alpha s}\left(\mathbf{X}_m\right)\right)\right]\right\|=o(M^{-1/2})$$

and

$$\mathbb{E}\left[\left\|\mathbf{D}^i_{\alpha m}\right\|^2\left\|\sum_{k=1}^{K}\int \psi^i_{\alpha}(\mathbf{z}^i,x_k)K_h\left(\mathbf{Z}^i_m-\mathbf{z}^i\right)d\mathbf{z}^i\mathbb{1}\left(X_m=x_k\right)-\psi^i_{\alpha}\left(\mathbf{X}_m\right)\right\|^2\right]=o(1).$$

With some slight abuse of notation, the first equality holds as its left hand side could be rewritten as[2]

$$\left\|\left[\sum_{s=0,1}\left(\sum_{k=1}^{K}\int\int \psi^i_{\alpha s}(\mathbf{z}^i,x_k)K_h\left(\mathbf{z}^i_m-\mathbf{z}^i\right)d\mathbf{z}^i\mathbb{1}\left(x_m=x_k\right)\gamma^i_{\alpha[s]}(\mathbf{x}_m)d\mathbf{x}_m-\int \psi^i_{\alpha s}\left(\mathbf{x}_m\right)\gamma^i_{\alpha[s]}(\mathbf{x}_m)d\mathbf{x}_m\right)\right]\right\|.$$

[2]Note that here each $\gamma^i_{\alpha[s]}(\mathbf{x}_m)$ already contains the density for $\mathbf{x}_m$ by construction. Also note that here expectation with respect to observation m is expressed using an integral with respect to $\mathbf{x}_m$ only to save space as no calculation is done in this step. In the following lines wherever explicit calculation is performed, this integral is written out in more details: integration with respect to the continuous variables $\mathbf{z}^i_m$ and summation with respect to the discrete variable x_m.

This expression could be further rewritten as

$$\begin{aligned}
& \left\| \sum_{l=1}^{K} \left[\sum_{s=0,1} \left(\int \int \psi^i_{\alpha s}(\mathbf{z}^i, x_l) K_h \left(\mathbf{z}^i_m - \mathbf{z}^i \right) d\mathbf{z}^i \gamma^i_{\alpha[s]}(\mathbf{z}^i_m, x_l) d\mathbf{z}^i_m - \int \psi^i_{\alpha s} \left(\mathbf{z}^i_m, x_l \right) \gamma^i_{\alpha[s]}(\mathbf{z}^i_m, x_l) d\mathbf{z}^i_m \right) \right] \right\| \\
= & \left\| \sum_{l=1}^{K} \left[\sum_{s=0,1} \left(\int \int \psi^i_{\alpha s}(\widetilde{\mathbf{z}}^i, x_l) K(u) (\gamma^i_{\alpha[s]}(\widetilde{\mathbf{z}}^i - uh, x_l) - \gamma^i_{\alpha[s]}(\widetilde{\mathbf{z}}^i, x_l)) du d\widetilde{\mathbf{z}}^i \right) \right] \right\| \\
\leq & \sum_{l=1}^{K} \int \left\| \psi^i_{\alpha}(\widetilde{\mathbf{z}}^i, x_l) \right\| \left\| \left[\int K(u) (\gamma^i_{\alpha}(\widetilde{\mathbf{z}}^i - uh, x_l) - \gamma^i_{\alpha}(\widetilde{\mathbf{z}}^i, x_l)) du \right] \right\| d\widetilde{\mathbf{z}}^i \\
= & \; o(M^{-1/2})
\end{aligned}$$

where the first equality holds by change of variable, and the inequality holds by triangle inequality and Cauchy-Schwarz inequality, and the last equality holds by Assumptions 2.5.1, 2.5.3 and 2.5.4.

The second claim that the second moment converges to zero holds as under Assumptions 2.5.1, 2.5.4 and 2.5.5, simplifying the expression and applying dominated convergence theorem delivers $\mathbb{E}\left[\left\|\sum_{k=1}^{K} \int \psi^i_{\alpha}(\mathbf{z}^i, x_k) K_h \left(\mathbf{Z}^i_m - \mathbf{z}^i\right) d\mathbf{z}^i \mathbb{1}\left(X_m = x_k\right) - \psi^i_{\alpha}\left(\mathbf{X}_m\right)\right\|^4\right] = \mathbb{E}\left[\left\|\int \psi^i_{\alpha}(\mathbf{z}^i, X_m) K_h \left(\mathbf{Z}^i_m - \mathbf{z}^i\right) d\mathbf{z}^i - \psi^i_{\alpha}\left(\mathbf{X}_m\right)\right\|^4\right] \to 0$. The claim then follows by Cauchy-Schwarz inequality.

Combining the established results from *Step 1* and *Step 2*, it holds that

$$\frac{1}{\sqrt{M}} \sum_{m=1}^{M} q^i_{\alpha}(\mathbf{X}_m, \widehat{\gamma}^i_{\alpha}) = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \left(q^i_{\alpha}(\mathbf{X}_m, \gamma^i_{\alpha}) + \nu^i_{\alpha}\left(\mathbf{X}_m\right) \right) + o_p(1),$$

and normality follows by central limit theorem.

B.8 Proof of Theorem 2.9

Without loss of generality we prove the root-n normality of $\widehat{\theta}^i_1$ in the case of $\alpha^i \leq 0$. The proof for other cases and other estimated structural parameters are similar. Define $\widehat{\Sigma}^i_1 =$

$\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\widehat{v}_1^i(X_m)^\top W_{2Lm}^i}{M}$, and $\Sigma_1^i = \mathbb{E}[v_1^i\,(X_m)\,v_1^i\,(X_m)^\top\,W_m^i]$. First note that

$$\begin{aligned}
\widehat{\theta}_1^i &= -\left(\widehat{\Sigma}_1^i\right)^{-1}\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}-\widehat{v}_1^i(X_m)^\top\theta_1^i+\widehat{v}_1^i(X_m)^\top\theta_1^i\right)W_{2Lm}^i}{M}\\
&= \theta_1^i-\left(\widehat{\Sigma}_1^i\right)^{-1}\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}+\widehat{v}_1^i(X_m)^\top\theta_1^i\right)W_{2Lm}^i}{M}
\end{aligned}$$

Thus

$$\sqrt{M}\left(\widehat{\theta}_1^i-\theta_1^i\right) = -\left(\widehat{\Sigma}_1^i\right)^{-1}\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}+\widehat{v}_1^i(X_m)^\top\theta_1^i\right)W_{2Lm}^i}{\sqrt{M}}$$

Thus it suffices to show the normality of $\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}+\widehat{v}_1^i(X_m)^\top\theta_1^i\right)W_{2Lm}^i}{\sqrt{M}}$. Also as

$$\begin{aligned}
&\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}+\widehat{v}_1^i(X_m)^\top\theta_1^i\right)W_{2Lm}^i}{\sqrt{M}}\\
=\;&\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}\right)W_{2Lm}^i}{\sqrt{M}}+\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\widehat{v}_1^i(X_m)^\top\theta_1^iW_{2Lm}^i}{\sqrt{M}}
\end{aligned}$$

and the second term is asymptotically normal, it suffices to show the normality of

$$\frac{\sum_{m=1}^{M}\widehat{v}_1^i(X_m)\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}\right)W_{2Lm}^i}{\sqrt{M}}.$$

Similar to Lewbel and Tang (2015), we have a first-step estimator that converges at root-N (in our problem, it's the estimated belief, while in their paper it's the estimated strategic coefficient). The sketch of the proof is thus similar to that used in Lewbel and Tang (2015). Define

$$q_{1\theta}^i\left(\mathbf{x}_m,\widehat{p}_1^{-i},\widehat{\gamma}_\theta^i\right) = \begin{bmatrix} X_m \\ \widehat{P}_1^{-i}(X_m)\end{bmatrix}\left(\frac{D_{00m}^i+\mathbb{1}\left(Z_{1m}^i\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^i|Z_{2m}^i,X_m\right)}\right)W_{2Lm}^i,$$

and define

$$q_{1\theta}^{i}\left(\mathbf{x}_m, \widehat{\gamma}_{\theta}^{i}\right) = v_1^i(X_m)\left(\frac{D_{00m}^{i} + \mathbb{1}\left(Z_{1m}^{i} \geq 0\right) - 1}{\widehat{f}\left(Z_{1m}^{i}|Z_{2m}^{i}, X_m\right)}\right) W_{2Lm}^{i}.$$

Let $J_{1\theta p}(\widehat{\gamma}_0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \left(\frac{D_{00m}^{i}+\mathbb{1}\left(Z_{1m}^{i}\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^{i}|Z_{2m}^{i},X_m\right)}\right) W_{2Lm}^{i}$. It holds that:

$$q_{1\theta}^{i}\left(\mathbf{X}_m, \widehat{p_1}^{-i}, \widehat{\gamma}_{\theta}^{i}\right) = q_{1\theta}^{i}\left(\mathbf{X}_m, \widehat{\gamma}_{\theta}^{i}\right) + J_{1\theta p}(\widehat{\gamma}_0)(\widehat{P}_1^{-i}(X_m) - P_1^{-i}(X_m)).$$

Notice that the second term could be rewritten as

$$J_{1\theta p}(\widehat{\gamma}_0)(\widehat{P}_1^{-i}(X_m) - P_1^{-i}(X_m)) = \sum_{k=1}^{K} J_{1\theta p}(\widehat{\gamma}_0, x_k)\mathbb{1}(X_m = x_k)(\widehat{P}_1^{-i}(x_k) - P_1^{-i}(x_k)),$$

where $J_{1\theta p}(\widehat{\gamma}_0, x_k)$ is defined as $J_{1\theta p}(\widehat{\gamma}_0)$ holding $X_m = x_k$. Thus

$$\begin{aligned} \frac{1}{\sqrt{M}}\sum_{m=1}^{M} q_{1\theta}^{i}\left(\mathbf{X}_m, \widehat{p_1}^{-i}, \widehat{\gamma}_{\theta}^{i}\right) &= \frac{1}{\sqrt{M}}\sum_{m=1}^{M} q_{1\theta}^{i}\left(\mathbf{X}_m, p_1^{-i}, \widehat{\gamma}_{\theta}^{i}\right) \\ + \sum_{k=1}^{K}[\frac{1}{M}\sum_{m=1}^{M} J_{1\theta p}(\widehat{\gamma}_0, x_k)\mathbb{1}(X_m = x_k)]&\sqrt{M}(\widehat{P}_1^{-i}(x_k) - P_1^{-i}(x_k)), \end{aligned}$$

by consistency of $\widehat{P}_1^{-i}(X_m)$, uniform consistency and positiveness of $\widehat{\gamma}_0$ and law of large numbers, it holds that $\frac{1}{M}\sum_{m=1}^{M} J_{1\theta p}(\widehat{\gamma}_0, x_k)\mathbb{1}(X_m = x_k) \xrightarrow{p} J_{1\theta pk}$, which is a constant vector. Thus the linear representation for the second term as a whole is written as:

$$\begin{aligned} r_1^i(\mathbf{x}_m) &= \sum_{k=1}^{K} J_{1\theta pk}[\frac{1}{Pr(X = x_k)}\mathbb{1}(D_m^{-i} = (1,1) \text{ or } (1,0), X_m = x_k)) \\ &- \frac{Pr(D^{-i} = (1,1) \text{ or } (1,0), X = x_k)}{Pr(X = x_k)^2}\mathbb{1}(X_m = x_k))] + o_p(N^{-1/2}). \end{aligned}$$

It remains to find the asymptotic linear representation of $\frac{1}{\sqrt{M}}\sum_{m=1}^{M} q_{1\theta}^{i}\left(\mathbf{x}_m, \widehat{\gamma}_{\theta}^{i}\right)$.[3] The reasoning in the proof is similar to that of Lemma 2.5.1, and thus we will not lay out all the details.

---

[3]Note that once we get this asymptotic linear representation, the asymptotic linear representation of $\overset{\circ}{q}{}_{1\theta}^{i}\left(\mathbf{x}_m, \widehat{\gamma}_{\theta}^{i}\right) \equiv \left(\frac{D_{00m}^{i}+\mathbb{1}\left(Z_{1m}^{i}\geq 0\right)-1}{\widehat{f}\left(Z_{1m}^{i}|Z_{2m}^{i},X_m\right)}\right) W_{2Lm}^{i}$ could then be found by removing $v_1^i(X_m)$ from each term.

Step 1. We show that $\frac{1}{M}\sum_{m=1}^{M}\left(q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i\right)-q_{1\theta}^{i}\left(\mathbf{x}_m,\gamma_\theta^{i*}\right)\right)=\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)dF_{\mathbf{X}}^{*}+o_p\left(M^{-1/2}\right)$. This could be established by showing

$$\frac{1}{M}\sum_{m=1}^{M}\left(q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i\right)-q_{1\theta}^{i}\left(\mathbf{x}_m,\gamma_\theta^{i*}\right)\right)=\frac{1}{M}\sum_{m=1}^{M}Q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)+o_p\left(M^{-1/2}\right) \quad \text{(B.31)}$$

and

$$\frac{1}{M}\sum_{m=1}^{M}Q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)=\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)dF_{\mathbf{X}}^{*}+o_p\left(M^{-1/2}\right). \quad \text{(B.32)}$$

To show equation (B.31), note that

$$\left\|q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i\right)-q_{1\theta}^{i}\left(\mathbf{x}_m,\gamma_\theta^{i*}\right)-Q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)\right\|\leq b_\theta^i\left(\gamma_\theta^{i*},\mathbf{x}_m\right)\sup_{\mathbf{x}\in\mathcal{H}}\left\|\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right\|^2,$$

where $\mathcal{H}$ is a compact subset of the support of $\mathbf{x}$. Under Assumption 2.5.1, 2.5.3 and 2.5.4, it holds that $\sup_{\mathbf{x}\in\mathcal{H}}\left\|\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right\|^2=o_p\left(M^{-1/2}\right)$. Equation (B.31) then holds by triangle inequality and weak law of large numbers.

To show equation (B.32), notice that

$$\begin{aligned}
&\frac{1}{M}\sum_{m=1}^{M}Q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)-\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)dF_{\mathbf{X}}^{*}\\
=&\frac{1}{M}\sum_{m=1}^{M}Q_{1\theta}^{i}\left(\mathbf{x}_m,\widehat{\gamma}_\theta^i-\overline{\gamma}_\theta^i\right)-\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_\theta^i-\overline{\gamma}_\theta^i\right)dF_{\mathbf{X}}^{*}\\
+&\frac{1}{M}\sum_{m=1}^{M}Q_{1\theta}^{i}\left(\mathbf{x}_m,\overline{\gamma}_\theta^i-\gamma_\theta^{i*}\right)-\int Q_{1\theta}^{i}\left(\mathbf{x},\overline{\gamma}_\theta^i-\gamma_\theta^{i*}\right)dF_{\mathbf{X}}^{*}.
\end{aligned}$$

Under Assumption 2.5.3 and 2.5.4 the first difference on the RHS is $o_p(M^{-1/2})$ by Lemma 8.4 in Newey and McFadden (1994) and the second difference is $o_p(M^{-1/2})$ by Chebyshev inequality.

Step 2. We show that $\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_\theta^i-\gamma_\theta^{i*}\right)dF_{\mathbf{X}}^{*}=\frac{1}{M}\sum_{m=1}^{M}\nu_{1\theta}^{i}\left(\mathbf{x}_m\right)+o_p\left(M^{-1/2}\right)$ for some function $\nu_{1\theta}^{i}\left(\mathbf{x}_{m}\right)$. This could be established by showing

$$\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_{\theta}^{i}-\gamma_{\theta}^{i*}\right)dF_{\mathbf{X}}^{*}=\int\left(\psi_{1\theta0}^{i}\left(\mathbf{x}\right)-\mu_{1\theta}^{i}\right)d\widehat{F}_{Y_{\theta}^{i}},\tag{B.33}$$

and

$$\begin{aligned}&\int\left(\psi_{1\theta0}^{i}\left(\mathbf{x}\right)-\mu_{1\theta}^{i}\right)d\widehat{F}_{Y_{\theta}^{i}}\\=\;&\frac{1}{M}\sum_{m=1}^{M}\left(\psi_{1\theta0}^{i}\left(\mathbf{x}_{m}\right)-\mu_{1\theta}^{i}\right)+o_{p}\left(M^{-1/2}\right)\end{aligned}\tag{B.34}$$

where $\widehat{F}_{Y_{\theta}^{i}}$ is a kernel-smoothed version of the empirical distribution of Y_{θ}^{i}. Then set $\psi_{1\theta0}^{i}\left(\mathbf{x}_{m}\right)-\mu_{1\theta}^{i}=\nu_{1\theta}^{i}\left(\mathbf{x}_{m}\right)$ gives the desired result.

To show equation (B.33), first we show that for generic γ_{θ}^{i}, it holds that

$$\int Q_{1\theta}^{i}(\mathbf{x},\gamma_{\theta}^{i})dF_{\mathbf{x}}^{*}=\int\psi_{1\theta0}^{i}\left(\mathbf{x}\right)\gamma_{0}^{i}d\mathbf{x},\tag{B.35}$$

for some function $\psi_{1\theta0}^{i}\left(\mathbf{x}\right)$ and $\psi_{1\theta2}^{i}\left(\mathbf{x}\right)$. Let $J_{1\theta0}^{i}$ denote the derivative of $q_{1\theta}^{i}\left(\mathbf{x},\gamma_{\theta}^{i}\right)$ with respect to γ_{0}^{i} evaluated at γ_{θ}^{i*}. Let $c^{i}(\mathbf{x})=v_{1}^{i}(x)w_{2L}^{i}(D_{00}^{i}+\mathbb{1}(z_{1}^{i}\geq0)-1)$. Note that

$$\int J_{1\theta0}^{i}\gamma_{0}^{i}dF_{\mathbf{x}}^{*}=\int\left(\mathbb{E}\left(\frac{c_{1}^{i}(\mathbf{x})}{\gamma_{0}^{i*}}|z_{2}^{i},x\right)\widetilde{\gamma}_{0}^{i*}(z_{2}^{i},x)-\frac{c_{1}^{i}(\mathbf{x})\int\gamma_{0}^{i*}(\mathbf{x})dt_{1}}{\gamma_{0}^{i*}}\right)\gamma_{0}^{i}d\mathbf{x},$$

where $\widetilde{\gamma}_{0}^{i*}(z_{2}^{i},x)$ denotes the true joint density of (Z_{2}^{i},X). Thus $\psi_{1\theta0}^{i}(\mathbf{x})=v_{1}^{i}(x)w_{2L}^{i}\frac{-\gamma_{2}^{i*}}{f(z_{1}^{i}|z_{2}^{i},x)}+\mathbb{E}\left(\frac{c_{1}^{i}(\mathbf{x})}{\gamma_{0}^{i*}}|z_{2}^{i},x\right)\widetilde{\gamma}_{0}^{i*}(z_{2}^{i},x)-\frac{c_{1}^{i}(\mathbf{x})\int\gamma_{0}^{i*}(\mathbf{x})dt_{1}}{\gamma_{0}^{i*}}$. Equation (B.35) implies that

$$\int Q_{1\theta}^{i}(\mathbf{x},\widehat{\gamma}_{\theta}^{i}-\gamma_{\theta}^{i*})dF_{\mathbf{x}}^{*}=\int\psi_{1\theta0}^{i}\left(\mathbf{x}\right)(\widehat{\gamma}_{\theta}^{i}-\gamma_{\theta}^{i*})d\mathbf{x}.\tag{B.36}$$

Plugging the expression for $\widehat{\gamma}_{\theta}^{i}$ and use the definition of $\mu_{1\theta}^{i}$ gives

$$\int Q_{1\theta}^{i}\left(\mathbf{x},\widehat{\gamma}_{\theta}^{i}-\gamma_{\theta}^{i*}\right)dF_{\mathbf{X}}^{*}=\int\left(\psi_{1\theta0}^{i}\left(\mathbf{x}\right)-\mu_{1\theta}^{i}\right)d\widehat{F}_{Y_{\theta}^{i}},\tag{B.37}$$

where we define

$$\int a(y_{\theta}^{i})d\widehat{F}_{Y_{\theta}^{i}}=\frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{K}\int a(d_{00m}^{i},\mathbf{z}_{m}^{i},x_{m})K_{h}(\mathbf{z}_{m}^{i}-\mathbf{z}^{i})d\mathbf{z}^{i}\mathbb{1}(x_{m}=x_{k}).\tag{B.38}$$

To show equation (B.34), notice that

$$\begin{aligned}
&\int \left(\psi^i_{1\theta 0}\left(\mathbf{x}\right) - \mu^i_{1\theta}\right) d\widehat{F}_{Y^i_\theta} - \frac{1}{M}\sum_{m=1}^{M}\left(\psi^i_{1\theta 0}\left(\mathbf{x}_m\right) - \mu^i_{1\theta}\right) \\
= \;& \frac{1}{M}\sum_{m=1}^{M}\left(\left(\sum_{k=1}^{K}\int \psi^i_{1\theta s}(\mathbf{z}^i, x_k) K_h\left(\mathbf{z}^i_m - \mathbf{z}^i\right) d\mathbf{z}^i \mathbb{1}\left(x_m = x_k\right) - \psi^i_{1\theta 0}\left(\mathbf{x}_m\right)\right)\right),
\end{aligned}$$

where $d^i_{\theta[0]m} = 1$ and $d^i_{\theta[2]m} = d^i_{00m}$. So it suffices to show that

$$\left\|\mathbb{E}\left[\left(\sum_{k=1}^{K}\int \psi^i_{1\theta 0}(\mathbf{z}^i, x_k) K_h\left(\mathbf{Z}^i_m - \mathbf{z}^i\right) d\mathbf{z}^i \mathbb{1}\left(X_m = x_k\right) - \psi^i_{1\theta 0}\left(\mathbf{X}_m\right)\right)\right]\right\| = o(M^{-1/2})$$

and

$$\mathbb{E}\left[\left\|\sum_{k=1}^{K}\int \boldsymbol{\psi}^i_{1\theta 0}(\mathbf{z}^i, x_k) K_h\left(\mathbf{Z}^i_m - \mathbf{z}^i\right) d\mathbf{z}^i \mathbb{1}\left(X_m = x_k\right) - \boldsymbol{\psi}^i_{1\theta 0}\left(\mathbf{X}_m\right)\right\|^2\right] = o(1).$$

Both equations hold under Assumptions 2.5.1, 2.5.3 and 2.5.4. Combining the established results from *Step 1* and *Step 2*, it holds that

$$\frac{1}{\sqrt{M}}\sum_{m=1}^{M} q^i_{1\theta}(\mathbf{X}_m, \widehat{\gamma}^i_\theta) = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(q^i_{1\theta}(\mathbf{X}_m, \gamma^i_\theta) + \nu^i_{1\theta}\left(\mathbf{X}_m\right)\right) + o_p(1).$$

Similarly, there exists $\mathring{q}^i_{1\theta}(\mathbf{X}_m, \gamma^i_\theta)$ and $\mathring{\nu}^i_{1\theta}\left(\mathbf{X}_m\right)$ such that

$$\frac{1}{\sqrt{M}}\sum_{m=1}^{M} \mathring{q}^i_{1\theta}(\mathbf{X}_m, \widehat{\gamma}^i_\theta) = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(\mathring{q}^i_{1\theta}(\mathbf{X}_m, \gamma^i_\theta) + \mathring{\nu}^i_{1\theta}\left(\mathbf{X}_m\right)\right) + o_p(1).$$

Note that

$$\begin{aligned}
&\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\overset{\circ}{q}{}^{i}_{1\theta}(\mathbf{X}_m,\widehat{\gamma}^i_\theta)\\
=\;&\frac{1}{\sqrt{M}}\sum_{k=1}^{K}\sum_{m=1}^{M}\overset{\circ}{q}{}^{i}_{1\theta}(\mathbf{Z}^i_m,x_k,\widehat{\gamma}^i_\theta)\mathbb{1}(X_m=x_k)\\
=\;&\frac{1}{\sqrt{M}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left(\overset{\circ}{q}{}^{i}_{1\theta}(\mathbf{Z}^i_m,x_k,\gamma^i_\theta)+\overset{\circ}{\nu}{}^{i}_{1\theta}\left(\mathbf{Z}^i_m,x_k\right)\right)\mathbb{1}(X_m=x_k)\\
=\;&\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(\overset{\circ}{q}{}^{i}_{1\theta}(\mathbf{X}_m,\gamma^i_\theta)+\overset{\circ}{\nu}{}^{i}_{1\theta}\left(\mathbf{X}_m\right)\right)+o_p(1).
\end{aligned}$$

Let $E^i_{1k} = \mathbb{E}\left(\left(\overset{\circ}{q}{}^{i}_{1\theta}(\mathbf{Z}^i_m,x_k,\gamma^i_\theta)+\overset{\circ}{\nu}{}^{i}_{1\theta}\left(\mathbf{Z}^i_m,x_k\right)\right)\mathbb{1}(X_m=x_k)\right)$, $E^i_1 = \mathbb{E}\left(\left(\overset{\circ}{q}{}^{i}_{1\theta}(\mathbf{X}_m,\gamma^i_\theta)+\overset{\circ}{\nu}{}^{i}_{1\theta}\left(\mathbf{X}_m\right)\right)\right)$.

$$\sqrt{M}(\widehat{\theta}^i_1-\theta^i_1)=(\Sigma^i_1)^{-1}\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(q^i_{1\theta}(\mathbf{X}_m,\gamma^i_\theta)+\nu^i_{1\theta}\left(\mathbf{X}_m\right)+r^i_1(\mathbf{X}_m)+v^i_1(X_m)v^i_1(X_m)^\top W^i_{2Lm}\theta^i_1\right)+o_p(1).$$

Thus

$$\sqrt{M}\left(\widehat{\theta}^i_1-\theta^i_1\right)\xrightarrow{d}N\left(0,\Omega_{\theta^i_1}\right),$$

with

$$\Omega_{\theta^i_1}=(\Sigma^i_1)^{-1}Var\left(q^i_{1\theta}(\mathbf{X}_m,\gamma^i_\theta)+\nu^i_{1\theta}\left(\mathbf{X}_m\right)+r^i_1(\mathbf{X}_m)+v^i_1(X_m)v^i_1(X_m)^\top W^i_{2Lm}\theta^i_1\right)(\Sigma^i_1)^{-1}.$$

Regarding the root-n normality of $\widehat{\alpha}^i$, notice that α^i is estimated as

$$\frac{\sum_{m=1}^{M}\frac{D^i_{11m}-\mathbb{1}\left(Z^i_{1m}\geq 0\right)}{\widehat{f}\left(Z^i_{1m}|Z^i_{2m},X_m\right)}W^i_{2Um}}{\sum_{m=1}^{M}W^i_{2Um}}+\frac{\sum_{m=1}^{M}\frac{D^i_{00m}+\mathbb{1}\left(Z^i_{1m}\geq 0\right)-1}{\widehat{f}(Z^i_{1m}|Z^i_{2m},X_m)}W^i_{2Lm}}{\sum_{m=1}^{M}W^i_{2Lm}}.$$

Use argument similar to the above, we can show that both terms in this estimator are root-n consistent and asymptotically normal, and our desired result holds.

Appendix C

IDENTIFICATION AND ESTIMATION OF BINARY GAMES OF INCOMPLETE INFORMATION UNDER SYMMETRY OF THE UNOBSERVED PRIVATE INFORMATION

C.1 Proof of Theorem 3.3.2

proof. The proof follows a similar argument as the proof of Theorem 2.1 in Lewbel et al. (2020). We need to replace their utility function with our expected payoff function, and replace their identifying restriction based on pdf with our identifying restriction based on cdf. To be self-contained, we provide the major steps of the proof here. If $\boldsymbol{\beta}^i = \boldsymbol{\beta}_0^i$, then there are two cases, either $(-z^i + 2\mathbf{x}_p^i\boldsymbol{\beta}_0^i, x) \in \Omega_{(Z^i,X)}$ implying $d^i(\boldsymbol{\beta}_0^i; z^i, x) = 0$, or $(-z^i + 2\mathbf{x}_p^i\boldsymbol{\beta}_0^i, x) \notin \Omega_{(Z^i,X)}$ and by definition $\Pr((Z^i, X) \in \mathcal{D}^i(\boldsymbol{\beta}^i)) = 0$. Showing that $\Pr((Z^i, X) \in \mathcal{D}^i(\boldsymbol{\beta}^i)) = 0$ implying $\boldsymbol{\beta}^i = \boldsymbol{\beta}_0^i$ is equivalent to showing that $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$ implies $\Pr((Z^i, X) \in \mathcal{D}^i(\boldsymbol{\beta}^i)) > 0$. Recall that $\mathscr{X}(\boldsymbol{\beta}^i)$ is the set of x such that $\mathbf{x}_p^i(\boldsymbol{\beta}^i - \boldsymbol{\beta}_0^i) \neq \mathbf{0}$ and also recall that $\mathscr{S}(\boldsymbol{\beta}^i) \equiv \{(z^i, x) : z^i \in \mathcal{S}_{Z^i|x},\ x \in \mathscr{X}(\boldsymbol{\beta}^i)\}$. By Assumption 3.3.5 (i), $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$ implies $\Pr(X \in \mathscr{X}(\boldsymbol{\beta}^i)) > 0$. Thus, as long as the parameter deviates from the true value, it's always possible to find x such that $\mathbf{x}_p^i(\boldsymbol{\beta}^i - \boldsymbol{\beta}_0^i) \neq \mathbf{0}$. By Assumption 3.3.5 (iii), it holds that $\Pr((Z^i, X) \in \mathscr{S}(\boldsymbol{\beta}^i)) > 0$. It remains to show that $\Pr(d^i(\boldsymbol{\beta}^i; Z^i, X) \neq 0 | (Z^i, X) \in \mathscr{S}(\boldsymbol{\beta}^i)) > 0$. To see it, recall that the distance function written in terms of the cdf of the error term is

$$
\begin{aligned}
& F_{\epsilon^i|x}(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i) - (1 - F_{\epsilon^i|x}(z^i - 2\mathbf{x}_p^i\boldsymbol{\beta}^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i)) \\
= \;& F_{\epsilon^i|x}(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i) - (1 - F_{\epsilon^i|x}(z^i - \mathbf{x}_p^i\boldsymbol{\beta}_0^i - 2\mathbf{x}_p^i(\boldsymbol{\beta}^i - \boldsymbol{\beta}_0^i)).
\end{aligned}
$$

If the claim is not true, so that we have $\Pr(d^i(\boldsymbol{\beta}^i; Z^i, X) \neq 0 | (Z^i, X) \in \mathscr{S}(\boldsymbol{\beta}^i)) = 0$, then this implies that $F_{\epsilon^i|x}(-z^i + \mathbf{x}_p^i\boldsymbol{\beta}_0^i) = 1 - F_{\epsilon^i|x}(z^i - \mathbf{x}_p^i\boldsymbol{\beta}_0^i - 2\mathbf{x}_p^i(\boldsymbol{\beta}^i - \boldsymbol{\beta}_0^i))$ even though $\mathbf{x}_p^i(\boldsymbol{\beta}^i - \boldsymbol{\beta}_0^i) \neq 0$, which is a contradiction under Assumption 3.3.6.

C.2 *Proof of Theorem 3.4.2*

proof. The proof follows a similar argument as the proof of Theorem 3.1 in Lewbel et al. (2020). We need to replace their utility function with our expected payoff function, and replace their identifying restriction based on pdf with our identifying restriction based on cdf. $Q^i(\boldsymbol{\beta}^i) = \frac{1}{2}\mathbb{E}[(\tau(Z^i, X)d^i(\boldsymbol{\beta}^i; Z^i, X))]^2 \geq 0$ by construction. To show $Q^i(\boldsymbol{\beta}_0^i) = 0$, notice that when $\boldsymbol{\beta}^i = \boldsymbol{\beta}_0^i$, if $(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i,X)}$, then by Assumption 3.4.1, $\tau^i(z^i, x) > 0$, and for such (z^i, x), it holds that $d^i(\boldsymbol{\beta}^i; z^i, x) = 0$. If $(z^i, x) \notin \mathcal{S}^{i,NT}_{(Z^i,X)}$, then by Assumption 3.4.1, $\tau^i(z^i, x) = 0$. To show $Q^i(\boldsymbol{\beta}^i) > 0$ for $\boldsymbol{\beta}^i \neq \boldsymbol{\beta}_0^i$. For such $\boldsymbol{\beta}^i$, by the proof of Theorem 3.3.2, we can always find $\mathscr{S}(\boldsymbol{\beta}^i) \equiv \{(z^i, x) : z^i \in \mathcal{S}_{Z^i|x},\ x \in \mathscr{X}(\boldsymbol{\beta}^i)\}$ such that $\Pr((Z^i, X) \in \mathscr{S}(\boldsymbol{\beta}^i)) > 0$. By Assumption 3.4.1, we can always find the set $\mathscr{S}(\boldsymbol{\beta}^i \bigcap \mathcal{S}^{i,NT}_{(Z^i,X)}$ which has strictly positive probability. On this set $\tau^i(z^i, x) > 0$ and $d^i(\boldsymbol{\beta}^i; z^i, x) \neq 0$, which implies $Q^i(\boldsymbol{\beta}^i) > 0$.

C.3 *Proof of Theorem 3.5.1*

The following lemma is Lemma 8.10 from Newey and McFadden (1994) (or Lemma S.A.3 in Lewbel et al. (2020)) used to control the rate of convergence of the first step nonparametric estimates:

Lemma C.3.1. *Under Assumptions 3.3.1, 3.5.1-3.5.3, for $t = 0, 1$ and $i = 1, 2$, it holds that*

$$(i). \quad \sup_{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}} |\widehat{f}^{i,(t)}(z^i, x) - f^{i,(t)}(z^i, x)| = O_p(\sqrt{\frac{logN}{Nh^{2t+L}}} + h^s) = o_p(N^{-1/4}),$$

$$(ii). \quad \sup_{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}} |\widehat{g}^{i,(t)}(z^i, x) - g^{i,(t)}(z^i, x)| = O_p(\sqrt{\frac{logN}{Nh^{2t+L}}} + h^s) = o_p(N^{-1/4}),$$

$$(iii). \quad \sup_{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}} |\widehat{P}^{i,(t)}(z^i, x) - P^{i,(t)}(z^i, x)| = O_p(\sqrt{\frac{logN}{Nh^{2t+L}}} + h^s) = o_p(N^{-1/4}),$$

$$(iv). \sup_{\substack{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}\\ p\in[0,1]}} |\widehat{f}^{i,(t)}(-z^i+2[x,p]\boldsymbol{\beta}^i_0,x)-f^{i,(t)}(-z^i+2[x,p]\boldsymbol{\beta}^i_0,x)| = O_p(\sqrt{\frac{logN}{Nh^{2t+L}}}+h^s) = o_p(N^{-1/4}),$$

$$(v). \sup_{\substack{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}\\ p\in[0,1]}} |\widehat{g}^{i,(t)}(-z^i+2[x,p]\boldsymbol{\beta}^i_0,x)-g^{i,(t)}(-z^i+2[x,p]\boldsymbol{\beta}^i_0,x)| = O_p(\sqrt{\frac{logN}{Nh^{2t+L}}}+h^s) = o_p(N^{-1/4}),$$

$$(vi). \sup_{\substack{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}\\ p\in[0,1]}} |\widehat{P}^{i,(t)}(-z^i+2[x,p]\boldsymbol{\beta}^i_0,x)-P^{i,(t)}(-z^i+2[x,p]\boldsymbol{\beta}^i_0,x)| = O_p(\sqrt{\frac{logN}{Nh^{2t+L}}}+h^s) = o_p(N^{-1/4}),$$

where superscript (t) denotes the $t-th$ order derivative with respect to z^i.

Remark C.3.1. *Notice that Lemma C.3.1 implies that under Assumptions 3.5.1-3.5.3, both $\widehat{d}^i_{-n}(\boldsymbol{\beta}^i_0; Z^i_n, X_n)$ and $\nabla_{\boldsymbol{\beta}^i}\widehat{d}^i_{-n}(\boldsymbol{\beta}^i_0; Z^i_n, X_n)$ are uniformly consistent on $\mathcal{S}^{i,NT}_{(Z^i,X)}\times[0,1]$ at faster than $N^{-1/4}$ rate.*

Proof of Theorem 3.5.1: Define $\widetilde{Q}^i_N(\boldsymbol{\beta}^i)$ as the sample objective function in which true belief instead of estimated belief is used. Applying a similar argument to the one used in the proof of Theorem 3.2 of Lewbel et al. (2020) gives $\sup_{\boldsymbol{\beta}^i\in\mathcal{B}^i}|\widetilde{Q}^i_N(\boldsymbol{\beta}^i) - Q^i(\boldsymbol{\beta}^i)| = o_p(1)$. Thus, it suffices to show $\sup_{\boldsymbol{\beta}^i\in\mathcal{B}^i}|Q^i_N(\boldsymbol{\beta}^i) - \widetilde{Q}^i_N(\boldsymbol{\beta}^i)| = o_p(1)$. To show this, define $\overline{Q}^i_N(\boldsymbol{\beta}^i)$ as the sample objective function in which estimated distance $\widehat{d}^i_{-n}(\boldsymbol{\beta}^i; Z^i_n, X_n)$ is replaced by true distance $d^i(\boldsymbol{\beta}^i; Z^i_n, X_n)$, and the estimated belief is replaced by true belief. An argument similar to the one showing the asymptotic negligibility of the first term on the right hand side in equation (S.A.3) of Lewbel et al. (2020) implies $\sup_{\boldsymbol{\beta}^i\in\mathcal{B}^i}|\widetilde{Q}^i_N(\boldsymbol{\beta}^i) - \overline{Q}^i_N(\boldsymbol{\beta}^i)| = o_p(1)$. Thus it suffices to show that $\sup_{\boldsymbol{\beta}^i\in\mathcal{B}^i}|Q^i_N(\boldsymbol{\beta}^i) - \overline{Q}^i_N(\boldsymbol{\beta}^i)| = o_p(1)$, which could be established as follows:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |\widetilde{Q}^i_N(\boldsymbol{\beta}^i) - \overline{Q}^i_N(\boldsymbol{\beta}^i)| \\
= \quad & \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |\frac{1}{2N} \sum_{n=1}^{N} [\tau^i(Z^i_n, X_n)]^2 [\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; Z^i_n, X_n) + d^i(\boldsymbol{\beta}^i; Z^i_n, X_n)][\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; Z^i_n, X_n) - d^i(\boldsymbol{\beta}^i; Z^i_n, X_n)]| \\
\leq \quad & C \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} \sup_{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)}} |\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; z^i, x) - d^i(\boldsymbol{\beta}^i; z^i, x)|.
\end{aligned}
$$

Let $\widehat{d^i}(\boldsymbol{\beta}^i; Z^i_n, X_n)$ be the true distance function evaluated at estimated belief instead of true belief, so that $\widehat{d^i}(\boldsymbol{\beta}^i; Z^i_n, X_n)$ and $\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; Z^i_n, X_n)$ are evaluated at the same values. Thus

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} \sup_{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)}} |\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; z^i, x) - d^i(\boldsymbol{\beta}^i; z^i, x)| \\
\leq \quad & \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} \sup_{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)}} |\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; z^i, x) - \widehat{d^i}(\boldsymbol{\beta}^i; z^i, x)| \\
+ \quad & \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} \sup_{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)}} |\widehat{d^i}(\boldsymbol{\beta}^i; z^i, x) - d^i(\boldsymbol{\beta}^i; z^i, x)| \\
\leq \quad & \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} \sup_{\substack{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)} \\ p \in [0,1]}} |\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; z^i, x) - \widehat{d^i}(\boldsymbol{\beta}^i; z^i, x)| \\
+ \quad & \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} \sup_{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)}} |\widehat{d^i}(\boldsymbol{\beta}^i; z^i, x) - d^i(\boldsymbol{\beta}^i; z^i, x)| \\
= \quad & o_p(1),
\end{aligned}
$$

where the first term is $o_p(1)$ as (iii) in Lemma C.3.1 implies that $\sup_{\substack{(z^i, x) \in \mathcal{S}^{i,NT}_{(Z^i, X)} \\ p \in [0,1]}} |\widehat{d^i}_{-n}(\boldsymbol{\beta}^i; z^i, x) - \widehat{d^i}(\boldsymbol{\beta}^i; z^i, x)| = o_p(1)$, and the second term is $o_p(1)$ by the smoothness of the true distance and consistency of estimated belief. Thus it holds that

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |Q_N^i(\boldsymbol{\beta}^i) - Q^i(\boldsymbol{\beta}^i)| \\
= \;&\sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |Q_N^i(\boldsymbol{\beta}^i) - \overline{Q}^i(\boldsymbol{\beta}^i) + \overline{Q}^i(\boldsymbol{\beta}^i) - \widetilde{Q}^i(\boldsymbol{\beta}^i) + \widetilde{Q}^i(\boldsymbol{\beta}^i) - Q^i(\boldsymbol{\beta}^i)| \\
\leq \;&\sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |Q_N^i(\boldsymbol{\beta}^i) - \overline{Q}^i(\boldsymbol{\beta}^i)| + \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |\overline{Q}^i(\boldsymbol{\beta}^i) - \widetilde{Q}^i(\boldsymbol{\beta}^i)| + \sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |\widetilde{Q}^i(\boldsymbol{\beta}^i) - Q^i(\boldsymbol{\beta}^i)| = o_p(1)
\end{aligned}
$$

□

C.4 *Proof of Theorem 3.5.2*

The proof builds on a series of lemmas that establishes consistency of the sample hessian matrix and gives asymptotic linear representation of the sample score. The following lemma gives consistency of the sample hessian matrix.

Lemma C.4.1. *Under Assumptions 3.2.1-3.5.3, it holds that* $H_N^i(\widetilde{\boldsymbol{\beta}}^i) \xrightarrow{p} H^i$.

Proof. Let $H^i(\boldsymbol{\beta}^i)$ denote the true hessian evaluated at some generic $\boldsymbol{\beta}^i$. Let $\widetilde{H}_N^i(\boldsymbol{\beta}^i)$ denote the sample hessian with estimated belief replaced by true belief. An argument similar to the proof of Lemma S.A.4 in Lewbel et al. (2020) implies $\sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |\widetilde{H}_N^i(\boldsymbol{\beta}^i) - H^i(\boldsymbol{\beta}^i)| = o_p(1)$. Then an argument similar to the one used in the proof of Theorem 3.5.1 establishes $\sup_{\boldsymbol{\beta}^i \in \mathcal{B}^i} |H_N^i(\boldsymbol{\beta}^i) - H^i(\boldsymbol{\beta}^i)| = o_p(1)$, which together with the continuity of $H^i(\boldsymbol{\beta}^i)$ at $\boldsymbol{\beta}_0^i$ implies the claim. □

Recall that by definition,

$$
q_N^i(\boldsymbol{\beta}_0^i) = \frac{1}{N} \sum_{n=1}^{N} [\tau^i(Z_n^i, X_n)]^2 \widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) \nabla_{\boldsymbol{\beta}^i} \widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n).
$$

We first do the following decomposition as in Lewbel et al. (2020), and show the second term in this decomposition will be $\sqrt{N}$ negligible uniformly over all evaluation points $(z^i, x) \in \mathcal{S}_{(Z^i,X)}^{i,NT}$ (note that this term is also uniformly negligible in belief p, as we have seen that any belief results in an evaluation point that belongs to $\mathcal{S}_{(Z^i,X)}^{i,NT}$).

$$\begin{aligned} q_N^i(\boldsymbol{\beta}_0^i) &= \frac{1}{N}\sum_{n=1}^N [\tau^i(Z_n^i, X_n)]^2 \widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)\nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) \\ &+ \frac{1}{N}\sum_{n=1}^N [\tau^i(Z_n^i, X_n)]^2 \widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)\left(\nabla_{\boldsymbol{\beta}^i}\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)\right) \\ &\equiv q_{N,1}^i(\boldsymbol{\beta}_0^i) + q_{N,2}^i(\boldsymbol{\beta}_0^i) \end{aligned}$$

Lemma C.4.2. *Under Assumptions 3.3.1, 3.3.5, 3.4.1-3.5.3, it holds that* $q_{N,2}^i(\boldsymbol{\beta}_0^i) = o_p(N^{-1/2})$.

Proof.

$$\begin{aligned} & q_{N,2}^i(\boldsymbol{\beta}_0^i) \\ =\ & \frac{1}{N}\sum_{n=1}^N [\tau^i(Z_n^i, X_n)]^2 \widehat{d}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)\left(\nabla_{\boldsymbol{\beta}^i}\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)\right) \\ +\ & \frac{1}{N}\sum_{n=1}^N [\tau^i(Z_n^i, X_n)]^2 (\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \widehat{d}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n))\left(\nabla_{\boldsymbol{\beta}^i}\widehat{d}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)\right). \end{aligned}$$

Note that under the stated assumptions, it holds that $\widehat{d}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)(\nabla_{\boldsymbol{\beta}^i}\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)) = o_p(N^{-1/2})$ as the first term in the product is $o_p(N^{-1/4})$ (by uniform consistency of estimated belief, smoothness of the true distance function, and the fact that true distance function equals zero at true belief), and the second term in the product is also $o_p(N^{-1/4})$ (by uniform consistency of $\nabla_{\boldsymbol{\beta}^i}\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)$). $(\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \widehat{d}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n))(\nabla_{\boldsymbol{\beta}^i}\widehat{d}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n) - \nabla_{\boldsymbol{\beta}^i} d^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)) = o_p(N^{-1/2})$ as the first term in this product is $o_p(N^{-1/4})$ by uniform consistency of estimated belief and by smoothness of the true distance function, and the second term in the product is also $o_p(N^{-1/4})$ by uniform consistency of $\nabla_{\boldsymbol{\beta}^i}\widehat{d}_{-n}^i(\boldsymbol{\beta}_0^i; Z_n^i, X_n)$. Thus $q_{N,2}^i(\boldsymbol{\beta}_0^i) = o_p(N^{-1/2})$ holds.

□

Define

$$\begin{aligned}
q^i_{N,1}(\boldsymbol{\beta}^i_0) &= \frac{1}{N}\sum_{n=1}^{N}[\tau^i(Z^i_n,X_n)]\nabla_{\boldsymbol{\beta}^i}d^i(\boldsymbol{\beta}^i_0;Z^i_n,X_n)\widehat{d}^i_{-n}(\boldsymbol{\beta}^i_0;Z^i_n,X_n)\\
&\equiv \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z^i_n,X_n,-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\widehat{d}^i_{-n}(\boldsymbol{\beta}^i_0;Z^i_n,X_n)\\
&= \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z^i_n,X_n,-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\left(\widehat{\varphi}^i_{o,-n}(Z_n,X_n)-\widehat{\varphi}^i_{cs,-n}(Z_n,X_n,\boldsymbol{\beta}^i_0)\right)\\
&\equiv q^i_{N,1,o}(\boldsymbol{\beta}^i_0)-q^i_{N,1,cs}(\boldsymbol{\beta}^i_0).
\end{aligned}$$

Notice that the estimated equilibrium belief enters $q^i_{N,1,cs}(\boldsymbol{\beta}^i_0)$, but not $q^i_{N,1,o}(\boldsymbol{\beta}^i_0)$. For notational convenience, we define some notational shortcuts: $\varphi^i_o \equiv P^i(Z^i_n,X_n)$, $f^i_o \equiv f^i(Z^i_n,X_n)$, $g^i_o \equiv P^i(Z^i_n,X_n)f^i(Z^i_n,X_n)$, $\xi^i \equiv \xi^i(Z^i_n,X_n,-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)$, $\varphi^i_{cs} \equiv 1-P^i(-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0,X_n)$, $f^i_{cs} \equiv f^i(-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0,X_n)$, $g^i_{cs} \equiv P^i(-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0,X_n)f^i(-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0,X_n)$. Let $\widehat{f}^i_{o,-n}$, $\widehat{g}^i_{o,-n}$ denote the corresponding leave-one-out estimators for f^i_o and g^i_o. Let $\widehat{f}^i_{cs,-n}$, $\widehat{g}^i_{cs,-n}$ denote the corresponding leave-one-out estimators for f^i_{cs} and g^i_{cs} (evaluated at estimated belief). In the following proof, when there is no confusion, we will use these notations and their corresponding shortcuts versions interchangeably. Let $\gamma^i_o=[g^i_o,f^i_o]$ and $\widehat{\gamma}^i_{o,-n}=[\widehat{g}^i_{o,-n},\widehat{f}^i_{o,-n}]$. Let $\gamma^i_{cs}=[g^i_{cs},f^i_{cs}]$ and $\widehat{\gamma}^i_{cs,-n}=[\widehat{g}^i_{cs,-n},\widehat{f}^i_{cs,-n}]$. In the following proof, the asymptotic linearity of $q^i_{N,1,o}(\boldsymbol{\beta}^i_0)$ will be shown directly. The asymptotic linearity of $q^i_{N,1,cs}(\boldsymbol{\beta}^i_0)$ will be shown based on a linear representation that holds uniformly for any belief and a Taylor expansion with respect to the estimated belief. The derivation of asymptotic linear representation of $q_{N,1,o}(\boldsymbol{\beta}^i_{10})$ uses the following lemma for the Hoeffding decomposition of a U-statistic, which is Lemma D.1 in Chen et al. (2016) with $J=2$.

Lemma C.4.3. *Let $U_N \equiv \binom{N}{2}^{-1}\sum_{\mathcal{P}_{m,n}} p_N(w_m,w_n)$, where $p_N(w_m,w_n)$ is the kernel. Let*

$$r_{1N}(W_m)=\mathbb{E}[p_N(W_m,W_n)|W_m],$$

$$r_{2N}(W_n)=\mathbb{E}[p_N(W_m,W_n)|W_n],$$

$$\theta_N=\mathbb{E}[p_N(W_m,W_n)]=\mathbb{E}[r_N(W_n)],$$

$$\textit{and } PU_N = \theta_N + \frac{1}{N}\sum_{n=1}^{N}[r_{1N}(W_n) - \theta_N] + \frac{1}{N}\sum_{n=1}^{N}[r_{2N}(W_n) - \theta_N].$$

If $\mathbb{E}[||p_N(W_m, W_n)||^2] = o(N)$, *then* $U_N = PU_N + o_p(N^{-1/2})$.

Lemma C.4.4 (Asymptotic Linear Representation of $q^i_{N,1,o}(\boldsymbol{\beta}^i_0)$)**.** *Under Assumptions 3.3.1, 3.3.5, 3.4.1, and 3.5.1-3.5.3, it holds that* $q^i_{N,1,o}(\boldsymbol{\beta}^i_0) = \frac{1}{N}\sum_{n=1}^{N} r^i_o(W_n) + o_p(N^{-1/2})$, *where* $r^i_o(W_n) = \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)P^i(Z^i_n, X_n) + \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)(D^i_n - P^i(Z^i_n, X_n))$.[1]

Proof. Following Newey and McFadden (1994), Hu (2008), linearization of $\widehat{\varphi}^i_{o,-n}(Z^i_n, X_n)$ with respect to $\widehat{\gamma}^i_{o,-n}$ gives

$$\begin{aligned}\widehat{\varphi}^i_{o,-n}(Z^i_n, X_n) &= \varphi^i_o(Z^i_n, X_n) \\ &+ \sum_{l=1}^{2} J^i_{l,o}\left(\widehat{\gamma}^i_{o,-n}[l] - \gamma^i_o[l]\right) + \sum_{l=1}^{2}\sum_{m=1}^{2}\frac{\widetilde{H}^i_{l,m,o}}{2}\left(\widehat{\gamma}^i_{o,-n}[l] - \gamma^i_o[l]\right)\left(\widehat{\gamma}^i_{o,-n}[m] - \gamma^i_o[m]\right),\end{aligned}$$

where $J^i_{1,o} = \frac{1}{f^i_o}$, $J^i_{2,o} = -\frac{g^i_o}{(f^i_o)^2}$, and the hessian is evaluated at $\widetilde{\gamma} = \gamma_o(1-\widetilde{t}) + \widetilde{t}\widehat{\gamma}_{o,-n}$ for some $\widetilde{t} \in (0,1)$.[2] The quadratic remainder satisfies

$$\begin{aligned}&|\sum_l\sum_m \frac{\widetilde{H}^i_{l,m,o}}{2}\left(\widehat{\gamma}^i_{o,-n}[l] - \gamma^i_o[l]\right)\left(\widehat{\gamma}^i_{o,-n}[m] - \gamma^i_o[m]\right)| \\ \leq\ & C \sup_{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}} ||\widehat{\gamma}^i_{o,-n} - \gamma^i_o||^2 = o_p(N^{-1/2}),\end{aligned}$$

where the last equality holds by Lemma C.3.1. Rearranging the leading term gives:

$$\begin{aligned}&\widehat{\varphi}^i_{o,-n}(Z^i_n, X_n) \\ =\ & P^i(Z^i_n, X_n) + \frac{1}{f^i(Z^i_n, X_n)}\frac{1}{(N-1)}\sum_{m=1,m\neq n}^{N}[D^i_m - P^i(Z^i_n, X_n)] \\ \times\ & k_h(Z^i_m - Z^i_n)\boldsymbol{K}_h(X_{m,c} - X_{n,c})\mathbb{1}(X_{m,d} = X_{n,d}) + o_p(N^{-1/2}). \qquad \text{(C.1)}\end{aligned}$$

[1] Here we do not cancel out the first term $\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)P^i(Z^i_n, X_n)$ in $r^i_o(W_n)$, and use it later to cancel out the first term in $r^i_{cs}(W_n)$.

[2] See Hu (2008) for an illustration of functional derivative defined using pathwise derivative as well as its relationship with ordinary derivative.

Multiplying $\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)$ on both sides and averaging gives $q_{N,1,o}^i(\boldsymbol{\beta}_0^i) = \Phi_{o,N}^i + o_p(N^{-1/2})$, where $\Phi_{o,N}^i$ is defined as

$$\frac{1}{N}\sum_{n=1}^{N}\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)P^i(Z_n^i, X_n) + \frac{1}{N}\sum_{n=1}^{N}[\frac{\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)}{f^i(Z_n^i, X_n)}$$
$$\times\frac{1}{(N-1)}\sum_{m=1, m\neq n}^{N}[D_m^i - P^i(Z_n^i, X_n)]k_h(Z_m^i - Z_n^i)\boldsymbol{K}_h(X_{m,c} - X_{n,c})\mathbb{1}(X_{m,d} = X_{n,d})].$$

Note that the second term in the above expression could be written as a U-statistic of order 2, with (non-symmetrized) kernel defined as follows:

$$\begin{aligned}
& \psi_{N,o,1}^i(W_n^i, W_m^i) \\
= \; & \frac{\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)}{f^i(Z_n^i, X_n)}[D_m^i - P^i(Z_n^i, X_n)]k_h(Z_m^i - Z_n^i)\boldsymbol{K}_h(X_{m,c} - X_{n,c})\mathbb{1}(X_{m,d} = X_{n,d}).
\end{aligned}$$

Its second moment is calculated as:

$$\begin{aligned}
& \mathbb{E}[||\psi_{N,o,1}^i(W_n^i, W_m^i)||^2] \\
= \; & \sum_{x_{n,d}}\sum_{x_{m,d}}\int_{x_{n,c}^i}\int_{z_n^i}\int_{x_{m,c}^i}\int_{z_m^i}||\xi^i(z_n^i, x_{n,c}, x_{n,d}, -z_n^i + 2\mathbf{x}_{pn}^i\boldsymbol{\beta}_0^i)||^2\frac{1}{f^i(z_n^i, x_{n,c}, x_{n,d})^2}[P^i(z_m^i, x_{m,c}, x_{m,d})^2 \\
+ \; & P^i(z_n^i, x_{n,c}, x_{n,d})^2 - 2P^i(z_m^i, x_{m,c}, x_{m,d})P^i(z_n^i, x_{n,c}, x_{n,d})]k_h^2(z_m^i - z_n^i)\boldsymbol{K}_h^2(x_{m,c} - x_{n,c})\mathbb{1}(x_{m,d} = x_{n,d}) \\
\times \; & f^i(z_m^i, x_{m.c}, x_{m,d})dz_m^i dx_{m,c} f^i(z_n^i, x_{n.c}, x_{n,d})dz_n^i dx_{n,c} \\
= \; & h^{-L_c-1}\sum_{x_{n,d}}\int_{x_{n,c}^i}\int_{z_n^i}\int_{u_x}\int_{u_z}||\xi^i(z_n^i, x_{n,c}, x_{n,d}, -z_n^i + 2\mathbf{x}_{pn}^i\boldsymbol{\beta}_0^i)||^2\frac{1}{f^i(z_n^i, x_{n,c}, x_{n,d})^2} \\
\times \; & [P^i(z_n^i + u_z h, x_{n,c} + u_x h, x_{n,d})^2 + P^i(z_n^i, x_{n,c}, x_{n,d})^2 \\
- \; & 2P^i(z_n^i + u_z h, x_{n,c} + u_x h, x_{n,d})P^i(z_n^i, x_{n,c}, x_{n,d})]k^2(u_z)\boldsymbol{K}^2(u_x) \\
\times \; & f^i(z_n^i + u_z h, x_{n,c} + u_x h, x_{n,d})du_z du_x f^i(z_n^i, x_{n.c}, x_{n,d})dz_n^i dx_{n,c} \\
= \; & O(h^{-L_c-1}) = O(NN^{-1}h^{-L_c-1}) = o(N), \qquad \text{(C.2)}
\end{aligned}$$

where the last equality holds by Assumption 3.5.3. Thus Lemma C.4.3 applies with the corresponding projections calculated as follows:

$$
\begin{aligned}
& \mathbb{E}[\psi^i_{N,o,1}(W^i_n, W^i_m)|W_n] \\
= \; & \sum_{x_{m,d}} \int_{z^i_m} \int_{x_{m,c}} \frac{\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(Z^i_n, X_n)} [P^i(z^i_m, x_{m,c}, x_{m,d}) - P^i(Z^i_n, X_n)] k_h(z^i_m - Z^i_n) \boldsymbol{K}_h(x_{m,c} - X_{n,c}) \\
\times \; & \mathbb{1}(x_{m,d} = X_{n,d}) f^i(z^i_m, x_{m,c}, x_{m,d}) dx_{m,c} dz^i_m \\
= \; & \int_{u_z} \int_{u_x} \frac{\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(Z^i_n, X_n)} [P^i(Z^i_n + u_z h, X_{n,c} + u_x h, X_{n,d}) - P^i(Z^i_n, X_n)] k(u_z) \boldsymbol{K}(u_x) \\
\times \; & f^i(Z^i_n + u_z h, X_{n,c} + u_x h, X_{n,d}) du_x du_z \\
= \; & O_p(h^s) = o_p(N^{-1/2})
\end{aligned}
$$

$$
\begin{aligned}
& \mathbb{E}[\psi^i_{N,o,1}(W^i_n, W^i_m)|W_m] \\
= \; & \sum_{x_{n,d}} \int_{z^i_n} \int_{x_{n,c}} \frac{\xi^i(z^i_n, x_{n,c}, x_{n,d}, -z^i_n + 2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(z^i_n, x_{n,c}, x_{n,d})} [D^i_m - P^i(z^i_n, x_{n,c}, x_{n,d})] k_h(Z^i_m - z^i_n) \boldsymbol{K}_h(X_{m,c} - x_{n,c}) \\
\times \; & \mathbb{1}(X_{m,d} = x_{n,d}) f^i(z^i_n, x_{n,c}, x_{n,d}) dx_{n,c} dz^i_n \\
= \; & \int_{z^i_n} \int_{x_{n,c}} \frac{\xi^i(Z^i_m + hu_z, X_{m,c} + hu_x, X_{m,d}, -Z^i_m - hu_z + 2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0 + B^i(X_m) hu_x + O(h^2))}{f^i(Z^i_m + hu_z, X_{m,c} + hu_x, X_{m,d})} \\
\times \; & [D^i_m - P^i(Z^i_m + hu_z, X_{m,c} + hu_x, x_{n,d})] k(u_z) \boldsymbol{K}(u_x) f^i(Z^i_m + u_z h, X_{m,c} + hu_x, X_{m,d}) du_x du_z \\
= \; & \xi^i(Z^i_m, X_m, -Z^i_m + 2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0)[D^i_m - P^i(Z^i_m, X_m)] + O_p(h^s) \\
= \; & \xi^i(Z^i_m, X_m, -Z^i_m + 2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0)[D^i_m - P^i(Z^i_m, X_m)] + o_p(N^{-1/2}),
\end{aligned}
$$

where $B^i(X_m) = \frac{\partial P^{-i}(X_m)}{\partial x_{m,c}}$. Both equations imply that $\mathbb{E}[\psi^i_{N,o,1}(W^i_n, W^i_m)] = o(N^{-1/2})$ by the law of iterated expectations, and Hoeffding decomposition gives:

$$
\begin{aligned}
& \Phi^i_{o,N} \\
= \; & \frac{1}{N} \sum_{n=1}^{N} \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0) P^i(Z^i_n, X_n) + \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)(D^i_n - P^i(Z^i_n, X_n)) \\
& + o_p(N^{-1/2}) \\
\equiv \; & \frac{1}{N} \sum_{n=1}^{N} r^i_o(W_n) + o_p(N^{-1/2}).
\end{aligned}
$$

We keep the term $\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)P^i(Z_n^i, X_n)$ in $r_o^i(W_n)$ to cancel it out with the first term in $r_{cs}^i(W_n)$ later.

$\square$

The next lemma gives asymptotic linear representation of $q_{N,1,cs}^i(\boldsymbol{\beta}_0^i)$.

Lemma C.4.5 (asymptotic linear representation of $q_{N,1,cs}^i(\boldsymbol{\beta}_0^i)$)**.** *Under Assumptions 3.2.1-3.5.3, it holds that* $q_{N,1,cs}^i(\boldsymbol{\beta}_0^i) = \frac{1}{N}\sum_{n=1}^N r_{cs}^i(W_n) + o_p(N^{-1/2})$*, where*

$$\begin{aligned}
r_{cs}^i(W_n) &= \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i) - \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) \\
&- 2\beta_{20}^i C^i(X_n)(D_n^{-i} - P^{-i}(X_n)) \\
&- \frac{\xi_i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n, Z_n^i)f(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)} \\
&\times [D_n^i - P^i(Z_n^i, X_n)].
\end{aligned}$$

Proof. As $q_{N,1,cs}^i(\boldsymbol{\beta}_0^i) = \frac{1}{N}\sum_{n=1}^N \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i) - \frac{1}{N}\sum_{n=1}^N \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)\widehat{P}_{-n}^i(-Z_n^i + 2\widehat{\mathbf{X}}_{pn}^i\boldsymbol{\beta}_0^i, X_n)$, we can focus on deriving a linear representation for the second term. For $\widehat{P}_{-n}^i(-Z_n^i + 2\widehat{\mathbf{X}}_{pn}^i\boldsymbol{\beta}_0^i, X_n)$, the smoothness of kernel and uniform consistency of estimated belief allow the following expansion:

$$\begin{aligned}
&\widehat{P}_{-n}^i(-Z_n^i + 2\widehat{\mathbf{X}}_{pn}^i\boldsymbol{\beta}_0^i, X_n) \\
=\ &\widehat{P}_{-n}^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) + \nabla_p[\widehat{P}_{-n}^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)](\widehat{P}^{-i}(X_n) - P^{-i}(X_n)) + o_p(N^{-1/2}) \\
=\ &\widehat{P}_{-n}^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) + \nabla_p[P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)](\widehat{P}^{-i}(X_n) - P^{-i}(X_n)) + \\
+\ &\nabla_p[\widehat{P}_{-n}^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) - P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)](\widehat{P}^{-i}(X_n) - P^{-i}(X_n)) + o_p(N^{-1/2}) \\
=\ &\widehat{P}_{-n}^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) + \nabla_p[P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)](\widehat{P}^{-i}(X_n) - P^{-i}(X_n)) + o_p(N^{-1/2}),
\end{aligned}$$

where the last equality holds by the fact that under Lemma C.3.1, the following inequality holds

$$
\begin{aligned}
& |\nabla_p[\widehat{P}^i_{-n}(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n) - P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)](\widehat{P}^{-i}(X_n) - P^{-i}(X_n))| \\
\leq \;& 2|\beta_{20}| \sup_{\substack{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)} \\ p\in[0,1]}} |\frac{\partial \widehat{P}^i(-z^i + 2[x,p]\boldsymbol{\beta}^i_0, x)}{\partial z^i} - \frac{\partial P^i(-z^i + 2[x,p]\boldsymbol{\beta}^i_0, x)}{\partial z^i}| \\
\times \;& \sup_{(z^i,x)\in\mathcal{S}^{i,NT}_{(Z^i,X)}} |\widehat{P}^{-i}(x) - P^{-i}(x)| \\
= \;& o_p(N^{-1/2}).
\end{aligned}
$$

Define $\widetilde{\Phi}^i_{cs,N} \equiv \frac{1}{N}\sum_{n=1}^N \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\widehat{P}^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)$, and $\Delta^i_{cs,N} \equiv \frac{1}{N}\sum_{n=1}^N \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\nabla_p[P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)](\widehat{P}^{-i}(X_n) - P^{-i}(X_n))$. We will find the linear representation of $\widetilde{\Phi}^i_{cs,N}$ first. By the same logic as in the proof of Lemma C.4.4, we could first linearize $\widetilde{\Phi}^i_{cs,N}$ in terms of unknown functions:

$$
\begin{aligned}
\widetilde{\Phi}^i_{cs,N} \;=\;& \frac{1}{N}\sum_{n=1}^N \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n) \\
+ \;& \sum_{l=1}^2 \frac{1}{N}\sum_{n=1}^N \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\widetilde{J}^i_{l,cs}\widetilde{\gamma}^i_{cs,-n}[l] \\
- \;& \sum_{l=1}^2 \frac{1}{N}\sum_{n=1}^N \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\widetilde{J}^i_{l,cs}\gamma^i_{cs}[l] + o_p(N^{-1/2}) \\
= \;& \frac{1}{N}\sum_{n=1}^N \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n) \\
+ \;& \frac{1}{N(N-1)}\sum_{n=1}^N \frac{\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)} \sum_{m=1, m\neq n}^N [(D^i_m - P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)) \\
\times \;& K_h(Z^i_m + Z^i_n - 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\boldsymbol{K}_h(X_{m,c} - X_{n,c})\mathbb{1}(X_{m,d} = X_{n,d})] + o_p(N^{-1/2}),
\end{aligned}
$$

where the second term is a U-statistic with kernel $\psi^i_{N,cs}(W_m, W_n)$ defined as:

$$
\begin{aligned}
& \frac{\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}[(D^i_m - P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n))K_h(Z^i_m + Z^i_n - 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\boldsymbol{K}_h(X_{m,c} - X_{n,c}) \\
\times \;& \mathbb{1}(X_{m,d} = X_{n,d})].
\end{aligned}
$$

The projection $\mathbb{E}(\psi^i_{N,cs}(W_m,W_n)|W_m)$ is given by:

$$\sum_{x_{n,d}}\int_{z^i_n}\int_{x_{n,c}}\frac{\xi^i(z^i_n,x_{n,c},x_{n,d},-z^i_n+2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(-z^i_n+2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0,x_{n,c},x_{n,d})}[(D^i_m-P^i(-z^i_n+2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0,x_{n,c},x_{n,d}))K_h(Z^i_m+z^i_n-2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0)$$
$$\times\boldsymbol{K}_h(x_{n,c}-X_{m,c})\times\mathbb{1}(X_{m,d}=x_{n,d})]f(z^i_n,x_{n,c},x_{n,d})dz^i_ndx_{n,c}.$$

Do the following change of variable: $Z^i_m+z^i_n-2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0=hu_z$ and $x_{n,c}-X_{m,c}=hu_x$, and plug in back we get

$$\begin{aligned}
&\int_{z^i_n}\int_{z_{n,c}}\frac{\xi^i(z^i_n,x_{n,c},X_{m,d},Z^i_m-hu_z)}{f^i(-z^i_n+2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0,x_{n,c},X_{m,d})}[(D^i_m-P^i(-z^i_n+2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0,x_{n,c},X_{m,d}))K_h(Z^i_m+z^i_n-2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0)\\
\times\;&\boldsymbol{K}_h(x_{n,c}-X_{m,c})\times]f(z^i_n,x_{n,c},X_{m,d})dz^i_ndx_{n,c}\\
=\;&\int_{u_z}\int_{u_x}\frac{\xi^i(-Z^i_m+2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0+u_zh+B^i(X_m)u_xh+O(h^2),X_{m,c}+u_xh,X_{m,d},Z^i_m-hu_z)}{f^i(Z^i_m-u_zh,X_{m,c}+u_xh,X_{m,d})}[(D^i_m\\
-\;&P^i(Z^i_m-u_zh,X_{m,c}+u_xh,X_{m,d}))k(u_z)\\
\times\;&\boldsymbol{k}(u_x)]f(-Z^i_m+2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0+u_zh+B^i(X_m)u_xh+O(h^2),X_{m,c}+u_xh,X_{m,d})du_zdu_x\\
=\;&\frac{\xi^i(-Z^i_m+2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0,X_{m,c},X_{m,d},Z^i_m)f(-Z^i_m+2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0,X_{m,c},X_{m,d})}{f^i(Z^i_m,X_{m,c},X_{m,d})}[(D^i_m-P^i(Z^i_m,X_{m,c},X_{m,d}))]\\
+\;&O_p(h^s)\\
=\;&\frac{\xi^i(-Z^i_m+2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0,X_m,Z^i_m)f(-Z^i_m+2\mathbf{X}^i_{pm}\boldsymbol{\beta}^i_0,X_m)}{f^i(Z^i_m,X_m)}[D^i_m-P^i(Z^i_m,X_m)]\\
+\;&o_p(N^{-1/2}),
\end{aligned}$$

The projection $\mathbb{E}(\psi^i_{N,cs}(W_m,W_n)|W_n)$ is given by:

$$\begin{aligned}
&\sum_{x_{m,d}}\int_{z^i_m}\int_{x_{m,c}}\frac{\xi^i(Z^i_n,X_n,-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)}{f^i(-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0,X_n)}[(P^i(z^i_m,x^i_{m,c},x^i_{m,d})-P^i(-Z^i_n+2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0,X_n))\\
\times\;&K_h(z^i_m+Z^i_n-2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\boldsymbol{K}_h(x_{m,c}-X_{n,c})\times\mathbb{1}(x_{m,d}=X_{n,d})]dz^i_mdx^i_{m,c}.
\end{aligned}$$

Change of variable $x_{m,c} - X_{n,c} = hu_x$ and $z_m^i + Z_n^i - 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i = u_z h$ gives

$$\begin{aligned}
& \int_{z_m^i}\int_{x_{m,c}} \frac{\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)}{f^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}[(P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i + u_x h, X_{n,c} + u_x h, X_{n,d}) \\
- & \; P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n))k(u_z)\boldsymbol{k}(u_x)]du_z du_x \\
= & \; O_p(h^s) = o_p(N^{-1/2}).
\end{aligned}$$

Both projections by law of iterated expectation implies that $\mathbb{E}(\psi_{N,cs}^i(W_m, W_n)) = o(N^{-1/2})$. Also notice that calculation similar to that in equation (C.2) gives $\mathbb{E}[||\psi_{N,cs}^i(W_m, W_n)||^2] = o(N)$. Thus we could apply Lemma C.4.3 and get

$$\begin{aligned}
& \frac{1}{N(N-1)}\sum_{n=1}^{N}\frac{\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)}{f^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}\sum_{m=1,m\neq n}^{N}[(D_m^i - P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)) \\
\times & \; K_h(Z_m^i + Z_n^i - 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)\boldsymbol{K}_h(X_{m,c} - X_{n,c})\mathbb{1}(X_{m,d} = X_{n,d})] \\
= & \; \frac{1}{N}\sum_{n=1}^{N}\frac{\xi^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n, Z_n^i)f^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)} \\
\times & \; [D_n^i - P^i(Z_n^i, X_n)] + o_p(N^{-1/2}).
\end{aligned}$$

This gives us the following linear representation for $\widetilde{\Phi}_{cs,N}^i$:

$$\begin{aligned}
\widetilde{\Phi}_{cs,N}^i = & \; \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) \\
+ & \; \frac{1}{N}\sum_{n=1}^{N}\frac{\xi^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n, Z_n^i)f^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)} \\
\times & \; [D_n^i - P^i(Z_n^i, X_n)] + o_p(N^{-1/2}).
\end{aligned}$$

Regarding $\Delta_{cs,N}^i$, first notice that by a similar reasoning as in the derivation of equation (C.1), $\widehat{P}^{-i}(X_n) - P^{-i}(X_n)$ admits the following linearization:

$$\widehat{P}^{-i}(X_n) - P^{-i}(X_n) = \frac{1}{f(X_n)}\frac{1}{N}\sum_{l=1}^{N}(D_l^{-i} - P^{-i}(X_n))\boldsymbol{K}_h(X_{l,c} - X_{n,c})\mathbb{1}(X_{l,d} = X_{n,d}) + o_p(N^{-1/2}).$$

Plug in this linearization gives:

$$\begin{aligned}
&\Delta^i_{cs,N} \\
= \;& \frac{1}{N}\sum_{n=1}^{N} 2\beta^i_{20}\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\frac{\partial P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}{\partial z^i}\frac{1}{f(X_n)}\frac{1}{N}\sum_{l=1}^{N}(D^{-i}_l - P^{-i}(X_n)) \\
\times \;& \boldsymbol{K}_h(X_{l,c} - X_{n,c})\mathbb{1}(X_{l,d} = X_{n,d}) + o_p(N^{-1/2}).
\end{aligned}$$

The leading term in this expression is a U-statistic of order 2 with kernel

$$\begin{aligned}
\delta^i_{cs,N}(W_n, W_l) \;&\equiv\; 2\beta^i_{20}\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\frac{\partial P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}{\partial z^i}\frac{1}{f(X_n)}(D^{-i}_l - P^{-i}(X_n)) \\
&\times\; \boldsymbol{K}_h(X_{l,c} - X_{n,c})\mathbb{1}(X_{l,d} = X_{n,d}).
\end{aligned}$$

The projections are calculated as follows:

$$\begin{aligned}
\mathbb{E}[\delta^i_{cs,N}(W_n, W_l)|W_n] \;&=\; \sum_{x_{l,d}}\int_{z^i_l}\int_{x_{l,c}} 2\beta^i_{20}\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)\frac{\partial P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}{\partial z^i}\frac{1}{f(X_n)} \\
&\quad \times(P^{-i}(x_{l,c}, x_{l,d}) - P^{-i}(X_n))\boldsymbol{K}_h(x_{l,c} - X_{n,c})\mathbb{1}(X_{l,d} = X_{n,d})f^i(z^i_l, x_{l,c}, x_{l,d})dx_{l,c}dz^i_1 \\
&=\; O_p(h^s) = o_p(N^{-1/2}).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\delta^i_{cs,N}(W_n, W_l)|W_l] \;&=\; \sum_{x_{n,d}}\int_{z^i_n}\int_{x_{n,c}} 2\beta^i_{20}\xi^i(z^i_n, x_{n,c}, x_{n,d}, -z^i_n + 2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0)\frac{\partial P^i(-z^i_n + 2\mathbf{x}^i_{pn}\boldsymbol{\beta}^i_0, x_{n,c}, x_{n,d})}{\partial z^i} \\
&\times\; \frac{1}{f(x_{n,c}, x_{n,d})}(D^{-i}_l - P^{-i}(x_{n,c}, x_{n,d})) \\
&\times\; \boldsymbol{K}_h(X_{l,c} - x_{n,c})\mathbb{1}(X_{l,d} = x_{n,d})f^i(z^i_n, x_{n,c}, x_{n,d})dx_{n,c}dz^i_n \\
&=\; 2\beta^i_{20}\int_{t^i}\xi^i(t^i, X_l, -t^i + 2\mathbf{X}^i_{pl}\boldsymbol{\beta}^i_0)\frac{\partial P^i(-t^i + 2X^i_{pl}\boldsymbol{\beta}^i_0, X_l)}{\partial z^i}f^i(t^i|X_l)dt^i(D^{-i}_l - P^{-i}(X_l)) \\
&\equiv\; 2\beta^i_{20}C^i(X_l)(D^{-i}_l - P^{-i}(X_l)) + O_p(h^s) \\
&=\; 2\beta^i_{20}C^i(X_l)(D^{-i}_l - P^{-i}(X_l)) + o_p(N^{-1/2})
\end{aligned}$$

Law of iterated expectations gives $\mathbb{E}[\delta^i_{cs,N}(W_n, W_l)] = o(N^{-1/2})$, and calculation similar to that in equation (C.2) gives $\mathbb{E}[\delta^i_{cs,N}(W_n, W_l)^2] = o(N)$ and Hoeffding decomposition gives:

$$\Delta^i_{cs,N} = \frac{1}{N}\sum_{n=1}^{N} 2\beta^i_{20} C^i(X_n)(D^{-i}_n - P^{-i}(X_n)) + o_p(N^{-1/2}). \tag{C.3}$$

This implies that

$$\begin{aligned}
q^i_{N,1,cs}(\boldsymbol{\beta}^i_0) &= \frac{1}{N}\sum_{n=1}^{N} \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0) \\
&- \frac{1}{N}\sum_{n=1}^{N} \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0) P^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n) \\
&- \frac{1}{N}\sum_{n=1}^{N} 2\beta^i_{20} C^i(X_n)(D^{-i}_n - P^{-i}(X_n)) \\
&- \frac{1}{N}\sum_{n=1}^{N} \frac{\xi^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n, Z^i_n) f^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}{f^i(Z^i_n, X_n)} \\
&\times [(D^i_n - P^i(Z^i_n, X_n))] + o_p(N^{-1/2}).
\end{aligned}$$

□

The last step in the proof is to combine Lemma C.4.4, Lemma C.4.5 and the definition of the true distance function to obtain the asymptotic linear representation of $q^i_{N,1}(\boldsymbol{\beta}^i_0)$.

$$
\begin{aligned}
q^i_{N,1}(\boldsymbol{\beta}^i_0) &= q^i_{N,1,o}(\boldsymbol{\beta}^i_0) - q^i_{N,1,cs}(\boldsymbol{\beta}^i_0) \\
&= \frac{1}{N}\sum_{n=1}^{N}(\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)P^i(Z^i_n, X_n)) \\
&+ \xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)(D^i_n - P^i(Z^i_n, X_n)) \\
&- \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0) \\
&+ \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)P^i(-Z^i_n + 2\mathbf{X}^{i\top}_{pn}\boldsymbol{\beta}^i_0, X_n) \\
&+ \frac{1}{N}\sum_{n=1}^{N}2\beta^i_{20}C^i(X_n)(D^{-i}_n - P^{-i}(X_n)) \\
&+ \frac{1}{N}\sum_{n=1}^{N}\frac{\xi^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n, Z^i_n)f^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}{f^i(Z^i_n, X_n)} \\
&\times [(D^i_n - P^i(Z^i_n, X_n))] + o_p(N^{-1/2}) \\
&= \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z^i_n, X_n, -Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0)(D^i_n - P^i(Z^i_n, X_n)) \\
&+ \frac{1}{N}\sum_{n=1}^{N}2\beta^i_{20}C^i(X_n)(D^{-i}_n - P^{-i}(X_n)) \\
&+ \frac{1}{N}\sum_{n=1}^{N}\frac{\xi^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n, Z^i_n)f^i(-Z^i_n + 2\mathbf{X}^i_{pn}\boldsymbol{\beta}^i_0, X_n)}{f^i(Z^i_n, X_n)} \\
&\times [D^i_n - P^i(Z^i_n, X_n)] + o_p(N^{-1/2})
\end{aligned}
$$

where the last equality holds by the fact that

$$\begin{aligned}
& \frac{1}{N}\sum_{n=1}^{N}(\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)P^i(Z_n^i, X_n)) - \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i) \\
+ & \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n) \\
= & \frac{1}{N}\sum_{n=1}^{N}\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)(P^i(Z_n^i, X_n) - 1 + P^i(-Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)) = 0,
\end{aligned}$$

where inside the bracket is our population distance function evaluated at true parameter value. Define $t_n^i \equiv \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)(D_n^i - P^i(Z_n^i, X_n)) + \frac{\xi^i(-Z_n^i+2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n, Z_n^i)f^i(-Z_n^i+2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)}(D_n^i - P^i(Z_n^i, X_n)) + 2\beta_{20}^i C^i(X_n)(D_n^{-i} - P^{-i}(X_n))$, part (i) of Theorem 3.5.2 immediately follows from the expression for $q_{N,1}^i(\boldsymbol{\beta}_0^i)$. To show part (ii) of this theorem, given that $\mathbb{E}(t_n^i) = 0$, it suffices to show that $\mathbb{E}(||t_n^i||^2)$ is finite. Define $t_{n1}^i \equiv \xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)(D_n^i - P^i(Z_n^i, X_n))$, $t_{n2}^i \equiv \frac{\xi^i(-Z_n^i+2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n, Z_n^i)f^i(-Z_n^i+2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i, X_n)}{f^i(Z_n^i, X_n)}(D_n^i - P^i(Z_n^i, X_n))$ and $t_{n3}^i \equiv 2\beta_{20}^i C^i(X_n)(D_n^{-i} - P^{-i}(X_n))$. Triangle inequality and Cauchy-Schwarz inequality implies that $\mathbb{E}\,||t_n^i||^2$ is bounded above by

$$\mathbb{E}\,||t_{n1}^i||^2 + \mathbb{E}\,||t_{n2}^i||^2 + \mathbb{E}\,||t_{n3}^i||^2 + 2\,\mathbb{E}\,||t_{n1}^i||^2\,\mathbb{E}\,||t_{n2}^i||^2 + 2\,\mathbb{E}\,||t_{n2}^i||^2\,\mathbb{E}\,||t_{n3}^i||^2 + 2\,\mathbb{E}\,||t_{n1}^i||^2\,\mathbb{E}\,||t_{n3}^i||^2.$$

It's obvious that $\mathbb{E}\,||t_{n1}^i||^2$ and $\mathbb{E}\,||t_{n2}^i||^2$ are bounded under Assumption 3.5.1. Regarding $\mathbb{E}\,||t_{n3}^i||^2$, notice that each element of t_{n3}^i equals zero outside $\mathcal{S}_{(Z^i,X)}^{i,NT}$. Also notice that Assumption 3.5.1 implies that $\xi^i(Z_n^i, X_n, -Z_n^i + 2\mathbf{X}_{pn}^i\boldsymbol{\beta}_0^i)$ and $\frac{1}{f^i(Z_n^i, X_n)}$ are uniformly bounded on $\mathcal{S}_{(Z^i,X)}^{i,NT}$. Thus $\mathbb{E}\,||t_{n3}^i||^2 = \sum_{x_d}\int_{z^i}\int_{x_c}\frac{||\xi^i(-z^i+2\mathbf{x}_p^i\boldsymbol{\beta}_0^i, x_c, x_d, z^i)||^2(f^i(-z^i+2\mathbf{x}_p^i\boldsymbol{\beta}_0^i, x_c, x_d))^2}{f^i(z^i, x_c, x_d)}\mathbb{E}[(D_n^i - P^i(z^i, x_c, x_d))^2|z^i, x_c, x_d]dz^i dx_c < \infty$. The conclusion of part (ii) then follows by Lindeberg-Levy central limit theorem.

C.5 Proof of Proposition 3.7.1

Proof. Under Assumption 3.7.3-3.7.5, the number of equilibria is identified by Lemma 1 in Xiao (2018), and equilibrium beliefs given any ω and x are identified by Lemma 3 in Xiao (2018). Given the identified equilibrium beliefs, the proof of identification follows the same argument as in the proof of Theorem 3.3.2.