Identification and Estimation of Sequential Games of Incomplete Information with Multiple Equilibria

Jangsu Yoon*
University of Wisconsin-Milwaukee†

July 15, 2020

Abstract

This paper discusses the identification and estimation of game-theoretic models, mainly focusing on sequential games of incomplete information. In most empirical games, researchers cannot observe the exact order of actions played in the game and rely on the assumption of simultaneous actions. My structural modeling generalizes an empirical game to encompass simultaneous and sequential actions as special cases. I specify a sequential game allowing for multiple players in each stage and multiple Perfect Bayesian Nash Equilibria, showing that the structural parameters, including the payoff function parameters, the order of actions, and equilibrium selection mechanism, are separately identified. The excluded regressor that affects the variation of payoff functions but does not affect the order of actions is helpful to attain point identification of structural parameters. Next, I consider a Sieve Minimum Distance (SMD) estimator of Ai and Chen (2003) for estimating structural parameters and verify its asymptotic properties. The Monte Carlo simulations evaluate the performance of the proposed estimator and provide numerical evidence of potential bias under the misspecified order of actions. The empirical application with an entry game of Walmart and Kmart shows that retailers compete sequentially in a significant portion of markets.

Keywords: Sequential games, Discrete games of incomplete information, Perfect Bayesian Nash Equilibria, Semiparametric estimation, Sieve Minimum Distance estimator, Multiple equilibria.

JEL Classification: C14, C35, C62, C73

*I am deeply indebted to my advisor Jack Porter for his valuable comments and constant supports that improve this research. I would like to appreciate Bruce Hansen, Xiaoxia Shi, and Joachim Freyberger for their helpful suggestions and opinions. I also benefited from feedbacks from Hiroaki Kaido, Xun Tang, Roger Moon and seminar participants at UW-Madison Econometrics Seminar, CMES 2019, MEG 2019, and discussions with my colleagues. I gratefully acknowledge the financial support from Yoshio Niho Faculty Excellence Fund. All errors are mine.

†Address: Jangsu Yoon, Department of Economics, University of Wisconsin-Milwaukee, 3210 North Maryland Avenue, Bolton Hall #882, Milwaukee, WI 53211, USA. Contact Information: yoon22@uwm.edu

1 Introduction

1.1 Motivation

This paper is an extension of the econometric approach to game-theoretic models, mainly focusing on sequential games. I consider a classic sequential move game that a later mover can observe the action of previous movers. Sequential games have been widely discussed in classical game theory, especially in extensive form games (Kreps and Wilson (1982), Milgrom and Roberts (1982), Fudenberg and Tirole (1991)). However, there are few empirical applications based on the model of sequential games in previous literature. Instead, almost all empirical models assuming strategic interactions among rational, forward-looking players use either a simultaneous game or a dynamic discrete game. In this paper, I specify a structural model based on sequential games, which share some properties with both simultaneous and dynamic discrete games but are fundamentally distinct from these games.

The model I develop is a generalized sequential game, which is distinguished from other types of games. First, the game is not a simultaneous game because due to asymmetric information sets among players. Players consist of multiple groups, and sequentially choose their actions while each group of players simultaneously make decisions. In a two-player sequential game, the second mover decides after observing the first mover's action while the first mover does not have information on the rival's response. Second, the game is not a dynamic discrete game because each player may not necessarily have multiple decision nodes. In a dynamic discrete game, simultaneous actions made by a group of players repeat for multiple periods. The sequential game model of the current paper includes a one-shot game in which each player has a distinct decision node or a sequential bargaining game between two groups of players as a special case. The setup, therefore, allows us to see the effect of commitment in some particular types of empirical games. Each player maximizes the expected profit based on the information set after observing the full history of actions made by previous groups.[1]

The sequential game model I consider is also different from other empirical game models because the structure of decision nodes is not necessarily known to the researchers. In a sequential entry game between two rivals, e.g. Walmart and Kmart, the order of actions may change across games. The researcher can observe the final entry decisions made by firms but cannot precisely know whether a game is simultaneous or not and who the first/second mover is. The probability distribution of the order of actions conditional on observable covariates is a component of structural parameters in my model.

Under similar assumptions as simultaneous games literature, I provide identification and estimation of structural parameters, including payoff function parameters, the equilibrium selection mechanism, and the distribution of the order of actions, under the econometric specification of sequential games. The asymptotic distribution of the functional of structural parameters is derived

[1] Note that the dynamic structure captured in this paper is up to a finite horizon dynamic game. The previous literature on estimating dynamic games allow for infinite horizon: see Aguirregabiria and Mira (2002, 2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), and Pesendorfer and Schmidt-Dengler (2008).

based on Ai and Chen (2003) and Chen and Pouzo (2009, 2015). Monte Carlo simulations and the empirical application to the entry game of Walmart and Kmart highlight that the conventional estimators without considering the order of actions may have a potential loss of precision.

The current paper contributes to the existing literature in several ways. First, I establish semiparametric identification and estimation of structural parameters for a sequential game of incomplete information considering both multiple equilibria and unobserved order of actions. The structural model provides a general extension of the identification and estimation in simultaneous games with incomplete information. The structural parameters include not only the payoff function of players but also unobserved heterogeneity, including the distribution of the order of actions and the equilibrium selection mechanism. The difficulty of identification comes from the unknown order of actions that is observable to players but unobserved by econometricians. I show that the number of heterogeneity types is finite, then separately identify the order of actions and the equilibrium selection mechanism as well as the parametric payoff function parameters. To the best of my knowledge, this paper is the first econometric approach to incomplete information sequential games considering unobserved heterogeneity.

Second, the paper accommodates discrete type unobserved heterogeneity into the semiparametric estimation and inference on empirical games. I propose a Sieve Minimum Distance (SMD) estimator of Ai and Chen (2003) for structural parameters using conditional moment restrictions derived from the model. The structural parameters consist of parametric payoff function parameters and nonparametric probability distribution function of unobserved heterogeneity. Estimating the parameters regarding unobserved heterogeneity is particularly essential to empirical games because the unobserved order of actions affects the estimates of payoff functions and the subsequent counterfactual outcomes. I verify consistency and asymptotic normality of the estimator and derive a valid confidence interval for parameters of interest.

Third, Monte Carlo simulations and an empirical application to the entry game between Walmart and Kmart present a potential implementation of the model to various economic illustrations with strategic interactions. I conduct numerical experiments to show the finite sample performance of the suggested estimator. The analyses also highlight numerical evidence of asymptotic bias under the misspecified order of actions. The result implies the benefit of using the proposed estimator that is robust to the unknown order of actions. The empirical application revisits the entry game between Walmart and Kmart discussed in Jia (2008). The new estimates based on the theoretical results of this paper show that Walmart and Kmart's entry decisions are more closely correlated with each other. The estimated order of actions, which is a function of regional dummies, verifies that each player is more likely to become the first mover at the counties in the vicinity of its headquarters.

The suggested structural model of sequential games can also be adapted to various theoretical and empirical topics, including firm entry, bargaining, matching, and dynamic discrete choice problems. In particular, the game-theoretic approach to analyze an oligopolistic market is common in the industrial organization literature. The entry game is a classic example discussed by Bresnahan and Reiss (1990, 1991). More empirical applications of entry games include the competition among retail and chain stores (Pinkse, Slade, and Brett (2002), Smith (2004), Holmes (2001, 2011), Davis (2006), Jia (2008), Ellickson, Houghton, and Timmins (2013), Aradillas-lopez and Gandhi (2016)), airline companies (Berry (1992), Ciliberto and Tamer (2009), Blevins (2015)), and video stores (Seim (2006)). Note that most of the previous literature mentioned above provide results under the simultaneous game assumption. However, a sequential game specification is sometimes more natural because information between players may not be symmetric: players may not simultaneously observe the rival's action without any difference in timing.

The applicability of sequential games is not limited to topics in entry games. The model can also be applied to the industry-specific bargaining literature, including Gal-Or (1997), Ho (2009), and Crawford and Yurukoglu (2012). The bargaining game between hospitals and insurers (Ho (2009)) is one example in which the strategic interactions between agents are essential to explain the market structure. Ho (2009) suggests a sequential game between hospital groups and insurance plans. More empirical topics include international relations, a tax/tariff competition between nations, or a price competition between firms.

1.2 Previous Literature

There have been many empirical works based on game-theoretic models. Most of these papers focused on identifying and estimating the payoff function in a simultaneous game, without specifying the order of actions or asymmetric information between players. Bresnahan and Reiss (1990, 1991) worked on the entry/exit model in oligopolistic markets. Berry (1992) applied the entry game to the airline industry to figure out the effect of strategic interactions between firms on the decision of airport presence. The following seminal papers include Tamer (2003), Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010), and Kline (2015), though these papers belong to the complete information game literature.

The current paper relates to the literature on incomplete information games where each player has private information not observed by other players. A flourishing line of research includes Aradillas-lopez (2010), Bajari, Hong, Krainer, and Nekipelov (2010), Tang (2010), De Paula and Tang (2012), Lewbel and Tang (2015), Wan and Xu (2014), Xiao (2018), and Aguirregabiria and Mira (2019). Each of these papers studied the identification and estimation of incomplete information games under various specifications. Most of the documents focused on nonparametric identification of the payoff function parameters. There are also various empirical applications assuming an incomplete information game, e.g., Seim (2006), Sweeting (2009), Vitorino (2012). Though this paper is also based on an incomplete information game, the model specification is distinct from previous literature as I consider both simultaneous moves and sequential moves.

There are several papers considering the effect of unobserved heterogeneity in models of empirical games. Kasahara and Shimotsu (2009), Arcidiacono and Miller (2011), and Hu and Shum (2012) provided nonparametric identification of dynamic models with unobserved heterogeneity. Based on the identification strategy, Grieco (2014), Igami and Yang (2016), and Aguirregabiria and Mira (2019) derived identification of empirical games with discrete type unobserved heterogeneity in player's payoff functions. Particularly Aguirregabiria and Mira (2019) showed how to separately identify payoff function parameters, equilibrium selection mechanism, and the distribution of unobserved heterogeneity in a simultaneous game. The current paper shares a similar model specification as Igami and Yang (2016) and Aguirregabiria and Mira (2019) in that the source of unobserved heterogeneity has discrete support. But the identification strategy in this paper is different from previous literature because the finite mixture approach used in Kasahara and Shimotsu (2009) does not hold. A sequential game model violates the assumption of first-order Markov property as the later mover's action depends on the full history of previous mover's actions.

The motivation of this paper is closely related to Einav (2010) and Blevins (2015). Both articles discussed the identification and estimation of sequential games. Einav (2010) studied the identification of a sequential game under the assumption of incomplete information and applied it to analyze a release date timing game in the movie industry. Blevins (2015) derived identification of a complete information sequential game when the unique Subgame Perfect Nash Equilibrium exists, and the order of actions is unobservable to econometricians. In addition to these inspirational works, the current paper's setup involves multiple equilibria and a more generalized order of actions. The valid estimation and inference method for structural parameters and verifying asymptotic properties of the estimator are new in the sequential game literature.

In the estimation of structural parameters, I apply the nonparametric sieve method to estimate the equilibrium selection mechanism and the distribution of the order of actions without relying on parametric assumptions. The SMD estimator of Ai and Chen (2003) and the Sieve Wald statistic of Chen and Pouzo (2015) fit the model setup. This paper verifies that the suggested estimation and inference methods are valid for handling models in empirical games.

1.3 Summary of Contents

In Section 2, I introduce a general model that represents a sequential game and define an equilibrium concept based on Perfect Bayesian Nash Equilibria (PBNE). Section 3 provides identification results. A necessary and sufficient condition of identification and some exclusion restrictions that help satisfy the identification condition are discussed. Section 4 suggests a SMD estimator of the structural parameters, following the identification result in Section 3. The asymptotic properties of the suggested estimator are discussed in the same section. Section 5 is about Monte Carlo simulations and an empirical application to the entry game of Walmart and Kmart. The Monte Carlo simulation part shows how sensitive the identification and estimation results are under misspecified order of actions. Section 6 sums up everything with conclusion.

2 The General Model

2.1 Basic Setting: Player, Action, and Payoff

Denote a sequential game by $\mathcal{G} \equiv \left\{ I_t, \mathcal{A}, \{u_i\}_{i \in I_t} \right\}_{t=1}^{T_o}$, which consists of a number of stage games indexed by $t \in \mathcal{T}_o \equiv \{1, \dots, T_o\}$. The set of players is given by $\mathcal{I} \equiv \{1, \dots, N\}$. For sequential actions, I suppose T_o groups of players allocated by a specific order of actions $o : \mathcal{I} \to \mathcal{T}_o$, which may not be observed by econometricians. Under the given mapping o, the group t players are assigned by $I_t \equiv \{i \in \mathcal{I} | o(i) = t\}$ with $\bigcup_{t=1}^{T_o} I_t = \mathcal{I}$, and $O = (o(1), \dots, o(N))'$ is a vector of assigned groups. $n_t \in \{1, 2, \dots, N\}$ is the number of players in group t. Assume that each player has only one decision node. This section describes the game under a given order $O \in \mathcal{O}$, where $\mathcal{O}$ includes in total N_o possible orders.[2]

The action set of players is $\mathcal{A} = \{a_0, a_1, \dots, a_L\}$, which is finite and discrete. Every player has the same $L+1$ possible actions. $\mathcal{A}^N$ defines the whole game's action space. The action set fixes $a_0 = 0$ as a baseline action or an outside option. For example, the action space of an entry game is $\mathcal{A} = \{0, 1\}$, and every player decides whether or not to enter the market.

Next, let $s_i \in \mathcal{A}$ be the action of player i, and $s_{-i} \in \mathcal{A}^{N-1}$ be the set of actions without s_i. The set of actions selected by all players is $s \equiv (s_1, \dots, s_N)' \in \mathbb{R}^N$. I use notation of subscripts $-$ and $+$ to highlight the timing of decision making. Let $s_-^{o(i)}$ be a $n_-^{o(i)} \equiv \sum_{t=1}^{o(i)-1} n_t$ dimensional history vector that consists of $\{s_j | o(j) < o(i)\}$, and similarly $s_+^{o(i)}$ includes $n_+^{o(i)} \equiv \sum_{t=o(i)}^{T_o} n_t - 1$ actions of $\{s_j | o(j) \geq o(i)\} \setminus \{s_i\}$ that are not realized when player i makes a decision. The history vector is a common subset of the information set for all players in group t. For each player i, $s = \left(s_-^{o(i)'}, s_i, s_+^{o(i)'} \right)'$.

Player i's payoff is

$$u_i(s, X, O, \epsilon_i) = \pi_i \left(s_i, s_+^{o(i)}, \mathcal{J}_{o(i)}; \beta \right) + \epsilon_i(s_i),$$

where π_i is the structural part of the payoffs that are known up to a finite dimensional parameter $\beta \in \mathcal{B} \subseteq \mathbb{R}^{d_\beta}$, and $\epsilon_i(s_i) : \mathcal{A} \to \mathbb{R}$ is player i's private information depending on her action s_i. The realized value of $\epsilon_i \equiv (\epsilon_i(a_0), \dots, \epsilon_i(a_L))'$ is only observable to player i but the distribution of ϵ_i is common knowledge for other players and econometricians. $\mathcal{J}_{o(i)} \equiv \left(s_-^{o(i)}, X, O \right)$ is the information set of player i, and X is a vector of observable covariates. The history vector $s_-^{o(i)}$ is the main difference from a simultaneous game as the first-mover's action is observed by the following movers. The model encompasses the conventional approach to simultaneous games where $T = 1$ as a special case, and the sequential game model of Einav (2010) where $T_o = N$ for $t = 1, \dots, N$ as well.

Remark 2.1. The specification of $n_t > 1$ not only generalizes simultaneous games but also enables various economic applications. The bargaining games discussed by Ho (2009), Collard-Wexler,

[2] If the game has two players ($N = 2$), there are three possible orders: two sequential actions $O = (1, 2)'$ or $O = (2, 1)'$, and one simultaneous actions $O = (1, 1)'$. In this case $N_O = 3$. In general, the possible number of orders follows the number of weak orderings for N players, or the ordered Bell numbers.

Gowrisankaran, and Lee (2019) can be adapted to the generalized setup in my paper. For example, Ho (2009) considered two groups of players: hospitals and insurance plans. The hospitals make take-it-or-leave-it offers in the first stage, and insurance plans respond in the second stage. The case of $n_t > 1$ is considered because an empirical case may not be certainly declared as a pure simultaneous or sequential game. For example, in entry games, researchers may observe the opening dates for stores, but the observed order may not reflect the decision timing of players. The real-world example is potentially located in the middle of a simultaneous and sequential game, and such a case requires a more flexible specification of strategic interactions.

Remark 2.2. The information set $\mathcal{J}_{o(i)}$ assumes the incumbent's action to be fully observed by all entrants. The assumption does not necessarily hold in some economic applications: sequential auctions in which entrants can only observe the winning bid (Brendstrup and Paarsch (2005)). The assumption that players observe the full action profile can be relaxed in several ways. First, each player's information set contains only up to some recent history. For example, a player in group t can observe actions played in group $t-1$ and $t-2$ but not others from group 1 to $t-3$. This finite memory assumption is not unusual in the game theory literature (Bhaskar, Mailath, and Morris (2013), Sperisen (2018)), but is uncommon in the empirical research. Second, one may assume that players simultaneously submit their action profile with commitment, but possibly modify their strategy with some cost. I leave the extensions to future research.

2.2 Equilibrium of the Game

This section introduces the *Perfect Bayesian Nash Equilibrium* (PBNE) as the equilibrium concept of the game. The equilibrium concept is required to infer the payoff function parameters from the player's action since a researcher can observe only covariates X and the realized action profile s. Define the strategy function S_i by a mapping $s_i = S_i\left(\mathcal{J}_{o(i)}, \epsilon_i\right)$ for $i \in \mathcal{I}$. Then the probability of player i to choose an action a conditional on her information set $\mathcal{J}_{o(i)}$ is

$$P\left(s_i = a | \mathcal{J}_{o(i)}\right) = \int 1\left\{S_i\left(\mathcal{J}_{o(i)}, \epsilon_i\right) = a\right\} dF\left(\epsilon_i | \mathcal{J}_{o(i)}\right), \tag{1}$$

and the expected profit function of player i for an action $s_i = a$ is derived as follows:

$$\begin{aligned}\Pi_i\left(a, \mathcal{J}_{o(i)}, \epsilon_i\right) &= \sum_{a_+^{o(i)}} \pi_i\left(a, a_+^{o(i)}, \mathcal{J}_{o(i)}; \beta_0\right) P\left(s_+^{o(i)} = a_+^{o(i)} | \mathcal{J}_{o(i)}\right) + \epsilon_i(a) \\ &\equiv \bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right) + \epsilon_i(a),\end{aligned}$$

where $a_+^{o(i)}$ denotes any realized action of $s_+^{o(i)}$. The expected payoff consists of two parts: $\bar{\pi}_i$ is the expected mean payoff, and ϵ_i is a private shock which does not depend on other players' actions. The functional form of π_i is known to the econometrician up to the finite dimensional parameter

β_0. Player i chooses the optimal action s_i by solving

$$s_i = \arg\max_{a \in \mathcal{A}} \Pi_i \left(a, \mathcal{J}_{o(i)}, \epsilon_i \right).$$

The sequential behavioral assumption leads to a market outcome, which corresponds to a Bayesian Nash Equilibrium in the simultaneous game literature. I use the Perfect Bayesian Nash Equilibrium concept to describe the market outcome.

Definition 2.1. (Perfect Bayesian Nash Equilibrium) A *Perfect Bayesian Nash Equilibrium* (PBNE) is a set of strategies $\left\{ S_i \left(\mathcal{J}_{o(i)}, \epsilon_i \right) \right\}_{i=1}^{N}$ such that for every player $i \in \mathcal{I}$,

$$S_i \left(\mathcal{J}_{o(i)}, \epsilon_i \right) = \arg\max_{a \in \mathcal{A}} \Pi_i \left(a, \mathcal{J}_{o(i)}, \epsilon_i \right), \tag{2}$$

where $\mathcal{J}_{o(i)} = \left(s_-^{o(i)}, X, O \right)$ and $s_-^{o(i)}$ is the history of equilibrium actions.

The conditions for PBNE are satisfied by Definition 2.1. First, the sequential rationality condition is satisfied by the equation (2). Second, the consistent belief condition holds because the distribution of private shocks conditional on history is common knowledge to all players. The definition requires the whole set of equilibrium strategies S to be a subset of *Bayesian Nash Equilibrium* (BNE).[3]

The equation (2) implies choice probabilities and payoff functions. For $i \in \mathcal{I}$ and $a \in \mathcal{A}$,

$$\begin{aligned} & S_i \left(\mathcal{J}_{o(i)}, \epsilon_i \right) = a \\ \iff \quad & \Pi_i \left(a, \mathcal{J}_{o(i)}, \epsilon_i \right) \geq \Pi_i \left(a', \mathcal{J}_{o(i)}, \epsilon_i \right) \text{ for all } a' \in \mathcal{A} \backslash \{a\}, \end{aligned}$$

so that

$$\begin{aligned} P \left(s_i = a | \mathcal{J}_{o(i)} \right) &= P \left(\Pi_i \left(a, \mathcal{J}_{o(i)}, \epsilon_i \right) \geq \Pi_i \left(a', \mathcal{J}_{o(i)}, \epsilon_i \right), \forall a' \neq a \right) \\ &= P \left(\epsilon_i \left(a' \right) - \epsilon_i \left(a \right) \leq \bar{\pi}_i \left(a, \mathcal{J}_{o(i)}; \beta_0 \right) - \bar{\pi}_i \left(a', \mathcal{J}_{o(i)}; \beta_0 \right), \forall a' \neq a \right) \\ &= F_a \left(\bar{\pi}_i \left(a, \mathcal{J}_{o(i)}; \beta_0 \right) - \bar{\pi}_i \left(a', \mathcal{J}_{o(i)}; \beta_0 \right), \forall a' \neq a \right), \end{aligned} \tag{3}$$

where F_a is the CDF of $\{\epsilon_i (a') - \epsilon_i (a), \forall a' \neq a\}$ conditional on $\mathcal{J}_{o(i)}$.

Remark 2.3. A simple static model with no history clarifies the difference of the current paper from the previous literature. Under $s_-^t = \emptyset$ (no history) and $T = 1$ (simultaneous actions), the CCP is defined by

$$P \left(s_i = a | X \right) = F_a \left(\bar{\pi}_i \left(a, X; \beta_0 \right) - \bar{\pi}_i \left(a', X; \beta_0 \right), \forall a' \neq a \right),$$

so that a simultaneous game is a special case that $s_-^{o(i)} = \emptyset$ is fixed and known value for all players.

[3]The concept of PBNE is a generally used equilibrium concept for incomplete information simultaneous games. The PBNE is weaker than MPE that is a commonly used equilibrium concept in dynamic games. The definition of MPE follows Maskin and Tirole (2001), and Bhaskar, Mailath, and Morris (2013).

The following assumptions establish the one-to-one mapping between $P\left(s_i = a|\mathcal{J}_{o(i)}\right)$ and $\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right)$.

Assumption 2.1. *(Distributional Property of ϵ_i) For all $i \in \mathcal{I}$ and $t \in \mathcal{T}_o$, the distribution of ϵ_i satisfies the followings:*

1. *$\epsilon_i(s_i)$ are i.i.d. across players and actions conditional on (X, O).*
2. *The CDF of ϵ_i is absolutely continuous with respect to the Lebesgue measure and common knowledge to both players and econometricians.*

Assumption 2.2. *(Normalization) For all $i \in \mathcal{I}$, $\pi_i\left(0, s_+^{o(i)}, \mathcal{J}_{o(i)}; \beta_0\right) = 0$ for all values of $s_+^{o(i)}$ and $\mathcal{J}_{o(i)}$.*

Assumption 2.1$-$1 states that a private shock is independently realized for every player conditional on observable covariates and the given order of actions.[4] Assumption 2.1$-$2 implies that the distribution function of private shocks is known and continuous almost everywhere. Assumption 2.2 is to normalize the payoff for the default action $s_i = 0$. The normalization enables to recover the expected payoff functions with $s_i \neq 0$. The specification is natural in entry models where no-entry gives no profit. The setup is commonly used in binary choice models with $\mathcal{A} = \{0, 1\}$. The Assumptions 2.1$-$2.2 derive two subsequent lemmas as below.

Lemma 2.1. *Under Assumption 2.1, a PBNE is defined by a vector of equilibrium probabilities*

$$\mathcal{P}(X, O) \equiv \left\{P\left(s_i = a|\mathcal{J}_{o(i)}\right) \middle| a \in \mathcal{A}, a_-^{o(i)} \in \mathcal{A}^{n_-^{o(i)}}, i \in \mathcal{I}\right\}$$

that satisfy the equation (3) for all $a \in \mathcal{A}$, $a_-^{o(i)} \in \mathcal{A}^{n_-^{o(i)}}$, $i \in \mathcal{I}$, and the number of PBNE is finite.

Proof. Appendix A.1. □

Hereafter define the equilibrium set $\mathcal{E}(X, O)$ by the set of PBNEs $\mathcal{P}(X, O)$. Denote $B(X, O)$ by the number of PBNE conditional on X and O, while the number depends on the true payoff parameter β. Define $\left\{\tau_{X,O,1}, \ldots, \tau_{X,O,B(X,O)}\right\}$ by equilibrium types in $\mathcal{E}(X, O)$. The next lemma is based on Hotz and Miller (1993), verifying the one-to-one mapping between each of equilibrium probabilities $\mathcal{P}(X, O)$ and the corresponding set of expected payoffs

$$\bar{\Pi}(X, O) \equiv \left\{\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right) \middle| a \in \mathcal{A}, a_-^{o(i)} \in \mathcal{A}^{n_-^{o(i)}}, i \in \mathcal{I}\right\}.$$

Lemma 2.2. *Suppose Assumptions 2.1$-$2.2 hold. Denote that $\mathcal{P}^{(k)}(X, O)$ and $\bar{\Pi}^{(k)}(X, O)$ are equilibrium probabilities and the expected payoffs when the equilibrium type is $\tau_{X,O,k}$. Then $\mathcal{P}^{(k)}(X, O)$ and $\bar{\Pi}^{(k)}(X, O)$ have one-to-one correspondence for $k = 1, \ldots, B(X, O)$.*

[4] A similar type of this assumption is also observed in Seim (2006), Sweeting (2009), Aradillas-Lopez (2010), Bajari, Hong, Krainer, and Nekipelov (2010), De Paula and Tang (2012), Lewbel and Tang (2015), Aguirregabiria and Mira (2019), and other incomplete information game literature.

Proof. Appendix A.2. □

The lemma implies that the mappings F_a in equation (3) are invertible. Then for a given value of the information set, the equilibrium probability $P\left(s_i = a|\mathcal{J}_{o(i)}\right)$ can be represented by a function of the payoff function parameters β_0. In the following sections, I assume that $\mathcal{P}(X, O)$ is identified up to a finite dimensional parameter β_0 from $\bar{\Pi}(X, O)$ and discuss conditions to identify structural parameters of the model: the payoff function parameters β, the equilibrium selection mechanism, and the distribution of the order of actions.

Example 2.1. (Type-I Extreme Distribution) Consider an example of Lemma 2.2 in the context of the multinomial Logit model. Suppose that $\epsilon_i(s_i)$ follows an i.i.d. standard Type-I extreme distribution. Then

$$P\left(s_i = a|\mathcal{J}_{o(i)}\right) = \frac{\exp\left(\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right)\right)}{\sum_{a' \in \mathcal{A}} \exp\left(\bar{\pi}_i\left(a', \mathcal{J}_{o(i)}; \beta_0\right)\right)},$$

for all $a \in \mathcal{A} \backslash \{0\}$, $a_-^{o(i)} \in \mathcal{A}^{n_-^{o(i)}}$, and $i \in \mathcal{I}$. The system of nonlinear equations consist of $L \times (L+1)^{n_-^{o(i)}}$ conditional probabilities for each i and $L \times (L+1)^{n_-^{o(i)}}$ expected payoff functions. The system has a unique solution of $\bar{\Pi}(X, O)$ by Lemma 2.2 for a given equilibrium probabilities $\mathcal{P}(X, O)$.

3 Identification

This section takes steps to identify structural parameters in the sequential game. The main goal is to figure out how to identify the structural parameters $\left\{\beta_0, \lambda_{\tau|X,O}, \rho_{O|X}\right\}$ where $\lambda_{\tau|X,O}$ is the probability of an equilibrium type τ conditional on X and O (i.e. equilibrium selection mechanism) and $\rho_{O|X}$ is the distribution of the unobserved order of actions O conditional on X.

I start with a simplified setup without multiple equilibria, focusing on the unknown order of actions. Then I generalize the model to encompass an arbitrary equilibrium selection mechanism. The challenging part is to consider both multiple equilibria and unknown order of actions. The multiple equilibria issue is well-known in simultaneous games literature while identifying the correct order of actions is an additional challenge in sequential games. I provide a necessary and sufficient condition of identification by exploiting the property that the type of equilibria and the possible order of actions are finite. Then I suggest an exclusion restriction that is helpful to achieve identification in the general case.

In the previous section, Lemma 2.2 verifies a connection of expected payoffs $\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right)$ and $P\left(s_i = a|\mathcal{J}_{o(i)}\right)$ for all $a \in \mathcal{A} \backslash \{0\}$, $a_-^{o(i)} \in \mathcal{A}^{n_-^{o(i)}}$, and $i \in \mathcal{I}$. Denote $P^{(k)}\left(s_i = a|\mathcal{J}_{o(i)}; \beta_0\right)$ for $k \in \left\{\tau_{X,O,1}, \ldots, \tau_{X,O,B(X,O;\beta_0)}\right\}$ by the type-k equilibrium probability of player i to choose an action $a \in \mathcal{A}$. $B(X, O; \beta_0)$ highlights the number of PBNE depending on the parameter value β_0.

Consider an action profile $\alpha = (\alpha_1, \ldots, \alpha_N)' \in \mathcal{A}^N$, all possible orders $\mathcal{O} \equiv \{O_1, \ldots, O_{N_O}\}$, and all types of equilibria $\left\{\tau_{X,O,1},\ldots,\tau_{X,O,B(X,O;\beta_0)}\right\}$, then for a joint action α,

$$P\left(s=\alpha|X\right)=\sum_{l=1}^{N_O}\sum_{b=1}^{B(X,O;\beta_0)}\prod_{i=1}^{N}P^{\left(\tau_{X,O_l,b}\right)}\left(s_i=\alpha_i|\mathcal{J}_{o_l(i)};\beta_0\right)\lambda_{\tau|X,O}\left(\tau_{X,O_l,b}|X,O_l\right)\rho_{O|X}\left(O_l|X\right), \tag{4}$$

where each $P^{\left(\tau_{X,O_l,b}\right)}\left(s_i=\alpha_i|\mathcal{J}_{o_l(i)};\beta_0\right)$ is an element of $\mathcal{P}\left(X,O_l\right)\in\mathcal{E}\left(X,O_l\right)$. If the equilibrium probability $P^{\left(\tau_{X,O,b}\right)}\left(s_i=a|\mathcal{J}_{o(i)};\beta_0\right)$ and the weight function $\lambda_{\tau|X,O}\rho_{O|X}$ are separately identified from $P\left(s=\alpha|X\right)$, the expected payoff function $\bar{\pi}_i\left(a,\mathcal{J}_{o(i)};\beta_0\right)$ is also identified by Lemma 2.2. Then the payoff function parameter β_0 is identified by solving $\bar{\pi}_i\left(a,\mathcal{J}_{o(i)};\beta_0\right)$ with $\mathcal{P}\left(X,O_l\right)$. The nonparametric weight components $\lambda_{\tau|X,O}\left(\tau_{X,O,b}|X,O\right)$ and $\rho_{O|X}\left(O|X\right)$ are separately identified from the product function $\lambda_{\tau|X,O}\left(\tau_{X,O,b}|X,O\right)\rho_{O|X}\left(O|X\right)$.

Definition 3.1. (Identification) Under the known distribution $P\left(s=\alpha|X\right)$ for all $\alpha\in\mathcal{A}^N$, a sequential game model is identified if and only if there exists a unique value of structural parameters $\left\{\beta_0,\lambda_{\tau|X,O},\rho_{O|X}\right\}$ that solves the equation (4).

The next subsection introduces identification steps in a simplified setup. The identification for a general setup follows in later sections, provided with additional assumptions.

3.1 Identification without Multiple Equilibria

Consider a simplified model without multiple equilibria, while the order of actions is not observable to researchers. Many empirical works of simultaneous games with incomplete information suffer from the potential multiple equilibria, but there are several ways to avoid multiple equilibria using additional assumptions. The first is to restrict the possible order of actions to fully sequential actions. The source of multiple equilibria in a sequential game comes from players who make decisions simultaneously. Einav (2010) suggested a pseudo-backward-induction method under the assumption of fully sequential actions. The second is to assume the degenerate equilibrium selection. The assumption that the same equilibrium is played every time may not be realistic, but showing the identification procedure without multiple equilibria is useful to clarify the identification strategy with multiple equilibria.

Suppose a unique equilibrium, $B\left(X,O\right)=1$ for all $X\in\mathcal{X}$ and $O\in\mathcal{O}$ regardless of $\beta\in\mathcal{B}$. Then

$$P\left(s=\alpha|X\right)=\sum_{l=1}^{N_o}\prod_{i=1}^{N}P\left(s_i=\alpha_i|\mathcal{J}_{o_l(i)};\beta_0\right)\rho_{O|X}\left(O_l|X\right), \tag{5}$$

where $P\left(s_i=\alpha_i|\mathcal{J}_{o_l(i)};\beta_0\right)=P^{\left(\tau_{X,O,1}\right)}\left(s_i=\alpha_i|\mathcal{J}_{o_l(i)};\beta_0\right)$ for simplicity. The equilibrium probability $P\left(s_i=\alpha_i|\mathcal{J}_{o(i)};\beta_0\right)$ for $i=1,\ldots,N$ belongs to the equilibrium set $\mathcal{E}\left(X,O\right)$. $P\left(s_i=\alpha_i|\mathcal{J}_{o(i)};\beta_0\right)$ is a function of the payoff function parameter β_0 by the equation (3). Starting by the last group order $i\in I_{T_o}$, $\left\{P\left(s_i=\alpha_i|\mathcal{J}_{o(i)};\beta_0\right)\right\}_{i\in I_{T_o}}$ are uniquely identified by solving n_{T_o} equations of (3). $\left\{P\left(s_i=\alpha_i|\mathcal{J}_{o(i)};\beta_0\right)\right\}_{i\in I_{T_o-1}},\ldots,\left\{P\left(s_i=\alpha_i|\mathcal{J}_{o(i)};\beta_0\right)\right\}_{i\in I_1}$ are also sequentially identified up to the parameter vector β_0 following the spirit of Einav (2010)'s pseudo-backward-induction method.

The distribution function of order $\left\{\rho_{O|X}\left(O_l|X\right)\right\}_{l=1}^{N_o}$ in the equation (5) belongs to a discrete probability function space

$$\mathcal{P}_O \equiv \left\{\rho_{O|X} \middle| \sum_{l=1}^{N_o} \rho_{O|X}\left(O_l|X\right) = 1,\ \rho_{O|X}\left(O_l|X\right) \geq 0 \text{ for } l = 1, \ldots, N_o\right\},$$

and there are $N_o - 1$ nonparametric components. Thus a necessary order condition of identification is $(L+1)^N \geq N_o$ so that the number of conditional moments is as many as the number of possible orders.

Define a matrix of equilibrium probabilities and a vector of conditional choice probabilities for all joint actions

$$\begin{aligned} \mathcal{P}\left(X;\beta_0\right) &\equiv \left[P\left(X,O_1;\beta_0\right), \ldots, P\left(X,O_{N_o};\beta_0\right)\right] \in \mathfrak{M}_{(L+1)^N \times N_o} \\ Q\left(X\right) &\equiv \left(P\left(s=\alpha^1|X\right), \ldots, P\left(s=\alpha^{(L+1)^N-1}|X\right), 1\right)' \in \mathbb{R}^{(L+1)^N} \end{aligned}$$

where $P\left(X,O;\beta_0\right) = \left(P\left(s=\alpha^1|X,O;\beta_0\right), \ldots, P\left(s=\alpha^{(L+1)^N-1}|X,O;\beta_0\right), 1\right)' \in \mathbb{R}^{(L+1)^N}$ is a vector of equilibrium probabilities for all joint actions except $\alpha = (0, \ldots, 0)$. The structural parameters are identified if there is a unique solution $\left(\beta_0, \rho_{O|X}\right)$ of the equation (5). The following Assumption 3.1 provides a sufficient rank condition for identification.

Assumption 3.1. *(Rank Condition) $\mathcal{P}\left(X;\beta_0\right)$ has full column rank N_o and the rank of the augmented matrix $\left[\mathcal{P}\left(X;\beta_0\right), Q\left(X\right)\right]$ is N_o for almost all $X \in \mathcal{X}$ only at the true parameter $\beta_0 \in \mathcal{B}$.*

Theorem 3.1. *Suppose Assumptions 2.1–2.2, and 3.1 hold. Then the structural parameters $\left(\beta_0, \rho_{O|X}\right)$ are identified on the parameter space $\beta_0 \in \mathcal{B}$ and $\rho_{O|X} \in \mathcal{P}_O$.*

Proof. Appendix B.1. □

Remark 3.1. Under the number of weak orderings on the set of players, the necessary order condition holds if $(L+1)^2 \geq 3$ for two players, $(L+1)^3 \geq 13$ for three players, and $(L+1)^4 \geq 75$ for four players. The condition allows a binary action set $\{a_0, a_1\}$ only for two-player games. The action set must include at least 3 actions for three and four-player games, 4 actions for five-player games, 5 actions for six and seven-player games, and 6 actions for eight-player games. Note that this computation still requires the unique equilibrium assumption, which is not realistic for orders with simultaneous moves.

Remark 3.2. Suppose a fully sequential game in which only one player makes a decision in each stage. Then multiple equilibria do not exist, and the order condition stated above is $(L+1)^N \geq N!$. The reduced number of possible orders allows a binary action set for two-player and three-player games, since $2^2 \geq 2!$ and $2^3 \geq 3!$. The action set should have at least three different actions for four-player to six-player games.

Example 3.1. (Two-player Entry Game) Suppose there are two players and the potential order of actions are given by $O=(1,2)'$ or $O=(2,1)'$. The simultaneous move $O=(1,1)'$ does not exist in this example; thereby, there are not multiple equilibria. The payoff structure of each player is

$$u_i\left(s,X,O,\epsilon_i\right)=\begin{cases}X_i'\beta_i-s_{-i}\delta_i-\epsilon_i & \text{if } s_i=1\\ 0 & \text{if } s_i=0.\end{cases}$$

Then if the order of actions is $O=(1,2)'$,

$$\begin{aligned}P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)&=\begin{cases}F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right) & \text{if } s_1=1\\ F_{\epsilon_2}\left(X_2'\beta_2\right) & \text{if } s_1=0\end{cases}\\ P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)&=F_{\epsilon_1}\left(X_1'\beta_1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right),\end{aligned}$$

and similarly if $O=(2,1)'$,

$$\begin{aligned}P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)&=\begin{cases}F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right) & \text{if } s_2=1\\ F_{\epsilon_1}\left(X_1'\beta_1\right) & \text{if } s_2=0\end{cases}\\ P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)&=F_{\epsilon_2}\left(X_2'\beta_2-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right).\end{aligned}$$

The resulting probability distribution of $s=(s_1,s_2)'$ conditional on the possible orders can be summarized as below.

$$\begin{aligned}P\left(s=(1,1)'|X,O=(1,2)'\right)&=F_{\epsilon_1}\left(X_1'\beta_1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right)F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\\ P\left(s=(1,0)'|X,O=(1,2)'\right)&=F_{\epsilon_1}\left(X_1'\beta_1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right)\left(1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\right)\\ P\left(s=(0,1)'|X,O=(1,2)'\right)&=\left(1-F_{\epsilon_1}\left(X_1'\beta_1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right)\right)F_{\epsilon_2}\left(X_2'\beta_2\right)\\ P\left(s=(0,0)'|X,O=(1,2)'\right)&=\left(1-F_{\epsilon_1}\left(X_1'\beta_1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right)\right)\left(1-F_{\epsilon_2}\left(X_2'\beta_2\right)\right),\end{aligned}$$

and

$$\begin{aligned}P\left(s=(1,1)'|X,O=(2,1)'\right)&=F_{\epsilon_2}\left(X_2'\beta_2-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right)F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\\ P\left(s=(1,0)'|X,O=(2,1)'\right)&=\left(1-F_{\epsilon_2}\left(X_2'\beta_2-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right)\right)F_{\epsilon_1}\left(X_1'\beta_1\right)\\ P\left(s=(0,1)'|X,O=(2,1)'\right)&=F_{\epsilon_2}\left(X_2'\beta_2-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right)\left(1-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\right)\\ P\left(s=(0,0)'|X,O=(2,1)'\right)&=\left(1-F_{\epsilon_2}\left(X_2'\beta_2-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right)\right)\left(1-F_{\epsilon_1}\left(X_1'\beta_1\right)\right).\end{aligned}$$

The equilibrium selection mechanism is not specified since there are not multiple equilibria. For example,

$$
\begin{aligned}
P\left(s=(1,1)'\,|X\right) =& P\left(O=(1,2)'\,|X\right) F_{\epsilon_1}\left(X_1'\beta_1 - F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right) F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right) \\
&+ P\left(O=(2,1)'\,|X\right) F_{\epsilon_2}\left(X_2'\beta_2 - F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right) F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right), \quad (6)
\end{aligned}
$$

and the payoff function parameters $\beta_0 = (\beta_1, \beta_2, \delta_1, \delta_2)$ and the probability distribution function of orders $P\left(O=(1,2)'\,|X\right)$, $P\left(O=(2,1)'\,|X\right) = 1 - P\left(O=(1,2)'\,|X\right)$ are parameter of interests in the model. The model specification satisfies the necessary order condition of identification because there are three conditional moments $P\left(s=(1,1)'\,|X\right)$, $P\left(s=(1,0)'\,|X\right)$, $P\left(s=(0,1)'\,|X\right)$ and one nonparametric component $P\left(O=(1,2)'\,|X\right)$. The parameters $\rho(X) = P\left(O=(1,2)'\,|X\right)$ and β are identified unless there is another pair of $\left(\bar{\rho}(X), \bar{\beta}\right)$ that satisfies the equation (6) and the equations for $P\left(s=(1,0)'\,|X\right)$ and $P\left(s=(0,1)'\,|X\right)$ for X with probability one.

The primary source of identification is the dimension of the action set. The number of players affects not only the number of conditional moments but also the number of possible orders. If the variation of the action set is not enough, I suggest an exclusion restriction that may help identify the nonparametric order functions without relying on the variation in actions.

Assumption 3.2. *(Exclusion Restriction) There exists a strict subset X_s of X such that $\rho_{O|X}(\cdot|X) = \rho_{O|X}(\cdot|X_s)$.*

The Assumption 3.2 implies the existence of instrumental variables $X_v = X \backslash X_s$. The variation of X_v does not affect the order distribution function $\rho_{O|X}$ but the conditional choice probability $P(s=\alpha|X)$ varies with X_v. Then a sufficient variation of X_v contributes to satisfy the order condition of identification. Suppose X_v has a discrete support $\left\{x_v^1, \ldots, x_v^{N_v}\right\}$ and define a matrix $\mathcal{P}(X_s;\beta_0) \equiv \left[\mathcal{P}\left((X_s, x_v^1);\beta_0\right)', \ldots, \mathcal{P}\left(\left(X_s, x_v^{N_v}\right);\beta_0\right)'\right]'$. The parameter β_0 and $\rho_{O|X}$ are separately identified from Theorem 3.1.

Corollary 3.1. *Suppose Assumptions 2.1–2.2, and 3.2 hold. If $\mathcal{P}(X_s;\beta_0)$ has full column rank N_o and rank $\left(\left[\mathcal{P}(X_s;\beta_0), Q(X)\right]\right) = N_o$ for almost all X_s only at the true parameter $\beta_0 \in \mathcal{B}$, then the structural parameters $\left(\beta_0, \rho_{O|X}\right)$ are identified on the parameter space $\beta_0 \in \mathcal{B}$ and $\rho_{O|X} \in \mathcal{P}_O$.*

Proof. Appendix B.2. □

Remark 3.3. Another simple but not a practical setup is to assume that $\rho_{O|X}$ is fully known to the researcher. Then the structural parameters are reduced to $\left\{\beta_0, \lambda_{\tau|X,O}\right\}$ so that only equilibrium types are considered. Under the additional assumption of identification at infinity, the only parameter of interest is β_0. However, there are two concerns in practice. First, the order of actions is usually not provided in many sources of data. Second, the estimated payoff function parameter can be biased if the order of actions is misspecified. I report the inconsistency problem of misspecification and provide a numerical evidence in Section 5.1.

3.2 Identification with Multiple Equilibria

In this section, I derive identification of structural parameters in a general setting with multiple equilibria. Multiple equilibria may appear if the order of actions contains any simultaneous moves. For example, $O=(1,1)'$ in a two-player game and $O=(1,1,2)'$ or $O=(2,1,1)'$ in a three-player game have potential multiple equilibria. Denote $P(\kappa=k|X)=\lambda_{\tau|X,O}(\tau|X,O)\,\rho_{O|X}(O|X)$ where the index κ contains both the equilibrium type τ and the order of actions O. For example, $\kappa=1$ if $\tau=\tau_1$ and $O=O_1$ and $\kappa=2$ if $\tau=\tau_2$ and $O=O_1$. Since the number of PBNEs and the possible order of actions are finite, the values of κ can be arranged by finite integers. $\lambda_{\tau|X,O}$ and $\rho_{O|X}$ are separately identified after the distribution of product unobservables $P(\kappa=k|X)$ is identified.

With a slight abuse of notation, I denote $P^{(k)}\left(s_i=\alpha_i|\mathcal{J}_{o_k(i)};\beta_0\right)$ by the equilibrium probability of player i to choose an action α_i when $\kappa=k$. Rewrite the conditional choice probabilities by

$$P(s=\alpha|X)=\sum_{k=1}^{N_\kappa}\prod_{i=1}^{N}P^{(k)}\left(s_i=\alpha_i|\mathcal{J}_{o_k(i)};\beta_0\right)P(\kappa=k|X), \tag{7}$$

as a generalization of the equation (5). $N_\kappa=\sum_{l=1}^{N_o}B(X,O_l;\beta_0)$ is the total number of possible κs as a combination of an equilibrium type and an order of actions.

Define a matrix of equilibrium probabilities

$$\mathcal{P}'(X;\beta_0)\equiv\left[P^{(1)}(X,O_1;\beta_0),\ldots,P^{(N_\kappa)}(X,O_{N_\kappa};\beta_0)\right]\in\mathfrak{M}_{(L+1)^N\times N_\kappa},$$

where $P^{(k)}(X,O;\beta_0)=\left(P^{(k)}\left(s=\alpha^1|X,O;\beta_0\right),\ldots,P^{(k)}\left(s=\alpha^{(L+1)^N-1}|X,O;\beta_0\right),1\right)'\in\mathbb{R}^{(L+1)^N}$. Then a similar rank condition as Assumption 3.1 works for identification.

Theorem 3.2. *Suppose Assumptions 2.1–2.2 hold. If $\mathcal{P}'(X;\beta_0)$ has full column rank N_κ and rank $([\mathcal{P}'(X;\beta_0),Q(X)])=N_\kappa$ for almost all $X\in\mathcal{X}$ only at the true parameter $\beta_0\in\mathcal{B}$, then the structural parameters of the sequential game $\left\{\beta_0,\lambda_{\tau|X,O},\rho_{O|X}\right\}$ are identified.*

Proof. Appendix B.3. □

The necessary order condition of identification is $(L+1)^N\geq N_\kappa$. The condition implies that the number of conditional choice probabilities is at least as many as the number of unobserved types. The order condition may not hold if there are too many possible orders or equilibrium types. The additional exclusion restriction (Assumption 3.2) or the assumption of degenerate equilibrium selection rule can be used to relax the order condition.

First, $P(\kappa=k|X)=\lambda_{\tau|X,O}(\tau|X,O)\,\rho_{O|X}(O|X_s)$ under Assumption 3.2. Then for $N!$ fully sequential orders, $P(\kappa=k|X)=\rho_{O|X}(O|X_s)$ and the variation of X_v does not influence $P(\kappa=k|X)$. The other $N_o-N!$ orders may have multiple equilibria. In this case, the order condition is relaxed by $(L+1)^N\geq N_\kappa-N!$. Second, the degenerate equilibrium selection assumes $\lambda_{\tau|X,O}(\tau|X,O)=1$ for a specific $\tau\in\left\{\tau_{X,O,1},\ldots,\tau_{X,O,B(X,O;\beta_0)}\right\}$. Then the model is identified if there is a unique $\mathcal{P}(X;\beta_0)$ to satisfy the rank condition for almost all $X\in\mathcal{X}$ out of $\prod_{l=1}^{N_o}B(X,O;\beta_0)$ possible equilibrium probability combinations. The order condition is $(L+1)^N \geq N_o$, the same as the one without multiple equilibria.

Example 3.2. (Two-player Entry Game, continued) Suppose two players and the potential order of actions are given by $O=(1,2)'$, $O=(2,1)'$, and $O=(1,1)'$. The structure of the game follows a sequential game if $O=(1,2)'$ or $O=(2,1)'$, while in some markets two player simultaneously move. The payoff structure of each player is

$$u_i\left(s, X, O, \epsilon_i\right)=\begin{cases}X_i'\beta_i - s_{-i}\delta_i - \epsilon_i & \text{if } s_i=1\\ 0 & \text{if } s_i=0.\end{cases}$$

Then if $O=(1,1)'$,

$$\begin{aligned}P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right) &= F_{\epsilon_1}\left(X_1'\beta_1 - P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\delta_1\right)\\ P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right) &= F_{\epsilon_2}\left(X_2'\beta_2 - P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)\delta_2\right),\end{aligned} \tag{8}$$

where $\mathcal{J}_{o(1)}$ is equivalent to $\mathcal{J}_{o(2)}$. By definition of the PBNE, the conditional choice probabilities $P\left(s_1=1|\mathcal{J}_{o(1)}\right)$ and $P\left(s_2=1|\mathcal{J}_{o(2)}\right)$ solve the equation (8) under the assumption that ϵ_1 and ϵ_2 are independent conditional on $\mathcal{J}_{o(i)}$. The resulting probability distribution of $s=(s_1,s_2)'$ can be summarized as below.

$$\begin{aligned}P\left(s=(1,1)'|X, O=(1,1)'\right) &= P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\\ P\left(s=(1,0)'|X, O=(1,1)'\right) &= P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)\left(1-P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\right)\\ P\left(s=(0,1)'|X, O=(1,1)'\right) &= \left(1-P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)\right)P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\\ P\left(s=(0,0)'|X, O=(1,1)'\right) &= \left(1-P\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)\right)\left(1-P\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\right).\end{aligned}$$

The solution of the equation (8) may not be unique. Suppose there are three PBNEs at the true parameter values β_0 and denote equilibrium probabilities by

$$\left\{\left(P^{(\tau)}\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right), P^{(\tau)}\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\right)\right\}_{\tau=1}^{3}.$$

Then for a joint action $s=(1,1)'$,

$$\begin{aligned}P\left(s=(1,1)'|X\right) =& P\left(O=(1,2)'|X\right)F_{\epsilon_1}\left(X_1'\beta_1 - F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right)F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\\ &+P\left(O=(2,1)'|X\right)F_{\epsilon_2}\left(X_2'\beta_2 - F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right)F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\\ &+P\left(O=(1,1)'|X\right)P\left(s_1=1, s_2=1|X, O=(1,1)'\right),\end{aligned}$$

where

$$
\begin{aligned}
P\left(s=(1,1)'\,|X,O=(1,1)'\right) &= \sum_{\tau=1}^{3}\lambda\left(\tau|X,O=(1,1)'\right)P^{(\tau)}\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)P^{(\tau)}\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\\
\sum_{\tau=1}^{3}\lambda\left(\tau|X,O=(1,1)'\right) &= 1\\
\lambda\left(1|X,O=(1,2)'\right) &= \lambda\left(1|X,O=(2,1)'\right)=1,
\end{aligned}
$$

and the parameter of interests in the model are (1) the payoff function parameters $\beta_0=(\beta_1,\beta_2,\delta_1,\delta_2)$, (2) the order of actions $P\left(O=(1,2)'\,|X\right)$, $P\left(O=(2,1)'\,|X\right)$, $P\left(O=(1,1)'\,|X\right)$, and (3) the equilibrium selection mechanism $\lambda\left(\tau|X,O=(1,1)'\right)$ for $\tau=1,2,3$. Note that $\lambda=1$ for sequential actions $O=(1,2)'$ and $(2,1)'$ thanks to the unique PBNE.

Define a new index κ to describe the product of $P\left(O|X\right)$ and $\lambda\left(\tau|X,O\right)$. I show that the identification of the product function $P\left(O|X\right)\lambda\left(\tau|X,O\right)$ is sufficient to identify $P\left(O|X\right)$ and $\lambda\left(\tau|X,O\right)$ separately. In this example,

$$
\begin{aligned}
P\left(\kappa=1|X\right) &= P\left(O=(1,2)'\,|X\right)\lambda\left(1|X,O=(1,2)'\right)\\
P\left(\kappa=2|X\right) &= P\left(O=(2,1)'\,|X\right)\lambda\left(1|X,O=(2,1)'\right)\\
P\left(\kappa=3|X\right) &= P\left(O=(1,1)'\,|X\right)\lambda\left(1|X,O=(1,1)'\right)\\
P\left(\kappa=4|X\right) &= P\left(O=(1,1)'\,|X\right)\lambda\left(2|X,O=(1,1)'\right)\\
P\left(\kappa=5|X\right) &= P\left(O=(1,1)'\,|X\right)\lambda\left(3|X,O=(1,1)'\right),
\end{aligned}
$$

and $\sum_{l=1}^{5}P\left(\kappa=l|X\right)=1$. If each of $P\left(\kappa=l|X\right)$ is identified,

$$
\begin{aligned}
P\left(O=(1,2)'\,|X\right) &= P\left(\kappa=1|X\right)\\
P\left(O=(2,1)'\,|X\right) &= P\left(\kappa=2|X\right)\\
P\left(O=(1,1)'\,|X\right) &= P\left(\kappa=3|X\right)+P\left(\kappa=4|X\right)+P\left(\kappa=5|X\right)\\
\lambda\left(1|X,O=(1,1)'\right) &= \frac{P\left(\kappa=3|X\right)}{P\left(\kappa=3|X\right)+P\left(\kappa=4|X\right)+P\left(\kappa=5|X\right)}\\
\lambda\left(2|X,O=(1,1)'\right) &= \frac{P\left(\kappa=4|X\right)}{P\left(\kappa=3|X\right)+P\left(\kappa=4|X\right)+P\left(\kappa=5|X\right)}\\
\lambda\left(3|X,O=(1,1)'\right) &= \frac{P\left(\kappa=5|X\right)}{P\left(\kappa=3|X\right)+P\left(\kappa=4|X\right)+P\left(\kappa=5|X\right)}.
\end{aligned}
$$

The identification of $P\left(\kappa=l|X\right)$ follows the similar steps as the previous example. The conditional choice probability $P\left(s=(1,1)'\,|X\right)$ is

$$\begin{aligned}P\left(s=(1,1)'\,|X\right)=&P\left(\kappa=1|X\right)F_{\epsilon_1}\left(X_1'\beta_1-F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\delta_1\right)F_{\epsilon_2}\left(X_2'\beta_2-\delta_2\right)\\&+P\left(\kappa=2|X\right)F_{\epsilon_2}\left(X_2'\beta_2-F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\delta_2\right)F_{\epsilon_1}\left(X_1'\beta_1-\delta_1\right)\\&+P\left(\kappa=3|X\right)P^{(1)}\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)P^{(1)}\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\\&+P\left(\kappa=4|X\right)P^{(2)}\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)P^{(2)}\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right)\\&+P\left(\kappa=5|X\right)P^{(3)}\left(s_1=1|\mathcal{J}_{o(1)};\beta_0\right)P^{(3)}\left(s_2=1|\mathcal{J}_{o(2)};\beta_0\right).\end{aligned}$$

In a binary game with two players, the conditional distribution function of κ on X is not nonparametrically identified since there are three conditional moments for three joint actions $P\left(s=(1,1)'\,|X\right)$, $P\left(s=(1,0)'\,|X\right)$, and $P\left(s=(0,1)'\,|X\right)$ but four unknown functions $P\left(\kappa=l|X\right)$ for $l=1,\ldots,4$. The necessary order condition does not hold under multiple equilibria.

In this example, the exclusion restriction of Assumption 3.2 still works for point identification of structural parameters. There is a single equilibrium under $O=(1,2)'$ and $O=(2,1)'$ so that $P\left(\kappa=1|X\right)=P\left(\kappa=1|X_s\right)$ and $P\left(\kappa=2|X\right)=P\left(\kappa=2|X_s\right)$. $P\left(\kappa=3|X\right)$, $P\left(\kappa=4|X\right)$, and $P\left(\kappa=5|X\right)$ are still functions of X_v because the number of PBNEs depends on the payoff-relevant covariates X. Since $P\left(\kappa=5|X\right)=1-\sum_{l=1}^{4}P\left(\kappa=l|X\right)$, there are two nonparametric components $P\left(\kappa=3|X\right)$ and $P\left(\kappa=4|X\right)$ conditional on X_s. The additional variation of X_v is a source of identification because a value of X_v provides $P\left(s=\alpha|X\right)$ for four joint actions but $P\left(\kappa=l|X\right)$ for at most three equilibria. The necessary order condition is satisfied since the number of conditional moments exceeds the number of unknown parameters.

The degenerate equilibrium selection rule applies that $\lambda\left(\tau|X,O=(1,1)'\right)=1$ for one of $\tau\in\{1,2,3\}$ and $\lambda\left(\tau|X,O=(1,1)'\right)=0$ for others. Define

$$\mathcal{P}^{(\tau)}\left(X;\beta_0\right)\equiv\left[P\left(X,O=(1,2)';\beta_0\right),P\left(X,O=(2,1)';\beta_0\right),P^{(\tau)}\left(X,O=(1,1)';\beta_0\right)\right],$$

for $\tau\in\{1,2,3\}$. The structural parameters are identified if $\mathcal{P}^{(\tau)}\left(X;\beta_0\right)$ is full rank for almost all $X\in\mathcal{X}$ only at $\beta_0\in\mathcal{B}$ and the selected τ. The order condition is the same as the one without multiple equilibria because only one equilibrium type is assigned to each order of actions.

4 Estimation and Inference

In this section, I suggest an estimator for the structural parameters based on Theorem 3.2 and figure out asymptotic properties of the estimator. I start by the conditional moment restrictions of the model and apply the SMD estimator of Ai and Chen (2003). The asymptotic properties of the SMD estimator are verified based on Ai and Chen (2003), Newey and Powell (2003), and Chen and Pouzo (2015). The smoothness assumptions for the equilibrium selection rule and the order of actions provide sufficient conditions for consistency and asymptotic normality of the proposed estimator.

4.1 Estimation of Structural Parameters

Suppose that a researcher observes n independent games. The actions $s^m = (s_{1,m}, \ldots, s_{N,m})'$ and observable characteristics X^m for each game $m = 1, \ldots, n$ are available. The representation in equation (7) describes the relation of conditional choice probabilities and structural parameters. For each joint action α, there is a conditional moment regarding structural parameters $\beta \in \mathcal{B}$ and $h_k(X) \in \mathcal{H}_k$ such that $\sum_{k=1}^{N_\kappa} h_k(X) = 1$:

$$E\left[1\{s^m = \alpha\} - \sum_{k=1}^{N_\kappa} \prod_{i=1}^{N} P^{(k)}\left(s_{i,m} = \alpha_i | \mathcal{J}^m_{o_k(i)}; \beta_0\right) h_k(X^m) \middle| X^m\right] = 0. \tag{9}$$

Hereafter I denote $h_{0,k}(X) \equiv P(\kappa = k|X)$.

There are in total $(L+1)^N$ conditional moments, including $E\left[1 - \sum_{k=1}^{N_\kappa} h_k(X^m) | X^m\right] = 0$. Denote the conditional moments by $E[\ell(Z^m, \theta) | X^m] = 0$, where $Z^m = \left(s^{m\prime}, X^{m\prime}\right)'$ and $\theta = (\beta, h_1, \ldots, h_{N_\kappa}) \in \mathcal{B} \times \mathcal{H}_1 \times \cdots \times \mathcal{H}_{N_\kappa}$. $\ell(Z^m, \theta) = \left(\ell_1(Z^m, \theta), \ldots, \ell_{(L+1)^N}(Z^m, \theta)\right)'$ with

$$\ell_j(Z^m, \theta) = 1\left\{s^m = \alpha^j\right\} - \sum_{k=1}^{N_\kappa} \prod_{i=1}^{N} P^{(k)}\left(s_{i,m} = \alpha_i^j | \mathcal{J}^m_{o_k(i)}; \beta\right) h_k(X^m),$$

for $j = 1, \ldots, (L+1)^N - 1$ and $\ell_{(L+1)^N}(Z^m, \theta) = 1 - \sum_{k=1}^{N_\kappa} h_k(X^m)$.

Assume that the number of unknown types N_κ is known to the researcher. Note that N_κ depends on the payoff relevant parameter β in general, while the upper bound of N_κ is still available. The additional assumptions also provide information on N_κ. $N_\kappa = N_o$ under the unique equilibrium or degenerate equilibrium selection assumption, and $N_\kappa = N!$ under the completely sequential order of actions.

I use a linear sieve estimator suggested by Ai and Chen (2003) and follow-up nonparametric sieve estimation literature to estimate nonparametric functions $h_1, \ldots, h_{N_\kappa}$. Define $\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_{N_\kappa}$ and a corresponding sieve space $\mathcal{H}_n = \mathcal{H}_{1,n} \times \cdots \times \mathcal{H}_{N_\kappa,n}$. Consider a sequence of known basis functions $\{p_j(X) | j = 1, 2, \ldots\}$, for example, tensor-product B-splines of order of order $\gamma' > d_x/2 + 1$. Denote $p^{J_n}(X) = (p_1(X), \ldots, p_{J_n}(X))'$ and $P^{J_n} = \left(p^{J_n}(X^1), \ldots, p^{J_n}(X^n)\right)'$. Let $\Theta = \mathcal{B} \times \mathcal{H}$ be the parameter space and $\theta_0 = (\beta_0', h_0')' = (\beta_0', h_{0,1}, \ldots, h_{0,N_\kappa})'$ be the true parameter vector. Then the SMD estimator $\hat{\theta}_n = \left(\hat{\beta}_n', \hat{h}_n'\right)' = \left(\hat{\beta}_n', \hat{h}_{1,n}, \ldots, \hat{h}_{N_\kappa,n}\right)'$ solves

$$\min_{\theta \in \mathcal{B} \times \mathcal{H}_n} \left(\sum_{m=1}^{n} \ell(Z^m, \theta) \otimes p^{J_n}(X^m)\right)' \left(I \otimes \left(P^{J_n\prime} P^{J_n}\right)^{-1}\right) \left(\sum_{m=1}^{n} \ell(Z^m, \theta) \otimes p^{J_n}(X^m)\right), \tag{10}$$

where I is the $(L+1)^N \times (L+1)^N$ dimensional identity weight matrix.

I provide assumptions for consistency and asymptotic normality of $\hat{\theta}_n$ below.

Assumption 4.1. *(Data Generating Process)*

1. *The data* $s^m = (s_{1,m}, \ldots, s_{N,m})$ *and* $X^m = \left(X'_{1,m}, \ldots, X'_{N,m}\right)'$ *for* $m = 1, \ldots, n$ *are i.i.d..*
2. $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ *is compact with nonempty interior, and the probability density function of* X *is bounded and bounded away from zero.*
3. *The smallest and the largest eigenvalues of* $E\left[p^{J_n}(X)\, p^{J_n}(X)'\right]$ *are bounded and bounded away from zero for all* J_n.

The first set of assumptions is common in the nonparametric sieve estimation. The next assumption imposes smoothness to the nonparametric components of the structural parameters. I assume that every nonparametric function $P(\kappa = k|X)$ is a smooth function on $\mathcal{X}$ so that linear sieves, including power series and splines, can approximate $P(\kappa = k|X)$. Assumption 4.2 implies that each CCP and $P(\kappa = k|X)$ has $\lfloor\gamma\rfloor$ bounded derivatives, where $\lfloor\gamma\rfloor$ is the largest integer less than γ.

Assumption 4.2. *(Smoothness Condition) The CCPs* $P(s = \alpha|X)$ *for* $\alpha \in \mathcal{A}^N$ *and the product functions* $h_{0,k}(X) = P(\kappa = k|X)$ *for* $k = 1, \ldots, N_\kappa$ *belong to a Hölder space with smoothness parameter* $\gamma > d_x/2$.

The next assumption is about continuity of equilibrium probability functions. The assumption holds if Assumption 2.1 holds and the player's payoff function is a continuous function of β. Examples 3.1 and 3.2 show that the equilibrium probability is continuous for a linear payoff function.

Assumption 4.3. *(Equilibrium Probability) The equilibrium probability* $P^{(k)}\left(s = \alpha|\mathcal{J}_{o_k(i)}; \beta\right)$ *for* $k = 1, \ldots, N_\kappa$ *is continuously differentiable at* $\beta \in \mathcal{B}$ *and* $X \in \mathcal{X}$ *almost surely.* $\mathcal{B}$ *is a compact and convex set in Euclidean space.*

The last assumption for consistency is the choice of smoothing parameters. Define K_n by the dimension of the sieve space $\mathcal{H}_{k,n}$ for $k = 1, \ldots, N_\kappa$. The SMD estimator works properly when the number of unconditional moments generated by a smoothing parameter J_n is at least as many as the number of unknowns. The unknown parameters include a d_β-dimensional β and N_κ probability mass functions approximated by K_n-dimensional sieve basis.

Assumption 4.4. *(Smoothing Parameters)* $(L+1)^N J_n \geq d_\beta + N_\kappa K_n$, $K_n \to \infty$ *and* $J_n/n \to 0$ *as* $n \to \infty$.

The next set of assumptions are additional conditions to derive asymptotic normality of the proposed SMD estimator. Ai and Chen (2003) provided conditions for asymptotic normality of $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$, focusing on the parametric component of the parameters. The result is applied to inference on the payoff relevant parameter β. I also verify regularity conditions in Chen and Pouzo (2015) to show inference on functionals of other structural parameters $\lambda_{\tau|X,O}$ and $\rho_{O|X}$. For example, I derive asymptotic normality of $\sqrt{n}\left(\hat{h}_{k,n}(x) - h_{0,k}(x)\right)$ for some $k \in \{1, \ldots, N_\kappa\}$ and $x \in \mathcal{X}$.

Assumption 4.5. *(Conditions for Asymptotic Normality of $\hat{\beta}_n$)*

1. *θ_0 is an interior point of Θ.*

2. *The equilibrium probability $P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)$ for $k=1,\dots,N_\kappa$ is twice continuously differentiable at $\beta\in\mathcal{B}$ and $X\in\mathcal{X}$ almost surely.*

3. *$J_n^{-\gamma/d_x}=o\left(n^{-1/4}\right)$, $K_n^{-\gamma/d_x}=o\left(n^{-1/4}\right)$, and $J_nK_n\log n=o\left(n^{1/2}\right)$.*

The first part of Assumption 4.5 is a standard non-boundary condition. The second to the last parts are additional smoothness assumptions to derive asymptotic normality of $\sqrt{n}\left(\hat{\beta}_n-\beta_0\right)$.

4.2 Asymptotic Properties of the Estimator

This subsection derives asymptotic properties of the proposed SMD estimator, based on the assumptions provided in the previous subsection. First, I verify the consistency of the SMD estimators $\hat{\beta}_n$ and $\hat{h}_n$. Second, I derive asymptotic normality of the SMD estimators of regular and irregular functionals. Particularly the asymptotic distributions of $\sqrt{n}\left(\hat{\beta}_n-\beta_0\right)$ and the sieve t statistic $\sqrt{n}\,\|\hat{\nu}_n^*\|_{sd,h}^{-1}\left(\hat{h}_{k,n}(x)-h_{0,k}(x)\right)$ for some $x\in\mathcal{X}$ are discussed. The closed form of the sieve variance $\|\hat{\nu}_n^*\|_{sd,h}^2$ is provided in the context of Ai and Chen (2003) and Chen and Pouzo (2009, 2015).

The first main result is consistency of $\hat{\theta}_n$ to θ_0. Define a metric $\|\cdot\|_s$ by

$$\|\theta\|_s=\|\beta\|_E+\max_{k=1,\dots,N_\kappa}\sup_{x\in\mathcal{X}}|h_k(x)|,$$

where $\|\cdot\|_E$ is the Euclidean metric. The following theorem presents consistency of $\hat{\theta}_n$ to θ_0 under $\|\cdot\|_s$.

Theorem 4.1. *(Consistency) Suppose that θ_0 is identified by Theorems 3.1 and 3.2. Under Assumptions 4.1–4.4, $\left\|\hat{\theta}_n-\theta_0\right\|_s=o_p(1)$.*

Proof. Appendix C.1. □

The asymptotic distribution of $\hat{\theta}_n$ is derived with additional assumptions introduced in the previous subsection. Define

$$D_n=E\left[\Delta_n(X)'\Delta_n(X)\right]$$
$$\Psi_n=E\left[\Delta_n(X)'\ell(Z,\theta_0)\ell(Z,\theta_0)'\Delta_n(X)\right],$$

where $\Delta_n(X)$ is a $(L+1)^N\times(d_\beta+N_\kappa K_n)$ dimensional sieve gradient matrix,

$$\Delta_n(X)\equiv\left[\Delta_{\beta_0}(X),\Delta_{h_0,n}(X)\right]$$
$$\Delta_{\beta_0}(X)\equiv\sum_{k=1}^{N_\kappa}\frac{\partial}{\partial\beta'}P^{(k)}(X,O;\beta_0)\,h_{0,k}(X)$$
$$\Delta_{h_0,n}(X)\equiv\left[P^{(1)}(X,O;\beta_0)\otimes p^{K_n'}(X),\dots,P^{(N_\kappa)}(X,O;\beta_0)\otimes p^{K_n'}(X)\right].$$

Assumption 4.6. *(Asymptotic Variance) The components of the asymptotic variance of* $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$ *are well-defined.*

1. D_n *is a positive definite matrix.*

2. $\Sigma_0(X) = E\left[\ell(Z, \theta_0)\,\ell(Z, \theta_0)'\,|X\right]$ *is positive definite and the smallest and the largest eigenvalues of* $\Sigma_0(X)$ *are bounded and bounded away from zero uniformly in* $X \in \mathcal{X}$.

Based on the provided assumptions above, I propose the following asymptotic normality results. The details of the proof verify sufficient conditions of asymptotic normality in Ai and Chen (2003) and Chen and Pouzo (2015). The sieve asymptotic variances of the estimators $\hat{\beta}_n$ and $\hat{h}_{k,n}$ have closed form expressions by applying the functional Delta method to the variance-covariance matrix $D_n^{-1}\Psi_n D_n^{-1}$. Denote that I_k is a $k \times k$ dimensional identity matrix and $\mathbb{O}_{j\times k}$ is a $j \times k$ dimensional zero matrix.

Theorem 4.2. *(Asymptotic Normality of* $\hat{\beta}_n$*) Suppose that assumptions for Theorem 4.1 are satisfied. Under the additional Assumptions 4.5–4.6,*

$$\sqrt{n}\,\|\nu_n^*\|_{sd,\beta}^{-1}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{d} N\left(0, I_{d_\beta}\right),$$

where

$$\begin{aligned}\|\nu_n^*\|_{sd,\beta}^2 &= G_\beta' D_n^{-1}\Psi_n D_n^{-1} G_\beta \\ G_\beta &\equiv \left[I_{d_\beta}, \mathbb{O}_{d_\beta\times(N_\kappa-1)K_n}\right]'.\end{aligned}$$

Proof. Appendix C.2. □

Note that the asymptotic variance the SMD estimator $\hat{\beta}_n$ is the limiting value of $\|\nu_n^*\|_{sd,\beta}^2$ as $n \to \infty$. I provide a closed form of the asymptotic variance in Appendix. The next theorem states the limiting distribution of the functional of SMD estimator $\hat{h}_n$. Suppose that $h_{0,k}(x)$ for some $x \in \mathcal{X}$ is the parameter of interest. Chen and Pouzo (2015) showed a closed form expression for the sieve variance $\|\nu_n^*\|_{sd,h}^2$, which is applied to inference on structural parameters $h_{0,k}(x)$.

Theorem 4.3. *(Asymptotic Normality of* $\hat{h}_{k,n}$*) Suppose that assumptions for Theorem 4.1 are satisfied. Under the additional Assumptions 4.5–4.6,*

$$\sqrt{n}\frac{\left(\hat{h}_{k,n}(x) - h_{0,k}(x)\right)}{\|\nu_n^*\|_{sd,h}} \xrightarrow{d} N(0,1),$$

for some $x \in \mathcal{X}$ *and* $k \in \{1, \ldots, N_\kappa\}$*, where*

$$\begin{aligned}\|\nu_n^*\|_{sd,h}^2 &= G_{h_k}' D_n^{-1}\Psi_n D_n^{-1} G_{h_k} \\ G_{h_k} &= \left[\mathbb{O}_{1\times\left(d_\beta+(k-1)K_n\right)}, p^{K_n'}(x), \mathbb{O}_{1\times(N_\kappa-k)K_n}\right]'.\end{aligned}$$

Proof. Appendix C.3. $\square$

Theorems 4.2 and 4.3 respectively show asymptotic normality of the parametric and nonparametric components of the structural parameter. The application of asymptotic normality to empirical illustrations requires a consistent estimator for $\|\nu_n^*\|_{sd,\beta}^2$ and $\|\nu_n^*\|_{sd,h}^2$. Both sieve variances $\|\nu_n^*\|_{sd,\beta}^2$ and $\|\nu_n^*\|_{sd,h}^2$ share the same components D_n and Ψ_n and nonstochastic G_β and G_{h_k} so that I can propose a plug-in sieve variance estimator using consistent estimators of D_n and Ψ_n. Provided that $\ell(Z^m,\theta)$ is pointwise smooth by Assumptions 4.2 and 4.3, I derive asymptotic normality of the structural parameter with a consistent variance estimator.

Define the plug-in estimators $\hat{D}_n$ and $\hat{\Psi}_n$:

$$\hat{D}_n = \frac{1}{n}\sum_{m=1}^{n} \hat{\Delta}_n(X^m)' \hat{\Delta}_n(X^m)$$
$$\hat{\Psi}_n = \frac{1}{n}\sum_{m=1}^{n} \hat{\Delta}_n(X^m)' \hat{\ell}\left(Z^m,\hat{\theta}_n\right) \hat{\ell}\left(Z^m,\hat{\theta}_n\right)' \hat{\Delta}_n(X^m),$$

where

$$\hat{\Delta}_n(X^m) \equiv \left[\hat{\Delta}_{\beta_0}(X^m), \hat{\Delta}_{h_0,n}(X^m)\right]$$
$$\hat{\Delta}_{\beta_0}(X^m) \equiv \sum_{k=1}^{N_\kappa} \frac{\partial}{\partial\beta'} P^{(k)}\left(X^m,O;\hat{\beta}_n\right)\hat{h}_n(X^m)$$
$$\hat{\Delta}_{h_0,n}(X^m) \equiv \left[P^{(1)}\left(X^m,O;\hat{\beta}_n\right)\otimes p^{K_n'}(X^m),\ldots,P^{(N_\kappa)}\left(X^m,O;\hat{\beta}_n\right)\otimes p^{K_n'}(X^m)\right].$$

The equilibrium probabilities $P^{(k)}\left(X^m,O;\hat{\beta}_n\right)$ are evaluated by solving the equation (3) at X^m and $\hat{\beta}_n$, and a numerical derivative $\frac{\partial}{\partial\beta'}P^{(k)}\left(X^m,O;\hat{\beta}_n\right)$ is computed at $\hat{\beta}_n$.

Theorem 4.4. *(Consistent Variance Estimator) Suppose that assumptions for Theorem 4.2 and 4.3 are satisfied. Then*

$$\sqrt{n}\,\|\hat{\nu}_n^*\|_{sd,\beta}^{-1}\left(\hat{\beta}_n-\beta_0\right)\xrightarrow{d} N\left(0,I_{d_\beta}\right)$$
$$\sqrt{n}\frac{\left(\hat{h}_{k,n}(x)-h_{0,k}(x)\right)}{\|\hat{\nu}_n^*\|_{sd,h}}\xrightarrow{d} N(0,1),$$

for some $x\in\mathcal{X}$ and $k\in\{1,\ldots,N_\kappa\}$, where

$$\|\hat{\nu}_n^*\|_{sd,\beta}^2 = G_\beta'\hat{D}_n^{-1}\hat{\Psi}_n\hat{D}_n^{-1}G_\beta$$
$$\|\hat{\nu}_n^*\|_{sd,h}^2 = G_{h_k}'\hat{D}_n^{-1}\hat{\Psi}_n\hat{D}_n^{-1}G_{h_k}.$$

Proof. Appendix C.4. $\square$

Note that the results in Theorems 4.2–4.4 can be applied to inference on other functionals of interest. Chen and Pouzo (2015) not only discussed inferences on linear functionals of the structural parameters but also nonlinear ones. The asymptotic distributions for other linear and nonlinear functionals, including $\rho_{O|X}(O|X)$ and $\lambda_{\tau|X,O}(\tau|X,O)$, are direct applications of the established theoretical results.

5 Simulation and Empirical Application

5.1 Simulation Results

This section presents Monte Carlo simulations to evaluate the performance of estimation and inference results under the finite sample. I also numerically illustrate the effect of misspecified order to the estimation results. The simulation not only verifies that the proposed SMD estimator is valid, but also indicates that the estimation of an empirical game without considering the order of actions can be misleading.

The simulation experiments consist of games with linear payoff functions.

$$u_1 = \begin{cases} X_1\beta_1 + \Delta_1 S_2 - \epsilon_1 & \text{if } S_1 = 1 \\ 0 & \text{if } S_1 = 0 \end{cases}$$

$$u_2 = \begin{cases} X_2\beta_2 + \Delta_2 S_1 - \epsilon_2 & \text{if } S_2 = 1 \\ 0 & \text{if } S_2 = 0, \end{cases}$$

where ϵ_i follows a Type 1 Extreme distribution for $i = 1, 2$. The parameter values are $(\beta_1, \beta_2) = (3, 2)$ and $(\Delta_1, \Delta_2) = (-6, -4)$. For simplicity, the components of the game have discrete supports: $X_i \in \{0.5, 1.0, 1.5, 2.0\}$, and follows a discrete uniform distribution on each support point. The game has only two players so that there are at most three PBNE under the simultaneous move. When there are multiple equilibria, the probability assigned for each equilibrium is fixed by 1/3.

In the first experiment, we consider a two-player game with two sequential actions $O_1 = (1, 2)'$ and $O_2 = (2, 1)'$. The probability of the order is given by $\rho_{O|X}(O_1|X) = (X_1 + X_2)/10$. For example, the player 1 is the first mover and the player 2 is the second mover with probability 0.2 for the market with $X_1 = 1$ and $X_2 = 1$. The possible orders are fully sequential so that there are not multiple equilibria. The number of simulations is 1000, and the sample size (the number of markets) varies by $n = 500$, 1000, 5000.

Table 1 summarizes the estimates and standard error of the SMD estimator. I provide the estimates of payoff parameters (β_1, β_2), (Δ_1, Δ_2), and some of the selected probability mass functions $P\left(O = (1, 2)' |X\right)$. There are 16 combinations of (X_1, X_2) since each X_i has four different values, and Table 1 only reports four values. The performance of the estimator for unreported outcome is similar as the reported results. The estimates are getting closer to the true parameter values and the corresponding standard errors are getting smaller as the number of games n increases.

In the second experiment, the set of possible orders consists of two sequential actions in the first experiment and the simultaneous move: $O_1 = (1, 2)'$, $O_2 = (2, 1)'$, and $O_3 = (1, 1)'$. $X_i \in \{1, 2, 3, 4\}$

	$n = 500$		$n = 1000$		$n = 5000$	
	Estimates	Std. Err.	Estimates	Std. Err.	Estimates	Std. Err.
β_1	3.0224	1.0244	2.9317	0.6819	2.9851	0.2848
β_2	3.2107	1.6446	2.5848	0.9277	2.1032	0.2727
Δ_1	-6.0939	1.9919	-5.8906	1.3065	-5.9763	0.5412
Δ_2	-6.4342	3.3420	-5.1746	1.8622	-4.2078	0.5303
$P\left(O=(1,2)' \mid X=(1,1)\right)$	0.2477	0.1061	0.2351	0.0810	0.2066	0.0422
$P\left(O=(1,2)' \mid X=(2,2)\right)$	0.4482	0.1593	0.4314	0.1241	0.4075	0.0567
$P\left(O=(1,2)' \mid X=(1,2)\right)$	0.3680	0.1992	0.3350	0.1384	0.3084	0.0507
$P\left(O=(1,2)' \mid X=(2,1)\right)$	0.3815	0.1980	0.3563	0.1597	0.3122	0.0775

Table 1: Estimation of the Incomplete Information Game Theoretic Model (Sequential Actions)

Note: n is sample size, and the true value of parameter is $(\beta_1, \beta_2) = (3,2)$, $(\Delta_1, \Delta_2) = (-6,-4)$, $P\left(O=(1,2)' \mid X\right) = (X_1+X_2)/10$. The number of simulations is 1000. The estimates are computed by the sample mean and the sample standard error of 1000 estimated values.

with a discrete uniform distribution. The probability of the order follows $\rho_{O|X}(O_1|X) = X_1/10$, $\rho_{O|X}(O_2|X) = X_2/10$, and $\rho_{O|X}(O_3|X) = 1-(X_1+X_2)/10$. There are three equilibria when $(X_1, X_2) = (1,1)$, and there is a unique equilibrium for other (X_1, X_2) values. The equilibrium selection mechanism assigns the same probability to each of the equilibrium when there are multiple equilibria: $\lambda_{\tau|X,O}(\tau|X=(1,1), O_2) = 1/3$ for every τ. I assume that the equilibrium selection probability is known to the econometrician in this example, because the number of joint actions should be as many as the number of equilibrium outcomes for identification. The number of games varies with $n = 500$, 1000, 5000.

The simulation results for the second experiment are presented in Table 2. The suggested SMD estimator works well under the correct order of actions and the multiple potential equilibria, and the performance of the estimator improves as the sample size n increases. The following empirical application in Section 5.2 uses the model specifications in the first and second experiments.

The next simulation experiments provide numerical evidence of estimation bias under the misspecified order of actions. I preserve the basic setup used in the second experiment and consider the situation as a completely simultaneous or sequential game. The order of actions depends on $\rho_{O|X}(O_1|X)$ and $\rho_{O|X}(O_2|X)$ so that $\rho_{O|X}(O_1|X) = \rho_{O|X}(O_2|X) = 0$ implies a simultaneous game and $\rho_{O|X}(O_1|X) = 1$ implies a sequential game. Assume that all games are simultaneous games, and estimate the model in two different ways. The first one is a correctly specified case: the data generating process follows a simultaneous game, and the model is estimated assuming a simultaneous game. The second one is a misspecified case: the DGP still follows a simultaneous game but apply the estimation method for sequential games. Similar experiments are conducted by assuming that all games are sequential games. The following table compares the estimates of payoff function parameters for these two cases.

Table 3 shows that the misspecified case does not capture the true parameter values $(\beta_1, \beta_2) =$

	$n = 500$		$n = 1000$		$n = 5000$	
	Estimates	Std. Err.	Estimates	Std. Err.	Estimates	Std. Err.
β_1	4.0194	1.7968	3.4223	1.1530	3.2314	0.5008
β_2	2.8799	1.6547	2.3853	1.0219	2.1518	0.3570
Δ_1	-8.0242	3.8142	-6.8062	2.4815	-6.4407	1.0980
Δ_2	-5.8660	3.5192	-4.7713	2.2413	-4.3089	0.7824
$P\left(O = (1,2)' \mid X = (1,2)\right)$	0.1423	0.1742	0.1206	0.1187	0.0995	0.0293
$P\left(O = (1,2)' \mid X = (2,3)\right)$	0.3139	0.3148	0.2870	0.2987	0.2265	0.1763
$P\left(O = (2,1)' \mid X = (2,1)\right)$	0.1639	0.2082	0.1310	0.1344	0.1050	0.0413
$P\left(O = (2,1)' \mid X = (2,2)\right)$	0.2196	0.1604	0.2164	0.1394	0.1998	0.0487

Table 2: Estimation of the Incomplete Information Game Theoretic Model (Sequential Actions and Simultaneous Action)

Note: n is sample size, and the true value of parameter is $(\beta_1, \beta_2) = (3,2)$, $(\Delta_1, \Delta_2) = (-6,-4)$, $P\left(O = (1,2)' \mid X\right) = X_1/10$, $P\left(O = (2,1)' \mid X\right) = X_2/10$. The number of simulations is 1000. The estimates are computed by the sample mean and the sample standard error of 1000 estimated values.

$(3,2)$ and $(\Delta_1, \Delta_2) = (-6,-4)$ within a 95% confidence interval in spite of a sufficient sample size. For example, the strategic interaction parameter of player 1 is $\Delta_1 = -6$, but the estimate under the misspecified order is -3.07. The bias exists in other parameters of interest according to Table 3. The result is similar when all games are simultaneous games. The proposed SMD estimator provides estimates that are close to the true parameter values under the correct specification, but the misspecified order of actions leads to a significant asymptotic bias. The main finding from the Table 3 informs a potential risk of misspecified order of actions. If the game in the market is a sequential game but an econometrician applies the estimation methods for a simultaneous game, the estimates can be misleading and may derive not plausible results.

5.2 Empirical Application: Entry Game

This section presents an empirical application of the game-theoretic model, considering the sequential order of actions. I apply the model to an entry game between Walmart and Kmart using the dataset of Jia (2008).[5] The empirical model of Jia (2008) is a three-stage game that captures the pre-chain period competition among small retailers, the simultaneous entry decisions of Walmart and Kmart, and the following market adjustment. The simplified model I develop is an incomplete information entry game of Walmart and Kmart without considering local retailers. The parameters of interest are the payoff function of players and the probability distribution of the order of actions.

The sources of the dataset are the Chain Store Guide (1988−1997), U.S. Census Bureau, and the Missouri State Census Data Center. The dataset contains the entry decision of Walmart and Kmart across 2065 counties in the United States observed in two periods, 1988 and 1997. There

[5] The data file is available at the following link: https://barwick.economics.cornell.edu/Data.html.

	$n = 500$		$n = 1000$		$n = 5000$	
Simultaneous (Correctly Specified)	Estimates	Std. Err.	Estimates	Std. Err.	Estimates	Std. Err.
β_1	3.6114	2.6312	3.4164	2.2967	3.1269	0.9769
β_2	2.4729	1.6111	2.3835	1.4545	2.1512	0.6344
Δ_1	-5.9525	4.4415	-5.4472	3.7776	-5.9944	1.7528
Δ_2	-4.3551	3.3789	-3.5876	2.5196	-3.9883	1.1988
Simultaneous (Misspecified)						
β_1	1.9210	0.8322	1.9710	0.4596	2.0053	0.1838
β_2	1.1123	0.6155	0.9349	0.4636	0.8546	0.3436
Δ_1	-2.8356	1.4094	-2.9900	0.8617	-3.0692	0.1795
Δ_2	-1.5654	1.4947	-1.1462	1.1220	-0.9185	0.4424
Sequential (Correctly Specified)						
β_1	3.4217	1.2058	3.1900	0.6493	3.0256	0.2261
β_2	2.1242	0.4740	2.0437	0.2370	2.0056	0.0931
Δ_1	-6.8553	2.5389	-6.3981	1.3835	-6.0523	0.4973
Δ_2	-4.2373	0.9610	-4.0790	0.4943	-4.0112	0.1977
Sequential (Misspecified)						
β_1	2.1169	1.3935	1.8801	1.3197	1.7695	0.5624
β_2	1.2128	0.7455	1.0883	0.5503	0.9933	0.2993
Δ_1	-1.1987	0.7330	-1.0990	0.7195	-0.9853	0.2997
Δ_2	-1.9240	1.2763	-1.7248	0.9508	-1.6221	0.5074

Table 3: Estimation under Misspecification versus Correct Specification

Note: n is sample size, and the true value of parameter is $(\beta_1, \beta_2) = (3, 2)$, $(\Delta_1, \Delta_2) = (-6, -4)$. $P\left(O = (1,1)' \,|X\right) = 1$ is a simultaneous move game, and $P\left(O = (1,2)' \,|X\right) = 1$ is a sequential game. The first two tables are estimates when the Data Generating Process (DGP) follows a simultaneous move, and the last two tables are estimates when the DGP follows a sequential move. Simultaneous (Misspecified) means that a simultaneous game (DGP) is treated as a sequential game (in estimation), and Sequential (Misspecified) means that a sequential game (DGP) is treated as a simultaneous game (in estimation). The number of simulations is 1000. The estimates are computed by the sample mean and the sample standard error of 1000 estimated values.

are also county-specific variables, including county population, retail sales per capita, percentage of the urban population, regional dummies, the number of small stores, and the number of Walmart and Kmart stores in markets that are not part of the sample. The summary statistics are available in Table 2 of Jia (2008). The current paper only uses the most recent sample in 1997.

Use subscripts W and K to indicate Walmart and Kmart, and assume the following payoff structure:

Walmart\Kmart	Entry ($S_K = 1$)	Exit ($S_K = 0$)
Entry ($S_W = 1$)	$\begin{pmatrix} \beta_{0W} + X_W'\beta_{1W} + \Delta_W - \epsilon_W, \\ \beta_{0K} + X_K'\beta_{1K} + \Delta_K - \epsilon_K \end{pmatrix}$	$(\beta_{0W} + X_W'\beta_{1W} - \epsilon_W, 0)$
Exit ($S_W = 0$)	$(0, \beta_{0K} + X_K'\beta_{1K} - \epsilon_K)$	$(0, 0)$

Table 4: The Payoff Structure of Walmart and Kmart

The payoff function of player i is $u_i = \beta_{0i} + X_i'\beta_{1i} + S_{-i}\Delta_i - \epsilon_i$ for $i = W, K$, following a linear form as in Bresnahan and Reiss (1991), Tamer (2003), and Aradillas-Lopez (2010). For each county, two players Walmart and Kmart compete in markets by making decisions to enter the market or not. The payoff is linear in parameters $(\beta_{0i}, \beta_{1i}, \Delta_i)$ and is normalized to zero when $S_i = 0$. X_i includes the log of the county population, the log of county retail sales per capita, the percentage of the urban population, and the number of small discount stores as common variables. There is also a player-specific variable in X_i, the distance weighted number of player i branches in markets that are not part of the sample. The player-specific variable works as an exclusion restriction to identify the payoff function coefficients. The parameters $\beta_{1W}, \beta_{1K}, \Delta_W, \Delta_K$ are common knowledge among players but are not observed by econometricians. $\epsilon = (\epsilon_W, \epsilon_K)'$ represents private information of players in incomplete information games, and ϵ_i follows a standard Type-I extreme distribution.

I consider the three possible order of actions for two players: (1) Walmart is the first mover and Kmart is the second mover, (2) Kmart is the first mover and Walmart is the second mover, (3) Walmart and Kmart move simultaneously. The probability distribution $\rho_{O|X}(O_k|X)$ for $k = 1, 2, 3$ is another parameter of interest. For simplicity, the control variables X for $\rho_{O|X}$ are only regional indicators: Midwest and South. Based on the dataset, each county belongs to one of three regions, Midwest, South, or others. The Midwest dummy is 1 if a county is located in the Great Lakes, Plains, or Rocky Mountain region. The South dummy is 1 for Southwest or Southeast regions. Note that the Kmart Headquarters is in Midwest (Illinois), and the Walmart Headquarters is in South (Arkansas).

The following Table 5 summarizes the estimation results of three different specifications on the potential order of actions. The first column of Table 5 shows the estimates assuming the simultaneous move only. The second column presents the estimates under the sequential orders without considering the simultaneous move. The last column of Table 5 is the main result of this empirical application. The estimates in the last column are robust to all three possible order of actions and also to the regional level heterogeneity of the order distribution. The probability

	Simultaneous Move		Sequential Move		Simultaneous+Sequential	
	Walmart	Kmart	Walmart	Kmart	Walmart	Kmart
Log Population	2.0205	2.0422	2.6290	2.1567	2.5840	2.5544
	(0.2051)	(0.3739)	(0.2243)	(0.2062)	(0.2981)	(0.3393)
Log Retail Sales	0.9744	2.1261	1.4394	1.9640	1.4665	2.1796
per Capita	(0.1625)	(0.3034)	(0.1856)	(0.2599)	(0.2187)	(0.3094)
Percentage of	1.0858	0.9940	1.2483	1.0260	1.2476	1.1590
Urban Population	(0.2701)	(0.4312)	(0.3337)	(0.3533)	(0.3808)	(0.3936)
Small Stores	-0.0277	-0.0362	-0.0724	-0.0298	-0.0757	-0.0366
	(0.0270)	(0.0245)	(0.0325)	(0.0241)	(0.0352)	(0.0275)
# of other Walmarts	-3.4444		-3.7524		-3.3114	
	(1.1099)		(0.9668)		(0.9876)	
# of other Kmarts		-0.6699		-1.4352		-1.6872
		(0.9996)		(0.9091)		(1.0192)
Constant	-13.5967	-24.9788	-18.8790	-23.2951	-18.9547	-25.6431
	(1.5686)	(3.0105)	(1.6805)	(2.4164)	(2.2993)	(3.1450)
Interaction Effect	-0.9934	-0.6866	-1.8410	-1.4281	-2.2053	-2.6667
(Δ_W, Δ_K)	(0.5338)	(1.3232)	(0.2543)	(0.2245)	(0.6367)	(0.5309)
Midwest	Walmart → Kmart		0.0827 (0.1372)		0.0018 (0.0158)	
	Kmart → Walmart		0.9173 (0.1372)		0.4705 (0.1408)	
	Simultaneous				0.5278 (0.1390)	
South	Walmart → Kmart		0.9999 (0.0000)		0.2449 (0.2006)	
	Kmart → Walmart		0.0001 (0.0000)		0.0042 (0.0606)	
	Simultaneous				0.7509 (0.2051)	
Other Regions	Walmart → Kmart		0.0011 (0.0171)		0.0450 (0.0960)	
	Kmart → Walmart		0.9989 (0.0171)		0.6836 (0.1956)	
	Simultaneous				0.2714 (0.1956)	

Table 5: Estimation of the Entry Game between Walmart and Kmart

Note: A two player game has three order of actions: (1) Walmart → Kmart (Walmart is the first mover), (2) Kmart → Walmart (Kmart is the first mover), (3) Walmart and Kmart move simultaneously. The estimated probability mass of each order of action and the standard error in the parenthesis are provided in the second and third column.

distribution of the order significantly varies with the value of regional dummies.

The result in Table 5 presents two noteworthy points. First, the estimates of the strategic interaction (Δ_W, Δ_K) are significantly different depending on the specification about the order of actions. The interaction effects under the conventional simultaneous game specification are $(\Delta_W, \Delta_K) = (-0.9934, -0.6866)$, but the new estimates considering the mixed order of actions are $(\Delta_W, \Delta_K) = (-2.2053, -2.2667)$. The different estimates imply different counterfactual outcomes. For example, under the simultaneous game assumption, a 10% increase in Kmart's entry probability results in a 3.3% decrease in Walmart's entry probability at the mean level of county-specific regressors. However, the new estimates with the mixed order of actions imply a 6.8% decrease in Walmart's entry probability under the same situation. The example suggests that the strategic interaction effects between Walmart and Kmart can be larger than the effects predicted in the previous literature.

Second, the estimated order of actions shows that Walmart and Kmart compete with each other in sequential way for a significant portion of the markets. It is well-known that Kmart is a nationwide retailer with a longer history than Walmart. According to the estimates, Kmart is the first mover in many counties except southern counties, taking 47.05% of the midwestern counties and 68.36% of other regions except midwestern and southern counties. Walmart is the first mover in southern counties where its Headquarters is located, taking 24.49% of the markets in South. Both simultaneous and sequential type of markets account for nonnegligible portion of the markets. The result confirms that the structural estimation of an empirical game without considering the order of actions may cause a substantial bias in estimates and a distorted counterfactual analysis.

6 Conclusion

This paper discusses the identification and estimation of sequential games with incomplete information. I specify a sequential game model of incomplete information, assuming that the observed market outcome follows a PBNE. The structural model presents three main parameters of interest, including the player's payoff parameter, the order of actions, and the equilibrium selection mechanism.

The main finding of the paper consists of three parts. First, I provide some necessary and sufficient conditions for identifying the structural parameters, exploiting that the number of PBNE and the number of the possible order of actions are finite. Second, I propose the SMD estimator for the structural parameters based on identification. The consistency and asymptotic normality of the suggested estimator are verified in the context of Ai and Chen (2003) and Chen and Pouzo (2015). Third, Monte Carlo simulations and an empirical application to the Walmart-Kmart entry game highlight the importance of correct specification on the order of actions.

The future research regarding this paper is the applicability of these theoretical findings to empirical settings, including a sequential entry game or a bargaining game between groups of players. Furthermore, the sequential game structure can apply to models with dynamics so that the general setting of my model on sequential games can be extended to finite or infinite horizon dynamic games. I leave these topics for future research.

Appendix

A Proof of Lemmas

A.1 Proof of Lemma 2.1

By definition, a PBNE consists of equilibrium probabilities in equation (3). First, I show that the equation (3) is a function of $\mathcal{P}(X,O)$ and the model primitives. Next, I verify the existence of $\mathcal{P}(X,O)$ solving the equation (3) and the number of solution is finite. The Assumption 2.1−1 implies that the joint probability of ϵ_j for $j = 1, \ldots, N$ are independent conditional on common variables (X, O). The history vector $s_-^{o(i)}$ in the information set $\mathcal{J}_{o(i)} = \left(s_-^{o(i)}, X, O\right)$ fully depends on $\epsilon_j | X, O$ for $o(j) < o(i)$ so that $\epsilon_i | X, O$ is independently realized with respect to $s_-^{o(i)}$. Then $\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right)$ in equation (3) is

$$\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right) = \sum_{a_+^{o(i)}} \pi_i\left(a, a_+^{o(i)}, \mathcal{J}_{o(i)}; \beta_0\right) \prod_{j \neq i, o(j) \geq o(i)} P\left(s_j = a^j | \mathcal{J}_{o(j)}\right), \tag{11}$$

where $\prod_{j \neq i, o(j) \geq o(i)} P\left(s_j = a^j | \mathcal{J}_{o(j)}\right) = P\left(s_+^{o(i)} = a_+^{o(i)} | \mathcal{J}_{o(i)}\right)$. The joint equilibrium probability for players $j \neq i$ with $o(j) \geq o(i)$, $P\left(s_+^{o(i)} = a_+^{o(i)} | \mathcal{J}_{o(i)}\right)$ is represented by the components of $\mathcal{P}(X,O)$. The equations of $P\left(s_i = a | \mathcal{J}_{o(i)}\right)$ for all $i \in \mathcal{I}$ show that a PBNE is defined by $\mathcal{P}(X,O)$ under the given payoff structure β_0.

The existence of a PBNE is verified by a fixed point approach. Note that each equation (3) is a continuous mapping of equilibrium probabilities

$$\left\{P\left(s_i = a | \mathcal{J}_{o(i)}\right), \left\{P\left(s_j = a^j | \mathcal{J}_{o(j)}\right) \middle| j \neq i, o(j) \geq o(i), a^j \in \mathcal{A}\right\}\right\},$$

for $a \in \mathcal{A}$ and all possible history vectors in $\mathcal{J}_{o(i)}$ and $\mathcal{J}_{o(j)}$. The probabilities construct $L+1$ equations for each player and history vector, and each player i has $(L+1)^{\sum_{t=1}^{o(i)-1} n_t}$ possible history vectors conditional on X and O. The total number of equations are the same as the number of unknown equilibrium probabilities. By Assumption 2.1−2, $\mathcal{P}(X,O)$ is continuous on a closed ball of a Euclidean space so that the existence of the solution is guaranteed by the Brouwer's Fixed Point Theorem.

Note that equilibrium probabilities $\mathcal{P}(X,O)$ may not be singleton: there could be multiple equilibria. Under Assumption 2.1, Theorem in Section 4 of Haller and Lagunoff (2000) showed that the number of PBNE is finite for almost all games. They derived the finiteness of MPE on almost every stochastic games with a finite number of players, discrete and finite action sets. The sequential game in the current paper is a special case of a stochastic game discussed in Haller and Lagunoff (2000).

A.2 Proof of Lemma 2.2

The proof follows similar steps in Hotz and Miller (1993) by showing that the sequential game model in this paper satisfies Proposition 1 of Hotz and Miller (1993). Fix a player $i \in \mathcal{I}$ and her information set $\mathcal{J}_{o(i)}$. Then the system of equations is

$$\begin{pmatrix} P\left(s_i = a_1 | \mathcal{J}_{o(i)}\right) \\ \cdots \\ P\left(s_i = a_L | \mathcal{J}_{o(i)}\right) \end{pmatrix} = \begin{pmatrix} F_{a_1}\left(\bar{\pi}_i\left(a_1, \mathcal{J}_{o(i)}; \beta_0\right), \ldots, \bar{\pi}_i\left(a_L, \mathcal{J}_{o(i)}; \beta_0\right)\right) \\ \cdots \\ F_{a_L}\left(\bar{\pi}_i\left(a_1, \mathcal{J}_{o(i)}; \beta_0\right), \ldots, \bar{\pi}_i\left(a_L, \mathcal{J}_{o(i)}; \beta_0\right)\right) \end{pmatrix},$$

which has L equations with L unknowns, provided with continuous mappings $F_{a_1}, \ldots, F_{a_L}$. The mappings are continuous by Assumption 2.1−2. The mapping does not include $\bar{\pi}_i\left(0, \mathcal{J}_{o(i)}; \beta_0\right) = 0$ by Assumption 2.2. Define

$$\mathcal{P}_i(X, O) \equiv \left(P\left(s_i = a_1 | \mathcal{J}_{o(i)}\right), \ldots, P\left(s_i = a_L | \mathcal{J}_{o(i)}\right)\right)'$$
$$\bar{\Pi}_i(X, O) \equiv \left(\bar{\pi}_i\left(a_1, \mathcal{J}_{o(i)}; \beta_0\right), \ldots, \bar{\pi}_i\left(a_L, \mathcal{J}_{o(i)}; \beta_0\right)\right)',$$

and apply the Inverse Function Theorem (IFT), then it is sufficient to check whether $F_{a_1}, \ldots, F_{a_L}$ are differentiable with respect to each $\bar{\pi}_i\left(a, \mathcal{J}_{o(i)}; \beta_0\right)$. If all CDFs are differentiable almost everywhere, the IFT implies that the joint operator $\mathfrak{F} = (F_{a_1}, \ldots, F_{a_L})$ has a differentiable inverse operator $\mathfrak{F}^{-1}$ such that $\mathfrak{F} \circ \mathfrak{F}^{-1} = \mathfrak{F}^{-1} \circ \mathfrak{F} = \mathfrak{I}$ where $\mathfrak{I}$ is the identity operator. Then the uniqueness of the operator $\mathfrak{F}$ guarantees a unique bijection mapping between $\mathcal{P}_i(X, O)$ and $\bar{\Pi}_i(X, O)$. By Assumption 2.1, each F_a is differentiable almost everywhere.

The unique one-to-one correspondence can be extended for all $i \in \mathcal{I}$ and $\mathcal{J}_{o(i)}$ without loss of generality. Thus the unique mapping between $\mathcal{P}(X, O)$ and $\bar{\Pi}(X, O)$ holds for X and O almost everywhere.

B Identification

B.1 Proof of Theorem 3.1

For structural parameters $\left(\beta_0, \rho_{O|X}\right)$, recall equation (5),

$$P(s = \alpha | X) = \sum_{l=1}^{N_o} \prod_{i=1}^{n} P\left(s_i = \alpha_i | \mathcal{J}_{o_l(i)}; \beta_0\right) \rho_{O|X}(O_l | X).$$

The observed conditional choice probability $P(s = \alpha | X)$ is a function of $\left(\beta_0, \rho_{O|X}\right)$. By definition, $P(s = \alpha | X)$ is a weighted average of equilibrium probabilities $\prod_{i=1}^{n} P\left(s_i = \alpha_i | \mathcal{J}_{o_l(i)}; \beta_0\right)$ where each equilibrium probability $P\left(s_i = \alpha_i | \mathcal{J}_{o_l(i)}; \beta_0\right)$ is known up to the finite dimensional parameter β_0. Denote a vector of conditional choice probabilities for $(L+1)^N$ joint actions by

$$Q(X) \equiv \left(P\left(s=\alpha^1|X\right), \ldots, P\left(s=\alpha^{(L+1)^N-1}|X\right), 1\right)',$$

and the probability distribution of the order of actions by

$$\rho_0(X) \equiv \left(\rho_{O|X}(O_1|X), \ldots, \rho_{O|X}(O_{N_o}|X)\right) \in \mathcal{P}_O.$$

Then the equation (5) is a set of equations

$$Q(X) = \mathcal{P}(X;\beta_0)\,\rho_0(X),$$

where $\mathcal{P}(X;\beta_0)$ is a matrix of equilibrium probabilities

$$\mathcal{P}(X;\beta_0) \equiv \begin{bmatrix} P\left(s=\alpha^1|X,O_1;\beta_0\right) & \cdots & P\left(s=\alpha^1|X,O_{N_o};\beta_0\right) \\ P\left(s=\alpha^2|X,O_1;\beta_0\right) & \cdots & P\left(s=\alpha^2|X,O_{N_o};\beta_0\right) \\ \cdots & \cdots & \cdots \\ P\left(s=\alpha^{(L+1)^N-1}|X,O_1;\beta_0\right) & \cdots & P\left(s=\alpha^{(L+1)^N-1}|X,O_{N_o};\beta_0\right) \\ 1 & \cdots & 1 \end{bmatrix}.$$

The last row of the matrix $\mathcal{P}(X;\beta_0)$ implies a constraint that $\sum_{l=1}^{N_o}\rho_{O|X}(O_l|X)=1$. As far as the nonlinear system of equations $Q(X)=\mathcal{P}(X;\beta_0)\,\rho_0(X)$ has a unique solution at $\beta=\beta_0\in\mathcal{B}$ and $\rho_0(X)\in\mathcal{P}_O$, the model is identified. By Assumption 3.1, the system of equations $Q(X)$ has a unique solution $(\beta_0,\rho_0(X))$ for almost all X by Rouché-Capelli theorem. The full rank condition of $\mathcal{P}(X;\beta_0)$ implies a necessary condition of identification, $(L+1)^N \geq N_o$ where N_o is the number of possible orders of actions.

B.2 Proof of Corollary 3.2

Assumption 3.2 suggests an exclusion restriction on the order of actions. Suppose $(L+1)^N < N_o$ so that a necessary condition of identification is violated. The necessary order condition of identification is weakened as a result of the exclusion restrictions. The exclusion restriction implies that $\rho_{O|X}(O|X) = \rho_{O|X}(O|X_s)$. Then an instrumental variable X_v with a discrete support $\left\{x_v^1,\ldots,x_v^{N_v}\right\}$ generates additional equations

$$Q(X) = \mathcal{P}(X_s;\beta_0)\,\rho_O(X_s),$$

where $\mathcal{P}(X_s;\beta_0) = \left[\mathcal{P}\left(\left(X_s,x_v^1\right);\beta_0\right)',\ldots,\mathcal{P}\left(\left(X_s,x_v^{N_v}\right);\beta_0\right)'\right]' \in \mathfrak{M}_{N_v(L+1)^N\times N_o}$. The probability of unobserved heterogeneity $\rho_{O|X}(O|X)$ only depends on the subset of X. Then an additional variation of X still implies $(L+1)^N$ conditional moments, while the number of additional unknown components is fixed. Without the overlapped equations $\sum_{l=1}^{N_o}\rho\left(O_l|X_s,x_v^1\right) =$

$\sum_{l=1}^{N_o} \rho\left(O_l|X_s, x_v^{N_v}\right) = 1$, the necessary condition with active equations is $N_v\left(L+1\right)^N - N_v + 1 \geq N_o$.

For example, consider a three-player game with binary actions $\{0,1\}$. The number of all possible orders is $N_o = 13$. The necessary order condition in the general case is $\left(L+1\right)^N = 8 < 13 = N_o$ so that $\rho_0\left(X\right)$ cannot be identified. Under the exclusion restriction, the order condition becomes $7N_v \geq 12$ so that an instrumental variable X_v with at least two values satisfies the necessary condition of identification. Similarly, a four-player game has at most $N_o = 75$ different order of actions so that the order condition holds with $15N_v \geq 74$. The number of support points of X_v has to be $N_v \geq 5$.

In the cases of fully sequential actions that only one player is assigned for each stage, there are $N!$ order of actions, and the order condition becomes $N_v\left(\left(L+1\right)^N - 1\right) \geq N! - 1$. The fully sequential actions do not generate multiple equilibria since the backward induction with a finite number of players implies a unique equilibrium probability. Suppose a binary game $L = 1$, then the required variation of X_v should be $N_v \geq 2$ for a four-player game, $N_v \geq 4$ for a five-player game, and $N_v \geq 12$ for a six-player game. The order condition can be more relaxed by additional constraints on the possible order of actions, e.g., Player 2 cannot move earlier than Player 1.

B.3 Proof of Theorem 3.2

The proof is similar as Theorem 3.1 in which the number of unobserved heterogeneity types is N_κ instead of N_o. The Rouché-Capelli theorem implies that a necessary and sufficient condition of identification is $rank\left(\mathcal{P}'\left(X;\beta\right)\right) = rank\left(\left[\mathcal{P}'\left(X;\beta\right), Q\left(X\right)\right]\right) = N_\kappa$ for almost all $X \in \mathcal{X}$ only at $\beta = \beta_0 \in \mathcal{B}$.

Suppose β_0 and $P\left(\kappa = k|X\right)$ for $k = 1, \ldots, N_\kappa$ are identified. The next step is to show that $\lambda_{\tau|X,O}$ and $\rho_{O|X}$ are separately identified from $P\left(\kappa = k|X\right)$. Note that $\lambda_{\tau|X,O}$ and $\rho_{O|X}$ are discrete probability mass functions. Then,

$$\sum_{\left\{k:\tau\in\left\{\tau_{X,O_l,1},\ldots,\tau_{X,O_l,B(X,O_l)}\right\},O=O_l\right\}} P\left(\kappa = k|X\right) = \sum_{k=1}^{B(X,O_l)} \lambda_{\tau|X,O}\left(\tau_{X,O_l,k}\right)\rho_{O|X}\left(O_l|X\right) = \rho_{O|X}\left(O_l|X\right),$$

because $\sum_{k=1}^{B(X,O_l)} \lambda_{\tau|X,O}\left(\tau_{X,O_l,k}\right) = 1$. Next, let the kth point of unobserved heterogeneity indicate $\tau = \tau_{X,O_l,m}$ and $O = O_l$ for $m \in \left\{1, \ldots, \tau_{X,O_l,B(X,O_l)}\right\}$ and $l \in \{1, \ldots, N_o\}$.

$$\lambda_{\tau|X,O}\left(\tau_{X,O_l,m}\right) = \frac{\lambda_{\tau|X,O}\left(\tau_{X,O_l,m}\right)\rho_{O|X}\left(O_l|X\right)}{\rho_{O|X}\left(O_l|X\right)} = \frac{P\left(\kappa = k|X\right)}{\sum_{\left\{k':\tau\in\left\{\tau_{X,O_l,1},\ldots,\tau_{X,O_l,B(X,O_l)}\right\},O=O_l\right\}} P\left(\kappa = k'|X\right)},$$

provided that $\rho_{O|X}\left(O_l|X\right) > 0$.

C Estimation and Inference

C.1 Proof of Theorem 4.1

The consistency of $\hat{\theta}_n$ verifies assumptions for Lemma 3.1 of Ai and Chen (2003). I mainly show that the model with conditional moments in equation (9) satisfies Assumptions 3.1 to 3.7 of Ai and Chen (2003).

Assumptions 3.1 and 3.2−(i) of Ai and Chen (2003) matches to Assumption 4.1 of the current paper. The assumption includes standard assumptions for nonparametric sieve estimation. The observables (s^m, X^m) for $m = 1, \ldots, n$ are i.i.d., the support of X is compact, and the probability density function of X is bounded by Assumption 4.1. Assumption 3.2−(i) corresponds to Assumption 4.1−3. The bounded eigenvalues of $E\left[p^{J_n}(X)\, p^{J_n}(X)'\right]$ depend on the choice of sieve basis. Any orthonormal sieve basis function satisfies the condition as $E\left[p_j(X)\, p_k(X)\right] = 1$ if $j = k$ and $E\left[p_j(X)\, p_k(X)\right] = 0$ if $j \neq k$. For Assumption 3.2−(ii) I assume that $p^{J_n}(X)$ is a vector of tensor-product Fourier series or B-spline of order γ' satisfying $\gamma' > d_x/2 + 1$. Then Assumption 3.2−(ii) holds by section 2.3 of Chen (2007).

Assumption 3.3 of Ai and Chen (2003) states that θ_0 is identified. Assumptions 2.1−3.2 in Theorems 3.1−3.2 show conditions for identification of θ_0. Assumption 3.4 of Ai and Chen (2003) holds by specifying the identity weight matrix I for the sieve minimum distance criterion function.

Assumption 3.5 of Ai and Chen (2003) is satisfied by the smoothness condition on the nonparametric parameter of interest. $h_k \in \mathcal{H}_k$ for $k = 1, \ldots, N_\kappa$ is a Hölder space with smoothness parameter $\gamma > d_x/2$ by Assumption 4.2. Proposition 3.2 of Ai and Chen (2003) confirms that $\mathcal{H}_k$ is compact under the strong norm $\|\cdot\|_s$. Under the assumption that $p^{K_n}(X)$ is tensor-product Fourier series or B-spline, Section 2.3 of Chen (2007) also verifies that for any $\theta \in \Theta$, there exists $\theta_n \in \Theta_n = \mathcal{B} \times \mathcal{H}_n$ such that $\|\theta_n - \theta_0\|_s = \max_{k=1,\ldots,N_\kappa} \sup_{X\in\mathcal{X}} |h_{0,k}(X) - h_{k,n}(X)| = O\left(K_n^{-\gamma/d_x}\right) = o(1)$.

Assumption 3.6−(i) of Ai and Chen (2003) requires the boundedness of the conditional moments and Hölder continuity of $\ell(Z,\theta)$ in $\theta \in \times$. The conditional moment $E\left[\|\ell(Z,\theta_0)\|_E |X\right]$ is bounded as the conditional moment consists of probability mass functions.

$$
\begin{aligned}
E\left[|\ell_j(Z,\theta_0)|\, |X\right] &= \left| P\left(s = \alpha^j |X\right) - \sum_{k=1}^{N_\kappa} \prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta_0\right) h_k^1(X) \right| \\
&\leq \left| P\left(s = \alpha^j |X\right)\right| + \left| \sum_{k=1}^{N_\kappa} \prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta_0\right) h_k^1(X) \right| \\
&\leq 1 + 1 = 2,
\end{aligned}
$$

for $j = 1, \ldots, (L+1)^N$. The number of actions L and the number of players N are finite so that $E\left[\|\ell(Z,\theta_0)\|_E |X\right]$ is bounded as well. Next, $\ell(Z,\theta)$ is Hölder continuous since for any $\theta^1 =$

$\left(\beta^1, h_1^1, \ldots, h_{N_\kappa}^1\right), \theta^2 = \left(\beta^2, h_1^2, \ldots, h_{N_\kappa}^2\right) \in \Theta$,

$$
\begin{aligned}
&\left|\ell_j\left(Z, \theta^1\right) - \ell_j\left(Z, \theta^2\right)\right| \\
=& \left|\sum_{k=1}^{N_\kappa}\left(\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^1\right) h_k^1(X) - \prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^2\right) h_k^2(X)\right)\right| \\
\leq& \left|\sum_{k=1}^{N_\kappa}\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^1\right)\left(h_k^1(X) - h_k^2(X)\right)\right| \\
&+ \left|\sum_{k=1}^{N_\kappa}\left(\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^2\right) - \prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^1\right)\right) h_k^2(X)\right| \\
\leq& \sum_{k=1}^{N_\kappa}\left(\left|h_k^1(X) - h_k^2(X)\right| + \left|\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^2\right) - \prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta^1\right)\right|\right) \\
\leq& C_1\left\|\theta^1 - \theta^2\right\|_s,
\end{aligned}
$$

for some $C_1 < \infty$. The last inequality is from

$$
\sum_{k=1}^{N_\kappa}\left|h_k^1(X) - h_k^2(X)\right| \leq N_\kappa \max_{k=1,\ldots,N_\kappa} \sup_{X \in \mathcal{X}}\left|h_k^1(X) - h_k^2(X)\right|,
$$

and

$$
\sum_{k=1}^{N_\kappa}\left|\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta_2\right) - \prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta_1\right)\right| \leq C_2 N_\kappa\left\|\beta^1 - \beta^2\right\|_E,
$$

where $C_2 = \max_{k\in\{1,\ldots,N_\kappa\}} \sup_{\beta\in\mathcal{B},\mathcal{J}_{o_k(1)},\ldots,\mathcal{J}_{o_k(N)}} \frac{\partial}{\partial\beta}\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta\right) < \infty$. C_2 is uniformly bounded by Assumption 4.3 since the derivative of the equilibrium probability function, $\frac{\partial}{\partial\beta}\prod_{i=1}^{N} P^{(k)}\left(s_i = \alpha_i^j | \mathcal{J}_{o_k(i)}; \beta\right)$ is continuous on a compact set of Euclidean space. Let $C_1 = \max\{N_\kappa, C_2 N_\kappa\}$, then

$$
\left|\ell_j\left(Z, \theta^1\right) - \ell_j\left(Z, \theta^2\right)\right| \leq C_1\left\|\theta^1 - \theta^2\right\|_s,
$$

where the strong norm is defined by

$$
\left\|\theta^1 - \theta^2\right\|_s = \left\|\beta^1 - \beta^2\right\|_E + \max_{k=1,\ldots,N_\kappa} \sup_{X\in\mathcal{X}}\left|h_k^1(X) - h_k^2(X)\right|.
$$

The same proofs for $j = 1, \ldots, (L+1)^N$ implies that $\ell(Z, \theta)$ is Hölder continuous in $\theta \in \Theta$.

Assumption 3.7 of Ai and Chen (2003) compares the number of unconditional moments and that of unknown parameters. There are $(L+1)^N$ joint actions and each joint action is associated with J_n sieve basis functions. The number of unknown parameters includes the payoff function parameter $\beta_0 \in \mathbb{R}^{d_\beta}$ and N_κ nonparametric probability mass functions $h_k(X) = P(\kappa = k|X)$ for $k = 1, \ldots, N_\kappa$. Each $h_k(X)$ is approximated by K_n sieve basis functions. Thus $(L+1)^N J_n$ should be at least as large as $d_\beta + N_\kappa K_n$. The remaining parts in Assumption 3.7 are directly verified by

Assumption 4.4.

With the verification of Assumptions 3.1 to 3.7, Lemma 3.1 of Ai and Chen (2003) implies that $\left\|\hat{\theta}_n - \theta_0\right\|_s = o_p(1)$. $\hat{\theta}_n$ is a consistent estimator of θ_0 with respect to the strong norm $\|\cdot\|_s$.

C.2 Proof of Theorem 4.2

This section discusses asymptotic normality of the SMD estimator $\hat{\theta}_n$, focusing on the parametric part $\hat{\beta}_n$ of the structural parameters. Note that the asymptotic normality result provided in this section is a special case of Chen and Pouzo (2015): see Appendix A of Chen and Pouzo (2015). I verify the remaining assumptions for asymptotic normality results presented in Ai and Chen (2003). Assumptions 3.1−3.9 and 4.1−4.6 of Ai and Chen (2003) will be discussed.

The first set of assumptions, Assumptions 3.1−3.9, relates to the convergence rate of the SMD estimator. Assumptions 3.2−(iii) and 3.5−(iii) of Ai and Chen (2003) restricts that the number of sieves does not increase too fast. A function in a Hölder space with smoothness parameter $\gamma > d_x/2$ can be approximated by a tensor-product B-spline of order γ' with $\gamma' > d_x/2 + 1$. Assumption 4.5−3 guarantees the required rate of convergence $J_n^{-\gamma/d_x} = K_n^{-\gamma/d_x} = o\left(n^{-1/4}\right)$. Assumption 3.4−(iii) holds with an identity weight matrix.

Assumptions 3.6−(iii) and (iv) follow the established proof in in Section C.1. Each $\ell_j(Z,\theta)$ for $j = 1, \ldots, (L+1)^N$ is bounded by 2 for all Z and $\theta \in \mathcal{B} \times \mathcal{H}_n$. The smoothness condition in Assumption 4.2 applies to verify Assumption 3.6−(iv).

Assumption 3.7−(ii) is directly implied by Assumption 4.5−3. $\sup_{x\in\mathcal{X}} \left\|p^{J_n}(x)\right\|_E = J_n^{1/2}$ since the basis function is a tensor-product B-spline of order γ', thereby $K_n \times \log n \times \left(\sup_{x\in\mathcal{X}} \left\|p^{J_n}(x)\right\|_E\right)^2 \times n^{-1/2} = o(1)$ is equivalent to $J_n K_n \log n = o\left(n^{1/2}\right)$. For Assumption 3.8, define $N\left(\varepsilon, \Theta_n, \|\cdot\|_s\right)$ by the bracketing number, or the minimum number of ε-brackets to cover Θ_n under the strong norm $\|\cdot\|_s$. A Hölder space approximated by splines satisfies the entropy number $\log\left(N\left(\varepsilon, \Theta_n, \|\cdot\|_s\right)\right) = C_3 K_n \log(1/\varepsilon) \leq C_3 K_n \log(K_n/\varepsilon)$ for some positive constant C_3 so that Assumption 3.8 of Ai and Chen (2003) is also verified.

Assumption 3.9 of Ai and Chen (2003) holds by pathwise differentiability of $\ell(Z,\theta)$ at $\theta = \theta_0$. Note that ℓ is linear in h and nonlinear in β. I focus on a component of $\ell(Z,\theta)$ and fix a $\ell_j(Z,\theta)$. Denote $\theta = (\beta', h')' \equiv \left(\beta^1, \beta^2, \ldots, \beta^{d_\beta}, h_1, \ldots, h_{N_\kappa}\right)'$. The pathwise derivative of $\ell_j(Z,\theta)$ at the direction $[\theta - \theta_0]$ evaluated at θ_0 is defined by

$$
\begin{aligned}
\frac{d\ell_j(Z,\theta_0)}{d\theta}[\theta - \theta_0] &\equiv \left.\frac{d\ell_j(Z,(1-\tau)\theta_0 + \tau\theta)}{d\tau}\right|_{\tau=0} \\
&= \sum_{k=1}^{N_\kappa} \frac{\partial}{\partial \beta'} P^{(k)}\left(s = \alpha^j | \mathcal{J}_{o_k(i)}; \beta_0\right)(\beta_0 - \beta)\, h_{0,k}(X) \\
&\quad + \sum_{k=1}^{N_\kappa} P^{(k)}\left(s = \alpha^j | \mathcal{J}_{o_k(i)}; \beta_0\right)\left(h_{0,k}(X) - h_k(X)\right),
\end{aligned}
$$

provided that the equilibrium probability $P^{(k)}\left(s = \alpha^j | \mathcal{J}_{o_k(i)}; \beta_0\right)$ is continuously differentiable at

$\beta_0 \in \mathcal{B}$ by Assumption 4.3. Next, for a metric

$$\left\|\theta^1 - \theta^2\right\| \equiv \sqrt{E\left[E\left[\frac{d\ell\left(Z,\theta_0\right)}{d\theta}\left[\theta^1 - \theta^2\right]\middle|X\right]' E\left[\frac{d\ell\left(Z,\theta_0\right)}{d\theta}\left[\theta^1 - \theta^2\right]\middle|X\right]\right]}$$

where $\frac{d\ell(Z,\theta_0)}{d\theta}\left[\theta^1 - \theta^2\right] \equiv \frac{d\ell(Z,\theta_0)}{d\theta}\left[\theta^1 - \theta_0\right] - \frac{d\ell(Z,\theta_0)}{d\theta}\left[\theta^2 - \theta_0\right]$,

$$\|\theta - \theta_0\|^2 = E\left[E\left[\frac{d\ell\left(Z,\theta_0\right)}{d\theta}\left[\theta - \theta_0\right]\middle|X\right]' E\left[\frac{d\ell\left(Z,\theta_0\right)}{d\theta}\left[\theta - \theta_0\right]\middle|X\right]\right].$$

Note that $E\left[\ell\left(Z,\theta_0\right)|X\right] = 0$, hence

$$\begin{aligned}
& E\left[\ell\left(Z,\theta\right)|X\right] \\
&= E\left[\ell\left(Z,\theta\right) - \ell\left(Z,\theta_0\right)|X\right] \\
&= \sum_{k=1}^{N_\kappa} P^{(k)}\left(s = \alpha^j|\mathcal{J}_{o_k(i)};\beta\right) h_k\left(X\right) - \sum_{k=1}^{N_\kappa} P^{(k)}\left(s = \alpha^j|\mathcal{J}_{o_k(i)};\beta_0\right) h_{0,k}\left(X\right) \\
&= \sum_{k=1}^{N_\kappa} \frac{\partial}{\partial\beta'} P^{(k)}\left(s = \alpha^j|\mathcal{J}_{o_k(i)};\bar{\beta}\right)\left(\beta - \beta_0\right) h_k\left(X\right) + \sum_{k=1}^{N_\kappa} P^{(k)}\left(s = \alpha^j|\mathcal{J}_{o_k(i)};\beta_0\right)\left(h_{0,k}\left(X\right) - h_k\left(X\right)\right),
\end{aligned}$$

for some $\bar{\beta} = \left(1 - \bar{\tau}\right)\beta_0 + \bar{\tau}\beta$ with $\bar{\tau} \in [0,1]$. By definition of $\frac{d\ell_j(Z,\theta_0)}{d\theta}\left[\theta - \theta_0\right]$ and $E\left[\ell\left(Z,\theta\right)|X\right]$,

$$\begin{aligned}
c_1 E\left[E\left[\ell\left(Z,\theta\right)|X\right]' E\left[\ell\left(Z,\theta\right)|X\right]\right] &\leq \|\theta - \theta_0\|^2 \\
&\leq c_2 E\left[E\left[\ell\left(Z,\theta\right)|X\right]' E\left[\ell\left(Z,\theta\right)|X\right]\right],
\end{aligned}$$

holds for some $c_1, c_2 > 0$ if $\frac{\partial}{\partial\beta'} P^{(k)}\left(s = \alpha^j|\mathcal{J}_{o_k(i)};\bar{\beta}\right)$ is uniformly bounded. Assumptions 4.2–4.3 imply that $E\left[E\left[\ell\left(Z,\theta\right)|X\right]' E\left[\ell\left(Z,\theta\right)|X\right]\right]$ and $\|\theta - \theta_0\|^2$ are almost equivalent for all $\theta \in \Theta_n$ with $\|\theta - \theta_0\|_s^2 = o\left(1\right)$.

The next set of conditions, Assumptions 4.1–4.6 of Ai and Chen (2003), is relevant to the asymptotic distribution of the SMD estimator. Assumption 4.1 of Ai and Chen (2003) is directly verified by Assumption 4.6. The remaining parts of the proof rely on the smoothness conditions implied by Assumptions 4.2–4.3. The Hölder smoothness of Assumption 4.2 and the corresponding smoothness of functions in the sieve space proves Assumption 4.2. Denote the effective parameter space $\mathcal{N}_{os}$ and its sieve space $\mathcal{N}_{osn}$ by

$$\begin{aligned}
\mathcal{N}_{os} &\equiv \left\{\theta \in \Theta|\ \|\theta - \theta_0\| \leq M_n\delta_n, \|\theta - \theta_0\|_s \leq M_n\delta_{s,n}\right\} \\
\mathcal{N}_{osn} &\equiv \mathcal{N}_{os} \cap \Theta_n,
\end{aligned}$$

with $M_n = \min\left\{\log\log\left(n+1\right), \log\left(\delta_{s,n}^{-1} + 1\right)\right\}$ for some sequences $M_n\delta_n = o\left(1\right)$ and $M_n\delta_{s,n} = M_n O\left(K_n^{-\gamma/d_x}\right) = o\left(n^{-1/4}\right)$.

Define $\bar{\mathcal{W}} = \bar{\mathcal{H}} - \{h_0\}$ where $\bar{\mathcal{H}}$ is the closure of the linear span of $\mathcal{H}$. For each $j \in \{1, \ldots, d_\beta\}$

where d_β is the dimension of β_0, and $w^j = \left(w_1^j, \ldots, w_{N_\kappa}^j\right) \in \bar{\mathcal{W}}$ and $\beta = \left(\beta_1, \ldots, \beta_{d_\beta}\right)' \in \mathcal{B}$, let

$$w^{j*} = \arg\min_{w^j \in \bar{\mathcal{W}}} E\left[\Delta_j\left(X, w^j\right)' \Delta_j\left(X, w^j\right) | X\right]$$

where

$$\begin{aligned}
&\Delta_j\left(X, w^j\right) \\
&\equiv \begin{pmatrix} \sum_{k=1}^{N_\kappa}\left(P^{(k)}\left(s=\alpha^1|\mathcal{J}_{o_k(i)};\beta_0\right) w_k^j(X) - \frac{\partial}{\partial\beta_j} P^{(k)}\left(s=\alpha^1|\mathcal{J}_{o_k(i)};\beta_0\right) h_{0,k}(X)\right) \\ \cdots \\ \sum_{k=1}^{N_\kappa}\left(P^{(k)}\left(s=\alpha^{(L+1)^N-1}|\mathcal{J}_{o_k(i)};\beta_0\right) w_k^j(X) - \frac{\partial}{\partial\beta_j} P^{(k)}\left(s=\alpha^{(L+1)^N-1}|\mathcal{J}_{o_k(i)};\beta_0\right) h_{0,k}(X)\right) \\ \sum_{k=1}^{N_\kappa} w_k^j(X) \end{pmatrix}.
\end{aligned}$$

Denote

$$\begin{aligned}
w^* &= \left(w^{1*}, \ldots, w^{d_\beta *}\right) \\
D_{w^*}(X) &= \left[\Delta_1\left(X, w^{1*}\right), \ldots, \Delta_{d_\beta}\left(X, w^{d_\beta *}\right)\right],
\end{aligned}$$

then $D_{w^*}(X)$ is the gradient matrix used for asymptotic normality of $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$. Since the equilibrium probability $P^{(k)}\left(s=\alpha^1|\mathcal{J}_{o_k(i)};\beta_0\right)$ is twice continuously differentiable by Assumption 4.6 and w^{j*} is Hölder continuous by the selected sieve basis function, Assumption 4.3 holds with w^*. The envelope condition for $\Delta_j\left(X, w^j\right)$ is satisfied because for each element of $\Delta_j\left(X, w^j\right)$ with a fixed $j \in \left\{1, \ldots, (L+1)^N\right\}$,

$$\begin{aligned}
&\left|\sum_{k=1}^{N_\kappa}\left(P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right) w_k^{j*}(X) - \frac{\partial}{\partial\beta_j} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right) h_{0,k}(X)\right)\right| \\
&\le N_\kappa \max_{k\in\{1,\ldots,N_\kappa\}}\left(\left|P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right) w_k^{j*}(X)\right| + \left|\frac{\partial}{\partial\beta_j} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right) h_{0,k}(X)\right|\right) \\
&\le N_\kappa\left(1 + C_2\right) < \infty,
\end{aligned}$$

where C_2 is a finite number in Section C.1. For the next step, recall that $\frac{\partial}{\partial\beta'} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)$ and $\frac{\partial^2}{\partial\beta'\partial\beta} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)$ are uniformly bounded on $\beta \in \mathcal{B}$ by Assumption 4.6−2. Also by Assumption 4.6−3,

$$\begin{aligned}
&E\left[\frac{\partial}{\partial\beta'} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right) - \frac{\partial}{\partial\beta'} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)\right] \\
&= E\left[\frac{\partial}{\partial\beta'\partial\beta} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\bar{\beta}\right)(\beta - \beta_0)\right] \\
&\le E\left[\left\|\frac{\partial}{\partial\beta'\partial\beta} P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\bar{\beta}\right)\right\|_E\right] \|\beta - \beta_0\|_E \le C_3 M_n \delta_{s,n} = o\left(n^{-1/4}\right),
\end{aligned}$$

for some finite number C_3 and $\bar{\beta} = (1-\tau)\beta_0 + \tau\beta$. Assumption 4.4 of Ai and Chen (2003) follows that

$$
\begin{aligned}
& E\left[\sum_{k=1}^{N_\kappa}\left(\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)\beta h_k(X) - \frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)\beta_0 h_{0,k}(X)\right)\right] \\
& \quad + E\left[\sum_{k=1}^{N_\kappa}\left(P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)h_k(X) - P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)h_{0,k}(X)\right)\right] \\
& \leq N_\kappa \max_{k\in\{1,\ldots,N_\kappa\}} E\left[\left|\left(\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right) - \frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)\right)\beta\right| h_k(X)\right] \\
& \quad + N_\kappa \max_{k\in\{1,\ldots,N_\kappa\}} E\left[\left|\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)(\beta-\beta_0)\right| h_k(X)\right] \\
& \quad + N_\kappa \max_{k\in\{1,\ldots,N_\kappa\}} E\left[\left|\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)\beta_0\right| |h_k(X) - h_{0,k}(X)|\right] = o\left(n^{-1/4}\right),
\end{aligned}
$$

because $|h_k(X)| \leq 1$, $\|\beta-\beta_0\|_E \leq M_n\delta_{s,n} = o\left(n^{-1/4}\right)$, and $\sup_{x\in\mathcal{X}}|h_k(x) - h_{0,k}(x)| \leq M_n\delta_{s,n} = o\left(n^{-1/4}\right)$. Assumptions 4.5 and 4.6 hold similarly since the equilibrium probability is twice continuously differentiable by Assumption 4.6−2. For $\theta_0, \theta \in \mathcal{N}_{os}$, and $\bar{\theta} \in \mathcal{N}_{osn}$, it is sufficient to show that

$$
\begin{aligned}
& E\left[\left|\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)\left(\beta_0-\bar{\beta}\right)h_k(X) - \frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)\left(\beta_0-\bar{\beta}\right)h_{0,k}(X)\right|\right] \\
& + E\left[\left|\left(P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right) - P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta_0\right)\right)\left(h_{0,k}(X) - \bar{h}_k(X)\right)\right|\right] = o\left(n^{-1/2}\right), \quad (12)
\end{aligned}
$$

for any equilibrium probability with a joint action α and for any kth component of the unobserved heterogeneity.

The first term of equation (12) is bounded by

$$
\begin{aligned}
& E\left[\left\|\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\beta\right)\right\|_E \left\|\beta_0-\bar{\beta}\right\|_E \|h_k - h_{0,k}\|_\infty\right] \\
& + E\left[\left\|\frac{\partial}{\partial\beta'\partial\beta}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\bar{\bar{\beta}}\right)\right\|_E \left\|\beta_0-\bar{\bar{\beta}}\right\|_E \left\|\beta_0-\bar{\beta}\right\|_E\right] = o\left(n^{-1/2}\right),
\end{aligned}
$$

where $\bar{\bar{\beta}}$ is a convex combination of β_0 and $\bar{\beta}$, and $\|\cdot\|_\infty$ is the supremum norm. The second term of equation (12) is also bounded by

$$
E\left[\left\|\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\check{\beta}\right)\right\|_E \|\beta_0-\beta_0\|_E \left\|\bar{h}_k - h_{0,k}\right\|_\infty\right] = o\left(n^{-1/2}\right),
$$

since $\left\|\frac{\partial}{\partial\beta'}P^{(k)}\left(s=\alpha|\mathcal{J}_{o_k(i)};\check{\beta}\right)\right\|_E$ for some $\check{\beta} = (1-\tau)\beta_0 + \tau\beta$ is uniformly bounded by a constant.

The asymptotic distribution with a closed form asymptotic variance is by Theorem 4.1 of Ai and Chen (2003),

$$
\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{d} N(0, V),
$$

where

$$V = E\left[D_{w^*}(X)' D_{w^*}(X)\right]^{-1} E\left[D_{w^*}(X)' E\left[\rho(Z,\theta_0)\rho(Z,\theta_0)' | X\right] D_{w^*}(X)\right] E\left[D_{w^*}(X)' D_{w^*}(X)\right]^{-1}.$$

C.3 Proof of Theorem 4.3

This section verifies additional conditions of asymptotic normality in Chen and Pouzo (2015). I demonstrate Assumptions A.4−A.7 of Chen and Pouzo (2015) hold for the current paper, and verify Assumption 3.5 of Chen and Pouzo (2015). Assumptions 3.1−3.4 of Chen and Pouzo (2015) are implied by the conditions in Sections C.1 and C.2: see Corollaries A.1, A.2, and B.1 of Ai and Chen (2003). The parameter of interest is $h_{0,k}(x)$ for some $x \in \mathcal{X}$. Consider the inferential statistic

$$\frac{\sqrt{n}\left(\hat{h}_{k,n}(x) - h_{0,k}(x)\right)}{\|\nu_n^*\|_{sd,h}},$$

provided with a closed form sieve representer

$$\nu_n^* = H_{h_k}(\cdot)' D_n^{-1} G_{h_k}$$
$$H_{h_k}(\cdot) \equiv \left[1_{1\times d_\beta}, \underbrace{p^{K_n'}(\cdot), \ldots, p^{K_n'}(\cdot)}_{1\times N_\kappa K_n}\right]'.$$

Assumptions A.4 and A.5 of Chen and Pouzo (2015) are implied by Assumptions 4.1−4.6 of the current paper. The compactness of $\mathcal{X}$ and bounded eigenvalues of $E\left[p^{J_n}(X) p^{J_n}(X)'\right]$ are from Assumption 4.1. The uniform boundedness of $\sup_{x\in\mathcal{X}} |p_j(x)|$is by the specification of B-spline basis function defined on a compact set. The convergence rate $J_n \log(J_n) = o(n)$ holds with Assumption 4.5.

Next, for any $\delta > 0$, the smoothness assumption establishes $\ell(Z,\theta) - \ell(Z,\theta') = \frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)(\theta' - \theta)$ for some $\bar{\theta}$ which is a convex combination of θ and θ'. Note that $\left\|\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)(\theta'-\theta)\right\|_E \leq \left\|\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)(\theta'-\theta)\right\|_E \|\theta'-\theta\|_s \leq C\|\theta'-\theta\|_s$ for some positive constant $C_4 > 0$. Each $\ell_j(Z,\theta)$ is continuously differentiable and the derivative function is defined on a compact support, so there exists a uniform upper bound of $\left\|\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)(\theta'-\theta)\right\|_E$. Therefore,

$$E\left[\sup_{\theta\in\mathcal{N}_{osn}, \|\theta-\theta'\|_s\leq\delta} \|\ell(Z,\theta) - \ell(Z,\theta')\|_E^2 | X = x\right] \leq C_4^2 \|\theta'-\theta\|_s^2 \leq C_4^2\delta^2$$

satisfies the Condition A.5−(ii) of Chen and Pouzo (2015) by setting $\kappa = 1$ and $K(x) = C_4$. The Condition A.5−(iii), $n\delta_n^2 M_n \delta_{s,n} \max\left\{M_n\delta_{s,n}\sqrt{C_n}, M_n\right\} = o(1)$ where

$$\sqrt{C_n} \equiv \int_0^1 \sqrt{1 + \log\left(N\left(w\left(M_n\delta_{s,n}\right), \mathcal{L}_{on}, \|\cdot\|_{L^2}\right)\right)} dw,$$

with $\mathcal{L}_{on} \equiv \{\ell(\cdot,\theta) - \ell(\cdot,\theta_0) : \theta \in \mathcal{N}_{osn}\}$ also holds because $\sqrt{C_n} = o(n^{\alpha})$ for any $\alpha > 0$ by the sieve space specification used in Section C.2, $\delta_{s,n} = O\left(K_n^{-\gamma/d_x}\right) = o\left(n^{-1/4}\right)$, $M_n \approx \log\log(n+1)$, and $\delta_n = o\left(n^{-1/4}\right)$.

The remaining parts in Assumptions A.6 and A.7 of Chen and Pouzo (2015) are already verified in Sections C.1−C.2. Corollaries C.1−C.3 of Ai and Chen (2003) prove the conditions in Assumptions A.6 and A.7 focusing on regular functionals. In the case of regular functionals, $\|\nu_n^*\| \to \|\nu^*\| < \infty$. In this section, an irregular functional $h_{0,k}(x)$ is the parameter of interest so that the scaled sieve representer $u_n^* = \nu_n^* / \|\nu_n^*\|_{sd,h}$. Thus it is sufficient to show that $\|u_n^*\| = \|v_n^*\| / \|\nu_n^*\|_{sd,h} \to \|\nu^*\| / \|\nu^*\|_{sd,h} < \infty$ under the norms $\|\cdot\|$ and $\|\cdot\|_{sd,h}$. Compare the ratio

$$\|u_n^*\| = \frac{\|v_n^*\|}{\|\nu_n^*\|_{sd,h}} = \frac{G_{h_k}' D_n^{-1} D_n D_n^{-1} G_{h_k}}{G_{h_k}' D_n^{-1} \Psi_n D_n^{-1} G_{h_k}},$$

and note that $\Psi_n = E\left[\Delta_n(X)' \ell(Z,\theta_0) \ell(Z,\theta_0)' \Delta_n(X)\right] \leq 4E\left[\Delta_n(X)' \Delta_n(X)\right]$ since $\ell_j(Z,\theta)$ is uniformly bounded by 2 for all $j = 1, \ldots, (L+1)^N$, Z, and θ. $\Sigma_0(X) = E\left[\ell(Z,\theta_0) \ell(Z,\theta_0)' | X\right]$ also has the smallest eigenvalue uniformly bounded away from zero by Assumption 4.6−3, so that $\|u_n^*\|$ is always bounded and converges to a finite value.

Assumption 3.5 of Chen and Pouzo (2015) holds because the parameter of interest $h_{0,k}(x)$ is a linear functional of the structural parameters and each $h_{0,k}$ is well approximated by Assumption 4.2. Assumption 3.5−(iii) is simply derived by $\|h_{0,k}(x) - h_{k,n}(x)\|_{\infty} = o\left(n^{-1/4}\right)$ and $\|\nu_n^*\| = O\left(n^{1/4}\right)$ by Assumption 4.5−3. Lastly, the Lindeberg condition applies to Assumption 3.6−(ii). Define

$$\mathcal{S}_n^* = \left(\frac{d\ell(Z,\theta_0)}{d\theta}[v_n^*]\right)' \ell(Z,\theta_0),$$

then Assumptions 4.5−4.6 imply that $E\left[\left|\frac{\mathcal{S}_n^*}{\|\nu_n^*\|_{sd,h}}\right|\right]$ is uniformly bounded by a constant. For $\epsilon > 0$,

$$\limsup_{n\to\infty} E\left[\left(\frac{\mathcal{S}_n^*}{\|\nu_n^*\|_{sd,h}}\right)^2 1\left\{\left|\frac{\mathcal{S}_n^*}{\epsilon\sqrt{n}\,\|\nu_n^*\|_{sd,h}}\right| > 1\right\}\right] = 0$$

In result, Theorem 4.3 is proved.

C.4 Proof of Theorem 4.4

In this section I demonstrate a set of sufficient conditions for Theorem 4.4. Assumption 4.1 of Chen and Pouzo (2015) summarizes the required conditions in addition to the conditions verified in Sections C.2 and C.3. The current paper assumes a linear functional $h_{0,k}(x)$ for $x \in \mathcal{X}$ as a parameter of interest. The weight matrix of the proposed SMD estimator is an identity matrix. Thus Assumptions 4.1−(i) and (iii) directly match to the setup in this paper.

Define the plug-in estimator for the sieve gradient matrix $\hat{\Delta}_n(X)$ by

$$\hat{\Delta}_n(X) \equiv \left[\hat{\Delta}_{\beta_0}(X), \hat{\Delta}_{h_0,n}(X)\right]$$
$$\hat{\Delta}_{\beta_0}(X) \equiv \sum_{k=1}^{N_\kappa} \frac{\partial}{\partial \beta'} P^{(k)}\left(X, O; \hat{\beta}_n\right) \hat{h}_n(X)$$
$$\hat{\Delta}_{h_0,n}(X) \equiv \left[P^{(1)}\left(X, O; \hat{\beta}_n\right) \otimes p^{K_n'}(X), \ldots, P^{(N_\kappa)}\left(X, O; \hat{\beta}_n\right) \otimes p^{K_n'}(X)\right],$$

and $\bar{V}_n = \left(\mathcal{B} \times \bar{\mathcal{H}}_n\right) - \{\theta_0\}$ where $\bar{\mathcal{H}}_n$ is the closure of the linear span of $\mathcal{H}_n$. Denote $\bar{V}_n^1 \equiv \left\{v \in \bar{V}_n | \, \|v\| = 1\right\}$.

Firstly I prove Assumption 4.1−(ii) and (v) of Chen and Pouzo (2015):

$$\sup_{v_1, v_2 \in \bar{V}_n^1} \left|\langle v_1, v_2\rangle_{n,I} - \langle v_1, v_2\rangle_I\right| = o_p(1)$$
$$\sup_{v \in \bar{V}_n^1} \left|\langle v, v\rangle_{n,\Sigma} - \langle v, v\rangle_\Sigma\right| = o_p(1),$$

where $I = I_{n,m}$ is an identity matrix, $\Sigma_{n,m} = \ell\left(Z^m, \hat{\theta}_n\right) \ell\left(Z^m, \hat{\theta}_n\right)'$, $\Sigma = \ell(Z, \theta_0) \ell(Z, \theta_0)'$, and

$$\langle v_1, v_2\rangle_{n,\Pi} \equiv \frac{1}{n} \sum_{m=1}^n \left(\frac{d\ell\left(Z^m, \hat{\theta}_n\right)}{d\theta}[v_1]\right)' \Pi_{n,m} \left(\frac{d\ell\left(Z^m, \hat{\theta}_n\right)}{d\theta}[v_2]\right)$$
$$\langle v_1, v_2\rangle_\Pi \equiv E\left[\left(\frac{d\ell(Z, \theta_0)}{d\theta}[v]\right)' \Pi \left(\frac{d\ell(Z, \theta_0)}{d\theta}[v]\right)\right].$$

Denote that $\sup_{Z \in \mathcal{Z}, \theta' \in \Theta} \sup_{v \in \bar{V}_n^1} \left|\frac{d\ell(Z,\theta')}{d\theta}[v]\right| \leq C_6$ for a positive constant $C_6 > 0$ by the smoothness condition in Assumptions 4.2−4.3. Also for some $\bar{\theta}_n$ which is a convex combination of θ_0 and $\hat{\theta}_n$ and $C_7 > 0$,

$$\begin{aligned}
\left\|\frac{d\ell\left(Z, \hat{\theta}_n\right)}{d\theta}[v] - \frac{d\ell(Z, \theta_0)}{d\theta}[v]\right\|_E = \left\|\frac{d\ell\left(Z, \bar{\theta}_n\right)}{d\theta d\theta'}\left(\hat{\theta}_n - \theta_0\right)[v]\right\|_E &\leq \left\|\frac{d\ell\left(Z, \bar{\theta}_n\right)}{d\theta d\theta'}\right\|_E \|v\| \left\|\hat{\theta}_n - \theta_0\right\|_s \\
&\leq \left\|\frac{d\ell\left(Z, \bar{\theta}_n\right)}{d\theta d\theta'}\right\|_E \left\|\hat{\theta}_n - \theta_0\right\|_s \\
&\leq C_7 \left\|\hat{\theta}_n - \theta_0\right\|_s,
\end{aligned}$$

by the Hölder condition. Also for $\bar{\Sigma} = \ell(Z, \theta) \ell(Z, \theta)'$, $\sup_{Z,\theta} \left\|\bar{\Sigma}\right\|_E \leq C_8$ by Assumption (4.6) that the smallest and the largest eigenvalues of $\Sigma_0(X)$ are bounded and bounded away from zero.

For any $v_1, v_2 \in \bar{V}_n^1$,

$$\begin{aligned}
&\sup_{v_1,v_2\in\bar{V}_n^1}\left|\langle v_1,v_2\rangle_{n,I}-\langle v_1,v_2\rangle_I\right| \\
&\leq \sup_{v_1,v_2\in\bar{V}_n^1}\left|\langle v_1,v_2\rangle_{n,I}-\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_1]\right)'\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_2]\right)\right| \\
&\quad+\sup_{v_1,v_2\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_1]\right)'\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_2]\right)-\langle v_1,v_2\rangle_I\right|,
\end{aligned} \tag{13}$$

where the first term of equation (13) is bounded by

$$\begin{aligned}
&\sup_{v_1,v_2\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v_1]-\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_1]\right)'\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v_2]\right)\right| \\
&+\sup_{v_1,v_2\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_1]\right)'\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v_2]-\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_2]\right)\right| \\
&\leq C_6C_7\left\|\hat{\theta}_n-\theta_0\right\|_s+C_6C_7\left\|\hat{\theta}_n-\theta_0\right\|_s\leq 2C_6C_7\left\|\hat{\theta}_n-\theta_0\right\|_s=o_p\left(1\right),
\end{aligned}$$

and the second term of equation (13) also uniformly converges to zero by the Uniform Law of Large Numbers. Applying the Lemma A.1 of Ai and Chen (2003),

$$\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_1]\right)'\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v_2]\right)=\langle v_1,v_2\rangle_I+o_p\left(1\right),$$

uniformly over $v_1, v_2 \in \bar{V}_n^1$ under Assumption 4.1.

Next,

$$\begin{aligned}
&\sup_{v\in\bar{V}_n^1}\left|\langle v,v\rangle_{n,\Sigma}-\langle v,v\rangle_\Sigma\right| \\
&\leq \sup_{v\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v]\right)'\Sigma_{n,m}\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v]\right)-\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)'\Sigma\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)\right| \\
&\quad+\sup_{v\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)'\Sigma\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)-\langle v,v\rangle_\Sigma\right|,
\end{aligned} \tag{14}$$

where the first term of equation (14) is bounded by

$$\begin{aligned}
&\sup_{v\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v]\right)'\Sigma_{n,m}\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v]-\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)\right|\\
&+\sup_{v\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\hat{\theta}_n\right)}{d\theta}[v]-\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)'\Sigma_{n,m}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)\right|\\
&+\sup_{v\in\bar{V}_n^1}\left|\frac{1}{n}\sum_{m=1}^{n}\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)'\left(\Sigma_{n,m}-\Sigma\right)\left(\frac{d\ell\left(Z^m,\theta_0\right)}{d\theta}[v]\right)\right|\\
\le{}& C_6C_7C_8\left\|\hat{\theta}_n-\theta_0\right\|_s+C_6C_7C_8\left\|\hat{\theta}_n-\theta_0\right\|_s+C_6^2C_7^2\left\|\hat{\theta}_n-\theta_0\right\|_s^2=o_p\left(1\right).
\end{aligned}$$

The second term of equation (14) is also uniformly bounded by the uniform convergence in Lemma A.1 of Ai and Chen (2003). The proof is the same as the case of equation (13), provided that the smallest and largest eigenvalues of $E\left[\ell\left(Z,\theta_0\right)\ell\left(Z,\theta_0\right)'|X\right]$ are bounded and bounded away from zero. The smoothness condition of ℓ and the uniform law of large numbers imply the desired results.

The last part is Assumption 4.1−(iv) of Chen and Pouzo (2015). Each $\ell_j\left(Z,\theta\right)=\ell_j\left(Z,\theta_0\right)+\frac{\partial}{\partial\theta'}\ell_j\left(Z,\bar{\theta}^j\right)\left(\theta-\theta_0\right)$ where $\bar{\theta}^j$ is a convex combination of θ_0 and θ. The derivative $\frac{\partial}{\partial\theta'}\ell_j\left(Z,\bar{\theta}^j\right)$ exists for $\bar{\theta}^j\in\Theta$ as the equilibrium probability matrix is a smooth function by Assumption 4.3. Using the approximation for $\ell_j\left(Z,\theta\right)$ for $j=1,\ldots,\left(L+1\right)^N$, define $\ell\left(Z,\theta\right)=\ell\left(Z,\theta_0\right)+\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)\left(\theta-\theta_0\right)$. Since

$$\begin{aligned}
&\ell\left(Z,\theta\right)\ell\left(Z,\theta\right)'-\ell\left(Z,\theta_0\right)\ell\left(Z,\theta_0\right)'\\
&=\ell\left(Z,\theta\right)\left(\ell\left(Z,\theta\right)-\ell\left(Z,\theta_0\right)\right)'+\left(\ell\left(Z,\theta\right)-\ell\left(Z,\theta_0\right)\right)\ell\left(Z,\theta_0\right)'\\
&=\ell\left(Z,\theta\right)\left(\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)\left(\theta-\theta_0\right)\right)'+\left(\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)\left(\theta-\theta_0\right)\right)\ell\left(Z,\theta_0\right)',
\end{aligned}$$

the Euclidean norm of the above expression satisfies

$$\begin{aligned}
&\left\|\ell\left(Z,\theta\right)\left(\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)\left(\theta-\theta_0\right)\right)'+\left(\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)\left(\theta-\theta_0\right)\right)\ell\left(Z,\theta_0\right)'\right\|_E\\
&\le 2\left\|\ell\left(Z,\theta\right)\right\|_E\left\|\frac{\partial}{\partial\theta'}\ell\left(Z,\bar{\theta}\right)\right\|_E\left\|\theta-\theta_0\right\|_s\le C\left\|\theta-\theta_0\right\|_s.
\end{aligned}$$

Each $\ell_j\left(Z,\theta\right)$ is bounded by 2, and the derivative of $\ell_j\left(Z,\theta\right)$ is also uniformly bounded due to the smoothness of the equilibrium probability (Assumption 4.3). Then

$$\sup_{x\in\mathcal{X}}E\left[\sup_{\theta\in\mathcal{N}_{osn}}\left\|\ell\left(Z,\theta\right)\ell\left(Z,\theta\right)'-\ell\left(Z,\theta_0\right)\ell\left(Z,\theta_0\right)'\right\|_E|X=x\right]\le C\sup_{\theta\in\mathcal{N}_{osn}}\left\|\theta-\theta_0\right\|_E=o\left(1\right).$$

Theorem 4.2 of Chen and Pouzo (2015) provides that the test statistic in Theorem 4.4 with a consistent plug-in variance-covariance matrix estimator is asymptotically normal under the assumptions verified above.

References

[1] Aguirregabiria, V., & Mira, P. (2002). Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models. *Econometrica*, 70(4), 1519-1543.

[2] Aguirregabiria, V., & Mira, P. (2007). Sequential Estimation of Dynamic Discrete games. *Econometrica*, 75(1), 1-53.

[3] Aguirregabiria, V., & Mira, P. (2019). Identification of Games of Incomplete Information with Multiple Equilibria and Unobserved Heterogeneity. *Quantitative Economics*, 10(4), 1659-1701.

[4] Ai, C., & Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, 71(6), 1795-1843.

[5] Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 3099-3132.

[6] Aradillas-López, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics*, 157(2), 409-431.

[7] Aradillas-López, A., & Gandhi, A. (2016). Estimation of games with ordered actions: An application to chain-store entry. *Quantitative Economics*, 7(3), 727-780.

[8] Arcidiacono, P., & Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6), 1823-1867.

[9] Bajari, P., Benkard, C. L., & Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, 75(5), 1331-1370.

[10] Bajari, P., Hong, H., Krainer, J., & Nekipelov, D. (2010). Estimating static models of strategic interactions. *Journal of Business & Economic Statistics*, 28(4), 469-482.

[11] Bajari, P., Hong, H., & Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5), 1529-1568.

[12] Berry, S. T. (1992). Estimation of a Model of Entry in the Airline Industry. *Econometrica*: Journal of the Econometric Society, 889-917.

[13] Bhaskar, V., Mailath, G. J., & Morris, S. (2013). A foundation for Markov equilibria in sequential games with finite social memory. *The Review of Economic Studies*, 80(3), 925-948.

[14] Blevins, J. R. (2015). Structural Estimation of Sequential Games of Complete Information. *Economic Inquiry*, 53(2), 791-811.

[15] Brendstrup, B., & Paarsch, H. J. (2006). Identification and estimation in sequential, asymmetric, english auctions. *Journal of Econometrics*, 134(1), 69-94.

[16] Bresnahan, T. F., & Reiss, P. C. (1990). Entry in monopoly market. *The Review of Economic Studies*, 57(4), 531-553.

[17] Bresnahan, T. F., & Reiss, P. C. (1991). Empirical models of discrete games. *Journal of Econometrics*, 48(1), 57-81.

[18] Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. Handbook of Econometrics, 6, 5549-5632.

[19] Chen, X., & Pouzo, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, 152(1), 46-60.

[20] Chen, X., & Pouzo, D. (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica*, 83(3), 1013-1079.

[21] Collard-Wexler, A., Gowrisankaran, G., & Lee, R. S. (2019). "Nash-in-Nash" bargaining: A microfoundation for applied work. *Journal of Political Economy*, 127(1), 163-195.

[22] Crawford, G. S., & Yurukoglu, A. (2012). The Welfare Effects of Bundling in Multichannel Television Markets. *American Economic Review*, 102(2), 643-85.

[23] Davis, P. (2006). Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics*, 37(4), 964-982.

[24] De Paula, A., & Tang, X. (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica*, 80(1), 143-172.

[25] Einav, L. (2010). Not all rivals look alike: Estimating an equilibrium model of the release date timing game. *Economic Inquiry*, 48(2), 369-390.

[26] Ellickson, P. B., Houghton, S., & Timmins, C. (2013). Estimating network economies in retail chains: a revealed preference approach. *The RAND Journal of Economics*, 44(2), 169-193.

[27] Ericson, R., and Pakes, A. (1995). Markov-Perfect Industry Dynamics: A Framework for Empirical Work, *Review of Economic Studies*, 62, 53-83.

[28] Fudenberg, D., & Tirole, J. (1991). Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2), 236-260.

[29] Gal-Or, E. (1997). Exclusionary Equilibria in Health-Care Markets. *Journal of Economics & Management Strategy*, 6(1), 5-43.

[30] Grieco, P. L. (2014). Discrete games with flexible information structures: An application to local grocery markets. *The RAND Journal of Economics*, 45(2), 303-340.

[31] Haller, H., & Lagunoff, R. (2000). Genericity and Markovian behavior in stochastic games. *Econometrica*, 68(5), 1231-1248.

[32] Ho, K. (2009). Insurer-Provider Networks in the Medical Care Market. *The American Economic Review,* 99(1), 393-430.

[33] Holmes, T. J. (2001). Bar codes lead to frequent deliveries and superstores. *RAND Journal of Economics*, 708-725.

[34] Holmes, T. J. (2011). The Diffusion of Wal-Mart and Economies of Density. *Econometrica*, 79(1), 253-302.

[35] Hotz, V. J., & Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3), 497-529.

[36] Hu, Y., & Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, 171(1), 32-44.

[37] Igami, M., & Yang, N. (2016). Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada. *Quantitative Economics*, 7(2), 483-521.

[38] Jia, P. (2008). What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. *Econometrica*, 76(6), 1263-1316.

[39] Kasahara, H., & Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77(1), 135-175.

[40] Kline, B. (2015). Identification of complete information games. *Journal of Econometrics*, 189(1), 117-131.

[41] Kreps, D. M., & Wilson, R. (1982). Reputation and imperfect information. *Journal of economic theory*, 27(2), 253-279.

[42] Lewbel, A., & Tang, X. (2015). Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics*, 189(1), 229-244.

[43] Maskin, E., & Tirole, J. (2001). Markov perfect equilibrium: I. Observable actions. *Journal of Economic Theory*, 100(2), 191-219.

[44] Milgrom, P., & Roberts, J. (1982). Predation, reputation, and entry deterrence. *Journal of Economic Theory*, 27(2), 280-312.

[45] Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5), 1565-1578.

[46] Otsu, T., Pesendorfer, M., & Takahashi, Y. (2016). Pooling data across markets in dynamic Markov games. *Quantitative Economics*, 7(2), 523-559.

[47] Pakes, A., & McGuire, P. (1994). Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model. *The Rand Journal of Economics*, 555-589.

[48] Pakes, A., Ostrovsky, M., & Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics*, 38(2), 373-399.

[49] Pakes, A., Porter, J., Ho, K., & Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83(1), 315-334.

[50] Pesendorfer, M., & Schmidt-Dengler, P. (2008). Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies*, 75(3), 901-928.

[51] Pinkse, J., Slade, M. E., & Brett, C. (2002). Spatial price competition: a semiparametric approach. *Econometrica*, 70(3), 1111-1153.

[52] Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, 37(3), 619-640.

[53] Smith, H. (2004). Supermarket choice and supermarket competition in market equilibrium. *The Review of Economic Studies*, 71(1), 235-263.

[54] Sperisen, B. (2018). Bounded memory and incomplete information. *Games and Economic Behavior*, 109, 382-400.

[55] Sweeting, A. (2009). The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *The RAND Journal of Economics*, 40(4), 710-742.

[56] Tang, X. (2010). Estimating simultaneous games with incomplete information under median restrictions. *Economics Letters,* 108(3), 273-276.

[57] Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1), 147-165.

[58] Vitorino, M. A. (2012). Empirical entry games with complementarities: An application to the shopping center industry. *Journal of Marketing Research*, 49(2), 175-191.

[59] Wan, Y., & Xu, H. (2014). Semiparametric identification of binary decision games of incomplete information with correlated private signals. *Journal of Econometrics*, 182(2), 235-246.

[60] Xiao, R. (2018). Identification and estimation of incomplete information games with multiple equilibria. *Journal of Econometrics*, 203(2), 328-343.