Essays on Econometric Analysis of Game-theoretic Models

Paul Sungwook Koh

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2022

Abstract

Essays on Econometric Analysis of Game-theoretic Models

Paul Sungwook Koh

This dissertation studies econometric analysis of game-theoretic models. I develop novel empirical models and methodologies to facilitate robust and computationally tractable econometric analysis. In Chapter 1, I develop an empirical model for analyzing *stable outcomes* in the presence of incomplete information. Empirically, many strategic settings are characterized by stable outcomes in which players' decisions are publicly observed, yet no player takes the opportunity to deviate. To analyze such situations, I build an empirical framework by introducing a novel solution concept that I call *Bayes stable equilibrium*. The framework allows the researcher to be agnostic about players' information and the equilibrium selection rule. Furthermore, I show that the Bayes stable equilibrium identified set is always weakly tighter than the Bayes correlated equilibrium identified set; numerical examples show that the shrinkage can be substantial. I propose computationally tractable approaches for estimation and inference and apply the framework to study the strategic entry decisions of McDonald's and Burger King in the US.

In Chapter 2, I study identification and estimation of a class of dynamic games when the underlying information structure is unknown to the researcher. I introduce *Markov correlated equilibrium*, a dynamic analog of Bayes correlated equilibrium studied in Bergemann and Morris (2016), and show that the set of Markov correlated equilibrium predictions coincides with the set of Markov perfect equilibrium predictions that can arise when the players might

observe more signals than assumed by the analyst. I propose an econometric approach for estimating dynamic games with weak assumption on players' information using Markov correlated equilibrium. I also propose multiple computational strategies to deal with the non-convexities that arise in dynamic environments.

In Chapter 3, I propose an extremely fast and simple approach to estimating static discrete games of complete information under pure strategy Nash equilibrium and no assumptions on the equilibrium selection rule. I characterize an identified set of parameters using a set of inequalities that are expressed in terms of closed-form multinomial logit probabilities. The key simplifications arise from using a subset of all identifying restrictions that are particularly easy to handle. Under standard assumptions, the identified set is convex and its projections can be obtained via convex programs. Numerical examples show that the identified set is quite tight. I also propose a simple approach to construct confidence sets whose projections can be obtained via convex programs. I demonstrate the usefulness of the approach using real-world data.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank the members of my dissertation committee, Sokbae Lee, Bernard Salanié, Qingmin Liu, Gautam Gowrisankaran, and Matthew Backus. I would like to express my deepest gratitude to my principal advisors Sokbae Lee and Bernard Salanié for their continuous guidance and support. They have served as my role models and helped me shape my way of thinking as an economist. In particular, they taught me how economic theory and econometric methodologies can be rigorously combined to tackle interesting real-world questions, which led me to dive deeper into the topic of structural econometrics. They have also helped me grow as an independent researcher, giving me skills and confidence that will prove to be most valuable in the future. I am also extremely grateful to Qingmin Liu, who have taught me how to think like a theorist. I vividly remember numerous lessons he has given me throughout my PhD journey. He has always encouraged me to ask big questions, and interactions with him led to the development of the first two chapters in my dissertation. I am also grateful to Gautam Gowrisankaran and Matthew Backus for their generous help and feedback.

I have benefited enormously from interactions with numerous faculty members and colleagues at Columbia. In particular, I have received helpful feedback from various participants at the microeconomics colloquium and the econometrics colloquium. I would like to thank them for withstanding many presentations of mine that were unpolished. They have nevertheless stayed interested and engaged, and actively given me various comments that are reflected in my dissertation. Letting me engage in their presentations also helped me contemplate on how to form questions and answers as a scholar.

Finally, I am truly thankful to my dear friends and family for all the support they have given me throughout the past several years. I was able to withstand many tough moments in my PhD journey. It would have been impossible to finish my PhD without them.

Chapter 1: Stable Outcomes and Information in Games: An Empirical Framework

1.1 Introduction

In dynamic strategic settings where firms can react after observing their opponents' choices, our intuitions suggest that firms' actions would change over time. Interestingly, we often see firms reach a certain steady-state in which no firm changes its decision even when it can. For example, major exporters' decisions to export products to specific markets remain unchanged for a long period (Ciliberto and Jäkel, 2021). Airline firms' decisions to operate between cities tend to be persistent (Ciliberto and Tamer, 2009). Food-service retailers operate in a local market over a long horizon, knowing precisely the identities of the competitors operating nearby. In all these examples, each firm's action constitutes a best response to the *observed* actions of the opponents.

The prevalence of incomplete information in the real world makes the phenomenon particularly interesting. When the state of the world is unknown, firms will use all information available to them; this includes the information revealed from their opponents' decisions. For example, while a coffee chain's own research might report that a given neighborhood is an unattractive location, observing that Starbucks—a chain known to have leading market research technology—enter the market may make it think twice.[1] Thus, if there is no further revision of actions after they are realized and observed, it must be that each firm holds information refined by observing the opponents' decisions.

Although stable outcomes in the presence of information asymmetries are common in the real world, it is not straightforward to model the data generating process. The main

[1] According to Tom O'Keefe, the founder of Tully's Coffee, Tully's early business expansion strategy was to "open across the street from every Starbucks" because "they do a great job at finding good locations." (Goll, 2000).

difficulty arises from the requirement that the firms' beliefs and actions must be consistent with each other at the equilibrium situations. On the one hand, if the firms' realized actions are best responses to each other, there must be beliefs that rationalize the actions as optimal. On the other hand, each firm's beliefs must be consistent with its private information about the state of the world *as well as* the information extracted from observing its opponents' decisions. Static Bayes Nash equilibrium, which has been a popular choice for empirical analysis of games with incomplete information, is not applicable because it does not account for the possible information updating and revision of actions after the opponents' actions are observed. Modeling convergence to stable outcomes via a dynamic games framework may be feasible but likely non-trivial and reliant on ad hoc assumptions. In this paper, we aim to develop a tractable equilibrium notion that satisfies the consistency requirement and facilitates econometric analysis when the econometrician observes a cross-section of stable outcomes at some point in time.

We propose a solution concept dubbed *Bayes stable equilibrium* as a basis for analyzing stable outcomes in the presence of incomplete information and argue that it has attractive properties. Bayes stable equilibrium is described as follows. A *decision rule* specifies a distribution over action profiles for each realization of the state of the world and players' private signals. Suppose that, after the state of the world and private signals are realized, an action profile is drawn from the decision rule, and the action profile is *publicly* recommended to the players. The decision rule is a Bayes stable equilibrium if the players always find no incentives to deviate from the publicly recommended action profile after observing their private signals and the action profile.

We justify Bayes stable equilibrium using a version of rational expectations equilibrium à la Radner (1979). First, we argue that rational expectations equilibrium, appropriately defined for our setting, provides a simple approach to rationalizing stable outcomes under incomplete information. We define rational expectations equilibrium by introducing an "outcome function" that maps players' information to action profiles; this approach is motivated by Liu (2020), who uses a similar approach to define the notion of stability in two-sided markets with incomplete information. Next, we show that the set of Bayes stable equilibrium predictions (joint distributions on states, signals, and actions) coincides with the set of rational expectations equilibrium predictions that can arise when the players might have more information than assumed by the analyst. Thus, Bayes stable equilibrium provides a convenient tool for describing the implications of rational expectations equilibria when the analyst only knows the minimal information available to the players; Bayes stable equilibrium is "informationally robust" in the same sense as the Bayes correlated equilibrium of Bergemann and Morris (2016). The informational robustness property of Bayes stable equilibrium is attractive given that it is often difficult for the researcher to know the true information structure governing the data generating process.

Assuming that the econometrician observes a cross-section of stable outcomes, we characterize the identified set of parameters using Bayes stable equilibrium as a solution concept. The Bayes stable equilibrium identified set has a number of attractive properties. First, it is robust to unknown equilibrium selection rules and information structures: the identified set is valid for arbitrary equilibrium selection rules and the possibility that the players actually observed more information than assumed by the econometrician. We let the model be "incomplete" in the sense of Tamer (2003), and the parameters are typically partially identified. Second, when strong assumptions on information are made, the Bayes stable equilibrium identified set collapses to the pure strategy Nash equilibrium identified set studied in Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011). Third, everything else equal, the Bayes stable equilibrium identified set is (weakly) tighter than the Bayes correlated equilibrium identified set studied in Magnolfi and Roncoroni (2021). While Bayes stable equilibrium and Bayes correlated equilibrium both allow estimation of games with weak assumptions on players' information, the former is stronger as it leverages the assumption that opponents' actions are observed to each player at the equilibrium situations.

We propose a computationally tractable approach to estimation and inference. We show that checking whether a candidate parameter enters the identified set (asking whether we can find an equilibrium consistent with data) solves a linear program. Furthermore, we propose a simple approach to inference by combining this property with the insights from Horowitz and Lee (2021): checking whether a candidate parameter belongs to the confidence set amounts to solving a convex program.

As an empirical application, we use our framework to analyze the strategic entry decisions of McDonald's and Burger King in the US. We estimate the model parameters using Bayes stable equilibrium and explore the role of informational assumptions on identification. We also use the model to simulate the impact of increasing access to healthy food in Mississippi food deserts. Our results suggest that the assumptions on players' information that are often used in the literature may be too strong, as the corresponding identified set can be empty. On the other hand, making no assumption on players' information produces an identified set that is too large, indicating that some assumptions on information are necessary to produce informative results. We show that an informative identified set can be obtained under an intermediate assumption which is also credible; this specification assumes that McDonald's has accurate information about its payoff shocks while Burger King may minimally observe nothing. We also compute the identified sets under the Bayes correlated equilibrium assumption and find that the Bayes stable equilibrium identified sets are substantially tighter under the same assumptions on players' information. Comparing the volumes of the identified sets, measured as the product of the projected confidence intervals, the volume under Bayes stable equilibrium is at most 5% of that under Bayes correlated equilibrium.

Related Literature

Our work adds to the literature on the econometric analysis of game-theoretic models (see de Paula (2013) and Aradillas-López (2020) for recent surveys).[2] Its key contribution lies in

[2]In his survey on the econometrics of static games, Aradillas-López (2020) classifies existing papers around five criteria: (i) Nash equilibrium versus weaker solution concepts; (ii) the presence of multiple solutions;

designing a framework that applies to a class of situations characterized by stable outcomes. Specifically, our framework would be best applied to cases where (i) it is reasonable to assume that the realized decisions represent firms' best responses to the *observed* decisions of the opponents, (ii) the stability of outcomes are not driven by high costs of revising actions, and (iii) the econometrician observes cross-sectional data of firms' stable decisions at some point in time.[3]

Our framework differs from the usual Nash framework in that we explicitly assume opponents' actions are observed in equilibrium situations. Static Nash frameworks are generally not consistent with stable outcomes because players might regret their original actions after observing opponents' actions. Furthermore, we are not aware of dynamic models (e.g., frameworks based on Markov perfect equilibrium) that can straightforwardly handle stable outcomes in incomplete information environment. The empirical literature has been aware of this issue (see the discussions in, e.g., Draganska et al. (2008), Einav (2010), and Ellickson and Misra (2011)). Our work fills this gap by developing an equilibrium concept that facilitates econometric analysis.

Using Bayes stable equilibrium as a solution concept allows the researcher to relax the common informational assumptions made in the empirical literature. An early work in this dimension is Grieco (2014) who considers a parametric class of information structures that nest standard assumptions. Our work is most closely related to recent papers that use Bayes correlated equilibrium as a basis for informationally robust econometric analysis: Magnolfi and Roncoroni (2021) applies Bayes correlated equilibrium to static entry games (which are

(iii) complete- versus incomplete-information games; (iv) correct versus incorrect beliefs; (v) parametric versus nonparametric models. To place our work in these categories, this paper (i) develops a new solution concept that is weaker than complete information pure strategy Nash equilibrium but stronger than Bayes correlated equilibrium; (ii) admits a set of equilibria; (iii) allows a general form of incomplete information which accommodate standard assumptions as special cases; (iv) assumes that players have correct beliefs; (v) imposes parametric assumptions on the payoff functions and the distribution of unobservables.

[3] This idea behind cross-sectional analysis of games is accentuated in Ciliberto and Tamer (2009): "*The idea behind cross-section studies is that in each market, firms are in a long-run equilibrium. The objective of our econometric analysis is to infer long-run relationships between the exogenous variables in the data and the market structure that we observe at some point in time, without trying to explain how firms reached the observed equilibrium.*" (pp.1792-1793).

also considered in this paper), Syrgkanis, Tamer, and Ziani (2021) to auctions, and Gualdani and Sinha (2020) to static, single-agent models.[4]

We contribute to the literature on the econometrics of moment inequality models by proposing a simple approach to constructing confidence sets based on the idea of Horowitz and Lee (2021).[5] Our approach is new in the context of econometric analysis of game-theoretic models and applicable under alternative solution concepts such as pure strategy Nash equilibrium or Bayes correlated equilibrium.

Our work also relates to the game theory literature in two dimensions. First, we introduce a solution concept based on the idea of rational expectations pioneered by Radner (1979). Our approach to defining a rational expectations equilibrium in games (the "outcome function" approach) is largely motivated by Liu (2020), who used the same idea to define the notion of stability in two-sided markets with incomplete information. There are works that share similar motivations to ours—namely that the equilibrium concept should be robust to players refining their information after observing opponents' actions—and study solution concepts closely related to rational expectations in games (e.g., Green and Laffont (1987), Minehart and Scotchmer (1999), Minelli and Polemarchakis (2003), and Kalai (2004)). In contrast to these works, the key departure is that we do not assume that individual strategy mappings generate actions nor that players' types are revealed.

Second, we add to the recent literature that studies solution concepts with informational robustness properties (e.g., Bergemann and Morris (2013; 2016; 2017) and Doval and Ely (2020a)). Bergemann and Morris (2016) established the informational robustness property of Bayes correlated equilibrium by showing that Bayes correlated equilibrium captures the

[4]There is also a strand of literature that studies the possibility that firms might have biased beliefs (see Aguirregabiria and Magesan (2020) and Aguirregabiria and Jeon (2020) for a review). The main difference is that the works in this literature assume that the econometrician knows the true information structure of the game but firms may not have correct beliefs whereas we assume that firms have correct beliefs but the econometrician does not know the true information structure.

[5]Recent development in inference with moment inequality models has introduced many alternative approaches for constructing confidence sets (see Ho and Rosen (2017), Canay and Shaikh (2017), and Molinari (2020) for recent surveys). However, to the best of our knowledge, most are not directly applicable to our setup, primarily due to the presence of a high-dimensional nuisance parameter and a large number of inequalities.

implications of Bayes Nash equilibrium when the players may observe more information than initially assumed. It turns out that we can use the same arguments to define Bayes stable equilibrium and establish its informational robustness property when the underlying solution concept is rational expectations equilibrium.

Finally, our empirical application contributes to the literature on entry competition in the fast-food industry. Existing empirical works that study strategic entries among the top burger chains include Toivanen and Waterson (2005), Thomadsen (2007), Yang (2012), Gayle and Luo (2015), Igami and Yang (2016), Yang (2020), and Aguirregabiria and Magesan (2020). In particular, Yang (2020), who studies strategic entry decisions in the Canadian hamburger industry, shares a similar motivation that players extract information from the opponents' actions, but uses a dynamic games framework to explicitly model the learning process. Our empirical work is distinguished by the use of novel datasets and its focus on exploring the role of informational assumptions. Moreover, to the best of our knowledge, we are the first to study the impact of the local food environment on the burger chains' strategic entry decisions.[6]

The rest of the paper is organized as follows. Section 1.2 introduces the notion of Bayes stable equilibrium in a general finite game of incomplete information and studies its property. Section 1.3 sets up the econometric model and provides identification results. Section 1.4 provides econometric strategies for computationally tractable estimation and inference. Section 1.5 applies our framework to the entry game played by McDonald's and Burger King in the US. Section 1.6 concludes. All proofs are in Appendix A.1.

Notation. Throughout the paper, we will use the following notation to express discrete probability distributions in a compact manner. When $\mathcal{Y}$ is a finite set, and $p(y)$ denotes the probability of $y \in \mathcal{Y}$, we will use $p_y \equiv p(y)$. Similarly, $q_{y|x} \equiv q(y|x)$ will be used to denote conditional probability of y given x. We let $\Delta_y \equiv \Delta(\mathcal{Y})$ denote the probability simplex on $\mathcal{Y}$, so that $p \in \Delta_y$ if and only if $p_y \geq 0$ for all $y \in \mathcal{Y}$ and $\sum_{y \in \mathcal{Y}} p_y = 1$. Similarly, we let $\Delta_{y|x}$

[6] For a list of works in economics that study issues related to food deserts, see Allcott et al. (2019) and the references cited therein.

denote the set of all probability distributions on $\mathcal{Y}$ conditional on x, so that $q \in \Delta_{y|x}$ if and only if $q_{y|x} \geq 0$ for all y and $\sum_{y \in \mathcal{Y}} q(y|x) = 1$. We also use the convention that writes an action profile as $a = (a_1, ..., a_I) = (a_i, a_{-i})$.

1.2 Model

We consider empirical settings characterized by two properties. First, the setting is *dynamic* in the sense that players can revise their actions after observing the opponents' actions. Second, players' actions are readily and publicly observed by others. Our objective is to describe certain "steady-state" situations in which all players publicly observe the realized action profile, yet no deviation occurs even when the agents have the opportunity to do so. When conducting econometric analysis, we will assume that the econometrician observes a cross-section of stable outcomes.

In this section, we introduce Bayes stable equilibrium as a solution concept that solves the consistency problem and facilitates econometric analysis while allowing for weak assumptions on players' information. Throughout the paper, we assume that the state of the world remains persistent enough to abstract away from modeling the transition of states over time, and that the costs of revising actions are sufficiently low so that we can ignore them.[7] We formalize the idea in a general class of discrete games of incomplete information, following the notation of Bergemann and Morris (2016).

We proceed as follows. In Section 1.2.1, we lay out the game environment. In Section 1.2.2, we formalize the notion of stable outcomes and motivate our solution concept. In Section 1.2.3, we argue that rational expectations equilibrium à la Radner (1979) can be used as a baseline solution concept for rationalizing stable outcomes in the presence of incomplete information. In Section 1.2.4, we introduce Bayes stable equilibrium. Then, in Section 1.2.5,

[7] In the real world, the costs of revising actions are not zero. However, the relevant question is whether high adjustment costs are the main driver of stable outcomes. We assume that the adjustment costs are negligible compared to the long-run profits obtained at stable outcomes. This is in the same spirit as the empirical matching models surveyed in Chiappori and Salanié (2016); the stable matching condition abstracts away from the costs associated with entering into a marriage or divorce.

we show that Bayes stable equilibrium characterizes the implications of rational expectations equilibria when the players might observe more information than assumed by the analyst. Finally, in Section 1.2.6, we compare the proposed solution concept to pure strategy Nash equilibrium and Bayes correlated equilibrium.

1.2.1 Discrete Games of Incomplete Information

Let $\mathcal{I} = \{1, 2..., I\}$ be the set of players. The players interact in a finite game of incomplete information (G, S).[8] A *basic game* $G = \langle \mathcal{E}, \psi, (\mathcal{A}_i, u_i)_{i=1}^I \rangle$ is a tuple of payoff-relevant primitives: $\mathcal{E}$ is a finite set of unobserved states; $\psi \in \Delta(\mathcal{E})$ is a common prior distribution with full support; $\mathcal{A}_i$ is a finite set of actions available to player i, and $\mathcal{A} \equiv \times_{i=1}^I \mathcal{A}_i$ is the set of action profiles; $u_i : \mathcal{A} \times \mathcal{E} \to \mathbb{R}$ is player i's von Neumann–Morgenstern utility function. An *information structure* $S = \langle (\mathcal{T}_i)_{i=1}^I, \pi \rangle$ is a tuple of information-related primitives: $\mathcal{T}_i$ is a finite set of signals (or types), and $\mathcal{T} \equiv \times_{i=1}^I \mathcal{T}_i$ is the set of signal profiles; $\pi : \mathcal{E} \to \Delta(\mathcal{T})$ is a signal distribution (which allows players' signals to be arbitrarily correlated). The interpretation is that the realized state of the world $\varepsilon \in \mathcal{E}$ drawn from the prior ψ is not directly observed by the players, but each player observes a private signal $t_i \in \mathcal{T}_i$ that can be used to learn about ε based on the signal distribution π. The game is common knowledge to the players. As highlighted by Bergemann and Morris (2016), the separation between the basic game and the information structure facilitates the analysis on the role of information structures.

In empirical applications, there is a finite set of exogenous covariates $\mathcal{X}$. We can augment the notation and let (G^x, S^x) describe the game in markets with characteristics $x \in \mathcal{X}$. Indexing each game by x is justified by assuming that the realized x is common-knowledge to the players and the econometrician, and that the game primitives are functions of x. We suppress the dependence on x for now.

The following two-player entry game serves as a running example as well as a baseline

[8] Throughout this paper, we assume that the state space is finite. The assumption simplifies the notation. In addition, even though continuous state space can be used, we will eventually need to discretize the space for feasible estimation. Syrgkanis, Tamer, and Ziani (2021) and Magnolfi and Roncoroni (2021) take similar discretization approaches for estimation with Bayes correlated equilibria.

model for our empirical application.

Example 1.1 (Two-player entry game)**.** The basic game G is described as follows. The state of the world $\varepsilon \in \mathcal{E}$ is a vector of player-specific payoff shocks, $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$, where $\varepsilon \sim \psi$ for some distribution ψ. Firm i's action set is $\mathcal{A}_i = \{0, 1\}$ where $a_i = 1$ represents staying in the market and $a_i = 0$ represents staying out. The payoff function is $u_i(a_i, a_j, \varepsilon_i) = a_i(\beta_i + \kappa_i a_j + \varepsilon_i)$ where $\beta_i \in \mathbb{R}$ is the intercept and $\kappa_i \in \mathbb{R}$ is the "spillover effect" parameter which may be negative or positive depending on the nature of competition. Then, $\beta_i + \varepsilon_i$ is the monopoly profit, $\beta_i + \kappa_i + \varepsilon_i$ is the duopoly profit, and the profit from staying out is zero.

Next, we provide examples of information structures to which we will pay special attention in our empirical application:

- In $S^{complete}$, each player observes the realization of ε. Formally, we have $\mathcal{T}_i \equiv \mathcal{E}$ for all player i, and $\pi(t_1 = \varepsilon, t_2 = \varepsilon|\varepsilon) = 1$ for each ε;

- In $S^{private}$, ε_i is private information to player i. We have $\mathcal{T}_i \equiv \mathcal{E}_i$ for all player i, and $\pi(t_1 = \varepsilon_1, t_2 = \varepsilon_2|\varepsilon) = 1$ for each ε;

- In S^{1P}, player 1 observes ε_1, but player 2 observes nothing. We have $\mathcal{T}_1 \equiv \mathcal{E}_1$, $\mathcal{T}_2 \equiv \{0\}$, and $\pi(t_1 = \varepsilon_1, t_2 = 0|\varepsilon) = 1$ for each ε. Player 2's signal is uninformative;

- Finally, in S^{null}, both players observe nothing. We have $\mathcal{T}_1 \equiv \mathcal{T}_2 \equiv \{0\}$.

Note that the information structures described above can be ordered from the most informative to the least informative: $S^{complete}$, $S^{private}$, S^{1P}, S^{null}. For example, $S^{complete}$ is "more informative" than $S^{private}$ since each player is allowed to "observe more." We will formally define a partial order on the set of information structures following Bergemann and Morris (2016) in Section 1.2.5. ■

1.2.2 Stable Outcomes

Let us formalize the notion of stable outcomes and motivate our solution concept.[9] Suppose that, at some point in time, the state of the world is ε, the private signals are $t = (t_1, ..., t_I)$, and the players' decisions are $a = (a_1, ..., a_I)$. Assume that each player i observes her private signal t_i as well as the outcome a. What are the conditions for having no deviation at this situation? A necessary condition is that each player i holds a belief $\mu^i \in \Delta(\mathcal{E})$ that gives no incentive to deviate from the status quo outcome a unilaterally.

Definition 1.1 (Stable outcome)**.** An outcome $a = (a_1, ..., a_I)$ is *stable* with respect to a system of beliefs $\mu = (\mu^i)_{i=1}^{I}$ if, for each player $i = 1, ..., I$,

$$\mathbb{E}_{\varepsilon \sim \mu^i}\left[u_i(a, \varepsilon)\right] \geq \mathbb{E}_{\varepsilon \sim \mu^i}\left[u_i(a_i', a_{-i}, \varepsilon)\right] \tag{1.1}$$

for all $a_i' \in \mathcal{A}_i$.

But how do these beliefs arise? A sensible equilibrium would require that the equilibrium action profiles and the equilibrium beliefs be consistent with each other: (i) each player's action must be optimal with respect to his belief, and (ii) each player's belief must be consistent with his private information as well as the observed decisions of the opponents. It is easy to see that static Bayes Nash equilibrium will not satisfy these conditions in general; when the players observe the realized actions, they will update their beliefs and possibly find incentives to revise their original actions. While it is natural to ask whether we can use a noncooperative dynamic game to model convergence to a pair of stable decisions and stable beliefs, such route is likely to be non-trivial and dependent on ad hoc assumptions. In the following sections, we propose a simple and pragmatic approach to the problem.

[9] The term "stability" has been used in different ways in the theory literature depending on the context. Our notion of stability is the closest to the "stable matching" defined in Liu (2020) under incomplete information matching games (the canonical complete information stable matching is a special case). There is also "hindsight (or ex-post) stability" of Kalai (2004), whose motivation is very similar to ours but differs in that it also requires players' types to be revealed after the play. To the best of our knowledge, the term "Bayes stable equilibrium" has not been used in the literature.

1.2.3 Rational Expectations Equilibrium

Before introducing Bayes stable equilibrium, which will be the solution concept we take to econometric analysis, we argue that a version of rational expectations equilibrium à la Radner (1979), appropriately defined for our setting, offers a simple conceptual framework for rationalizing stable outcomes in the presence of incomplete information. After introducing the definition of Bayes stable equilibrium in the next section, we show that Bayes stable equilibrium characterizes the set of rational expectations equilibrium predictions when the analyst does not know the underlying information structure. Thus, Bayes stable equilibrium offers a tool for analyzing stable outcomes with weak assumptions on players' information.

To define rational expectations equilibrium in our setting, we follow Liu (2020) and use the "outcome function" approach described as follows.[10] Let a game (G, S) be given. Let $\delta : \mathcal{T} \to \Delta(\mathcal{A})$ be an *outcome function* in (G, S); an outcome function specifies a probability distribution over action profiles at each realization of players' signals. Assume that δ is common knowledge to the players. Suppose that, after the state of the world $\varepsilon \in \mathcal{E}$ and the signal profile $t \in \mathcal{T}$ are realized according to the prior distribution $\psi(\cdot)$ and the signal distribution $\pi(\cdot|\varepsilon)$, an action profile $a \in \mathcal{A}$ is drawn from the outcome function $\delta(\cdot|t)$, and the players publicly observe a. Each player i, having observed his private signal and the realized action profile (t_i, a_i, a_{-i}), updates his beliefs about the state of the world ε using Bayes' rule, and decides whether to adhere to the observed outcome (play a_i) or not (deviate to $a_i' \neq a_i$). If δ is such that the players always find the realized action profiles optimal, we call it a rational expectations equilibrium of (G, S). Let $\mathbb{E}_{\varepsilon}^{\delta}\left[u_i\left(a_i', a_{-i}, \varepsilon\right)|t_i, a_i, a_{-i}\right]$ denote the expected payoff to player i from choosing a_i' conditional on observing private signal t_i

[10] Liu (2020) observed that the traditional idea of rational expectations equilibrium à la Radner (1979) can be used to define stable matching in two-sided markets with incomplete information. Specifically, he uses an "outcome function" approach and introduces a *matching function* that maps players' types to an observable match. We follow his insights and adopt the logic of rational expectations equilibrium to handle a similar notion of stability in games with incomplete information. Minehart and Scotchmer (1999) and Minelli and Polemarchakis (2003) have made similar attempts to connect rational expectations equilibrium to games without price. While their definition of rational expectations equilibrium refers to strategy profiles, we apply the definition to outcomes functions that are not necessarily the product of individual strategy mappings.

and action profile (a_i, a_{-i}).

Definition 1.2 (Rational expectations equilibrium)**.** An outcome function δ is a *rational expectations equilibrium* for (G, S) if, for each $i = 1, ..., I$, $t_i \in \mathcal{T}_i$, $(a_i, a_{-i}) \in \mathcal{A}$ such that $\Pr^{\delta}(t_i, a_i, a_{-i}) > 0$, we have

$$\mathbb{E}_{\varepsilon}^{\delta}\left[u_i\left(a_i, a_{-i}, \varepsilon\right)|t_i, a_i, a_{-i}\right] \geq \mathbb{E}_{\varepsilon}^{\delta}\left[u_i\left(a_i', a_{-i}, \varepsilon\right)|t_i, a_i, a_{-i}\right] \tag{1.2}$$

for all $a_i' \in \mathcal{A}_i$.

One way to understand the concept is to interpret the outcome function $\delta : \mathcal{T} \to \Delta(\mathcal{A})$ as a reduced-form relationship between players' information and the outcome of the game. We are agnostic about the details on how δ came about. However, it is assumed that the players agree on a common δ, and use δ to infer opponents' information after observing the realized decisions. Thus, δ serves as the players' "model" for connecting the uncertainties to the observables.

There is nothing conceptually new; we simply connect the definition of rational expectations equilibrium to our setting. The rational expectations equilibrium in Radner (1979) refers to a *price function* (a mapping from agents' information to an observable price) such that every price on its support clears the market when the agents use the prices to not only calculate their budgets but also to refine their information by inferring others' information. In our setting, rational expectations equilibrium refers to an *outcome function* (a mapping from players' information to an action profile) such that every action profile on the its support gives no deviation incentives to the players when they use the realized action profile to infer opponents' information.

In a rational expectations equilibrium, outcomes and beliefs are determined simultaneously such that the stability condition (1.1) is satisfied. If the environment—the state of the world and the players' signals—stays unchanged and the outcomes are generated by a rational expectations equilibrium, the realized decisions persist over time. In the econometric analysis, we will assume that the econometrician observes these decisions at some point in time.

1.2.4 Bayes Stable Equilibrium

Let us introduce Bayes stable equilibrium. Let (G, S) be given. A *decision rule* in (G, S) is a mapping $\sigma : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ that specifies a probability distribution over action profiles at each realization of state and signals. Assume that σ is common knowledge to the players. Suppose the data generating process is described as follows. First, the state of the world $\varepsilon \in \mathcal{E}$ is drawn from $\psi(\cdot)$ and the profile of private signals $t \in \mathcal{T}$ is drawn from $\pi(\cdot|\varepsilon)$. Next, an action profile $a \in \mathcal{A}$ is drawn from $\sigma(\cdot|\varepsilon, t)$ and publicly observed by the players. Then, each player i, having observed her private signal and the realized action profile (t_i, a_i, a_{-i}), updates her beliefs about the state of the world ε using Bayes' rule, and decides whether to adhere to the observed outcome (play a_i) or not (deviate to $a_i' \neq a_i$). If the players always have no incentives to deviate from the realized action profiles, we call σ a Bayes stable equilibrium.

Definition 1.3 (Bayes Stable Equilibrium)**.** A decision rule σ is a Bayes stable equilibrium for (G, S) if, for each $i = 1, ..., I$, $t_i \in \mathcal{T}_i$, $(a_i, a_{-i}) \in \mathcal{A}$ such that $\Pr^{\sigma}(t_i, a_i, a_{-i}) > 0$, we have

$$\mathbb{E}_{\varepsilon}^{\sigma}\left[u_i(a_i, a_{-i}, \varepsilon)|t_i, a_i, a_{-i}\right] \geq \mathbb{E}_{\varepsilon}^{\sigma}\left[u_i(a_i', a_{-i}, \varepsilon)|t_i, a_i, a_{-i}\right] \tag{1.3}$$

for all $a_i' \in \mathcal{A}_i$.

One way to understand the definition is to interpret σ as the recommendation strategy of an omniscient mediator. The mediator commits to σ and announces it to the players at the beginning of the game. Then, after (ε, t) is realized and observed by the mediator, the mediator draws an action profile a from $\sigma(\cdot|\varepsilon, t)$ and publicly recommends it to the players. The Bayes stable equilibrium condition requires that the publicly recommended action profiles are always incentive compatible to the players.

Note that an outcome function δ does not depend on the state of the world ε whereas a decision rule σ can. The measurability of an outcome function with respect to players' information reflects the requirement that if any outcome is to be achieved, it must depend on what the players know, but cannot depend on what they do not know. On the other hand, a decision rule allows the realized action profiles to be correlated with the unobserved state of the world. In the next section, we show that the correlation arises because Bayes stable equilibrium captures the implications of rational expectations equilibria when the players might observe extra signals about the state of the world that are unknown to the analyst.

We can simplify the obedience condition (1.3) so that the decision rule enters the equilibrium conditions linearly. Given that player i observes signal t_i and recommendation (a_i, a_{-i}), the expected profit from choosing a_i' is

$$\begin{aligned}\mathbb{E}_{\varepsilon}^{\sigma}\left[u_i\left(a_i', a_{-i}, \varepsilon\right)|t_i, a_i, a_{-i}\right] &= \sum_{\varepsilon} u_i\left(a_i', a_{-i}, \varepsilon\right) \Pr^{\sigma}\left(\varepsilon | t_i, a_i, a_{-i}\right) \\ &= \sum_{\varepsilon} u_i\left(a_i', a_{-i}, \varepsilon\right) \left(\frac{\sum_{t_{-i}} \psi\left(\varepsilon\right) \pi\left(t_i, t_{-i}|\varepsilon\right) \sigma\left(a_i, a_{-i}|\varepsilon, t_i, t_{-i}\right)}{\sum_{\tilde{\varepsilon}, \tilde{t}_{-i}} \psi\left(\tilde{\varepsilon}\right) \pi\left(t_i, \tilde{t}_{-i}|\tilde{\varepsilon}\right) \sigma\left(a_i, a_{-i}|\tilde{\varepsilon}, t_i, \tilde{t}_{-i}\right)}\right).\end{aligned}$$

Then, after cancelling out the denominator, which is constant across all possible realizations of $\varepsilon \in \mathcal{E}, t_{-i} \in \mathcal{T}_{-i}$, the obedience condition (1.3) can be rewritten as follows:[11]

$$\sum_{\varepsilon, t_{-i}} \psi_{\varepsilon} \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i\left(a, \varepsilon\right) \geq \sum_{\varepsilon, t_{-i}} \psi_{\varepsilon} \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i\left(a_i', a_{-i}, \varepsilon\right), \quad \forall i \in \mathcal{I}, t_i \in \mathcal{T}_i, a \in \mathcal{A}, a_i' \in \mathcal{A}_i. \quad (1.4)$$

Since σ enters the expression linearly, finding a Bayes stable equilibrium solves a linear feasibility program; this feature will make econometric analysis computationally tractable.

1.2.5 Informational Robustness of Bayes Stable Equilibrium

In Section 1.2.3, we have argued that an analyst can use rational expectations equilibrium as a description of stable outcomes under incomplete information situations. More often than not, however, it is difficult to know the true information structure governing the data

[11] Using a similar argument, we can express the rational expectation equilibrium conditions for an outcome function δ in (G, S) as: $\sum_{\varepsilon, t_{-i}} \psi_{\varepsilon} \pi_{t|\varepsilon} \delta_{a|t} u_i\left(a, \varepsilon\right) \geq \sum_{\varepsilon, t_{-i}} \psi_{\varepsilon} \pi_{t|\varepsilon} \delta_{a|t} u_i\left(a_i', a_{-i}, \varepsilon\right), \ \forall i, t_i, a, a_i'$.

generating process in the real world. Clearly, attempts to characterize all feasible predictions (joint distribution on states, signals, and actions) of a model by a direct enumeration over all possible information structures are likely to be futile since the set of information structures is large. Nevertheless, it would be desirable to have a tractable way of characterizing the set of predictions without relying on a specific assumption on players' information.

We show that Bayes stable equilibrium provides a tractable characterization of all rational expectations equilibrium predictions that can arise when the players might observe more information than assumed by the analyst. Thus, Bayes stable equilibrium serves as a tool for analyzing stable outcomes with weak assumptions on players' information. This result closely resembles the informational robustness property of Bayes correlated equilibrium (established in Theorem 1 of Bergemann and Morris (2016)), namely that Bayes correlated equilibrium provides a shortcut to charactering all Bayes Nash equilibrium predictions that can arise when the players might observe more information than specified by the analyst.

We formalize the idea as follows. First, to capture the idea that players observe more information under one information structure than under another, we introduce the notion of *expansion* defined in Bergemann and Morris (2016).

Definition 1.4 (Expansion)**.** Let $S = (\mathcal{T}, \pi)$ be an information structure. $S^* = (\mathcal{T}^*, \pi^*)$ is an *expansion* of S, or $S^* \succsim_E S$, if there exists $\left(\tilde{\mathcal{T}}_i\right)_{i=1}^{I}$ and $\lambda : \mathcal{E} \times \mathcal{T} \to \Delta\left(\tilde{\mathcal{T}}\right)$ such that $\mathcal{T}_i^* = \mathcal{T}_i \times \tilde{\mathcal{T}}_i$ for all $i = 1, ..., I$ and $\pi^*\left(t, \tilde{t}|\varepsilon\right) = \pi\left(t|\varepsilon\right)\lambda\left(\tilde{t}|\varepsilon, t\right)$.

Intuitively, S^* is an expansion of S if each player is allowed to observe more signals under S^* than under S. In other words, in S, each player i observes a private signal t_i, whereas in S^*, each i gets to observe an additional signal $\tilde{t}_i$ generated by an augmenting signal distribution λ. The notion of expansion defines a partial order on the set of information structures which we represent as $S^* \succsim_E S$.

Example 1.2 (Continued)**.** Expansion defines a partial order on information structures $S^{complete}$, $S^{private}$, S^{1P}, and S^{null}. Clearly, $S^{complete} \succsim_E S^{private} \succsim_E S^{1P} \succsim_E S^{null}$. For example, to show $S^{private} \succsim_E S^{1P}$, take $\mathcal{T}_1^{private} = \mathcal{E}_1$, $\mathcal{T}_2^{private} = \mathcal{E}_2$, $\mathcal{T}_1^{1P} = \mathcal{E}_1$, $\mathcal{T}_2^{1P} = \{0\}$, $\tilde{\mathcal{T}}_1 = \{0\}$, $\tilde{\mathcal{T}}_2 = \mathcal{E}_2$, and $\lambda\left(\tilde{t}_1 = 0, \tilde{t}_2 = \varepsilon_2 | \varepsilon_2\right) = 1$, i.e., in $S^{private}$, Player 2 receives an extra signal that informs him the realization of ε_2. ■

Let $\mathcal{P}^{BSE}_{\varepsilon,t,a}(G,S)$ be the set of joint distributions on $\mathcal{E} \times \mathcal{T} \times \mathcal{A}$ that can arise in a Bayes stable equilibrium of (G,S). Let $\mathcal{P}^{REE}_{\varepsilon,t,a}(G,S)$ be defined similarly. Note that if $S^* \succsim_E S$, joint distributions on $\mathcal{E} \times \mathcal{T}^* \times \mathcal{A}$ induce marginals on $\mathcal{E} \times \mathcal{T} \times \mathcal{A}$. The following theorem states that by considering Bayes stable equilibrium of (G,S), we can capture all joint distributions on $\mathcal{E} \times \mathcal{T} \times \mathcal{A}$ that can arise in a rational expectations equilibrium under some information structure that is more informative than S.

Theorem 1.1 (Informational robustness)**.** *For any basic game G and information structure S, $\mathcal{P}^{BSE}_{\varepsilon,t,a}(G,S) = \bigcup_{S^* \succsim_E S} \mathcal{P}^{REE}_{\varepsilon,t,a}(G,S^*)$.*

The proof of the theorem closely follows that of Bergemann and Morris (2016) Theorem 1. The "$\subseteq$" direction is established by taking the equilibrium decision rule $\sigma : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ as an augmenting signal function which generates a "public signal" a that is commonly observed by the agents. We then construct a trivial outcome function δ that places unit mass on the recommended outcome, i.e., $\delta(\tilde{a}|a) = 1$ if and only if $\tilde{a} = a$. Then the rational expectations equilibrium condition for δ in a game with the augmented information structure is implied by the obedience condition for σ. Conversely, the "$\supseteq$" direction is established by integrating out the "extra signals" $\tilde{t}_i$ from the rational expectations equilibrium condition, which directly implies the obedience condition for the induced decision rule $\sigma(a|\varepsilon,t) \equiv \sum_{\tilde{t}} \lambda\left(\tilde{t}|\varepsilon,t\right)\delta\left(a|t,\tilde{t}\right)$.

Theorem 1.1 can be framed in terms of marginal distributions on the action profiles. This characterization is more relevant for econometric analysis; typical data only contain information on players' decisions but not the signals nor the state of the world. Let $\mathcal{P}^{BSE}_a(G,S)$ be the set of marginal distributions on $\mathcal{A}$ that can arise in a Bayes stable equilibrium of (G,S). Let $\mathcal{P}^{REE}_a(G,S)$ be defined similarly.

Corollary 1.1 (Observational equivalence)**.** *For any basic game G and information structure S, $\mathcal{P}^{BSE}_a(G,S) = \bigcup_{S^* \succsim_E S} \mathcal{P}^{REE}_a(G,S^*)$.*

1.2.6 Relationship to Other Solution Concepts

Bayes stable equilibrium is an empirically motivated notion that offers a simple approach to rationalizing stable outcomes while accounting for the informational feedback from players' observation of realized outcomes. Its focus is different from the traditional Nash frameworks that study the implications of non-cooperative assumptions. We provide a comparison in Appendix A.4 using a simple two-player entry game example familiar in the econometric literature on game-theoretic models.

In the rest of the section, we compare our solution concepts to pure strategy Nash equilibrium and Bayes correlated equilibrium. First, we show that our framework has pure strategy Nash equilibrium as a special case. Second, we show that Bayes stable equilibrium refines Bayes correlated equilibrium as the former imposes stronger restrictions than the latter.

1.2.6.1 Comparison to Pure Strategy Nash Equilibrium

The following theorem says that pure strategy Nash equilibrium arises as a special case of rational expectations equilibrium (or Bayes stable equilibrium) when strong assumptions on players' information are made.

Theorem 1.2 (Relationship to pure strategy Nash equilibrium)**.**

1. *Let G be an arbitrary basic game and let $S^{complete}$ be an information structure in which the state of the world ε is publicly observed by the players. An outcome function $\delta : \mathcal{E} \to \Delta(\mathcal{A})$ is a rational expectations equilibrium of $\left(G, S^{complete}\right)$ if and only if, for every $\varepsilon \in \mathcal{E}$, $\delta_{\tilde{a}|\varepsilon} > 0$ implies $\tilde{a}$ is a pure-strategy Nash equilibrium action profile at ε. Furthermore, δ is a rational expectations equilibrium of $\left(G, S^{complete}\right)$ if and only if it is a Bayes stable equilibrium of $\left(G, S^{complete}\right)$.*

2. *Suppose that the basic game G is such that $\varepsilon = (\varepsilon_1, ..., \varepsilon_I)$ and $u_i(a, \varepsilon) = u_i(a, \varepsilon_i)$, and let $S^{private}$ be an information structure in which each player i observes ε_i. Then an outcome function $\delta : \mathcal{E} \to \Delta(\mathcal{A})$ is a rational expectations equilibrium of $(G, S^{private})$*

if and only if it is a rational expectations equilibrium of $\left(G, S^{complete}\right)$. *Furthermore,* δ *is a rational expectations equilibrium of* $(G, S^{private})$ *if and only if it is a Bayes stable equilibrium of* $(G, S^{private})$.

Theorem 1.2.1 states that in any game with complete information structure, assuming rational expectations equilibrium is (observationally) equivalent to assuming pure strategy Nash equilibrium at each ε. A rational expectations equilibrium outcome function δ is just a coordination device (or a selection mechanism) over pure strategy Nash outcomes. It also implies that, when players have complete information, a rational expectations equilibrium exists if and only if there is at least one pure strategy Nash equilibrium action profile at each $\varepsilon \in \mathcal{E}$ (on the support of ψ).

Theorem 1.2.2 implies that in a class of games where the state of the world ε is simply a vector of player-specific payoff shocks (which is a common assumption for empirical models of discrete games), we can use weaker informational assumptions to rationalize pure strategy Nash outcomes. Intuitively, when each player i observes his ε_i and an outcome a in an equilibrium situation, opponents' types ε_{-i} are *payoff-irrelevant*. In a pure strategy Nash equilibrium framework, i uses its knowledge of ε_{-i} to *predict* a_{-i}. However, under the rational expectations equilibrium assumption, it is assumed that i *observes* a_{-i}, so ε_{-i} becomes irrelevant to i. Therefore, under the rational expectations equilibrium assumption, it is sufficient that player i observes ε_i in order to support pure strategy Nash outcomes.

Note that under the assumptions in the theorem, there is no difference between an outcome function and a decision rule because players' signals exhaust information about the state of the world. Hence, Bayes stable equilibrium and rational expectations equilibrium are identical in these circumstances.

1.2.6.2 Comparison to Bayes Correlated Equilibrium

Bayes stable equilibrium is a refinement of Bayes correlated equilibrium because the equilibrium conditions for the former are stronger than those for the latter. To describe Bayes correlated equilibrium, suppose that an omniscient mediator commits to a decision rule $\sigma : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ in (G, S) and announces it to the players so that σ is common knowledge to the players. After the state of the world ε is drawn from the prior $\psi(\cdot)$ and the signal profile t is drawn from the signal distribution $\pi(\cdot|\varepsilon)$, the mediator observes (ε, t) and draws an action profile a from the decision rule $\sigma(\cdot|\varepsilon, t)$. Then, the mediator *privately* recommends a_i to each player i. Each player i, having observed his private signal t_i and the privately recommended action a_i, decides whether to follow the recommendation (play a_i) or not (deviate to $a_i' \neq a_i$). If the players are always obedient, then the decision rule is a Bayes correlated equilibrium of (G, S).

Formally, a decision rule $\sigma : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ in (G, S) is a *Bayes correlated equilibrium* if for each $i \in \mathcal{I}$, $t_i \in \mathcal{T}_i$, and $a_i \in \mathcal{A}_i$, we have

$$\mathbb{E}^{\sigma}_{(\varepsilon, a_{-i})}\left[u_i(a_i, a_{-i}, \varepsilon) | t_i, a_i\right] \geq \mathbb{E}^{\sigma}_{(\varepsilon, a_{-i})}\left[u_i(a_i', a_{-i}, \varepsilon) | t_i, a_i\right]$$

for all $a_i' \in \mathcal{A}_i$ whenever $\Pr^{\sigma}(t_i, a_i) > 0$, or more compactly,

$$\sum_{\varepsilon, t_{-i}, a_{-i}} \psi_{\varepsilon} \pi_{t|\varepsilon} \sigma_{a|\varepsilon, t} u_i(a_i, a_{-i}, \varepsilon) \geq \sum_{\varepsilon, t_{-i}, a_{-i}} \psi_{\varepsilon} \pi_{t|\varepsilon} \sigma_{a|\varepsilon, t} u_i(a_i', a_{-i}, \varepsilon), \quad \forall i, t_i, a_i, a_i'. \tag{1.5}$$

The only difference between Bayes stable equilibrium and Bayes correlated equilibrium is that, after (a_i, a_{-i}) is drawn from the decision rule, the former assumes that the mediator informs each player i the entire action profile (a_i, a_{-i}) whereas the latter assumes that the mediator informs each player i only a_i but not a_{-i}. That is, compared to the Bayes correlated equilibrium conditions (1.5) which integrate out opponents' actions a_{-i} since each player i needs to anticipate a_{-i}, Bayes stable equilibrium conditions (1.4) condition on a_{-i} because it is assumed that all actions are publicly observed at the equilibrium situation. The following is immediate.

Theorem 1.3 (Relationship to Bayes correlated equilibrium)**.** *If a decision rule σ is a Bayes stable equilibrium of (G, S), it is a Bayes correlated equilibrium of (G, S).*

Action profiles on the equilibrium path of a Bayes correlated equilibrium may be subject to regret; a player who observes the realized decisions of the opponents might want to revise her action. In contrast, Bayes stable equilibrium explicitly requires that such regret not occur.

When information is complete, Bayes correlated equilibrium reduces to the canonical correlated equilibrium, whereas Bayes stable equilibrium reduces to pure strategy Nash equilibrium in the sense described in Theorem 1.2. When there is a single player, the two solution concepts are identical because there is no informational feedback from observing opponents' actions.

1.3 Econometric Model and Identification

In this section, we describe the econometric model which is based on a general class of discrete games of incomplete information. We characterize the identified set under the assumption that the data are generated by a Bayes stable equilibrium and discuss the properties of the identified set.

1.3.1 Setup

Let us denote market covariates as $x \in \mathcal{X}$ where $\mathcal{X}$ is a finite set; the covariates are common knowledge to the players and observed by the econometrician. Players interact in a set of games $\left(G^{x,\theta}, S^{x}\right)$, $x \in \mathcal{X}$, each indexed by a finite-dimensional parameter $\theta \in \Theta$; the game being played is common knowledge to the players. The basic game at x is $G^{x,\theta} = \langle \mathcal{E}, \psi^{x,\theta}, \left(\mathcal{A}_i, u_i^{x,\theta}\right)_{i=1}^{I} \rangle$ and the information structure at x is $S^{x} = \langle \left(\mathcal{T}_i\right)_{i=1}^{I}, \pi^{x} \rangle$.[12] We maintain the assumption that the set $\mathcal{E}$ is finite in order to make estimation feasible.[13] We assume that θ enters the prior distributions $\psi^{x,\theta} \in \Delta\left(\mathcal{E}\right)$ and the payoff functions

[12]It is without loss to assume that $\mathcal{E}$ and $\mathcal{T}$ do not depend on x because we can use $\mathcal{E} \equiv \cup_x \mathcal{E}^x$ and $\mathcal{T} \equiv \cup_x \mathcal{T}^x$. In general, we can also let θ enter the information structures, which would make the information structures be part of the objects the econometrician wants to identify. In this paper, however, we focus on identifying the parameters of the payoff functions and the distribution of the payoff shocks.

[13]If the benchmark distribution of unobservables is continuous, it will be discretized. Increasing the number of points in $\mathcal{E}$ can make the discrete approximation more accurate at the expense of increased computational

$u_i^{x,\theta} : \mathcal{A} \times \mathcal{E} \to \mathbb{R}$, and that the econometrician knows the prior and the payoff functions up to θ. As standard in the empirical literature, we assume that the state of the world is a vector of player-specific payoff shocks, i.e., $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_I)$ and $u_i^{x,\theta}(a, \varepsilon) = u_i^{x,\theta}(a, \varepsilon_i)$.

The data $\{(a_m, x_m)\}_{m=1}^n$ represent a cross-section of action profiles and covariates in markets $m = 1, ..., n$ that are independent from each other. Let $\phi^x \in \Delta(\mathcal{A})$ denote the *conditional choice probabilities* that represent the probability of observing each action profile conditional on covariate value x. We assume that the econometrician can identify ϕ^x at each $x \in \mathcal{X}$ as $n \to \infty$. The set of baseline assumptions for identification analysis is summarized below.

Assumption 1.1 (Baseline assumption for identification).

1. *The set of covariates $\mathcal{X}$ and the set of states $\mathcal{E}$ are finite.*
2. *The prior distribution $\psi^{x,\theta} \in \Delta(\mathcal{E})$ and the payoff functions $u_i^{x,\theta}(\cdot)$ are known up to a finite-dimensional parameter θ.*
3. *The state of the world is a vector of player-specific payoff shocks, i.e., $\varepsilon = (\varepsilon_1, ..., \varepsilon_I)$ and $u_i^{x,\theta}(a, \varepsilon) = u_i^{x,\theta}(a, \varepsilon_i)$.*
4. *Conditional choice probabilities $\phi^x \in \Delta(\mathcal{A})$, $x \in \mathcal{X}$, are identified from the data.*

Example. (Continued) In the baseline example, there are no covariates. The econometrician assumes that the prior distribution is $\varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$ (which will be discretized). The payoff function is $u_i^\theta(a_i, a_j, \varepsilon_i) = a_i(\kappa_i a_j + \varepsilon_i)$ where $\theta = (\kappa_1, \kappa_2) \in \mathbb{R}^2$ is the parameter of interest. The econometrician observes the conditional choice probabilities $\phi = \left(\phi_{(0,0)}, \phi_{(0,1)}, \phi_{(1,0)}, \phi_{(1,1)}\right)$ whose elements represent the probability of each action profile, e.g., $\phi_{(1,0)}$ is the probability that firm 1 enters ($a_1 = 1$) but firm 2 stays out ($a_2 = 0$). ■

Given Assumption 1.1, the identified set of parameters can be defined when the solution concept and the information structure are specified. For any game $\left(G^{x,\theta}, S^x\right)$, let

burden. See Appendix A.2 for the details on how we make discrete approximations to continuous distributions.

$\mathcal{P}_a^{SC}\left(G^{x,\theta}, S^x\right)$ be the set of feasible probability distributions on $\mathcal{A}$ (the conditional choice probabilities) under solution concept SC. The identified set of parameters is defined as follows.

Definition 1.5 (Identified set of parameters)**.** Given Assumption 1.1, a solution concept SC, and information structures $\tilde{S} = \left(\tilde{S}^x\right)_{x\in\mathcal{X}}$, the identified set of parameters is defined as:

$$\Theta_I^{SC}\left(\tilde{S}\right) \equiv \left\{\theta \in \Theta:\ \forall x \in \mathcal{X},\ \phi^x \in \mathcal{P}_a^{SC}\left(G^{x,\theta}, \tilde{S}^x\right)\right\}.$$

In words, a candidate parameter θ enters the identified set $\Theta_I^{SC}\left(\tilde{S}\right)$ if at each $x \in \mathcal{X}$, the observed conditional choice probabilities ϕ^x can arise under some equilibrium of the model.

1.3.2 Identification and Informational Robustness

We translate the observational equivalence between rational expectations equilibrium and Bayes stable equilibrium (Corollary 1.1) in terms of identified sets. Consider the following assumption.

Assumption 1.2 (Identification under rational expectations equilibrium)**.** *In each market with covariates $x \in \mathcal{X}$, the data are generated by a rational expectations equilibrium of $\left(G^{x,\theta_0}, \tilde{S}^{x,0}\right)$ for some information structure $\tilde{S}^{x,0}$ that is an expansion of S^x ($\tilde{S}^{x,0} \succsim_E S^x$).*

Assumption 1.2 says that there is a true parameter θ_0 underlying the data generating process, and that at each $x \in \mathcal{X}$, the true information structure is some $\tilde{S}^{x,0}$ that is an expansion of S^x. In practice, we will consider a scenario where the econometrician can only pin down S^x but not the true information structure $\tilde{S}^{x,0}$. In other words, the econometrician knows the *baseline information structure* S^x that describes the *minimal* information available to the players, but does not know whether the players actually had access to more signals than prescribed in S^x. Then, under Assumptions 1.1 and 1.2, the econometrician will have to admit all information structures that are expansions of the baseline information structure S^x. This approach contrasts with the traditional approaches that assume the econometrician knows the true information structure exactly. For example, if the econometrician sets the baseline information structure as $S^{private}$ (player i observes ε_i), then we effectively allow each i to have more information about ε_{-i} whereas the traditional approaches would prohibit this possibility.

However, directly working with Assumption 1.2 is computationally infeasible because it requires searching over the set of information structures which is large. We show that Assumption 1.2 can be replaced with the following assumption, which does not rely on unknown information structures.

Assumption 1.3 (Identification under Bayes stable equilibrium). *In each market with covariates $x \in \mathcal{X}$, the data are generated by a Bayes stable equilibrium of $\left(G^{x,\theta_0}, S^x\right)$.*

The following theorem is the consequence of Corollary 1.1; Assumption 1.2 and Assumption 1.3 are observationally equivalent.

Theorem 1.4 (Equivalence of identified sets). *The identified set under Assumptions 1.1 and 1.2 is equal to the identified set under Assumptions 1.1 and 1.3.*

Theorem 1.4 says that in order to compute the identified set when the data are generated by some rational expectations equilibrium under an unknown information structure, we can proceed as if the data are generated by a Bayes stable equilibrium with known information structure.

Magnolfi and Roncoroni (2021) and Syrgkanis, Tamer, and Ziani (2021) use similar results for econometric analysis, but with Bayes correlated equilibrium. They assume that the underlying data generating process is described by Bayes Nash equilibria, whereas we rely on rational expectations equilibria. Also see Gualdani and Sinha (2020) for a single-agent case.

Our identification results make no assumptions on the equilibrium selection rule. The Bayes stable equilibrium identified set under Assumptions 1.1 and 1.3 is valid even when the data are generated from a mixture of multiple equilibria. The convexity of the set of Bayes stable equilibria (readily verified from the equilibrium conditions (1.4) since σ enters the expression linearly) makes the single equilibrium assumption innocuous. For example, if the data are generated by two equilibria σ^1 and σ^2 with mixture probability λ and $(1-\lambda)$, then since $\sigma^\lambda \equiv \lambda\sigma^1 + (1-\lambda)\,\sigma^2$ is another equilibrium that generates the same joint distributions, it is as if the data were generated by a single equilibrium σ^λ.[14]

1.3.3 Relationship Between Identified Sets

Recall that in $S^{complete}$ each player i observes the realization of ε, and in $S^{private}$ each player i observes the realization of ε_i (see Example 1.1). We let $\Theta_I^{SC}\left(S^{complete}\right)$ denote the identified set when $S^x = S^{complete}$ at every $x \in \mathcal{X}$; $\Theta_I^{SC}\left(S^{private}\right)$ is defined similarly. Finally, $S^1 \succsim_E S^2$ if and only if $S^{1,x} \succsim_E S^{2,x}$ at every $x \in \mathcal{X}$. The following theorem shows the relationship between identified sets.

Theorem 1.5 (Relationship between identified sets)**.** *Suppose Assumption 1.1 holds.*

1. *If $S' \succsim_E S''$, then $\Theta_I^{BSE}\left(S'\right) \subseteq \Theta_I^{BSE}\left(S''\right)$.*
2. *$\Theta_I^{BSE}\left(S^{complete}\right) = \Theta_I^{PSNE}\left(S^{complete}\right) = \Theta_I^{BSE}\left(S^{private}\right)$.*
3. *For any information structure S, $\Theta_I^{BSE}\left(S\right) \subseteq \Theta_I^{BCE}\left(S\right)$.*

First, Theorem 1.5.1 says that a stronger assumption on information leads to a tighter identified set. The result is intuitive given that the feasible set of equilibria shrinks when more information is available to the players. A consequence of Theorem 1.5.1 is that we will have $\Theta_I^{BSE}\left(S^{complete}\right) \subseteq \Theta_I^{BSE}\left(\tilde{S}\right) \subseteq \Theta_I^{BSE}\left(S^{null}\right)$ for any $\tilde{S}$, i.e., the tightest identified set is obtained when $S^{complete}$ is assumed and the loosest identified set is obtained when S^{null} is assumed. Note that $\Theta_I^{BSE}\left(S^{null}\right)$ corresponds to the identified set that makes no assumption on players' information.

Second, Theorem 1.5.2 (which follows from Theorem 1.2) says that Bayes stable equilibrium and pure strategy Nash equilibrium are observationally equivalent when $S^{complete}$ is

[14] Also see Syrgkanis, Tamer, and Ziani (2021) Lemma 2 for a general argument on why it is without loss to assume that the data are generated by a single equilibrium if the set of predictions is convex.

assumed.[15] Furthermore, due to Assumption 1.1.3, Bayes stable equilibrium can deliver the same identified set under $S^{private}$ which is weaker than $S^{complete}$. Thus, if the researcher takes Bayes stable equilibrium (or rational expectations equilibrium) to be a reasonable notion for the given empirical setting, pure strategy Nash equilibrium outcomes can be rationalized with informational assumptions that are weaker than the complete information assumption.

Finally, Theorem 1.5.3 (which follows from Theorem 1.3) says that for any baseline assumption on players' information, the Bayes stable equilibrium identified set is a subset of the Bayes correlated equilibrium identified set.

1.3.4 Identifying Power of Informational Assumptions

We use a two-player entry game (our running example) to numerically illustrate the identifying power of various informational assumptions in the spirit of Aradillas-Lopez and Tamer (2008). We also compare the identifying power to that of Bayes correlated equilibrium studied in Magnolfi and Roncoroni (2021).

Each player's payoff function is $u_i^\theta(a_i, a_j, \varepsilon_i) = a_i(\kappa_i a_j + \varepsilon_i)$. We assume $(\varepsilon_1, \varepsilon_2)$ follows a bivariate normal distribution with zero mean, unit variance, and zero correlation. As a discrete approximation to the prior distribution, we use a grid of 30 points for each $\mathcal{E}_i$ and a Gaussian copula to put the appropriate probability mass on each grid point $(\varepsilon_1, \varepsilon_2)$.[16] We set $(\kappa_1, \kappa_2) = (-1.0, -1.0)$ and generate choice probabilities using the pure strategy Nash equilibrium assumption with arbitrary selection rule.[17]

To construct the identified sets, we take the distribution of unobservables as known, and collect all points (κ_1, κ_2) compatible with the given solution concept and informational assumptions. We plot the convex hulls of the identified sets in Figure 1.1.

Figure 1.1-(a) shows the BSE identified sets obtained under different baseline information

[15] When Assumption 1.1.3 is imposed, rational expectations equilibrium and Bayes stable equilibrium are identical under $S^{private}$ and $S^{complete}$. This is because a profile of players' signals is equal to the state of the world, so conditioning on players' information is equivalent to conditioning on the state of the world.

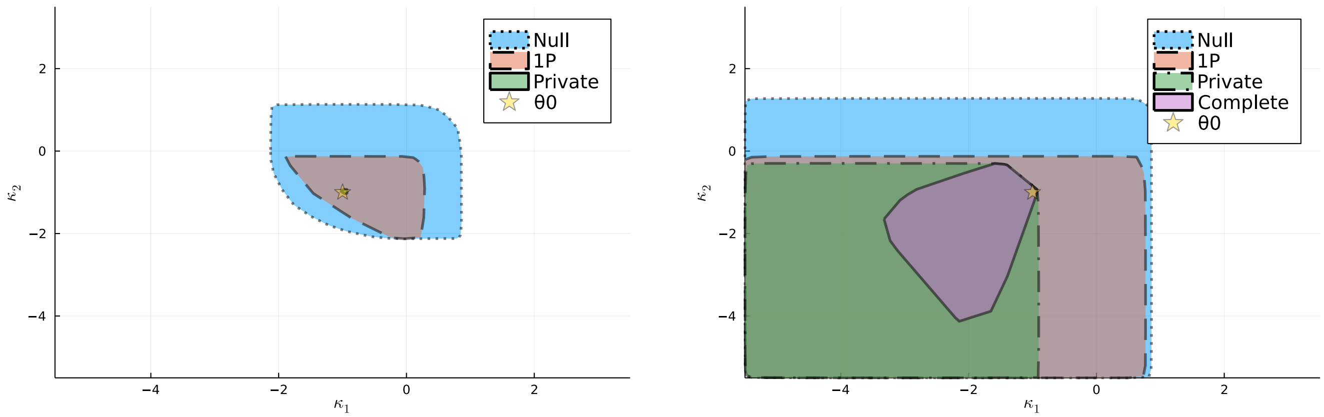
[16] Computational details can be found in Appendix A.2.

[17] Specifically, we generate population choice probability by finding a feasible $\sigma : \mathcal{E} \to \Delta(\mathcal{A})$ which satisfies the inequalities in (1.8) as described in Section 1.4.1.

Figure 1.1: Convex Hulls of Identified Sets

(a) BSE

(b) BCE



structures. The identified sets shrink as the informational assumptions get stronger. We omit the complete information case since $\Theta_I^{BSE}(S^{private}) = \Theta_I^{BSE}(S^{complete})$. Setting the baseline information structure as S^{null} (making no assumption on information) generates an identified set that is quite permissive while using $S^{private}$ generates a tight identified set (which corresponds to the PSNE identified set). Similarly, Figure 1.1-(b) shows the BCE identified sets obtained under different baseline information structures, and that stronger assumptions on information lead to tighter identified sets. The figures show that assumptions on players' information play a crucial role in determining the size of the identified set. In this sense, imposing strong assumption on players' information may be far from innocuous because it places strong restrictions for identification.

As stated in Theorem 1.5.3, comparing Figure 1.1-(a) and 1.1-(b) shows that, for any given baseline information structure, the corresponding BSE identified set is a subset of the corresponding BCE identified set. Our numerical example illustrates that under the same informational assumption, the BSE identified set can be substantially tighter than the BCE identified set. The numerical example illustrates the identifying power of incorporating observability of opponents' actions in the equilibrium conditions.

1.4 Estimation and Inference

We propose a computationally attractive approach for estimation and inference. In Section 1.4.1, we show that whether a candidate parameter enters the identified set can be determined by solving a single linear feasibility program. In Section 1.4.2, we show that this property can be combined with the insights from Horowitz and Lee (2021) to make construction of the confidence sets simple and computationally tractable: determining whether a candidate parameter enters the confidence set amounts to solving a convex feasibility program. Finally, in Section 1.4.3, we provide some practical suggestions for computational implementations.

1.4.1 A Linear Programming Characterization

The following proposition provides a computationally attractive characterization of the identified set. Let $\Theta_I \equiv \Theta_I^{BSE}(S)$ denote the sharp identified set. Let $\partial u_i^{x,\theta}(a_i', a, \varepsilon_i) \equiv u_i^{x,\theta}(a_i', a_{-i}, \varepsilon_i) - u_i^{x,\theta}(a_i, a_{-i}, \varepsilon_i)$ denote the gains from unilaterally deviating to a_i' from outcome (a_i, a_{-i}) given ε_i. Recall our notation: $\sigma^x \in \Delta_{a|\varepsilon,t}$ if and only if $\sigma^x_{a|\varepsilon,t} \geq 0$ for all a, ε, t and $\sum_{a\in\mathcal{A}} \sigma^x_{a|\varepsilon,t} = 1$.

Theorem 1.6 (Linear programming characterization)**.** *Under Assumptions 1.1 and 1.3, $\theta \in \Theta_I$ if and only if, for each $x \in \mathcal{X}$, there exists $\sigma^x \in \Delta_{a|\varepsilon,t}$ such that*

1. *(Obedience) For all $i \in \mathcal{I}$, $t_i \in \mathcal{T}_i$, $a \in \mathcal{A}$, $a_i' \in \mathcal{A}_i$,*

$$\sum_{\varepsilon\in\mathcal{E},t_{-i}\in\mathcal{T}_{-i}} \psi_\varepsilon^{x,\theta} \pi^x_{t|\varepsilon} \sigma^x_{a|\varepsilon,t} \partial u_i^{x,\theta}(a_i', a, \varepsilon_i) \leq 0. \tag{1.6}$$

2. *(Consistency) For all $a \in \mathcal{A}$,*

$$\phi_a^x = \sum_{\varepsilon\in\mathcal{E},t\in\mathcal{T}} \psi_\varepsilon^{x,\theta} \pi^x_{t|\varepsilon} \sigma^x_{a|\varepsilon,t}. \tag{1.7}$$

Theorem 1.6 says that for any candidate $\theta \in \Theta$, whether $\theta \in \Theta_I$ can be determined by solving a single linear feasibility program. The first condition (1.6) states that the nuisance parameter σ^x should be a decision rule that satisfies the Bayes stable equilibrium conditions. The second condition (1.7) states that the observed conditional choice probabilities must be consistent with those induced by the equilibrium decision rule. Given a candidate θ as fixed, $\psi_\varepsilon^{x,\theta}$, $\pi_{t|\varepsilon}^x$, $\partial u_i^{x,\theta}$, and ϕ_a^x are known objects. Also note that $\sigma^x \in \Delta_{a|\varepsilon,t}$ represent constraints that are linear in σ^x. Then, since the variables of optimization σ^x enter the constraints linearly, the program is linear.

Since our empirical framework obtains pure strategy Nash equilibrium as a special case, the complete information pure strategy Nash equilibrium identified set can be computed using linear programs as well. Let Θ_I^{PSNE} be the sharp identified set obtained under the pure strategy Nash equilibrium assumption and no assumption on the equilibrium selection rule. As a corollary to Theorem 1.5 and Theorem 1.6, whether $\theta \in \Theta_I^{PSNE}$ can also be determined via a single linear feasibility program. Thus, Bayes stable equilibrium identified sets embed the pure strategy Nash equilibrium identified set studied in Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011) as a special case.

Corollary 1.2 (Linear programming characterization of PSNE identified set)**.** *$\theta \in \Theta_I^{PSNE}$ if and only if, for each $x \in \mathcal{X}$, there exists $\sigma^x \in \Delta_{a|\varepsilon}$ such that*

1. *(Obedience) For all $i \in \mathcal{I}$, $\varepsilon_i \in \mathcal{E}_i$, $a \in \mathcal{A}$, $a_i' \in \mathcal{A}_i$,*

$$\sum_{\varepsilon_{-i} \in \mathcal{E}_{-i}} \psi_\varepsilon^{x,\theta} \sigma_{a|\varepsilon}^x \partial u_i^{x,\theta}(a_i', a, \varepsilon_i) \leq 0.$$

2. *(Consistency) For all $a \in \mathcal{A}$,*

$$\phi_a^x = \sum_{\varepsilon \in \mathcal{E}} \psi_\varepsilon^x \sigma_{a|\varepsilon}^x.$$

Example (Continued)**.** Suppose the econometrician wants to identify $\theta = (\kappa_1, \kappa_2) \in \mathbb{R}^2$ based on the population choice probabilities $\phi = \left(\phi_{(0,0)}, \phi_{(0,1)}, \phi_{(1,0)}, \phi_{(1,1)}\right) \in \mathbb{R}^4$. Then

$\theta \in \Theta_I^{PSNE}$ if and only if there exists $\sigma \in \Delta_{a|\varepsilon}$ such that

$$\sum_{\varepsilon_{-i}} \psi_\varepsilon \sigma_{a|\varepsilon} \left((a_i' - a_i) (\kappa_i a_{-i} + \varepsilon_i) \right) \leq 0, \quad \forall i, \varepsilon_i, a_i, a_{-i}, a_i' \tag{1.8}$$

$$\phi_a = \sum_{\varepsilon} \psi_\varepsilon \sigma_{a|\varepsilon}, \quad \forall a.$$

which is a linear feasibility program. ■

1.4.2 A Simple Approach to Inference

We leverage the insights from Horowitz and Lee (2021) and propose a simple approach to inference on the structural parameters.[18] The key idea behind our approach is summarized as follows. In discrete games, all information in the data is summarized by the conditional choice probabilities, as apparent in Theorem 1.6. The statistical sampling uncertainty arises only from the estimation of the unknown population conditional choice probabilities which are multinomial proportion parameters. Then, if we control for the sampling uncertainty associated with the estimation of the conditional choice probabilities, we will be able to do inference on the structural parameters of interest. This strategy is feasible given that the number of multinomial proportion parameters to estimate is small relative to the sample size. Thus, we construct a confidence set for the conditional choice probabilities, and translate inference on the conditional choice probabilities to inference on the structural parameters using the characterizations in Theorem 1.6.[19]

Let $\phi \equiv (\phi^x)_{x \in \mathcal{X}}$ be the population choice probabilities. Let us make the dependence of the identified set on ϕ explicit by writing

$$\Theta_I \equiv \Theta_I (\phi) .$$

[18] Horowitz and Lee (2021) describe methods for carrying out non-asymptotic inference when the partially identified parameters are solutions to a class of optimization problem. While we leverage the insights from their work, we focus on asymptotic inference with multinomial proportion parameters.

[19] A similar idea has been used by Kline and Tamer (2016) who propose a Bayesian method for inference. They leverage the idea that a posterior on the reduced-form parameters (the conditional choice probabilities) can be translated to posterior statements on θ using a known mapping between them.

In other words, the identified set is constructed by inverting the mapping from the structural parameters to the conditional choice probabilities; if we know ϕ accurately, then we can obtain the population identified set.

When there is a finite number of observations, ϕ is unknown. However, we are able to construct a confidence region for ϕ that accounts for the sampling uncertainty. Let $\alpha \in (0,1)$. We assume that the econometrician can construct a *convex* confidence set Φ_n^α that covers ϕ with high probability asymptotically.

Assumption 1.4 (Convex confidence set for CCP). *Let $\alpha \in (0,1)$. A set Φ_n^α such that*

$$\liminf_{n\to\infty} Pr(\phi \in \Phi_n^\alpha) \geq 1-\alpha$$

is available. Moreover, $\phi \in \Phi_n^\alpha$ can be expressed as a collection of convex constraints.

Leading examples of Φ_n^α are box constraints or ellipsoid constraints; the former will be characterized by constraints that are linear in ϕ^x and the latter will be characterized by those quadratic in ϕ^x. For example, we can construct simultaneous confidence intervals for each $\phi_a^x \in \mathbb{R}$ such that the probability of covering all $\{\phi_a^x\}_{a\in\mathcal{A},x\in\mathcal{X}}$ simultaneously is asymptotically no smaller than $1-\alpha$.

Define the confidence set for the identified set as

$$\widehat{\Theta}_I^\alpha \equiv \bigcup_{\tilde{\phi}\in\Phi_n^\alpha} \Theta_I\left(\tilde{\phi}\right). \tag{1.9}$$

By construction, if Φ_n^α covers ϕ with high probability, then $\widehat{\Theta}_I^\alpha$ covers Θ_I with high probability.

Theorem 1.7 (Inference). *Suppose Φ_n^α satisfies Assumption 1.4 and $\widehat{\Theta}_I^\alpha$ is constructed as (1.9).*

1. $\liminf_{n\to\infty} Pr\left(\Theta_I \subseteq \widehat{\Theta}_I^\alpha\right) \geq 1-\alpha$.

2. *For each θ, determining $\theta \in \widehat{\Theta}_I^\alpha$ solves a convex program.*

Theorem 1.7.1 follows directly from (1.9) and the assumption on Φ_n^α. To understand Theorem 1.7.2, note that $\theta \in \widehat{\Theta}_I^\alpha$ if and only if, for all $x \in \mathcal{X}$, there exist $\sigma^x : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ and $\phi^x \in \Delta(\mathcal{A})$ such that (1.6), (1.7), and $\phi \in \Phi_n^\alpha$ are satisfied. Compared to the population program described in Theorem 1.6 which treated ϕ as known constants, we make ϕ part of the optimization variables and impose convex constraints $\phi \in \Phi_n^\alpha$. Since all equality constraints are linear in (σ, ϕ) and inequality constraints are convex in (σ, ϕ), the feasibility program is convex (see Boyd and Vandenberghe (2004)). Note that the computational tractability comes from the fact that ϕ enters the restrictions in Theorem 1.6 in an additively separable manner; letting ϕ be part of the optimization variable does not disrupt the linearity of the constraints with respect to the variables of optimization.

Finally, we note that computation can be made faster by constructing Φ_n^α as linear constraints since then $\theta \in \widehat{\Theta}_I^\alpha$ can be determined via a linear program. In our empirical application, we construct Φ_n^α as simultaneous confidence intervals for the multinomial proportion parameters ϕ using the results in Fitzpatrick and Scott (1987).[20]

1.4.3 Implementation

We propose a practical routine for obtaining the confidence set $\widehat{\Theta}_I^\alpha$. Theorem 1.7 says that for any candidate θ, we can determine whether $\theta \in \widehat{\Theta}_I^\alpha$ by solving a convex (feasibility) program. This feature is attractive, but it only provides us a binary answer ("yes" or "no").

As commonly done in existing works on partially identified game-theoretic models (e.g., Ciliberto and Tamer (2009), Syrgkanis, Tamer, and Ziani (2021), Magnolfi and Roncoroni (2021)), we define a non-negative criterion function $\widehat{Q}_n^\alpha(\theta) \geq 0$ with the property that $\widehat{Q}_n^\alpha(\theta) = 0$ if and only if $\theta \in \widehat{\Theta}_I^\alpha$. The value of $\widehat{Q}_n^\alpha(\theta)$ for each θ can be obtained by solving a convex program. The advantage of using a criterion function is that the value of $\widehat{Q}_n^\alpha(\theta)$ gives us information on how "far" θ is from the identified set (which corresponds to the zero-level set). Moreover, the gradients of the criterion functions provide information on

[20]See Appendix A.2.2 for details. We also provide Monte Carlo evidence that the proposed method has desirable coverage probabilities even when $\mathcal{X}$ has many elements.

which directions to descend in order to spot a local minimum.

Let $\{w^x\}_{x\in\mathcal{X}}$ be the set of strictly positive weights for each bin $x \in \mathcal{X}$. The choice of weights can be arbitrary although we will choose values proportional to the number of observations at each bin x. Let $q^x \in \mathbb{R}$ and $q \equiv (q^x)_{x\in\mathcal{X}}$. Let $\widehat{Q}_n^\alpha(\theta)$ be the value of the following convex program.

$$\min_{q,\sigma,\phi} \sum_{x\in\mathcal{X}} w^x q^x \quad \text{subject to} \tag{1.10}$$

$$\sum_{\varepsilon,t_{-i}} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x \partial u_i^{x,\theta}(\tilde{a}_i, a, \varepsilon_i) \le q^x, \quad \forall i, x, t_i, a, \tilde{a}_i$$

$$\phi_a^x = \sum_{\varepsilon,t} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x, \quad \forall a, x$$

$$q^x \ge 0,\ \sigma^x \in \Delta_{a|\varepsilon,t},\ \phi^x \in \Delta_a, \quad \forall x$$

$$\phi \in \Phi_n^\alpha.$$

Intuitively, $q^x \ge 0$ measures the minimal violation of the inequalities necessary at bin x; when all equilibrium conditions can be satisfied, the solver will drive the value of q^x to zero.[21] The solution to (1.10) measures the weighted average of the minimal violations of the equilibrium conditions required to make θ compatible with data. Also note that the choice of weights do not affect the results if the researcher is only interested in the set of θ's whose criterion function values are exactly zero.

The following summarizes the properties of the criterion function approach.

Theorem 1.8 (Implementation).

1. *For any $\theta \in \Theta$, program (1.10) is feasible and convex.*
2. *$\widehat{Q}_n^\alpha(\theta) = 0$ if and only if $\theta \in \widehat{\Theta}_I^\alpha$.*
3. *If the gradient $\nabla\widehat{Q}_n^\alpha(\theta)$ exists at θ, it can be obtained as a byproduct to program (1.10) via the envelope theorem.*

[21] This formulation uses the fact that $\max\{z_1, ..., z_K\}$ can be obtained by solving $\min t$ subject to $z_k \le t$ for $k = 1, ..., K$.

In particular, Theorem 1.8.3 says that, due to the envelope theorem, we can obtain the gradients for free when we evaluate the criterion function at each point (assuming the analytic derivatives of $\psi^{x,\theta}$ and $u_i^{x,\theta}$ are available). In practice, we need to identify the minimizers of $\widehat{Q}_n^{\alpha}(\theta)$ in order to numerically approximate $\widehat{\Theta}_I^{\alpha}$. However, doing so by conducting an extensive grid search over the whole parameter space can be computationally costly especially when the dimension of θ is high. Due to Theorem 1.8.3, one can use gradient-based optimization algorithms to identify a minimizer of the criterion function.[22] The ability to quickly identify $\arg\min_\theta \widehat{Q}_n^{\alpha}(\theta)$ is advantageous since we can quickly test whether the identified set is empty, or restrict the search to points near the minimizer.

For our empirical application, we use a heuristic approach to approximate $\widehat{\Theta}_I^{\alpha}$. The idea is to identify a minimizer of the criterion function and run a random walk process starting from the minimizer in order to collect nearby points that have zero criterion function values. This way we avoid the need to evaluate points that are far from the identified set. See Appendix A.2.3 for details.

1.5 Empirical Application: Entry Game by McDonald's and Burger King in the US

We apply our framework to study the entry game by McDonald's and Burger King in the US using rich datasets. Entry competition in the fast food industry fits our framework well due to two stylized facts. First, the decisions on whether or not to operate outlets are highly persistent, indicating that the firms' decisions are publicly observed. Tables 1.1 and 1.2 report the three-year transition probability of the firms' decisions and the market outcomes (a_{MD}, a_{BK}) (where $a_i = 1$ if firm i is present in the market and $a_i = 0$ otherwise), measured for all urban census tracts (which correspond to our definition of markets) in the contiguous US over 1997-2019. For instance, the probability that McDonald's has an outlet in operation

[22] When program (1.10) has a manageable number of variables, then the nested minimization problem $\min_\theta \widehat{Q}_n^{\alpha}(\theta)$ can be solved more efficiently as a single joint minimization problem using a large-scale nonlinear solver (Su and Judd, 2012). We use this approach for our empirical application in the next section.

in a local market three years later conditional on it having an outlet in operation today is 0.95. Together with the assumption that the costs of revising decisions are sufficiently low, the evidence supports the claim that firms' decisions are best-responses to opponents' decisions that are readily observed.[23]

Table 1.1: Three-year Transition Probability of Decisions

McDonald's			Burger King		
$t \backslash t+3$	Out	In	$t \backslash t+3$	Out	In
Out	0.98	0.02	Out	0.99	0.01
In	0.05	0.95	In	0.08	0.92

Notes: Measured for urban tracts in the contiguous US, 1997-2019.

Table 1.2: Three-year Transition Probability of Market Outcomes (a_{MD}, a_{BK})

$t \backslash t+3$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	0.97	0.01	0.02	0.00
$(0,1)$	0.09	0.87	0.00	0.04
$(1,0)$	0.06	0.00	0.92	0.02
$(1,1)$	0.00	0.04	0.08	0.88

Notes: Measured for urban tracts in the contiguous US, 1997-2019.

Second, information asymmetries and information spillover from observing others' decisions are common features in the industry. It is well-documented that competitors take extra scrutiny over the locations where McDonald's opens new outlets in order to take advantage of McDonald's leading market research technology.[24] Our notion of equilibrium accounts for this phenomenon.

Using the proposed framework, we estimate the entry game under different baseline information structures in order to explore the role of informational assumptions on identification.

[23]Indeed, there are usually extra costs associated with opening a new outlet or closing an existing outlet. For example, franchisees (or franchisors) might be constrained by terms of contract or costs associated with reverting actions, at least in the short-run. We assume away these considerations because it seems unlikely that high adjustment costs are driving the decisions we observe in the data.

[24]See Ridley (2008) and Yang (2020) who provide anecdotal evidence on how competing firms learn about the profitability of a location from entries of leading firms such as McDonald's and Starbucks. For example, according to *The Wall Street Journal*, "In the past, many restaurants... plopped themselves next to a McDonald's to piggyback on the No. 1 burger chain's market research." (Leung, 2003)

We also compare our results to those obtained under Bayes correlated equilibrium which also allows estimation with weak assumptions on players' information. We then perform a policy exercise that studies how the market structures in Mississippi food deserts respond after increasing access to healthy food.

1.5.1 Data Description

We combine multiple datasets to construct the final dataset for structural estimation of the entry game. In the final dataset, the unit of observation is a market (urban tract). Each observation contains information on the firms' market entry decisions and the observable characteristics of the firms and the market.

Our primary dataset comes from Data Axle Historical Business Database, which contains a (approximately) complete list of fast-food chain outlets operating in the US between 1997 and 2019 at an annual level.[25] The advantage of this dataset is that it provides the address information of the burger outlets across all regions of the US. The use of this dataset to study strategic entry decisions is new.[26]

Although we use panel data to investigate the persistence of decisions over time, we use cross-section data to estimate the structural model. The idea is to illustrate that the econometrician can use cross-sectional data as a snapshot of the stable outcomes of the markets at some point in time.[27] We use the 2010 cross-section since it was the last year for

[25]This database contains location information for a detailed list of business establishments in the US from 1997 to 2019. The provider attempts to increase accuracy by using an internal verification procedure after collecting data from multiple sources. The dataset is approximately complete in the sense that the list is not free of error. However, we compare the number of burger outlets in the data and the number reported in external sources and confirm that the information is highly accurate for the case of burger chains. See Appendix A.3 for details.

[26]We are not the first to study the entry game between McDonald's and Burger King in the US. Gayle and Luo (2015) uses 2011 cross-sectional data hand-collected using the online restaurant locator on the brands' websites. However, they define a local market as an "isolated city" that is more than 10 miles away from the closest neighboring city, which is larger than our definition that uses a census tract. Moreover, they focus on examining assumptions on the order of entries.

[27]If we wanted to exploit the information available in panel data, we would need to model the dependence of observations across time. However, given that market environments usually seem to stay very stable over time, it is not clear how to leverage the information for structural estimation. For simplicity, we focus on analyzing a single cross-section (which also represents a typical dataset available to researchers).

which decennial census data were available. We describe the main features of our dataset below. Further details on data construction are provided in Appendix A.3.

Market Definition

Markets are defined as 2010 urban census tracts in the contiguous US. A census tract is classified as urban if its geographic centroid is in an *urbanized area* defined by the Census. The final data contain 54,944 markets. We code $a_i = 1$ if firm i had an outlet operating in the market.[28] The unconditional probabilities of market outcomes are $\left(\hat{\phi}_{00}, \hat{\phi}_{01}, \hat{\phi}_{10}, \hat{\phi}_{11}\right) = (0.74, 0.06, 0.15, 0.05)$ where $\hat{\phi}_a$ is the sample frequency of outcome $a = (a_{MD}, a_{BK})$.

Exclusion Restrictions

We use two firm-specific variables that have been used in existing works: distance to headquarters (Zhu et al. (2009), Zhu and Singh (2009), Yang (2012)) and own outlets in nearby markets (Toivanen and Waterson (2005), Igami and Yang (2016), Yang (2020)). Variable distance to headquarter measures the distance between the center of each market to the firms' respective headquarters. The associated exclusion restriction is valid if the cost of operating an outlet increases with its distance to own headquarter, but is unrelated to the distance to opponents' headquarters. Variable own outlets in neighboring markets is constructed by finding all outlets in tracts that are adjacent to a given tract. The underlying assumption is that an outlet's profit can be affected by an own-brand outlet in a neighboring market, but not by a competing brand's outlet in a neighboring market; competition with opponents occur only within each market.

Summary Statistics

Summary statistics are provided in Table 1.3. Continuous variables are discretized to binary variables by using cutoffs around their medians. Clearly, the entry probability of McDonald's

[28] McDonald's (resp. Burger King) has more than one outlets in 1.5% (resp. 0.3%) of the markets.

is higher. McDonald's is more likely to have an outlet present in adjacent markets. The distance to headquarter is higher for Burger King on average because Burger King has its headquarter in Florida while McDonald's has its headquarter in Chicago.

Table 1.3: Summary Statistics

	Mean	Std dev	Min	Max	N
Decision variables					
MD Entry	0.196	0.397	0.00	1.00	54940
BK Entry	0.106	0.307	0.00	1.00	54940
Firm-specific variables					
MD outlets present in nearby markets	0.720	0.449	0.00	1.00	54940
BK outlets present in nearby markets	0.483	0.500	0.00	1.00	54940
Long distance to MD HQ (>1.6K km)	0.285	0.451	0.00	1.00	54940
Long distance to BK HQ (>1.6K km)	0.712	0.453	0.00	1.00	54940
Market environment variables					
Many eating/drinking places (>7 stores)	0.465	0.499	0.00	1.00	54940
High income per capita (>25K dollars)	0.502	0.500	0.00	1.00	54940
Low access to healthy food	0.856	0.351	0.00	1.00	54940
Food desert	0.334	0.472	0.00	1.00	54940

Notes: All variables are binary. Each observation corresponds to urban census tracts.

Market environment variables control for the determinants of profitability that are common across firms. We obtain the following variables to describe market environments. First, we have an indicator for whether a tract has many eating or drinking places; the variable is obtained from the National Neighborhood Data Archive (NaNDA) which provides business activity information at the tract-level. Second, we have an indicator for whether a tract has high income per capita; the variable is from the census. Finally, from the Food Access Research Atlas, we obtain indicators for whether a tract has low access to healthy food and whether a tract is classified as a food desert. A tract is classified as having low access to healthy food if at least 500 or 33 percent of the population lives more than 1/2 mile from the nearest supermarket, supercenter, or large grocery store. A tract is classified as a food desert if it has low income and low access to healthy food, where the criteria for low-income are from the U.S. Department of Treasury's New Markets Tax Credit program.

The last rows of Table 1.3 shows that 85% of all urban census tracts are classified as having low access to healthy food and 33% are classified as food deserts. In the counterfactual analysis, we select food deserts in Mississippi and investigate the impact of increasing access to healthy food on the strategic entry decisions of the firms.

1.5.2 Preliminary Analysis

Before estimating the structural model, we examine the data patterns using simple probit regressions. Each market m contains binary decisions of each firm $a_{im} \in \{0, 1\}$ where $a_{im} = 0$ if firm i stays out in market m and $a_{im} = 1$ if i stays in. We pool the decisions of the firms in each market (so that the unit of observation is (i, m)) and regress the binary decisions on market characteristics. Table 1.4 reports the average marginal effects computed from the regression results.

Table 1.4: Average Marginal Effects from Simple Probit Models

	(1) In	(2) In	(3) In
Own-brand outlets present in nearby markets	-0.067 (0.002)	-0.076 (0.002)	-0.096 (0.002)
Long distance to HQ (> 1.6K km)	-0.083 (0.003)	-0.083 (0.003)	-0.010 (0.003)
Many eating/drinking places (>7)		0.203 (0.002)	0.203 (0.002)
High income per capita (>25K dollars)		-0.038 (0.002)	-0.037 (0.002)
Low access to healthy food		0.039 (0.004)	0.041 (0.004)
McDonald's			0.109 (0.002)
State Dummies	Yes	Yes	Yes
N	107,042	107,042	107,042

Notes: Each observation corresponds to a firm-market pair. Standard errors, which are given in the parentheses, are clustered at the market-level. All variables are binary.

Table 1.4 conveys three messages. First, the presence of own outlets in neighboring markets and distance to headquarter are negatively correlated with entry decisions. This appears to be consistent with our prior that these variables have a negative impact on potential profits. Second, the number of eating and drinking places strongly affects the burger chains' entries. This is presumably because districts with high concentration of food services are also places with high traffic of people who eat out. Finally, low access to healthy food is *positively* correlated with entry decisions. That is, the burger chains are more likely to enter a market when there are fewer healthy substitutes for food.

While Table 1.4 provides a helpful snapshot for what drives the chains' entry decisions, the estimates are likely to be biased since they ignore the fact that firms' decisions affect each other. Such consideration is crucial not only for estimating the parameters of the model but also for studying a policy experiment. In the next section, we estimate the entry game using Bayes stable equilibrium as a solution concept.

1.5.3 Entry Game Setup

We posit a canonical entry game that extends the running example to incorporate covariates in the payoff functions. Let us recall the notation. We use $i = 1, 2$ to denote McDonald's and Burger King respectively. In each market m, firm i can choose a binary action $a_{im} \in \{0, 1\}$ where $a_{im} = 1$ if i stays in and $a_{im} = 0$ if i stays out. The payoff function is specified as

$$u_i^{x_m,\theta}(a_{im}, a_{jm}, \varepsilon_{im}) = a_{im}\left(\beta_i^T x_{im} + \kappa_i a_{jm} + \varepsilon_{im}\right).$$

That is, the payoff from operating in the market is $\beta_i^T x_{im} + \kappa_i a_{jm} + \varepsilon_{im}$ where x_{im} represents market covariates, a_{jm} represents whether the opponent is present, and ε_{im} is the idiosyncratic payoff shock which includes determinants of payoffs that are unobserved by the econometrician, e.g., managerial ability. The payoff from staying out is normalized to zero. We model $(\varepsilon_{1m}, \varepsilon_{2m}) \in \mathbb{R}^2$ as being normally distributed with zero mean, unit variance, and correlation coefficient $\rho \in [0, 1)$. Our specification of the payoff functions is quite standard in the literature.[29]

We estimate the parameters under the baseline information assumptions specified previously in Example 1.1: S^{null}, S^{1P}, $S^{private}$. To recap, S^{null} is the information structure in which each player observes nothing; in S^{1P}, Player 1 observes (only) ε_1 whereas Player 2

[29] A more flexible specification might add a richer set of covariates or let the spillover effects κ_i be a function of the observable covariates as done in Ciliberto and Tamer (2009). We keep the specification parsimonious.

observes nothing; in $S^{private}$, Player 1 observes ε_1 and Player 2 observes ε_2.

Under the Bayes stable equilibrium assumption, the baseline information structures should be interpreted as specifying what the players *minimally* observe. Then estimating the model with S^{null} as the baseline information structure amounts to making no assumption on players' information. On the other hand, if the baseline information structure is set to $S^{private}$, then the identified set is robust to all cases in which the players observe at least their payoff shocks. Finally, setting the baseline information structure to S^{1P} amounts to assuming that McDonald's has good information about its payoff shocks whereas Burger King might minimally have no information about its payoff shock. This assumption relaxes the standard assumption on information (namely the information structure is fixed at either $S^{private}$ or $S^{complete}$) and is consistent with the anecdotal evidence that McDonald's is a leader in the market research technology.

1.5.4 Estimation Results

In order to keep the model parsimonious and reduce the computational burden, we take some steps before estimation, which are described as follows (see Appendix A.2 for further details). First, we assume that the coefficients for common market-level variables (eating places, income per capita, and low access to healthy food) are identical across the two players.[30] We also assume that the coefficients of the firm-specific variables (distance to headquarter and the presence of own-brand outlets in nearby markets) are non-positive. Second, while the benchmark distribution of the latent variables $(\varepsilon_{1m}, \varepsilon_{2m})$ is continuous, we use discretized normal distribution for feasible estimation. Third, we discretize each variable to binary bins; since there are 7 variables in the covariates, this gives $2^7 = 128$ discrete covariate bins. Conditional choice probabilities are non-parametrically estimated using the observations within each bin. Fourth, to construct confidence sets for the conditional choice probabilities, we

[30]This assumption is not without loss and can be refuted on the basis that each chain might react differently to market environment. However, we believe it is reasonable given that McDonald's and Burger King are close substitutes to each other.

used simultaneous confidence bands based on the method described in Fitzpatrick and Scott (1987); using simultaneous confidence bands makes the evaluation of the criterion function a linear program.

1.5.4.1 The Role of Informational Assumptions on Identification

Table 1.5: Bayes Stable Equilibrium Identified Sets

Baseline Information	S^{null}	S^{1P}	$S^{private}$
McDonald's Variables			
Spillover Effects	$[-1.83, 1.62]$	$[-0.89, -0.14]$	-
Constant	$[-1.64, 0.32]$	$[-1.46, -1.04]$	-
Nearby Outlets	$[-1.24, -0.00]$	$[-0.56, -0.25]$	-
Distance to HQ	$[-1.23, -0.00]$	$[-0.26, -0.00]$	-
Burger King Variables			
Spillover Effects	$[-1.81, 1.22]$	$[-1.19, -0.25]$	-
Constant	$[-2.38, 0.44]$	$[-1.48, -0.76]$	-
Nearby Outlets	$[-1.44, -0.00]$	$[-0.53, -0.00]$	-
Distance to HQ	$[-1.41, -0.00]$	$[-0.52, -0.00]$	-
Common Market-level Variables			
Eating Places	$[-0.31, 1.87]$	$[0.82, 1.21]$	-
Income Per Capita	$[-1.02, 0.75]$	$[-0.54, -0.18]$	-
Low Access	$[-0.71, 1.31]$	$[0.25, 0.54]$	-
Correlation parameter ρ	$[0.00, 0.99]$	$[0.42, 0.91]$	-
Number of Markets	54940	54940	54940

Notes: Table reports the projections of confidence sets obtained with nominal level $\alpha = 0.05$. The identified set for $S^{private}$ not reported because it is empty.

Table 1.5 reports projections of the 95% confidence sets obtained under the Bayes stable equilibrium assumption with different baseline information structures. There are three main findings related to the role of informational assumption. First, making no assumption on players' information leads to an uninformative identified set. The confidence set under S^{null} is quite large, and we cannot determine the signs of the parameters. Therefore, being utterly agnostic about players' information does not give us enough identifying power to draw meaningful conclusions.

Second, standard assumptions on information may be too strong. It is quite standard to assume that each player i observes (exactly) ε_i or $(\varepsilon_i, \varepsilon_{-i})$. Setting baseline information structure as $S^{private}$ nests all these cases. However, we find that the identified set under $S^{private}$

is empty, suggesting the possibility of misspecification.[31] Thus, assuming that each player observes at least their ε_i may be too strong. Since the Bayes stable equilibrium identified set under $S^{private}$ is equivalent to the pure strategy Nash equilibrium identified set (see Theorem 1.5.2), the pure strategy Nash equilibrium assumption would also be rejected.[32]

Third, we find that setting the baseline information structure to S^{1P} can produce an informative identified set. Recall that the identified set under S^{1P} makes the assumption that McDonald's has accurate information about its payoff shock, but Burger King's information can be arbitrary. This assumption is consistent with the anecdotal evidence that McDonald's has superior information on the potential profitability of each market, and Burger King tries to free-ride on McDonald's information by observing what McDonald's does. Table 1.5 shows that, even if we substantially relax the assumption on Burger King's information, we can determine the signs of the most parameters. For example, we can see that burger chains are more likely to enter in markets that have low access to healthy food. We can also learn that the firms' payoff shocks are highly correlated to each other.

In conclusion, we find that the informativeness of the identified set crucially depends on the underlying assumption on players' information. At least in our empirical application, it is difficult to draw a meaningful economic conclusion without making assumptions on players' information. On the other hand, under the maintained solution concept, the model rejects the popular assumptions made in the literature, namely that each firm i observes at least its ε_i. A credible intermediate case S^{1P}, which is consistent with our knowledge of the market research technology in the fast food industry, delivers strong identifying power.

Table 1.6: Bayes Correlated Equilibrium Identified Sets

Baseline Information	S^{null}	S^{1P}	$S^{private}$
McDonald's Variables			
Spillover Effects	$[-4.83, 1.92]$	$[-4.85, -0.17]$	$[-4.85, -2.11]$
Constant	$[-1.64, 0.34]$	$[-1.53, 0.29]$	$[-1.37, 0.31]$
Nearby Outlets	$[-1.33, -0.00]$	$[-1.11, -0.00]$	$[-0.97, -0.00]$
Distance to HQ	$[-1.35, -0.00]$	$[-1.10, -0.00]$	$[-0.88, -0.00]$
Burger King Variables			
Spillover Effects	$[-3.84, 3.33]$	$[-3.98, 0.72]$	$[-3.38, -1.03]$
Constant	$[-3.71, 0.61]$	$[-1.65, 0.62]$	$[-1.62, 0.44]$
Nearby Outlets	$[-1.71, -0.00]$	$[-1.23, -0.00]$	$[-1.11, -0.00]$
Distance to HQ	$[-1.70, -0.00]$	$[-1.03, -0.00]$	$[-0.86, -0.00]$
Common Market-level Variables			
Eating Places	$[-0.24, 1.98]$	$[0.51, 1.76]$	$[0.49, 1.68]$
Income Per Capita	$[-1.32, 0.84]$	$[-1.16, 0.14]$	$[-1.08, 0.11]$
Low Access	$[-0.59, 1.49]$	$[-0.37, 1.31]$	$[-0.28, 1.07]$
Correlation parameter ρ	$[0.00, 0.99]$	$[0.00, 0.99]$	$[0.00, 0.97]$
BSE volume/BCE volume	0.05036	0.00000	-
Number of Markets	54940	54940	54940

Notes: Table reports the projections of confidence sets obtained with nominal level $\alpha = 0.05$. BSE/BCE volume computed by taking products of projected intervals.

1.5.4.2 Comparison to Bayes Correlated Equilibrium Identified Sets

We compare the Bayes stable equilibrium identified sets to the Bayes correlated equilibrium identified sets studied in Magnolfi and Roncoroni (2021). The Bayes correlated equilibrium identified sets are reported in Table 1.6. We can readily see that the Bayes correlated equilibrium assumption produces a much larger set for each baseline information structure. Even when we set $S^{private}$ as the baseline information structure, it is not easy to learn the signs of many parameters. For example, we cannot determine whether low access to healthy food promotes or deters entries by the burger chains.

Comparing Tables 1.5 and 1.6 suggests that if the researcher is willing to accept the Bayes stable equilibrium assumption, it can add significant identifying power while providing the same kind of informational robustness as Bayes correlated equilibria. At least in the context

[31] Specifically, we consistently find that the minimum of the criterion function under $S^{private}$ is strictly greater than zero. This is also true even if we do not use sign constraints or reduce the nominal level to a very low level (e.g., $\alpha = 0.0001$).

[32] Of course, the emptiness of the identified set might be due to misspecification in payoff functions, distribution of errors, etc. Our statements are conditional on these specifications being correct.

of our empirical application, we believe it is reasonable to assume that McDonald's decisions that we observe in the data represent best-responses to the *observed* decisions of Burger King and vice versa.

1.5.5 Counterfactual Analysis: The Impact of Increasing Access to Healthy Food on Market Structure

We consider a policy experiment to predict changes in market structure in Mississippi food deserts after increasing access to healthy food.[33] Mississippi is often called one of the "hungriest" states in the US.[34] Mississippi had 664 census tracts in 2010, and 329 of them are classified as urban tracts, which correspond to our definition of markets. Out of 329 urban tracts, 185 tracts (approximately 56%) are classified as food deserts, according to the U.S. Department of Agriculture. According to the definition of food deserts, all of these tracts are classified as having low access to healthy food.

We conduct a policy experiment as follows. We select the 185 tracts classified as food deserts in Mississippi and then increase access to healthy food. This amounts to changing the low access indicator from one (low access) to zero (high access) in all these markets. In reality, such policy would correspond to increasing healthy food providers (grocery stores, supermarkets, or farmers' markets) by providing subsidies or tax breaks. We then recompute the equilibria in these markets and report the weighted average of the bounds associated with each measure of market structure.[35] See Appendix A.2.4 for computational details.

[33]Consumption of fast-food is determined by both supply-side factors (e.g., availability of healthy substitutes in the neighborhood) and demand-side factors (consumers' inherent preference for fast-food). Allcott et al. (2019) points out that consumers' eating habits may be largely driven by their preferences. They also find that increasing the number of supermarkets may be ineffective for promoting healthy eating by low-income households. However, while they consider supermarket entry within a 10-15 minute drive, the high-access indicator we consider is more stringent because 1/2-mile corresponds to less than a 10 minute walk. Thus, one can interpret our experiment as studying what would happen to the burger chains if people had providers of healthy food readily available around them. We believe our results are not mutually exclusive with the findings of Allcott et al. (2019).

[34]For example, Mississippi has been identified as the most food insecure state in the country since 2010 according to Feeding America. See https://mississippitoday.org/2018/05/04/mississippi-still-the-hungriest-state/.

[35]Our counterfactual analysis corresponds to a partial equilibrium analysis. We abstract away from considering how entry or exit in each market can affect the burger chains' decisions in neighboring markets and

Table 1.7: The Impact of Increasing Access to Healthy Food in Mississippi Food Deserts

		$BSE(S^{1P})$		$BCE(S^{1P})$	
	Data	Pre	Post	Pre	Post
Expected number of entrants	0.47	$[0.28, 1.01]$	$[0.15, 0.79]$	$[0.10, 1.18]$	$[0.03, 1.17]$
Probability of MD entry	0.30	$[0.11, 0.32]$	$[0.04, 0.23]$	$[0.00, 0.71]$	$[0.00, 0.67]$
Probability of BK entry	0.17	$[0.00, 0.84]$	$[0.00, 0.72]$	$[0.00, 1.00]$	$[0.00, 1.00]$
Probability of no entrant	0.64	$[0.15, 0.74]$	$[0.28, 0.85]$	$[0.00, 0.90]$	$[0.00, 0.97]$

Notes: Data column represents the sample estimates obtained using markets corresponding to Mississippi food deserts. Final bounds obtained by simulating equilibria at each parameter in the identified set, and then taken union over all bounds. Each number is obtained by taking a weighted average with weights proportional to the number of markets in each covariate bin.

We report the results of the counterfactual analysis in Table 1.7. The first column reports the estimates obtained from the data of the 185 markets corresponding to Mississippi food deserts. For example, the probability of observing McDonald's enter the market in Mississippi food deserts is 0.30, much larger than the unconditional probability obtained using all markets, which was around 0.20.

The second and third columns report the bounds obtained before ("Pre" has low access indicators set to one) and after the counterfactual policy ("Post" has low access indicators set to zero) using the S^{1P}-Bayes stable equilibrium identified set. The bounds are pretty wide because we have considered all parameters in the identified set and made no assumption on the equilibrium selection. However, they shift in the expected directions. For example, the bounds on the expected number of entrants shift from $[0.28, 1.01]$ to $[0.15, 0.79]$. Since the mean number of entrants in the data was 0.47 and the post-counterfactual bounds are $[0.15, 0.79]$, the maximal change we can expect is $0.15 - 0.47 = -0.32$. In some cases, we can make a stronger statement: while the unconditional probability of observing McDonald's enter in data was 0.30, the upper bound in the Post-regime decreases to 0.23, so we can expect that the probability of McDonald's enter to decrease by *at least* 0.07.

Our results suggest that meaningful counterfactual statements may be made even with weak assumptions on players' information. The bounds do not depend on specific assumptions

the responses of healthy food providers.

on equilibrium selection and admit all information structures that are expansions of the baseline information structure.[36] Hence our approach can also serve as a useful tool to conduct sensitivity analysis for researchers who want to see whether their predictions are driven by assumptions on equilibrium selection or what the players know.

For comparison, in the last two columns, we report the counterfactual results obtained using the S^{1P}-Bayes correlated equilibrium identified set. One can readily see that the bounds are pretty large compared to the Bayes stable equilibrium counterpart. For example, we cannot make any statement about the probability of Burger King's entry after the counterfactual policy is implemented. Table 1.7 shows that Bayes correlated equilibrium predictions can be too permissive, especially when no assumption is imposed on what equilibrium might be selected in the counterfactual world.

1.6 Conclusion

This paper presents an empirical framework for analyzing stable outcomes with weak assumptions on players' information. We propose Bayes stable equilibrium as a framework for analyzing stable outcomes which appear in various empirical settings. Our framework can be an attractive alternative to existing methods for practitioners who want to work with an empirical game-theoretic model and be robust to informational assumptions. Furthermore, we believe the proposed computational algorithms can also be helpful in similar settings, especially since reducing computational burden remains a fundamental challenge in the literature.

We believe there are many exciting avenues for future research. First, providing a non-cooperative foundation to our solution concepts remains an open question. While we can imagine a dynamic adjustment process that converges to stable outcomes, how to formalize this idea is yet unclear. Second, it will be interesting to find reasonable ways of imposing equilibrium selection. While Bayes stable equilibrium (or Bayes correlated equilibrium) has

[36] Our predictions are conservative because we do not make any assumptions on how the information structure or the equilibrium selection rule might change after the counterfactual policy.

the informational robustness property, the set of predictions may be too large, limiting our ability to make sharp predictions for counterfactual analysis. Finding ways to sharpen predictions without sacrificing robustness to information will be helpful. Third, our counterfactual analysis is limited to a partial equilibrium analysis. It will be interesting to think about ways to model the strategic interactions of healthy food providers and unhealthy food providers together.

Chapter 2: Estimating Dynamic Games with Unknown Information Structure

2.1 Introduction

This paper develops an empirical framework for estimating a popular class of dynamic games with weak assumptions on players' information. A series of seminal papers have developed econometric models for estimating dynamic discrete choice and dynamic games models in a computationally feasible manner (e.g., Rust (1994), Ericson and Pakes (1995), Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2008)), providing toolkits that can be applied to answer various economic questions that involve dynamics and strategic interactions (see Aguirregabiria and Mira (2010) and Aguirregabiria, Collard-Wexler, and Ryan (2021) for recent surveys). However, these empirical models commonly impose a restrictive form of information structure, namely that player-specific payoff shocks are private information to each player. While the assumption facilitates computational tractability, it may result in biased estimates due to misspecification. In fact, it is usually difficult to know the true information structure a priori or learn about it through data.

In this paper, building on the popular class of dynamic games, we propose an empirical model that allows the researcher to estimate the model parameters while admitting weak assumption on players' information. Specifically, contrary to the standard approach that assumes that the analyst knows the *exact* form of the information structure underlying the data generating process, we let the analyst specify the *minimal* information available to the players and be agnostic about the possibility that the players might have more information in a form that is *unknown* to the analyst. Thus, our framework allows the analyst to obtain a set of parameter estimates that are "informationally-robust" in the sense that the analyst does not need to know the true information structure of the game.

The main contributions of this paper are twofold. First, we develop a solution concept dubbed *Markov correlated equilibrium* as a tool for tractably capturing the set of Markov perfect equilibrium predictions that can arise when the players might observe more signals than assumed by the analyst. We show that Markov correlated equilibrium is a dynamic Markovian analog of Bayes correlated equilibrium of Bergemann and Morris (2013; 2016). Thus, Markov correlated equilibrium inherits the robustness property of Bayes correlated equilibrium.

Second, we propose computationally tractable approaches for estimating the model parameters. In general, existing estimation strategies (e.g., nested fixed-point or CCP-inversion) are not applicable since the parameters are set-identified given the multiplicity of Markov correlated equilibria (we do not assume a specific equilibrium selection rule). Furthermore, while computing a Bayes correlated equilibrium is easy as it amounts to solving a linear program, computing a Markov correlated equilibrium is more difficult due to the non-convexities that arise in a dynamic environment. To allow feasible estimation, we propose several computational strategies. The computational tractability stems from our formulation of the equilibrium conditions using the one-shot deviation principle. We then estimate the model using mathematical programming with equilibrium constraints (Su and Judd (2012) and Egesdal, Lai, and Su (2015a)). We illustrate the performance of our approach via numerical examples.

Our framework allows informationally-robust estimation of dynamic games which does not resort to ad hoc assumptions on players' information. The literature on the econometric analysis of games has witnessed burgeoning interest in developing empirical frameworks that are robust to misspecification of information structures (Pakes, Porter, Ho, and Ishii (2015), Gualdani and Sinha (2020), Syrgkanis, Tamer, and Ziani (2021), Magnolfi and Roncoroni (2021), Koh (2022)). While Doval and Ely (2020b) and Makris and Renou (2021) extend

Bayes correlated equilibrium to extensive form games and multi-stage games respectively, this paper limits attention to a class of dynamic games that has served as workhorse models for empirical analysis. In defining Markov correlated equilibrium, we exploit the special structure commonly imposed in this class of models—that the unobserved latent variables are independent over time—which makes the characterization of informationally robust predictions straightforward. To the best of our knowledge, this paper is the first to study informationally robust estimation in a dynamic environment.

The rest of the paper is organized as follows. In Section 2.2, we establish a connection between Markov correlated equilibria and Markov perfect equilibria in a class of dynamic games and discuss its properties. In Section 2.3, we introduce an econometric model and discuss informationally robust identification. Section 2.4 proposes several computational strategies for finding the identified set. Section 2.5 illustrates the performance of our approach via numerical examples. Section 2.6 concludes.

Notation. Throughout the paper, we will use the following notation. If Y is a random variable whose support is a finite set $\mathcal{Y}$, we let $p(y) \equiv p_y \equiv \Pr(Y = y)$ for $y \in \mathcal{Y}$. Similarly, $p(y|x) \equiv p_{y|x} \equiv \Pr(Y = y|X = x)$ denotes the conditional probability of $Y = y$ given $X = x$. We let $\Delta_y \equiv \Delta(\mathcal{Y})$ denote the probability simplex on $\mathcal{Y}$: $p \in \Delta_y$ if and only if $p_y \geq 0$ for all $y \in \mathcal{Y}$ and $\sum_{y\in\mathcal{Y}} p_y = 1$. Similarly, we let $p \in \Delta_{y|x}$ if and only if $p_{y|x} \geq 0$ for each $y \in \mathcal{Y}$ and $\sum_{y\in\mathcal{Y}} p_{y|x} = 1$. We also use the convention that writes an action profile as $a = (a_1, ..., a_I) = (a_i, a_{-i})$.

2.2 Model

We consider a class of dynamic Markovian games in infinite-horizon that have been used as a standard framework in the econometric models of dynamic games literature. We first introduce the concepts in a general environment. Then, we tailor the model when we consider econometric analysis.

2.2.1 Setup

Let $t = 1, 2, ..., \infty$ denote discrete time. A *stationary dynamic Markov game of incomplete information* is characterized by a pair of *basic game* G and *information structure* S. A basic game $G = \langle \mathcal{I}, (\mathcal{A}_i, u_i)_{i\in\mathcal{I}}, \mathcal{X}, \mathcal{E}, \psi, f, \delta \rangle$ specifies payoff-relevant primitives of the model: $i \in \mathcal{I} = \{1, 2, ..., I\}$ indexes players; $a_{it} \in \mathcal{A}_i$ denotes player i's action at time t where $\mathcal{A}_i$ is a finite set of actions available to player i; $a_t = (a_{1t}, ..., a_{It}) \in \mathcal{A} \equiv \times_{i\in\mathcal{I}}\mathcal{A}_i$ denotes an action profile; $x_t \in \mathcal{X}$ denotes a state variable that is publicly observed by the players, and $\mathcal{X}$ is assumed to be finite; $\varepsilon_t \in \mathcal{E}$ denotes a latent state variable that is not directly observed by the players, and $\mathcal{E}$ is assumed to be finite[1]; $\psi(\varepsilon_t|x_t)$ represents the players' prior belief about the probability ε_t conditional on x_t; $f(x_{t+1}|a_t, x_t, \varepsilon_t)$ specifies the probability of transitioning to state x_{t+1} conditional on $(a_t, x_t, \varepsilon_t)$; $u_i : \mathcal{A} \times \mathcal{X} \times \mathcal{E} \to \mathbb{R}$ is the payoff function of player i; $\delta \in [0, 1)$ is the common discount factor.

An information structure $S = \langle (\mathcal{T}_i)_{i=1}^{I}, \pi \rangle$ specifies information-relevant primitives; $\tau_{it} \in \mathcal{T}_i$ denotes player i's private signal at period t where $\mathcal{T}_i$ a finite set of private signals; $\tau_t = (\tau_{1t}, ..., \tau_{It}) \in \mathcal{T} = \times_{i\in\mathcal{I}}\mathcal{T}_i$ is a signal profile; $\pi : \mathcal{X} \times \mathcal{E} \to \Delta(\mathcal{T})$ is a signal distribution that maps the state of the world to a signal profile which allows for arbitrary correlation in the private signals. The interpretation is that each player i does not observe the latent state ε_t directly (x_t is common knowledge) but each player i receives a private signal τ_{it} that is informative about ε_t. Note that the informativeness of the signals can also depend on the public state x_t since π is also a function of x_t.

The game (G, S) is common knowledge to the players. The timing of the model is described as follows. At the beginning of period t, state $x_t \in \mathcal{X}$ is given and publicly observed. Conditional on x_t, the latent state $\varepsilon_t \in \mathcal{E}$ is drawn from the probability distribution $\psi(\cdot|x_t)$. Next, a profile of private signals $\tau_t = (\tau_{1t}, ..., \tau_{It}) \in \mathcal{T}$ is drawn from the signal distribution

[1] The assumption that the state space is finite is used for simplicity in notation and can be relaxed. There is no conceptual difficulty in extending the framework to a more general space. However, finite set assumption makes the connection to econometric analysis transparent because the state space will need to be discretized for feasible estimation.

$\pi(\cdot|x_t, \varepsilon_t)$. At this point, each player i observes (x_t, τ_{it}) which is used to infer the realized latent state ε_t via Bayes' rule. Then, the players simultaneously determine their actions $a_t = (a_{1t}, ..., a_{It})$, and each player i receives period t payoff $u_i(a_t, x_t, \varepsilon_t)$. Finally, given the realization of $(a_t, x_t, \varepsilon_t)$, the observable state x_t transitions to x_{t+1} according to probability distribution $f(x_{t+1}|a_t, x_t, \varepsilon_t)$. Given x_{t+1}, period $t+1$ begins.

The players are assumed to be rational and forward-looking. In each period t, each player i will choose actions to maximize the expected discounted inter-temporal payoffs

$$\mathbb{E}\left\{\sum_{s=t}^{\infty} \delta^{s-t} u_i(a_s, x_s, \varepsilon_s) |x_t, \tau_{it}\right\}.$$

The conditioning on (x_t, τ_{it}) reflects the assumption that i observes the public state x_t and her signal τ_{it} before choosing period t action.

Note that the primitives imply that the transition probability of state variables factors as $P_{x_{t+1},\varepsilon_{t+1}|a_t,x_t,\varepsilon_t} = \psi_{\varepsilon_{t+1}|x_{t+1}} f_{x_{t+1}|a_t,x_t,\varepsilon_t}$. The assumption that the latent variables are generated independently of the states in the previous period is standard in the empirical literature and is crucial for simplifying the computation of equilibria.[2]This assumption will also play a vital role in this paper by allowing us to directly leverage Theorem 1 of Bergemann and Morris (2016) in characterizing the set of robust predictions in a dynamic environment.

Example 2.1 (Two-player dynamic entry game)**.** In a two-player dynamic entry game, players' decisions are binary with $a_{it} = 0$ if firm i is not active in the market and $a_{it} = 1$ if active. The endogenous state variable is the lagged decision $z_{it} = a_{it-1}$ that determines the firm's incumbency status. Potential entrants pay an entry cost to enter if it enters the market. Incumbents do not pay an entry cost to stay active, but receives a scrap value (or pay exit costs) upon exiting. A firm receives profit by operating in the market each period, but the payoff may decrease if the opponent is also present in the market. The per-period payoff can

[2]When information asymmetries are persistent, computation of perfect Bayesian equilibria easily becomes computationally intractable. See Fershtman and Pakes (2012) for a discussion.

be summarized as

$$u_i\left(a_{it}, a_{jt}, z_{it}, \varepsilon_{it}\right)=\begin{cases} MP \times \mathbb{I}\left\{a_{jt}=0\right\}+DP \times \mathbb{I}\left\{a_{jt}=1\right\}+EC \times \mathbb{I}\left\{z_{it}=0\right\}+\varepsilon_{it} & \text{if } a_{it}=1 \\ SV \times \mathbb{I}\left\{z_{it}=1\right\} & \text{if } a_{it}=0 \end{cases}$$

where MP denotes monopoly profit, DP duopoly profit, EC entry cost, and SV scrap value. ε_{it} is the idiosyncratic shock to the operating profit. Here, the latent state is a two-dimensional vector $\varepsilon_t=(\varepsilon_{1t}, \varepsilon_{2t})$.

A common assumption used in the empirical literature is that ε_{it} is privately observed by i. We denote such information structure by S^{ε_i}. Formally, S^{ε_i} is described by a signal space $\mathcal{T}_i=\mathcal{E}_i$ and a signal distribution that places unit mass on the event that $\tau_i=\varepsilon_i$ for each player i. Other information structures can be considered. For example, S^{null} represents an information structure in which each player receives no (null) signal about the realization of ε_t. ■

2.2.2 Markov Perfect Equilibrium

From now on, we suppress the time subscripts unless necessary since the environment is stationary. A Markov strategy of player i is a mapping $\beta_i : \mathcal{X} \times \mathcal{T}_i \rightarrow \Delta\left(\mathcal{A}_i\right)$ that specifies a probability distribution over actions at each realization of observable state and private signal. A strategy profile $\beta=(\beta_1, ..., \beta_I)$ is a Markov perfect equilibrium if for each player i, any action on the support of β_i maximizes i's expected discounted sum of payoffs given that all players behave optimally now and in the future.

Definition 2.1 (Markov Perfect Equilibrium)**.** A strategy profile β is a *Markov perfect equilibrium* of (G, S) if there exists $\left\{\tilde{V}_i\left(x, \tau_i\right)\right\}_{(i,x,\tau_i)\in\mathcal{I}\times\mathcal{X}\times\mathcal{T}_i}$ such that

1. (Bellman) For all $i \in \mathcal{I}$, $x \in \mathcal{X}$, $\tau_i \in \mathcal{T}_i$,

$$\tilde{V}_i\left(x, \tau_i\right)=\max_{\tilde{a}_i \in \mathcal{A}_i} \mathbb{E}^{\beta}_{(a_{-i},\varepsilon)}\left\{u_i\left(\tilde{a}_i, a_{-i}, x, \varepsilon\right)+\delta \sum_{x', \tau_i'} P_{x', \tau_i' | \tilde{a}_i, a_{-i}, x, \varepsilon} \tilde{V}_i\left(x', \tau_i'\right) \middle| x, \tau_i\right\} \quad (2.1)$$

where the expectation is taken over a_{-i} and ε.

2. (Optimality) For all $i \in \mathcal{I}$, $x \in \mathcal{X}$, $\tau_i \in \mathcal{T}_i$, $\beta_i (a_i | x, \tau_i) > 0$ implies

$$a_i = \arg\max_{\tilde{a}_i \in \mathcal{A}_i} \mathbb{E}^{\beta}_{(a_{-i},\varepsilon)} \left\{ u_i (\tilde{a}_i, a_{-i}, x, \varepsilon) + \delta \sum_{x', \tau_i'} P_{x', \tau_i' | \tilde{a}_i, a_{-i}, x, \varepsilon} \tilde{V}_i (x', \tau_i') \middle| x, \tau_i \right\}. \quad (2.2)$$

Equation (2.1) says that each $\tilde{V}_i (x, \tau_i) \in \mathbb{R}$ represents i's value function at information set (x, τ_i) when the players are following β. Equation (2.2) says that each actions on the support of β_i maximizes i's discounted sum of payoffs computed using the value functions.

One-shot Deviation Principle

We will find it useful to restate Definition 2.1 using the one-shot deviation principle which states that β is an equilibrium strategy profile if and only if there exists no player and information set at which a one-shot deviation can be profitable.[3] First, by formulating Markov perfect equilibrium as a Bayes Nash equilibrium of a reduced normal form game induced by a strategy profile, we can leverage results from the static environment.[4] Second, the formulation facilitates computation.

Any strategy profile $\beta = (\beta_1, ..., \beta_I)$ induces an *ex-ante value function* $V_i^{\beta} \in \mathbb{R}^{|\mathcal{X}|}$ for each player i, which captures i's expected payoff at state x (before the realization of ε) when the players follow the prescriptions in β. For each i, V_i^{β} is a unique solution to

$$V_i^{\beta} (x) = \sum_{\varepsilon \in \mathcal{E}, t \in \mathcal{T}, a \in \mathcal{A}} \psi_{\varepsilon | x} \pi_{t | x, \varepsilon} \beta_{a | x, \tau} \left\{ u_i (a, x, \varepsilon) + \delta \sum_{x' \in \mathcal{X}} V_i^{\beta} (x') f_{x' | a, x, \varepsilon} \right\}, \quad \forall i \in \mathcal{I}, x \in \mathcal{X}. \quad (2.3)$$

[3]The one-shot deviation principle states that a strategy profile β is subgame perfect if and only if there are no profitable one-shot deviations (Mailath and Samuelson, 2006). A sufficient condition is that a game is "continuous at infinity" which says events in the distant future are relatively unimportant, a condition which is satisfied in the current environment because the overall payoffs are a discounted sum of per-period payoffs (Fudenberg and Tirole, 1991).

[4]Doraszelski and Escobar (2010) also uses the observation that, holding fixed the value of continued play, the strategic situation that the players face in a given state of a dynamic system is akin to a normal form "static" game.

where $\beta_{a|x,\tau} \equiv \prod_{j=1}^{I} \beta_j (a_j|x, \tau_j)$ denotes the conditional distribution over action profiles induced by the strategy profile β.

Define the *outcome-specific function* given β as

$$v_i^{\beta} (a, x, \varepsilon) \equiv u_i (a, x, \varepsilon) + \delta \sum_{x' \in \mathcal{X}} V_i^{\beta} (x') f_{x'|a,x,\varepsilon}.$$

The outcome-specific function represents the continuation payoff to player i is (a, x, ε) is realized today and all players behave according to β from tomorrow and onward.

A strategy profile β induces a *reduced normal form (basic) game* $G^{\beta} = \langle \mathcal{I}, \left(\mathcal{A}_i, v_i^{\beta} \right)_{i \in \mathcal{I}}, \mathcal{X}, \mathcal{E}, \psi \rangle$. Then $\left(G^{\beta}, S \right)$ describes a "static" game in which player i's payoff function is given by $v_i^{\beta} (a, x, \varepsilon)$.

Lemma 2.1. *A strategy profile β is a Markov perfect equilibrium of (G, S) if and only if β is a Bayes Nash equilibrium of $\left(G^{\beta}, S \right)$.*

2.2.3 Markov Correlated Equilibrium

Decision Rule

A stationary Markov decision rule in (G, S) is a mapping

$$\sigma : \mathcal{X} \times \mathcal{E} \times \mathcal{T} \to \Delta (\mathcal{A})$$

that specifies a probability distribution over action profiles at each realization of state of the world and players' signals.

It is instructive to think of σ as a recommendation strategy of an omniscient mediator who observes (x, ε, τ) and privately recommends actions to each player. Suppose that the mediator commits to a stationary recommendation strategy σ and announces it to the players. After the state and players' signals (x, ε, τ) are realized, an action profile $a = (a_1, ..., a_I)$ is drawn from the probability distribution $\sigma (\cdot|x, \varepsilon, \tau)$, and each a_i is privately recommended to each player i. Each player i, having observed (x, τ_i, a_i), decides whether to obey (play a_i) or not (deviate to $a_i' \neq a_i$). If σ is such that the players are always obedient, we call σ a Markov correlated equilibrium of (G, S).

Definition

To express the equilibrium conditions, let us introduce some notation. Similarly as before, let $V_i^\sigma \in \mathbb{R}^{|\mathcal{X}|}$ denote the ex-ante value function induced by σ; V_i^σ is obtained as the unique solution to

$$V_i^\sigma(x) = \sum_{\varepsilon\in\mathcal{E},\tau\in\mathcal{T},a\in\mathcal{A}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}\left\{u_i(a,x,\varepsilon) + \delta\sum_{x'\in\mathcal{X}} V_i^\sigma(x')\, f_{x'|a,x,\varepsilon}\right\}, \quad \forall x \in \mathcal{X}.$$

Also define

$$v_i^\sigma(a,x,\varepsilon) := u_i(a,x,\varepsilon) + \delta\sum_{x'\in\mathcal{X}} V_i^\sigma(x')\, f_{x'|a,x,\varepsilon}$$

as the outcome-specified function associated with σ; $v_i^\sigma(a,x,\varepsilon)$ represents the payoff to player i if (a,x,ε) is realized today and the players' actions are determined by σ in the future.

Let $\mathbb{E}^\sigma\left[v_i^\sigma(\tilde{a}_i, a_{-i}, x, \varepsilon)\,|x, \tau_i, a_i\right]$ be the expected payoff to player i from player $\tilde{a}_i$ when i observes (x, τ_i) and receives a private recommendation from the mediator to play a_i. The following definition states that σ is a Markov correlated equilibrium of (G, S) if the players do not have incentives to disobey the recommendations of the mediator.

Definition 2.2. A decision rule σ is a *Markov correlated equilibrium* of (G, S) if for each $i \in \mathcal{I}, x \in \mathcal{X}, \tau_i \in \mathcal{T}_i$, and $a_i \in \mathcal{A}_i$, we have

$$\mathbb{E}^\sigma\left[v_i^\sigma(a,x,\varepsilon)\,|x,\tau_i,a_i\right] \geq \mathbb{E}^\sigma\left[v_i^\sigma(\tilde{a}_i, a_{-i}, x, \varepsilon)\,|x,\tau_i,a_i\right] \tag{2.4}$$

for each $\tilde{a}_i \neq a_i$ whenever $\Pr^\sigma(x, t_i, a_i) > 0$.

Since

$$\mathbb{E}^{\sigma}\left[v_i\left(\tilde{a}_i, a_{-i}, x, \varepsilon\right)|x, \tau_i, a_i\right] = \sum_{\varepsilon, a_{-i}} v_i^{\sigma}\left(\tilde{a}_i, a_{-i}, x, \varepsilon\right) \Pr^{\sigma}\left(\varepsilon, a_{-i}|x, \tau_i, a_i\right)$$
$$= \sum_{\varepsilon, a_{-i}} v_i^{\sigma}\left(\tilde{a}_i, a_{-i}, x, \varepsilon\right)\left(\frac{\sum_{\tau_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}}{\sum_{\varepsilon,\tau_{-i},a_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}}\right),$$

after cancelling out the denominator, which is constant across all possible realizations of $(\varepsilon, \tau_{-i}, a_{-i})$, (2.4) can be rewritten as:

$$\sum_{\varepsilon,\tau_{-i},a_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}v_i^{\sigma}\left(a, x, \varepsilon\right) \geq \sum_{\varepsilon,\tau_{-i},a_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}v_i^{\sigma}\left(\tilde{a}_i, a_{-i}, x, \varepsilon\right), \quad \forall i, \tau_i, a_i, \tilde{a}_i. \tag{2.5}$$

As before, any arbitrary σ in (G, S) induces a reduced normal form game (G^{σ}, S) where $G^{\sigma} = \langle \mathcal{I}, (\mathcal{A}_i, v_i^{\sigma}), \mathcal{X}, \mathcal{E}, \psi \rangle$. Then (2.5) implies that if σ is a Markov correlated equilibrium of (G, S), it is a Bayes correlated equilibrium of the reduced-form game (G^{σ}, S) induced by σ.

Lemma 2.2. *A decision rule σ is a Markov correlated equilibrium of (G, S) if and only if σ is a Bayes correlated equilibrium of (G^{σ}, S).*

Comparison to Bayes correlated equilibrium

It is easy to see that when $\delta = 0$, Markov correlated equilibrium collapses to the Bayes correlated equilibrium:

$$\sum_{\varepsilon,\tau_{-i},a_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}u_i\left(a, x, \varepsilon\right) \geq \sum_{\varepsilon,\tau_{-i},a_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}u_i\left(\tilde{a}_i, a_{-i}, x, \varepsilon\right), \quad \forall i, \tau_i, a_i, \tilde{a}_i. \tag{2.6}$$

Thus, the former is a proper dynamic analog of the latter.

Unfortunately, however, contrary its static analog (2.6), the Markov correlated equilibrium conditions (2.5) are not linear with respect to the decision rule. The non-linearity introduces challenges that are absent in the static case. First, computation of a Markov correlated equilibrium requires solving a non-convex program whereas computation of a Bayes correlated equilibrium amounts to solving a linear program. Second, the set of Markov correlated equilibria is generally non-convex. In the static case, it is without loss to assume that the data are generated by a single equilibrium even when they are generated by a mixture of multiple equilibria. In contrast, in the dynamic case, the single equilibrium assumption is no longer without loss since a mixture of multiple equilibria does not necessarily correspond to a Markov correlated equilibrium.

2.2.4 Informational Robustness of Markov Correlated Equilibrium

Partial Ordering of Information Structures

To discuss the informational robustness property of Markov correlated equilibrium, we use a partial order on the set of information structures. We use expansion, defined in BM, which orders information structures based on the level of informativeness of the signals. Let $\omega \in \Omega \equiv \mathcal{X} \times \mathcal{E}$ denote the state of the world.

Definition 2.3 (Expansion)**.** Let $S = (\mathcal{T}, \pi)$ be an information structure. $S^* = (\mathcal{T}^*, \pi^*)$ is an expansion of S, or $S^* \succsim_E S$, if there exists $\left(\tilde{\mathcal{T}}_i\right)_{i=1}^{I}$ and $\lambda : \Omega \times \mathcal{T} \to \Delta\left(\tilde{\mathcal{T}}\right)$ such that $\mathcal{T}_i^* = \mathcal{T}_i \times \tilde{\mathcal{T}}_i$ for all $i = 1, ..., I$ and $\pi^*(\tau, \tilde{\tau}|\omega) = \pi(\tau|\omega)\,\lambda(\tilde{\tau}|\omega, \tau)$.

Intuitively, if $S^* \succsim_E S$, then the players, in addition to the signals in S, get to observe *extra* signals generated by λ.

Reduced Normal Form Games

Let $\beta = (\beta_i)_{i \in \mathcal{I}}$ be a strategy profile in (G, S^*) where $\beta_i : \mathcal{X} \times \mathcal{T}_i \times \tilde{\mathcal{T}}_i \to \Delta(\mathcal{A}_i)$. We say β *induce*s a decision rule σ for (G, S) if

$$\sigma(a|x, \varepsilon, \tau) = \sum_{\tilde{\tau} \in \tilde{\mathcal{T}}} \lambda(\tilde{\tau}|x, \varepsilon, \tau) \prod_{j=1}^{I} \beta_j(a_j|x, \tau_j, \tilde{\tau}_j), \quad \forall a, x, \varepsilon, \tau.$$

The following lemma states that if β induces σ, then the associated reduced-form games, G^β and G^σ, are identical.

Lemma 2.3. *If β induces σ, then the associated reduced normal form games, G^β and G^σ, are identical.*

Lemma 2.3 is intuitive because if β induces σ, then β and σ induce identical distributions over action profiles at each state of the world, so the associated ex-ante value functions must be identical to each other, i.e., $V_i^\beta = V_i^\sigma$. Then, the outcome-specific payoff functions are also identical to each other, i.e., $v_i^\beta = v_i^\sigma$, making all primitives of the basic game identical to each other.

Informational Robustness of Markov Correlated Equilibrium

Our main theorem states that the set of Markov correlated equilibrium of a game (G, S) captures the implications of Markov perfect equilibrium when the players might observe more signals. Our result is a dynamic Markovian analog of Theorem 1 of BM.

A solution concept generates a *prediction* defined as a probability distribution over actions at each realized state and signal. Let $\mathcal{P}^{MPE}_{a|x,\varepsilon,\tau}(G, S)$ denote the set of predictions that can be induced by a Markov perfect equilibrium in (G, S). Let $\mathcal{P}^{MCE}_{a|x,\varepsilon,\tau}(G, S)$ be defined similarly.

Theorem 2.1 (Informational Robustness)**.** *For any basic game G and information structure S, $\mathcal{P}^{MCE}_{a|x,\varepsilon,\tau}(G, S) = \bigcup_{S^* \succsim_E S} \mathcal{P}^{MPE}_{a|x,\varepsilon,\tau}(G, S^*)$.*

Theorem 2.1 is useful because $\mathcal{P}^{MCE}_{a|x,\varepsilon,\tau}(G, S)$ is far easier to characterize than $\bigcup_{S^* \succsim_E S} \mathcal{P}^{MPE}_{a|x,\varepsilon,\tau}$ since the latter requires searching over the set of information structures while the former does not.

In the econometric analysis, it is common to assume that the analyst observes the conditional choice probabilities. Thus, it is useful to characterize the implications of Theorem 2.1 in terms of conditional choice probabilities. Let $\mathcal{P}^{MPE}_{a|x}(G, S)$ denote the set of feasible CCPs that can arise under a Markov perfect equilibrium of (G, S). Let $\mathcal{P}^{MCE}_{a|x}(G, S)$ be defined similarly.

Corollary 2.1. *For any basic game G and information structure S, $\mathcal{P}_{a|x}^{MCE}(G,S) = \bigcup_{S^* \succsim S} \mathcal{P}_{a|x}^{MPE}(G,S^*)$.*

2.3 Econometric Model and Identification

We specialize the model by imposing popular assumptions designed to facilitate econometric analysis. We provide arguments for identification while being agnostic about the underlying information structure of the game.

2.3.1 Setup

Let $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_I)$ and assume that $\varepsilon_i \in \mathcal{E}_i$ only enters player i's payoff. Assume $u_i = u_i^\theta$ and $\psi = \psi^\theta$ where θ is a finite-dimensional parameter vector that contains payoff-parameters that the parameters that govern the distribution of the latent variables. We also assume $f_{x'|a,x,\varepsilon} = f_{x'|a,x}$, i.e., the transition probabilities of observable states are independent of the latent variable ε.

The econometrician observes players' actions and common knowledge state variables across M independent markets over T periods. The data is given as

$$\{a_{m,t}, x_{m,t} : \; m = 1, 2, ..., M, \; t = 1, 2, ..., T\}.$$

We assume that M is large. It is assume that the conditional choice probabilities $\phi : \mathcal{X} \to \Delta(\mathcal{A})$ and the transition probability function $f : \mathcal{A} \times \mathcal{X} \to \Delta(\mathcal{X})$ can be non-parametrically identified when the sample size is large. Furthermore, as standard in the literature, we assume that the common discount factor $\delta \in [0,1)$ is known to the researcher.

The following summarize the set of baseline assumptions used for econometric analysis.

Assumption 2.1 (Baseline assumptions for identification).

1. *The set of covariates $\mathcal{X}$ and the set of latent states $\mathcal{E}$ are finite.*
2. *The prior distribution ψ^θ and the payoff functions u_i^θ are known up to a finite-dimensional parameter θ.*

3. *The state of the world is a vector of player-specific payoff shocks, i.e.,* $\varepsilon = (\varepsilon_1, ..., \varepsilon_I)$ *and* $u_i^\theta(a, x, \varepsilon) = u_i^\theta(a, x, \varepsilon_i)$. *The transition probability of observable states is independent of the latent state, i.e.,* $f_{x'|a,x,\varepsilon} = f_{x'|a,x}$.

4. *The conditional choice probabilities* $\phi : \mathcal{X} \to \Delta(\mathcal{A})$ *and the transition probability function* $f : \mathcal{A} \times \mathcal{X} \to \Delta(\mathcal{A})$ *are identified from the data.*

The only "unconventional" assumption is that $\mathcal{E}$ is finite. One may understand this assumption as discretizing the space of latent variables for estimation to be feasible. The discretization of state space for feasible estimation is also used in Gualdani and Sinha (2020), Magnolfi and Roncoroni (2021), and Syrgkanis, Tamer, and Ziani (2021).

Identified Set

Let $\left(G^\theta, S\right)$ be a dynamic game of incomplete information. Given a solution concept SC, which is used to characterize the feasible probabilities over actions, the identified set of parameters is defined as follows.

Definition 2.4 (Identified set)**.** Given Assumption 2.1, a solution concept SC, and information structure $\tilde{S}$, the identified set of parameters is defined as:

$$\Theta_I^{SC}\left(\tilde{S}\right) \equiv \left\{\theta \in \Theta : \ \phi \in \mathcal{P}_{a|x}^{SC}\left(G^\theta, \tilde{S}\right)\right\}.$$

In words, a candidate parameter θ enters the identified set $\Theta_I^{SC}\left(\tilde{S}\right)$ if the observed CCPs ϕ can be rationalized by some equilibrium (defined by the given solution concept) of the model.

2.3.2 Informationally Robust Identified Set

The following assumption states that the data are generated by a Markov perfect equilibrium under some information structure that is an expansion of a baseline information structure denoted S.

Assumption 2.2 (Identification under Markov perfect equilibrium)**.** *The data are generated by a Markov perfect equilibrium of* $\left(G^{\theta_0}, \tilde{S}^0\right)$ *for some information structure* $\tilde{S}^0$ *that is an expansion of* S.

We will consider a scenario where the true information structure that generates the data is $\tilde{S}^0$, but the researcher only knows some S for which $\tilde{S}^0 \succsim_E S$. That is, the researcher knows some *minimal* information available to the players. For example, the researcher may know that each player i observes at least ε_i. In this case, the analyst has knowledge that $S = S^{\varepsilon_i}$ although she does not know whether the players have access to more information than prescribed by S^{ε_i}. Then, in principle, the analyst should check whether there exists *some* Markov perfect equilibrium under *some* information structure $\tilde{S} \succsim_E S^{\varepsilon_i}$ that can rationalize the data. Clearly, such task is computationally infeasible.

Fortunately, to accomplish the above task, we can replace Assumption 2.2 with the following assumption.

Assumption 2.3 (Identification under Markov correlated equilibrium)**.** *The data are generated by a Markov correlated equilibrium of* $\left(G^{\theta_0}, S\right)$.

Theorem 2.2 (Equivalence of identified sets)**.** *The identified set under Assumptions 2.1 and 2.2 is equal to the identified set under Assumptions 2.1 and 2.3.*

Theorem 2.2 implies that, if the researcher knows that the data are generated by a Markov perfect equilibrium but does not know the underlying information structure, the researcher can proceed by treating the data as if they were generated from a Markov correlated equilibrium. Similar results have been used by Gualdani and Sinha (2020), Magnolfi and Roncoroni (2021), and Syrgkanis, Tamer, and Ziani (2021) in static environments to leverage the informational robustness and computational tractability of Bayes correlated equilibrium. Theorem 2.2 shows that we can extend the result to a popular class of dynamic Markov games.

Equilibrium selection rule

Unlike in the static case in which Bayes correlated equilibrium is the working solution concept, the result is not robust to arbitrary selection over multiple equilibria. Although we do not assume a particular equilibrium selection rule, we must assume that the data are generated by a *single* Markov perfect equilibrium. Identification under arbitrary equilibrium selection rule is feasible but more complicated (e.g., one can use Beresteanu, Molchanov, and Molinari (2011)).

2.4 Computational Strategies

Computing the sharp identified set can be computationally challenging. We propose multiple computational strategies for obtaining the identified set.

2.4.1 Sharp Identified Set

Let $\Theta_I \equiv \Theta_I^{MCE}(S)$. Fix a candidate parameter $\theta \in \Theta$. Let $\partial v_i^\theta(\tilde{a}_i, a, x, \varepsilon_i) \equiv v_i^\theta(\tilde{a}_i, a_{-i}, x, \varepsilon_i) - v_i^\theta(a, x, \varepsilon_i)$ denote the profit from deviating to $\tilde{a}_i$ when the original action is a_i.

Theorem 2.3 (Sharp identified set)**.** *$\theta \in \Theta_I$ if and only if there exists $\sigma : \mathcal{X} \times \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ and $V = (V_{i,x})_{i \in \mathcal{I}, x \in \mathcal{X}} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{X}|}$ that satisfies the following conditions:*

$$\sum_{\varepsilon, \tau_{-i}, a_{-i}} \psi_{\varepsilon|x} \pi_{\tau|x,\varepsilon} \sigma_{a|x,\varepsilon,\tau} \partial v_i^\theta(\tilde{a}_i, a, x, \varepsilon_i) \le 0, \quad \forall x, \tau_i, a_i, \tilde{a}_i \tag{2.7}$$

$$v_i^\theta(a, x, \varepsilon_i) = u_i^\theta(a, x, \varepsilon_i) + \delta \sum_{x'} V_{i,x'} f_{x'|a,x}, \quad \forall i, a, x, \varepsilon_i \tag{2.8}$$

$$V_{i,x} = \sum_{\varepsilon, \tau, a} \psi_{\varepsilon|x} \pi_{\tau|a,x} \sigma_{a|x,\varepsilon,\tau} u_i^\theta(a, x, \varepsilon_i) + \delta \sum_{a, x'} \phi_{a|x} f_{x'|a,x} V_{i,x'}, \quad \forall i, x \tag{2.9}$$

$$\phi_{a|x} = \sum_{\varepsilon, \tau} \psi_{\varepsilon|x} \pi_{\tau|a,x} \sigma_{a|x,\varepsilon,\tau}, \quad \forall a, x \tag{2.10}$$

In the above, (2.7)–(2.9) represent the Markov correlated equilibrium conditions and (2.10) represents the consistency requirement that σ induces the CCPs observed in the data.

Note that (i) $\sigma : \mathcal{X} \times \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ translates to the requirement that $\sigma_{a|x,\varepsilon,\tau} \geq 0$ for all a, x, ε, τ and $\sum_a \sigma_{a|x,\varepsilon,\tau} = 1$ for all x, ε, τ in the optimization problem; (ii) $v_i^\theta(a, x, \varepsilon_i)$ is an auxiliary expression, and (iii) (2.9) is obtained by imposing (2.10). Thus, to test where $\theta \in \Theta_I$, the analyst solves a non-convex program whose variables of optimization are σ and V, and the non-convexity arises from the bilinearity of (2.7) with respect to σ and V. Theorem 2.3 shows that we can use an approach reminiscent of Su and Judd (2012) in order to test whether $\theta \in \Theta_I$. Furthermore, to find the projections of Θ_I in direction $p \in \mathbb{R}^{\dim(\theta)}$, one can solve

$$\min_{\theta,\sigma,V} p^T \theta$$

subject to the constraints in Theorem 2.3.

Example 2.2. When the baseline information structure is set to $S = S^{\varepsilon_i}$, the program in Theorem 2.3 reduces to: find $\sigma : \mathcal{X} \times \mathcal{E} \to \Delta(\mathcal{A})$ and $V_i \in \mathbb{R}^{|\mathcal{X}|}$, $i \in \mathcal{I}$, such that

$$\sum_{\varepsilon_{-i}, a_{-i}} \psi_{\varepsilon|x} \sigma_{a|x,\varepsilon} \partial v_i^\theta(\tilde{a}_i, a, x, \varepsilon_i) \leq 0, \quad \forall i, x, \varepsilon_i, a_i, \tilde{a}_i$$

$$v_i^\theta(a, x, \varepsilon_i) = u_i^\theta(a, x, \varepsilon_i) + \delta \sum_{x'} f_{x'|a,x} V_{i,x'}, \quad \forall i, a, x, \varepsilon_i$$

$$V_{i,x} = \sum_{\varepsilon, a} \psi_{\varepsilon|x} \sigma_{a|x,\varepsilon} u_i^\theta(a, x, \varepsilon_i) + \delta \sum_{a,x'} \phi_{a|x} f_{x'|a,x} V_{i,x'}, \quad \forall i, x$$

$$\phi_{a|x} = \sum_{\varepsilon} \psi_{\varepsilon|x} \sigma_{a|x,\varepsilon}, \quad \forall a, x.$$

■

A Criterion Function Approach

It is useful to take the following approach which computes a criterion function $Q : \Theta \mapsto \mathbb{R}_+$ such that $\theta \in \Theta_I$ if and only if $Q(\theta) = 0$.

$$Q(\theta): \quad \min_{t,V,\sigma} t \quad \text{subject to} \tag{2.11}$$

$$\sum_{\varepsilon,\tau_{-i},a_{-i}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}\sigma_{a|x,\varepsilon,\tau}\left(\partial u_i^\theta\left(\tilde{a}_i,a,x,\varepsilon_i\right)+\delta\sum_{x'}V_{i,x'}\partial f_{x'|\tilde{a}_i,a,x}\right)\leq t, \quad \forall i,x,\tau_i,a_i,\tilde{a}_i$$

$$V_{i,x}=\sum_{\varepsilon,\tau,a}\psi_{\varepsilon|x}\pi_{\tau|a,x}\sigma_{a|x,\varepsilon,\tau}u_i^\theta\left(a,x,\varepsilon_i\right)+\delta\sum_{a,x'}\phi_{a|x}f_{x'|a,x}V_{i,x'}, \quad \forall i,x$$

$$t\geq 0, \quad \phi=\phi\left(\sigma\right), \quad \sigma\in\Delta_{a|x,\varepsilon,\tau}$$

where $\phi=\phi(\sigma)$ is a shorthand for the consistency condition and $\sigma\in\Delta_{a|x,\varepsilon,\tau}$ is a shorthand for the condition that requires σ be proper conditional probability distributions.

Theorem 2.4 (Criterion function approach)**.** *For any $\theta\in\Theta$, (2.11) is feasible. Furthermore, $Q(\theta)=0$ if and only if $\theta\in\Theta_I$.*

2.4.2 Fully Robust Set

In the special case where S is set to S^{null}, i.e., when the players are assumed to receive no minimal information, determining whether θ enters the identified set is more tractable since it involves solving a linear program.

When $S=S^{null}$, the identification conditions collapse to

$$\sum_{\varepsilon,a_{-i}}\psi_{\varepsilon|x}\sigma_{a|x,\varepsilon}\left(\partial u_i^\theta\left(\tilde{a}_i,a,x,\varepsilon_i\right)+\delta\sum_{x'}V_{i,x'}\partial f_{x'|\tilde{a}_i,a,x}\right)\leq 0, \quad \forall i,x,a_i,\tilde{a}_i$$

$$V_{i,x}=\sum_{\varepsilon,a}\psi_{\varepsilon|x}\sigma_{a|x,\varepsilon}u_i^\theta\left(a,x,\varepsilon_i\right)+\delta\sum_{a,x'}V_{i,x'}\phi_{a|x}f_{x'|a,x}, \quad \forall i,x$$

$$\phi_{a|x}=\sum_{\varepsilon}\psi_{\varepsilon|x}\sigma_{a|x,\varepsilon}, \quad \forall a,x.$$

But the best-response condition can be rewritten as

$$\sum_{\varepsilon,a_{-i}}\psi_{\varepsilon|x}\sigma_{a|x,\varepsilon}\partial u_i^\theta\left(\tilde{a}_i,a,x,\varepsilon_i\right)+\delta\sum_{a_{-i}}\phi_{a|x}\left(\sum_{x'}V_{i,x'}\partial f_{x'|\tilde{a}_i,a,x}\right)\leq 0,\forall i,x,a_i,\tilde{a}_i.$$

Thus, determining $\theta \in \Theta_I^{MCE}\left(S^{null}\right)$ can be done by solving a linear program. This suggests that to test $\theta \in \Theta_I^{MCE}(S)$ for any S, one can start by testing $\theta \in \Theta_I^{MCE}\left(S^{null}\right)$, which is easy to compute.

2.5 Numerical Examples

We use numerical examples to test our computational strategies and investigate the identifying power of the proposed solution concept. The first experiment uses a simplified example of Aguirregabiria and Mira (2007), also used in Kasahara and Shimotsu (2012) and Egesdal, Lai, and Su (2015b). The second experiment modifies the example in Pesendorfer and Schmidt-Dengler (2008) and studies the role of excluded variables. The results show that a variation in player-specific excluded observable variables is a key source of identifying power as illustrated in Magnolfi and Roncoroni (2021) in the case of Bayes correlated equilibrium.

2.5.1 Experiment 1

The following game is a simplified version those used in Aguirregabiria and Mira (2007), Kasahara and Shimotsu (2012), and Egesdal, Lai, and Su (2015b). There are $I = 2$ players, and actions are binary, $a_{it} \in \mathcal{A}_i = \{0, 1\}$ where $a_{it} = 1$ denotes entry and $a_i = 0$ denotes staying out. The set of possible market size is $\mathcal{S} = \{2, 6, 10\}$. The per-period payoff functions are given as

$$u_i^\theta\left(a_{it}, a_{-it}, s_t, z_{it}, \varepsilon_{it}\right) = \begin{cases} \theta^{RS}\log\left(s_t\right) - \theta^{RN}a_{-it} - \theta^{FC} - \theta^{EC}\left(1 - z_{it}\right) + \varepsilon_{it}, & \text{if } a_{i,t} = 1 \\ 0 & \text{if } a_{i,t} = 0 \end{cases}$$

where $z_{it} = a_{it-1}$ represents the incumbency status. The common knowledge state is $x_t = (s_t, z_{1t}, z_{2t})$; there are $|\mathcal{X}| = |\mathcal{S}| \times |\mathcal{Z}_1| \times |\mathcal{Z}_2| = 3 \times 2 \times 2 = 12$ points in the support of $\mathcal{X}$. It is assumed that $\varepsilon_{it} \overset{\text{iid}}{\sim} \text{Logistic}\,(0, 1)$. The transition probability function of $s_t \in \mathbb{R}$ is given by the $|\mathcal{S}| \times |\mathcal{S}|$ matrix

$$f_{\mathcal{S}}\left(s_{t+1}|s_t\right) = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

where an (i, j)-th entry denotes $\Pr\left(s_{t+1} = j | s_t = i\right)$. There are four parameters to estimate, $\theta = \left(\theta^{RS}, \theta^{RN}, \theta^{FC}, \theta^{EC}\right)$. We fix the common discount factor at $\delta = 0.96$. The true parameter is set to $\theta_0 = (1.0, 1.4, 1.0, 1.0)$. Each $\mathcal{E}_i$ was discretized to 8 points using the method described in Kennan (2006). The conditional choice probability ϕ was generated by finding a Markov perfect equilibrium under the assumption that each player i observes ε_i but not ε_{-i}.

Projections of the identified set

Table 2.1 shows the projection intervals obtained by using the projection method.

Table 2.1: Projections of the Identified Set

		Projection Interval	
	True value	L	R
θ^{RS}	1.0	-0.818	2.553
θ^{RN}	1.4	-5.137	7.418
θ^{FC}	1.0	-0.120	1.395
θ^{EC}	1.0	0.144	1.681

Notes: The table entry represents the projection intervals of the identified set.

Table 2.1 shows that the projected intervals covers the true parameter as expected. However, the projection intervals show that the identified set can be quite large in the given setting. The fact that the Markov correlated equilibrium identified set is large points out that the common assumption that the underlying information structure of the game is known to the analyst is strong.

We find that since the dimensionality of the problem is large, the solver often converges to a local optimum or fails to converge. Hence it can be useful to use many starting points which is parallelizable. Also, we have found it useful start the search with outer bounds by relaxing the constraints by a small tolerance level and further tightening the constraints for sharper bounds.

Visualization of the Identified sets

To provide a visualization of the identified sets, Figure 2.1 plots subsets of the identified set obtained by finding the identified set of pairs of parameters on fine grids while hold the other parameters at their true values. For example, to obtain Figure 2.1-(a), we impose $\left(\theta^{FC}, \theta^{EC}\right) = (1.0, 1.0)$, and plot the identified set of $\left(\theta^{RS}, \theta^{RN}\right)$.

In Figure 2.1, we show two identified sets. The "sharp" set corresponds to $\Theta_I^{MCE}\left(S^{\varepsilon_i}\right)$ and the "outer" set corresponds to $\Theta_I^{MCE}\left(S^{null}\right)$. We have shown that $\Theta_I^{MCE}\left(S^{\varepsilon_i}\right) \subseteq \Theta_I^{MCE}\left(S^{null}\right)$ and that computing $\Theta_I^{MCE}\left(S^{null}\right)$ is easier since determining $\theta \in \Theta_I^{MCE}\left(S^{null}\right)$ amounts to solving a linear program. The figure demonstrates that while the outer sets can be quite loose compared to the sharp set, they help rule out points that do not enter the sharp set.

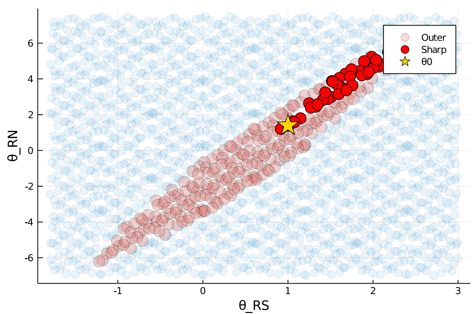
2.5.2 Experiment 2

This example modifies the simple example used in Pesendorfer and Schmidt-Dengler (2008) by adding excluded variables. In this example, there are $I = 2$ players whose per-period payoffs are given as

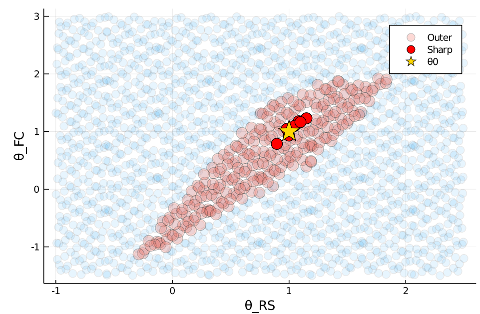
$$u_i\left(a_{i,t}, a_{-i,t}, z_{i,t}, w_i, \varepsilon_{i,t}\right) = \begin{cases} \theta^m + \theta^c a_{-i,t} + \theta^e\left(1 - z_{i,t}\right) + \theta^w w_i + \varepsilon_{i,t} & \text{if } a_{i,t} = 1 \\ \theta^\kappa z_{i,t} & \text{if } a_{i,t} = 0 \end{cases}$$

where $a_{i,t} \in \mathcal{A}_i = \{0, 1\}$ is the entry decision, $z_{i,t} \in \mathcal{Z}_i = \{0, 1\}$ denotes the incumbency status, and $w_i \in \mathcal{W}_i$ is a player-specific time-invariant excluded variable. There are five parameters $\theta = (\theta^m, \theta^c, \theta^e, \theta^w, \theta^\kappa)$; θ^m is the monopoly profit; θ^c is the competition effect; θ^e is the entry cost; θ^w is the parameter associated with the excluded variable; θ^κ is the scrap

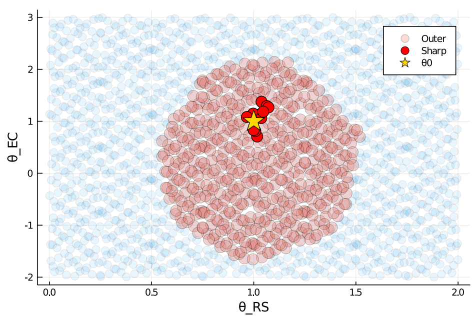
Figure 2.1: Subset of identified sets for Experiment 1



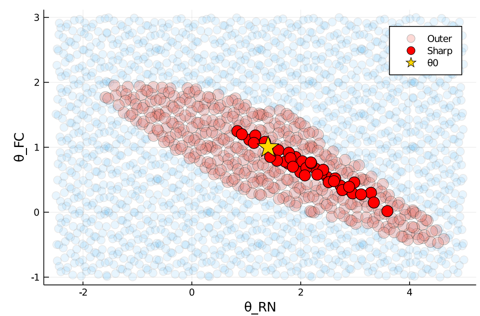
(a) Case 1: $\left(\theta^{RS}, \theta^{RN}\right)$



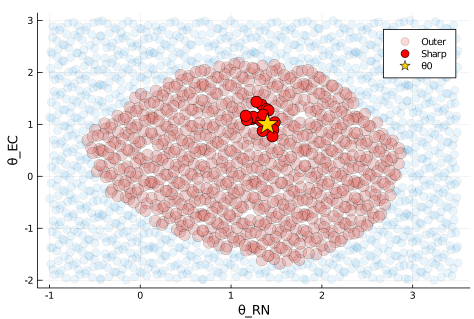
(b) Case 2: $\left(\theta^{RS}, \theta^{FC}\right)$



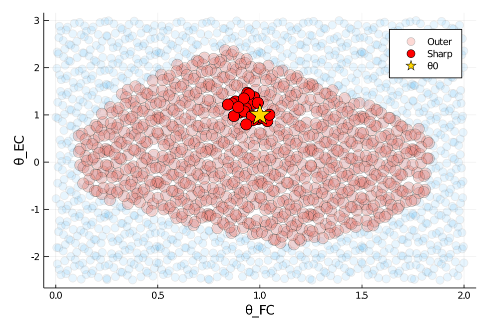
(c) Case 3: $\left(\theta^{RS}, \theta^{EC}\right)$



(d) Case 4: $\left(\theta^{RN}, \theta^{FC}\right)$



(e) Case 5: $\left(\theta^{RN}, \theta^{EC}\right)$



(f) Case 6: $\left(\theta^{FC}, \theta^{EC}\right)$

value.

We set $\delta = 0.9$ and $\theta_0 = (\theta_0^m, \theta_0^c, \theta_0^e, \theta_0^w, \theta_0^\kappa) = (1.2, -0.8, -0.5, 1.0, 0.3)$. We consider three sets of excluded covariates: $\mathcal{W}_i^0 = \{-0.5, 0.5\}$, $\mathcal{W}_i^1 = \{-1.0, 1.0\}$, and $\mathcal{W}_i^2 = \{-1.5, 1.5\}$. Following Pesendorfer and Schmidt-Dengler (2008), we assume θ_0^κ is known.

Table 2.2: Projection of Identified Set

		$\mathcal{W}_i^0 = \{-0.5, 0.5\}$		$\mathcal{W}_i^1 = \{-1, 1\}$		$\mathcal{W}_i^2 = \{-1.5, 1.5\}$	
	True value	L	R	L	R	L	R
θ^m	1.2	-5.88	6.89	0.49	3.41	0.27	2.35
θ^c	-0.8	-10.0	10.0	-4.37	0.18	-2.59	0.22
θ^e	-0.5	-1.56	0.85	-1.04	0.08	-1.05	0.42
θ^w	1.0	-0.30	2.08	0.71	1.39	0.69	1.19

Notes: In obtaining the identified set, each parameter was restricted to lie in the [-10,10] interval.

Table 2.2 shows that, as expected, larger variation in excluded covariates shrinks the identified set considerably.

2.6 Conclusion

In this paper, we have studied informationally robust econometric analysis in a class of dynamic Markov games. We have introduced Markov correlated equilibrium and showed that it allows tractable characterization of Markov perfect equilibrium predictions when the underlying information structure is unknown to the analyst. We have also proposed computational strategies for obtaining the identified sets.

Several tasks remain for future research. First, it is important to introduce computationally light approaches that will work well for problems of realistic size so that the proposed framework can be readily applied for various empirical works. Second, it is also important to propose tractable strategies for constructing the confidence set. Our conjecture is that using the approaches proposed by Kline and Tamer (2016) or Horowitz and Lee (2021) will be complementary to our estimation strategies. Third, we aim to apply the framework to study the dynamic entry game by coffee chains in the US. Coffee chain industry has witness rapid increase in the number of coffee stores by major players such as Starbucks over the past decades, indicating that the firms' strategic decisions ought to be analyzed using dynamic games framework. We plan to compare the estimation results with and without strong assumptions on players' information.

Chapter 3: Estimating Discrete Games of Complete Information: Bringing Logit Back in the Game

3.1 Introduction

With increasing access to rich data and computational capabilities, empirical analysis of game-theoretic models has become standard in economics, especially in the field of empirical industrial organization. However, estimation of games with multiple equilibria remains computationally challenging since the parameters are usually partially identified. Traditional statistical approaches such as the maximum likelihood estimation are not applicable when the model is "incomplete" à la Tamer (2003). While assumptions that "complete" the model (e.g., by assuming a particular equilibrium selection mechanism) can lead to point-identification, it is often difficult for the researcher to know the true selection mechanism that governs the data generating process, and using arbitrary assumptions can lead to biased estimates.[1]

A series of important works have proposed practical algorithms for estimating games while being agnostic on the unknown equilibrium selection rules (e.g., Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010b), Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2011), Henry, Méango, and Queyranne (2015), Pakes, Porter, Ho, and Ishii (2015), and Kline and Tamer (2016)). However, existing methods are often costly to implement because they require a combination of (i) simulation to approximate model-implied probabilities, (ii) repeatedly solving for the equilibria at each game, and/or (iii) a grid search over the parameter space. For example, a recent paper by Ciliberto and Jäkel (2021), which uses Ciliberto and Tamer (2009)'s routine to estimate the entry game by superstar exporters,

[1]See Ellickson and Misra (2011), de Paula (2013), Ho and Rosen (2017), and Aradillas-López (2020) for recent surveys of econometric methodologies for estimating static discrete games.

reports that the estimation algorithm took a week to run.[2]

This paper contributes to the literature on the econometric analysis of games by proposing a novel estimation algorithm which is extremely fast and simple to implement. Our approach applies to static discrete games of complete information with finite actions and finite players, assumes pure strategy Nash equilibrium, and makes no assumption on the equilibrium selection rule. The approach is fast because it requires neither simulation, nor solving for the equilibria, nor grid search. We characterize an identified set whose restrictions are expressed in terms of closed-form multinomial logit probabilities. The identified set is non-sharp, but numerical examples suggest that it is tight in practice.

Let us preview the main results. We show that, under standard assumptions, we can find an (non-sharp) identified set that can be expressed as

$$\Theta_I := \left\{ \theta \in \Theta : \ \phi\left(y|x\right) \leq \prod_{i=1}^{I} \frac{\exp\left(v_i\left(y,x;\theta\right)\right)}{\sum_{y_i' \in \mathcal{Y}_i} \exp\left(v_i\left(y_i', y_{-i}, x;\theta\right)\right)}, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \right\}$$

where θ is a finite-dimensional structural parameter; $i = 1, ..., I$ are players; $y_i \in \mathcal{Y}_i$ denotes player i's action where $\mathcal{Y}_i$ is finite, and $y = (y_1, ..., y_I) \in \times_{i=1}^{I} \mathcal{Y}_i$ denotes an action profile; $x \in \mathcal{X}$ denotes observable covariates where $\mathcal{X}$ is finite; $v_i\left(y,x;\theta\right)$ denotes the deterministic part of player i's payoff functions known up to θ; $\phi\left(y|x\right)$ denotes the conditional choice probabilities (the probability of observing action profile y at covariate bin x) that are identified by the data. The expression says that whether a candidate parameter enters the identified set can be determined using closed-form expressions. We can obtain the above characterization by assuming that the players' payoff functions are additively separable and that their "payoff shocks" independently and identically follow the type 1 extreme value distribution. Furthermore, when $v_i\left(y,x;\theta\right)$ is linear in θ, Θ_I is convex and its projections can be found using convex programs; the projection intervals corresponding to each component of θ can be found by solving two convex programs: one for the lower bound and one for the upper

[2]Computation time depends on multiple factors, e.g., method for conducting inference, the number of simulation draws, how long to explore the parameter space, etc. In any case, estimating games without equilibrium selection rules is generally regarded as computationally costly.

bound. Since the identified set uses a subset of identifying restrictions that characterize the sharp set, we have $\Theta_I^{sharp} \subseteq \Theta_I$, i.e., Θ_I is an outer set.

The main insight of our approach is to use a subset of identifying restrictions that are particularly easy to handle. Our characterization is based on Galichon and Henry (2011) who show that the sharp identified set under the pure strategy Nash equilibrium assumption can be characterized by a finite set of inequalities (whose expressions are not necessarily in closed-form). We show that focusing on a subset of restrictions (namely singleton classes) allows us to express the restrictions in closed-forms when combined with the assumption that the payoff shocks are independently and identically type 1 extreme value distributed. In games with binary actions, this amounts to computing the probabilities over rectangles that are merely the product of logistic cumulative probabilities.

We also propose a simple approach to constructing confidence sets for the identified sets. The proposed approach leverages the key insights from Horowitz and Lee (2021). The main advantage of the approach is that it is extremely fast and easy to implement. Finding the projections of the confidence sets amounts to solving a set of convex programs.

We illustrate the usefulness of our approach using real-world data. We apply our methodology to the empirical models in Kline and Tamer (2016) and Ellickson and Misra (2011). We download the datasets used in their papers and apply our methodology. We find that our method yields results that are qualitatively similar to those reported in the original papers. Moreover, in both examples, it takes less than one CPU second in total to obtain the projection of the confidence sets.

The main contribution of this paper is to provide a fast and simple approach that dramatically lowers the costs of estimating static games of complete information with no ad hoc assumptions on the equilibrium selection rule. To the best of our knowledge, no existing algorithm for estimating complete information games (without ad hoc assumption on the equilibrium selection rule) leverages the multinomial logit formula.[3] Thus, this paper shows

[3]While Bajari, Hong, and Ryan (2010b) use multinomial logit formula to model the equilibrium selection probabilities, their method requires separately identifying the set of all equilibria at each simulated draw of

how to bring logit "back in the game."

The rest of the paper is organized as follows. Section 3.2 introduces the model and propose an identified set that is characterizable with closed-form inequalities. Section 3.3 presents computational strategies for obtaining and summarizing the identified set via convex programs. Section 3.4 proposes a computationally tractable approach to constructing confidence sets. Section 3.5 shows the performance of our methodology when applied to a real-world dataset. Section 3.6 concludes. All proofs are in Appendix C.1.

3.2 Theory

3.2.1 Preliminaries

Basic Setup

Our framework follows the exposition in Galichon and Henry (2011) (henceforth GH). An econometric model is specified as

$$y \in G\left(x, \varepsilon; \theta\right).$$

$y \in \mathcal{Y}$ denotes outcome variables; $x \in \mathcal{X}$ denotes a vector of exogenous covariates observable to the researcher; $\varepsilon \in \mathcal{E}$ is a vector of latent variables, unobserved by the researcher. The distribution of ε is denoted as $\nu\left(\cdot|x;\theta\right)$; $\theta \in \Theta \subseteq \mathbb{R}^d$ is a vector of parameters that the econometrician wants to identify; $G : \mathcal{X} \times \mathcal{E} \times \Theta \rightrightarrows \mathcal{Y}$ represents an *incomplete econometric model*: given θ, the model specifies the set of possible outcomes at each economic state (x, ε). We assume that the econometrician observes $\{y_m, x_m\}_{m=1}^n$ where the observations are independent across m. We assume that as $n \to \infty$, the econometrician can identify the conditional choice probabilities (CCP) $\phi\left(y|x\right) \equiv \Pr\left(Y = y|X = x\right)$ for each $y \in \mathcal{Y}$ and $x \in \mathcal{X}$.

unobserved states. In contrast, our approach demonstrates that multinomial logit formula can be used to construct bounds on the conditional choice probabilities without resorting to simulation and enumeration of equilibria.

Static Discrete Games of Complete Information

Let us specialize to static discrete games of complete information with finite players and actions. Let $m = 1, ..., n$ be independent markets; we will suppress subscript m unless required. Let $i \in \mathcal{I} = \{1, ..., I\}$ be the players. Player i's action is denoted by $y_i \in \mathcal{Y}_i$ where $\mathcal{Y}_i$ is finite. $(y_1, y_2, ..., y_I) \in \mathcal{Y} \equiv \times_{i=1}^{I} \mathcal{Y}_i$ denotes an action profile, and we use the convention $y = (y_i, y_{-i})$ where $y_{-i} = (y_1, ..., y_{i-1}, y_{i+1}, ..., y_I)$ denotes the actions of i's opponents. Covariate vector is denoted by $x \in \mathcal{X}$ where $\mathcal{X}$ is finite; x collects observable traits of each market that are player-specific as well as common to the players.[4] Finally, $\varepsilon \in \mathcal{E}$ captures all latent variables that enter the players' payoff functions. As is standard in the empirical literature, we assume that $\varepsilon = (\varepsilon_1, ..., \varepsilon_I)$ and each ε_i enters only i's payoff. We assume that information is complete, i.e., the realized state of the world (x, ε) is common knowledge to the players. Each player's payoff functions are given by $u_i : \mathcal{Y} \times \mathcal{X} \times \mathcal{E}_i \times \Theta \to \mathbb{R}$. The relevant solution concept is pure strategy Nash equilibrium.[5] Then $G(\cdot|x, \varepsilon; \theta)$ describes the mapping from the exogenous state (x, ε) to set of pure strategy Nash equilibria outcomes that can be supported at (x, ε).

Galichon and Henry (2011)'s Characterization of Sharp Identified Sets

The identified set proposed in this paper exploits GH's characterization of the sharp identified sets. For clarity, let us review their main findings. For any subset $A \subseteq \mathcal{Y}$, assume that $G^{-1}(A|x; \theta) := \{\varepsilon \in \mathcal{E} : G(\varepsilon|x; \theta) \cap A \neq \emptyset\}$ is measurable so that the *generalized likelihood function* (also called *Choquet capacity*) $\mathcal{L}(A|x; \theta) = \nu(G^{-1}(A|x; \theta)|x; \theta)$ is well-defined.[6] In words, $G^{-1}(A|x; \theta)$ collects all state vector ε such that the economic model yields some y in A as a possible outcome, and $\mathcal{L}(A|x; \theta)$ measures its probability; put differently, $\mathcal{L}(A|x; \theta)$ represents the *maximal* probability of observing some outcome in A. Theorem 1 of GH states that the sharp identified set, denoted Θ_I^{sharp}, is equal to the set of θ's such that the

[4] The assumption that $\mathcal{X}$ is finite is not necessary for identification but simplifies the estimation procedure.
[5] We do not consider mixed strategy Nash equilibrium.
[6] See GH for details on the technical assumptions such as measurability.

true conditional choice probabilities are dominated by the generalized likelihood for every possible subset A of $\mathcal{Y}$ (or, more formally, true conditional choice probabilities lie in the core of the Choquet capacity $\mathcal{L}$).

Theorem 3.1 (Galichon and Henry (2011))**.** *The sharp identified set Θ_I^{sharp} is equal to the set of parameters such that the true distribution of the observable variables lies in the core of the generalized likelihood predicted by the model. Hence,*

$$\Theta_I^{sharp} = \left\{ \theta \in \Theta : \textit{ for every subset } A \textit{ of } \mathcal{Y}, \sum_{y \in A} \phi(y|x) \leq \mathcal{L}(A|x;\theta), \ x - a.s. \right\}. \tag{3.1}$$

The significance of Theorem 3.1 is that it provides an operational characterization of the sharp identified set: for any candidate parameter $\theta \in \Theta$, determining whether $\theta \in \Theta_I^{\text{sharp}}$ can be done using a finite number of inequalities without resorting to an (unknown) equilibrium selection rule which is an infinite-dimensional nuisance parameter.

However, numerically computing Θ_I^{sharp} using (3.1) can be costly because the researcher needs to (i) compute $\mathcal{L}(A|x;\theta)$ for each subset A which generally requires simulating a large number of ε's, (ii) solve for the equilibria at each realized draw (i.e., find the set $G(x, \varepsilon; \theta)$ at every point of x, ε, and θ), and (iii) iterate (i)-(ii) at a large number of points in the parameter space Θ. The high computational burden due to (i)-(iii) is a common feature of existing approaches for estimating complete information games (e.g., Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010b), Beresteanu, Molchanov, and Molinari (2011), and Henry, Méango, and Queyranne (2015)).

3.2.2 A Tractable Identified Set

We propose an (non-sharp) identified set that is characterized by a set of closed-form inequalities. The key idea is to use, out of all inequalities that define the sharp set in (3.1), only those based on the *singleton class* (subsets of $\mathcal{Y}$ that are singletons), and assume that the payoff shocks are additive and independently and identically follow the type 1 extreme value distribution.[7]

We make the following assumption which is quite standard.[8]

Assumption 3.1. *Each player's payoff function takes an additively separable form* $u_i(y_i, y_{-i}, x, \varepsilon_i; \theta) = v_i(y_i, y_{-i}, x; \theta) + \varepsilon_i(y_i)$*. Each* $\varepsilon_i(y_i) \in \mathbb{R}$ *independently and identically follows the type 1 extreme value distribution.*

We use a two-player entry game as a running example throughout the paper. The empirical applications in Section 3.5 are extensions of the running example.

Example 3.1 (Two-player entry game without covariates.)**.** Consider a simple two-player entry game without covariates. Each player $i = 1, 2$ can choose to enter $(y_i = 1)$ or stay out $(y_i = 0)$. The set of possible outcomes is $\mathcal{Y} \equiv \times_{i=1}^{2} \mathcal{Y}_i$ where $\mathcal{Y}_i = \{0, 1\}$. The exogenous state of the world is given by $\varepsilon = (\varepsilon_1, \varepsilon_2)$ where each $\varepsilon_i = (\varepsilon_i(0), \varepsilon_i(1)) \in \mathbb{R}^2$ describes action-specific payoff shocks. After observing the realization of $(\varepsilon_1, \varepsilon_2)$, the players choose actions simultaneously. Each player i receives the following payoff:

$$u_i(y_i, y_{-i}, \varepsilon_i; \theta) = \begin{cases} \beta_i + \Delta_i y_j + \varepsilon_i(1) & \text{if } y_i = 1 \\ \varepsilon_i(0) & \text{if } y_i = 0 \end{cases}$$

where $\beta_i \in \mathbb{R}$ is i's intercept and $\Delta_i \in \mathbb{R}$ is the "competition effect" parameter that captures how much j's presence increases or decreases i's profit when i is operating in the market. In this case, we have $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2)$ and $v_i(y_i, y_j; \theta) = y_i(\beta_i + \Delta_i y_j)$. ■

We claim that, under Assumption 3.1, the generalized likelihood functions $\mathcal{L}(A|x;\theta)$ have closed-form expressions when $A \subseteq \mathcal{Y}$ is a singleton. Let us abuse the notation and write $\mathcal{L}(y|x;\theta) \equiv \mathcal{L}(\{y\}|x;\theta)$. When we restrict attention to the singleton class, we obtain an

[7] Andrews, Berry, and Barwick (2004) also uses identifying restrictions that are equivalent to ours, but resorts to simulation when generating the moment inequalities.

[8] The assumption that the errors follow the type 1 extreme value distribution has been more common for incomplete information games. See, e.g., Bajari, Hong, Krainer, and Nekipelov (2010a).

outer set defined as[9]

$$\Theta_I := \{\theta \in \Theta : \ \phi(y|x) \leq \mathcal{L}(y|x;\theta) \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}\}. \tag{3.2}$$

Clearly, $\Theta_I^{\text{sharp}} \subseteq \Theta_I$. Although Θ_I is non-sharp, the following theorem shows that it can be attractive to practitioners because computing it is easy.

Theorem 3.2. *Under Assumption 3.1, the identified set in (3.2) can be rewritten as*

$$\Theta_I = \left\{\theta \in \Theta : \ \phi(y|x) \leq \prod_{i=1}^{I} \frac{\exp(v_i(y,x;\theta))}{\sum_{y_i' \in \mathcal{Y}_i} \exp(v_i(y_i', y_{-i}, x;\theta))}, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}\right\}. \tag{3.3}$$

Suppose that the deterministic part of the payoff function $v_i(y,x;\theta)$ is known up to θ, and the CCPs $\phi(y|x)$ are identified from the data. Theorem 3.2 shows that, for each candidate θ, determining whether $\theta \in \Theta_I$ can be done using closed-form expressions; the researcher (i) does not have to simulate ε's, and (ii) does not need to solve for the equilibria of the game even once. Checking whether $\theta \in \Theta_I$ is extremely easy even when the sets $\mathcal{Y}$ and $\mathcal{X}$ are large. Section 3.3 shows that, by adding an assumption, the projections of Θ_I can be obtained without conducting a grid search over the parameter space.

Illustration in a Two-player Entry Game

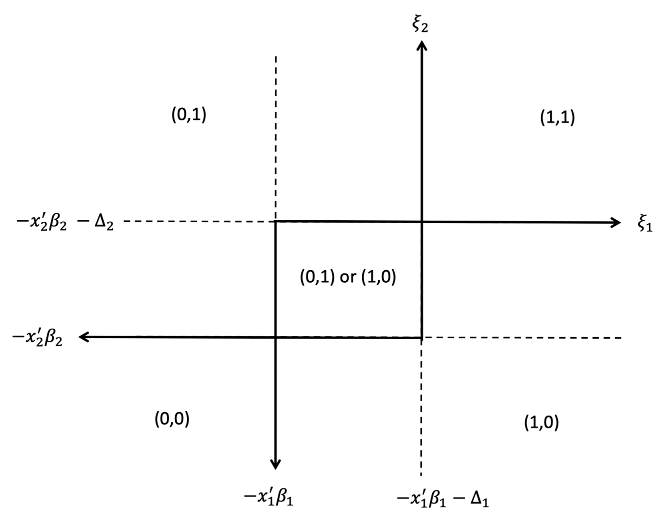
Let us illustrate the idea behind Theorem 3.2 using an example that extends Example 3.1 by introducing observable covariates. Suppose that observable covariates $x \in \mathcal{X}$ are available and that the payoff functions are

$$u_i(y_i, y_j, x;\theta) = \begin{cases} x_i^T \beta_i + \Delta_i y_j + \varepsilon_i(1) & \text{if } y_i = 1 \\ \varepsilon_i(0) & \text{if } y_i = 0 \end{cases}$$

[9] For simplicity in notation, we assume that $\mathcal{X}$ is finite and the associated random variable X has full support on $\mathcal{X}$ so that we can drop the "almost-surely" qualifier.

Then $y_i = 1$ if and only if $x_i^T \beta_i + \Delta_i y_j + \xi_i \geq 0$ where $\xi_i \equiv \varepsilon_i(1) - \varepsilon_i(0)$ follows the standard logistic distribution. When $\Delta_i < 0$ for $i = 1, 2$, it is well-known that the set of equilibria at each realization of (ξ_1, ξ_2) are given by Figure 3.1 (see, e.g., Tamer (2003), Ciliberto and Tamer (2009), or Aradillas-López (2020)). The ξ's at the center region admits two outcomes, $(0,1)$ and $(1,0)$, as Nash equilibria. The other region admits a unique equilibrium.

Figure 3.1: Structure of equilibria in two-player entry game

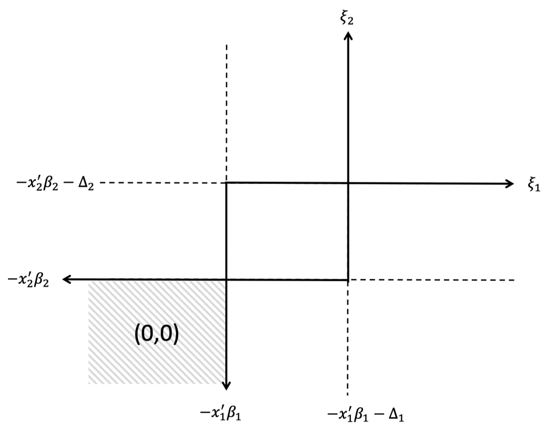


In Figure 3.2, we represent the regions in (ξ_1, ξ_2)-space corresponding to $G^{-1}(y|x;\theta)$ for each $y \in \mathcal{Y}$. For example, Figure 3.2-(a) shows the set of (ξ_1, ξ_2) that supports the outcome $y = (0,0)$ (both players do not enter) as a possible Nash equilibrium. Similarly, Figure 3.2-(b) represents the set of (ξ_1, ξ_2) which supports $(y_1, y_2) = (0,1)$ as a possible Nash equilibrium.

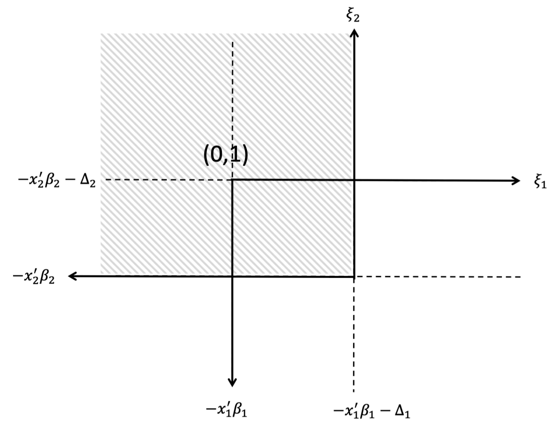
A common feature of $G^{-1}(y|x;\theta)$ for each y is that they are *rectangles* whose probabilities are extremely easy to compute since each $\xi_i \equiv \varepsilon_i(1) - \varepsilon_i(0)$ follows the standard logistic distribution.[10] It is straightforward to see that (3.3) is equivalent to: $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2) \in$

[10]Recall that if a random variable Z follows a standard logistic distribution, then, for each $z \in \mathbb{R}$, $\Pr(Z \leq z) = 1/(1 + e^{-z})$ and $\Pr(Z > z) = e^{-z}/(1 + e^{-z})$.

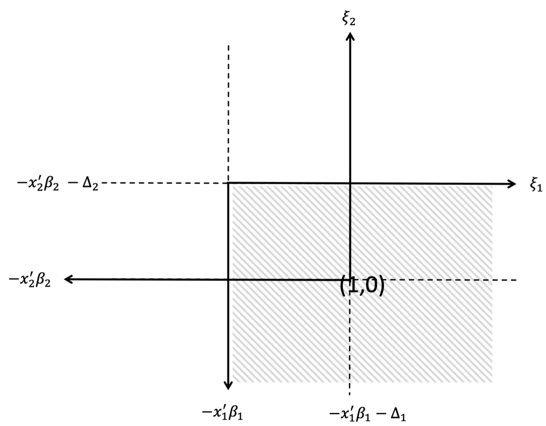
Figure 3.2: Computing probability of rectangles



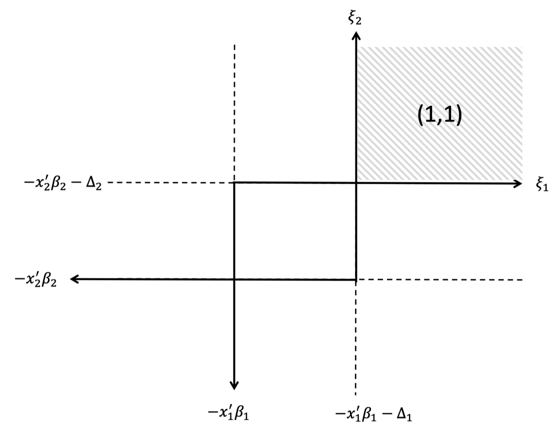
(a) $G^{-1}(0,0|x;\theta)$



(b) $G^{-1}(0,1|x;\theta)$



(c) $G^{-1}(1,0|x;\theta)$



(d) $G^{-1}(1,1|x;\theta)$

Θ_I if and only if for all $x \in \mathcal{X}$,

$$\phi(0,0|x) \leq F(\beta_1' x_1) F(\beta_2' x_2)$$

$$\phi(1,1|x) \leq (1 - F(\beta_1' x_1 + \Delta_1))(1 - F(\beta_2' x_2 + \Delta_2))$$

$$\phi(1,0|x) \leq (1 - F(\beta_1' x_1 + \Delta_1)) F(\beta_2' x_2)$$

$$\phi(0,1|x) \leq F(\beta_1' x_1)(1 - F(\beta_2' x_2 + \Delta_2))$$

where $F(z) \equiv 1/(1+e^{-z})$ is the cumulative distribution function of the standard logistic random variable. The above inequalities are intuitive given that the probabilities of the rectangles represent the maximal probabilities of observing each outcome.

3.2.3 Comparison to Sharp Identified Set

Using an outer set can result in unnecessary loss of information compared to the sharp set that exhausts all information available to the researcher. At the minimum, our identified set can help quickly rule out parameters that are rejected by the model.[11] However, it is natural to ask how loose the outer set can be. We examine this question using a numerical example.

Let us continue with Example 3.1. Set $\theta = (\beta_1, \beta_2, \Delta_1, \Delta_2) = (0.0, 0.0, -0.5, -0.5)$ and assume that the equilibrium selection rule is symmetric. The model yields the choice probability vector $(\phi_{00}, \phi_{10}, \phi_{01}, \phi_{11}) = (0.250, 0.304, 0.304, 0.142)$.

Table 3.1: Comparison of the Identified Sets

	Sharp Set	Outer Set
β_1	$[-0.214, 0.193]$	$[-0.217, 0.196]$
β_2	$[-0.214, 0.193]$	$[-0.217, 0.196]$
Δ_1	$[-0.936, -0.014]$	$[-0.945, -0.005]$
Δ_2	$[-0.936, -0.014]$	$[-0.945, -0.005]$

Table 3.1 reports the projections of the sharp set Θ_I^{sharp} and our outer set Θ_I. Comparison of the intervals reveals that although the outer set generates projection intervals that are slightly wider, the difference is small. Thus, we conclude that our identified set is quite tight in practice.

3.2.4 Convex Identified Set

Theorem 3.2 makes no assumption on the structure of $v_i(y, x; \theta)$; as long as the functional form of $v_i(y, x; \theta)$ is known up to θ, determining $\theta \in \Theta_I$ can be done by evaluating the closed-form inequalities. However, the researcher still needs to test $\theta \in \Theta_I$ over many candidate parameters in the parameter space.

[11] Kédagni, Li, and Mourifié (2021) discusses warnings against using outer sets, namely that outer sets can contradict each other when the model is misspecified. Their results suggest that the researcher should use the sharp characterization whenever possible. If the researcher shares the concern, the researcher can "sharpen" our outer set by applying methodologies designed to obtain the sharp set, e.g., Henry, Méango, and Queyranne (2015).

Fortunately, it turns out that adding a mild assumption can facilitate the computation even further by eliminating the need for a grid search. Specifically, we can show that finding the projections of Θ_I can be done via convex programs if the deterministic part of the payoff functions are linear in the parameters. Here, we establish that Θ_I is convex if $v_i(y, x; \theta)$ in linear in θ.

Assumption 3.2. *$v_i(y, x; \theta)$ is linear in θ.*

Let us rearrange the inequalities (3.3) in a way that allows us to leverage Assumption 3.2. Rewrite the inequalities in (3.3) as

$$\phi(y|x) \leq \prod_{i=1}^{I} \frac{\exp(v_i(y, x; \theta))}{\sum_{y_i' \in \mathcal{Y}_i} \exp(v_i(y_i', y_{-i}, x; \theta))}, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \tag{3.4}$$

$$\Leftrightarrow g_{y,x}(\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \tag{3.5}$$

where

$$g_{y,x}(\theta) := \log(\phi(y|x)) + \sum_{i=1}^{I} \left\{ \log\left(\sum_{y_i' \in \mathcal{Y}_i} \exp(v_i(y_i', y_{-i}, x; \theta)) \right) - v_i(y, x; \theta) \right\} \tag{3.6}$$

is obtained by taking logs on both sides of (3.4) and rearranging the terms. Thus, we have

$$\Theta_I \equiv \{\theta \in \Theta : \ g_{y,x}(\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}\}. \tag{3.7}$$

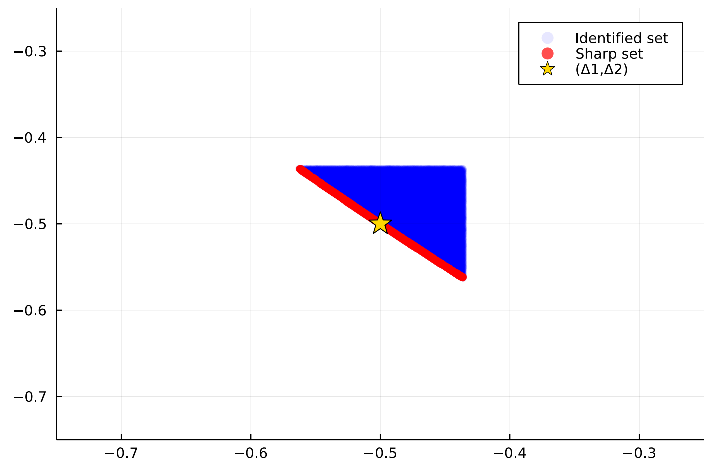
Lemma 3.1. *Under Assumptions 3.1 and 3.2, for each $(y, x) \in \mathcal{Y} \times \mathcal{X}$, $g_{y,x}(\theta)$ is convex in θ.*

From Lemma 3.1 and (3.7), the following result immediately follows.

Theorem 3.3. *Under Assumptions 3.1 and 3.2, Θ_I is convex.*

Example. (Example 3.1 continued) Let us continue to consider the example considered in Section 3.2.3, but now suppose that the researcher knows $\beta_i = 0$ for $i = 1, 2$. Thus, the researcher is interested in identifying the set of (Δ_1, Δ_2) given the observed choice probabilities. Figure 3.3 plots Θ_I in blue, Θ_I^{sharp} in red, and the location of the true parameter in star. Clearly, Θ_I is convex as implied by Theorem 3.3. Also note that, while Θ_I is much wider than Θ_I^{sharp}, they deliver identical projection intervals in each dimension. ■

Figure 3.3: Convex identified set



Remark 3.1. If the researcher is interested in non-parametric identification of $v_i(y, x)$, the restrictions

$$\log\left(\phi\left(y|x\right)\right) + \sum_{i=1}^{I}\left(\log\left(\sum_{y_i' \in \mathcal{Y}_i} \exp\left(v_i\left(y_i', y_{-i}, x\right)\right)\right) - v_i\left(y, x\right)\right) \leq 0, \quad \forall y, x$$

define a convex set of v_i's consistent with data under Assumption 3.1. ■

3.3 Computation

In this section, we discuss estimation algorithms that can be used in practice. Although Theorem 3.2 provides a closed-form characterization of the identified set, numerically approximating it may require a grid search over the parameter space. The number of points on the grid to numerically approximate the set with reasonable accuracy increases exponentially due to the curse of dimensionality. We show that we can avoid a naive grid search by formulating the problems as optimization problems that can be handled by state-of-the-art numerical solvers. Specifically, when Assumptions 3.1 and 3.2 are satisfied, finding the projections of Θ_I amounts to solving a set of convex programs.[12]

3.3.1 Projections of the Identified Set

A common way to summarize the information of a set-valued identified set is to report its projections at each dimension. Let $p \in \mathbb{R}^d$ be a (unit) vector that specifies the direction of a projection. The projections of Θ_I in direction p can be found by solving

$$\min_{\theta} p^T \theta \quad \text{subject to } \theta \in \Theta_I. \tag{3.8}$$

For example, if $p = (1, 0, ..., 0)$, then (3.8) identifies the lower bound of the projection interval corresponding to the first component of θ. On the other hand, using $p = (-1, 0, ..., 0)$ identifies the upper bound.

The following theorem states the projection problem (3.8) can be formulated as convex programs.

Theorem 3.4. *Suppose Assumptions 3.1 and 3.2 hold.*

1. *Θ_I is non-empty if and only if a solution to the following convex feasibility program exists.*

$$\min_{\theta} 1 \quad \textit{subject to } g_{y,x}(\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}. \tag{3.9}$$

2. *If Θ_I is non-empty, its projection in direction $p \in \mathbb{R}^d$ can be obtained by solving the following convex program.*

$$\min_{\theta} p^T \theta \quad \textit{subject to } g_{y,x}(\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}. \tag{3.10}$$

[12] This approach is akin to the mathematical program with equilibrium constraints (MPEC) approach à la Su and Judd (2012).

Theorem 3.4.1 says that whether Θ_I is non-empty can be checked by solving a convex feasibility program that involves minimizing a constant objective function subject to convex constraints. Theorem 3.4.2 says that, if Θ_I is non-empty, its projections can be obtained by solving convex programs. Thus, if θ is d-dimensional, and the researcher wants to report the projection intervals in each dimension, they can be obtained after solving $2d$ convex programs.

3.3.2 A Criterion Function Approach

In practice, Θ_I may be empty due to numerical errors or misspecification. However, the researcher may want to proceed and identify a point in the parameter space that minimizes the violations of the identifying inequalities, which can be taken as a pseudo-true parameter.[13] We propose an approach that can be used in these situations.

Let $\{w_x\}_{x\in\mathcal{X}}$ be a set of weights such that $w_x > 0$ for all $x \in \mathcal{X}$ and $\sum_{x\in\mathcal{X}} w_x = 1$. For example, we can choose w_x to be proportional to the number of observations at bin $x \in \mathcal{X}$. At each $\theta \in \Theta$, define the criterion function $Q : \Theta \to \mathbb{R}_+$ to be the value of the following program:

$$Q(\theta): \qquad \min_{\{t_{y,x}\}_{y,x}} \sum_{y,x} w_x t_{y,x} \tag{3.11}$$

$$g_{y,x}(\theta) \leq \log(1 + t_{y,x}), \quad \forall y, x \tag{3.12}$$

$$t_{y,x} \geq 0, \quad \forall y, x \tag{3.13}$$

Intuitively, $Q(\theta)$ measures the weighted average of the minimal violations of the identifying restrictions that are required to admit θ. The following lemma shows properties of the criterion function.

Lemma 3.2. *Let $Q(\theta)$ be the criterion function defined by (3.11)–(3.13).*

[13] Finding the point that minimizes the violations is a popular approach. See, e.g., Chernozhukov, Hong, and Tamer (2007) and Ciliberto and Tamer (2009). A recent paper by Andrews and Kwon (2019) discusses strategies for conducting inference with misspecified moment inequality models.

1. *For all $\theta \in \Theta$, the program (3.11)–(3.13) is convex and feasible.*

2. *$Q(\theta) \geq 0$ for all $\theta \in \Theta$.*

3. *$Q(\theta) = 0$ if and only if $\theta \in \Theta_I$.*

By definition, $\Theta_I \equiv \{\theta \in \Theta : \; Q(\theta) \leq 0\}$ and Θ_I is non-empty if and only if $\min_\theta Q(\theta) = 0$. In practice, however, we may have a situation where $\min_\theta Q(\theta) > 0$, i.e., there is no θ at which (3.5) holds *exactly*. This may be due to numerical errors, sampling error, or misspecification.

To make progress, we propose using a lower level set of the criterion function:

$$\Theta_I(c) := \{\theta \in \Theta : \; Q(\theta) \leq c\} \tag{3.14}$$

where $c \geq 0$ is a relaxation parameter. Clearly, $\Theta_I(0) \equiv \Theta_I$. The minimal relaxation parameter to make (3.14) non-empty is $c^* := \min_\theta Q(\theta)$.

The following theorem can be useful in practice. Let $h^w(t) := \sum_{y \in \mathcal{Y}, x \in \mathcal{X}} w_x t_{y,x}$.

Theorem 3.5. *Suppose Assumption 3.1 and 3.2 hold. Let $c^* \equiv \min_\theta Q(\theta)$.*

1. *$\Theta_I(c)$ is convex for all $c \geq 0$.*

2. *c^* can be obtained using the following convex program:*

$$\min_{\theta,t} h^w(t) \;\; \textit{subject to } (3.12),\, (3.13). \tag{3.15}$$

3. *For any $c \geq 0$, the projection of $\Theta_I(c)$ in the direction $p \in \mathbb{R}^d$ can be obtained using the following convex program*

$$\min_{\theta,t} p^T\theta \; \textit{subject to } (3.12),\, (3.13),\, \textit{and } h^w(t) \leq c. \tag{3.16}$$

Theorem 3.5 suggests the following algorithm:

- Step 1: Obtain the minimal relaxation $c^* \equiv \min_\theta Q(\theta)$ by solving (3.15).

- Step 2: Find the projections of $\Theta_I(c^*)$ by solving (3.16).

Following these steps requires solving $2d+1$ convex programs: one for obtaining c^*, and two for each $k = 1, ..., d$ to find the lower and upper bounds of the projection intervals.

Remark 3.2. Numerical optimization packages such as `JuMP` in Julia (Dunning et al. (2017)) use automatic differentiation techniques so that the user does not have to provide gradients. The above optimization problems can be plugged into the solver as it is without any need for problem reformulation (e.g., vectorization or stacking the coefficients to matrices) nor gradient/hessian information, making the implementation very simple and convenient. ■

3.3.3 Computation Time

We quantify the computational advantage of our approach by comparing it to a benchmark approach. We take the estimation algorithm of Ciliberto and Tamer (2009) (henceforth CT) as a benchmark for comparison. The main advantage of CT is that the algorithm is easily applicable to a large class of problems because it does not require explicit characterization of equilibria, and thus it is easy to implement. The CT algorithm has been successfully used in recent papers such as Ciliberto and Jäkel (2021) and Ciliberto, Murry, and Tamer (2021).[14]

The CT-identified set is defined as $\Theta_I^{CT} = \left\{\theta \in \Theta : \; Q^{CT}(\theta) \leq 0\right\}$ where $Q^{CT} : \Theta \to \mathbb{R}_+$ is a criterion function that quantifies the violation of inequalities

$$H_1(y|x;\theta) \leq \phi(y|x) \leq H_2(y|x;\theta), \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}.$$

Here, H_1 represents the probability that the model predicts a particular outcome as the unique equilibrium, and H_2 represents the probability mass of the region where there are

[14] Also see Henry, Méango, and Queyranne (2015) who propose a computationally fast approach to constructing confidence sets for partially identified incomplete structural models. The main advantage of their approach is that the algorithm provides a computationally tractable way of approximating the sharp identified set compared to previous approaches. However, as in other existing approaches for estimating complete information games, the approach requires (i) simulating latent variables, (ii) solving for all equilibria at each draw, and (iii) exploring the parameter space.

Table 3.2: Comparison of Computation Time

K	1	10	100	1000
Koh	0.0120	0.0331	0.2187	6.4299
CT	3.8e4	3.8e5	3.8e6	3.8e7

Notes: The entries represent the average CPU time to find projections for four parameters (100 simulations).

multiple equilibria. To evaluate $Q^{CT}(\theta)$ at each θ, we follow the algorithm described in the supplementary material of Ciliberto and Tamer (2009).

Our experiment extends Example 3.1 as follows. To capture increasing computational complexity with respect to the number of supports in the covariates, we draw common market shocks $\omega^k \sim U[0,1]$ for $k = 1, ..., K$. We then generate CCPs $\phi^k \in \mathbb{R}^4$ at each k assuming that the payoff functions are given by $u_i\left(y_i, y_j, \xi_i; \theta, \omega^k\right) = y_i\left(\beta_i + \Delta_i y_j + \omega^k + \xi_i\right)$ where ξ_i follows the standard logistic distribution. Finally, assuming that the values of ϕ^k and ω^k, $k = 1, ..., K$, are known to the researcher (take ω^k as observable covariates), we obtain the identified sets using each methodology.

Table 3.2 reports the average computational time for different values of K. The computational time for CT is estimated as follows.[15] Let τ^{CT} be the average time to evaluate $Q^{CT}(\theta)$ when $K = 1$; we use $R = 10,000$ draws of ξ's to obtain H_1 and H_2. Since the CT algorithm involves repeating the same routine at each $k = 1, ..., K$ and each point in the parameter space, we estimate the total computation time as $T^{CT} = \tau^{CT} \times K \times \#(\Theta)$ where $\#(\Theta)$ denotes the number of points on the grid that approximates the parameter space Θ . We assume that the researcher uses $\#(\Theta) = 100,000$.

Table 3.2 shows that our approach is dramatically faster than CT even in a very simple experiment. When the dimension of θ is larger, the computational advantage of our approach can be greater because grid search over the parameter space becomes exponentially harder.

[15] All simulations in the paper were performed in Julia (v1.6.1) on a MacBook Pro with Apple M1 Chip and 16GB RAM.

3.4 Inference

We propose a simple approach to inference that leverages the key insights from Horowitz and Lee (2021).[16] The main idea is to account for the sampling uncertainty by constructing a confidence set for the CCPs. We show that finding the projections of the confidence set also amounts to solving convex programs.

3.4.1 Theory

Let $\phi \equiv (\phi(y|x))_{y\in\mathcal{Y},x\in\mathcal{X}}$ be the population CCPs. Let $\Theta_I \equiv \Theta_I(\phi)$ to make the dependence of the identified set on ϕ explicit. Let Φ_n^α be the confidence set for ϕ with the property that the asymptotic probability of $\phi \in \Phi_n^\alpha$ is bounded below by $1-\alpha$.

Assumption 3.3. *Let $\alpha \in (0,1)$. A set Φ_n^α such that*

$$\liminf_{n\to\infty} Pr(\phi \in \Phi_n^\alpha) \geq 1-\alpha$$

is available. Moreover, $\phi \in \Phi_n^\alpha$ can be expressed as a collection of convex constraints.

We construct the confidence set as

$$\widehat{\Theta}_I^\alpha := \bigcup_{\tilde{\phi}\in\Phi_n^\alpha} \Theta_I\left(\tilde{\phi}\right). \tag{3.17}$$

The confidence set (3.17) has the property that it covers the population identified set $\Theta_I(\phi)$ with probability at least $1-\alpha$ asymptotically.

Theorem 3.6. *Let Φ_n^α satisfy Assumption 3.3. Then*

$$\liminf_{n\to\infty} Pr\left(\Theta_I \subseteq \widehat{\Theta}_I^\alpha\right) \geq 1-\alpha.$$

[16] A similar approach has been used in Koh (2022).

3.4.2 Implementation

We propose constructing Φ_n^α as simultaneous confidence intervals using Fitzpatrick and Scott (1987). Let $n^x \in \mathbb{Z}_+$ be the number of observations at covariate bin $x \in \mathcal{X}$ so that $n \equiv \sum_{x\in\mathcal{X}} n^x$.

Fitzpatric-Scott Simultaneous Confidence Sets

Let $\hat{\phi}(y|x)$ be the nonparametric estimates of the CCPs obtained via the frequency estimators. Let $\beta_\alpha = 1-(1-\alpha)^{1/|\mathcal{X}|}$. Let

$$\hat{L}(y|x) \equiv \max\left\{\hat{\phi}(y|x) - \frac{z(\beta_\alpha/4)}{2\sqrt{n^x}}, 0\right\}$$
$$\hat{U}(y|x) \equiv \min\left\{\hat{\phi}(y|x) + \frac{z(\beta_\alpha/4)}{2\sqrt{n^x}}, 1\right\}$$

where $z(\tau) \in \mathbb{R}$ denotes the upper $100(1-\tau)\%$ quantile of the standard normal distribution, and n^x denotes the number of observations at bin x. Let

$$\Phi_n^\alpha = \left\{\phi \in \Delta_{y|x} : \hat{L}(y|x) \leq \phi(y|x) \leq \hat{U}(y|x), \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}\right\} \tag{3.18}$$

where $\phi \in \Delta_{y|x}$ represents the condition that $\sum_{y\in\mathcal{Y}} \phi(y|x) = 1$ for all $x \in \mathcal{X}$, i.e., $\phi(\cdot|x)$ is an element of $(|\mathcal{Y}|-1)$-dimensional simplex at each $x \in \mathcal{X}$. It can be shown that the asymptotic probability of the event $\phi \in \Phi_n^\alpha$ is bounded below by $1-\alpha$. Thus, Φ_n^α in (3.18) satisfies Assumption 3.3.

Original Problem

Recall from (3.7) that $\theta \in \Theta_I$ if and only if

$$\log(\phi(y|x)) + q(y,x;\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \tag{3.19}$$

where $q(y,x;\theta) \equiv \sum_{i=1}^{I} \left\{ \log \left(\sum_{y_i' \in \mathcal{Y}_i} \exp \left(v_i \left(y_i', y_{-i}, x; \theta \right) \right) \right) - v_i \left(y, x; \theta \right) \right\}$. In the population problem, ϕ was assumed to known to the econometrician.

When the number of observation is finite, we account for the sampling uncertainty by constructing the confidence set Φ_n^α for ϕ. According to (3.17), $\theta \in \widehat{\Theta}_I^\alpha$ if and only if there exists $\phi \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{X}|}$ such that (3.19) and $\phi \in \Phi_n^\alpha$. Since Φ_n^α is given by (3.18), the condition $\phi \in \Phi_n^\alpha$ can be replaced with

$$\hat{L}(y|x) \leq \phi(y|x) \leq \hat{U}(y|x), \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \tag{3.20}$$

$$\sum_{y \in \mathcal{Y}} \phi(y|x) = 1, \quad \forall x \in \mathcal{X}. \tag{3.21}$$

Then, the projection of $\widehat{\Theta}_I^\alpha$ in direction p can be found by solving

$$\min_{\theta, \phi} p^T \theta \quad \text{subject to (3.19), (3.20), and (3.21).} \tag{3.22}$$

Transformation into a Convex Program

However, the program (3.22) is non-convex. To transform it into a convex form, change the variable of optimization to $\mu(y|x) \equiv \log(\phi(y|x))$ and replace the constraints with

$$\mu(y|x) + q(y,x;\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \tag{3.23}$$

$$\log \left(\hat{L}(y|x) \right) \leq \mu(y|x) \leq \log \left(\hat{U}(y|x) \right), \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \tag{3.24}$$

$$\sum_{y \in \mathcal{Y}} \exp(\mu(y|x)) = 1, \quad \forall x \in \mathcal{X}. \tag{3.25}$$

Finally, replace (3.25) with the linearized constraints

$$\sum_{y \in \mathcal{Y}} \left(\exp(\hat{\mu}(y|x)) \left(1 + \mu(y|x) - \hat{\mu}(y|x) \right) \right) = 1, \quad \forall x \in \mathcal{X} \tag{3.26}$$

Table 3.3: Projection Intervals of the Confidence Set

	True value	$n=\infty$	$n=200$	$n=500$	$n=2000$	$n=10000$
β_1	0.0	$[-0.22, 0.19]$	$[-0.43, 0.65]$	$[-0.42, 0.51]$	$[-0.37, 0.37]$	$[-0.30, 0.27]$
β_2	0.0	$[-0.22, 0.19]$	$[-0.43, 0.65]$	$[-0.42, 0.51]$	$[-0.37, 0.37]$	$[-0.30, 0.27]$
Δ_1	-0.5	$[-0.95, 0.00]$	$[-1.78, 0.00]$	$[-1.58, 0.00]$	$[-1.32, 0.00]$	$[-1.13, 0.00]$
Δ_2	-0.5	$[-0.95, 0.00]$	$[-1.78, 0.00]$	$[-1.58, 0.00]$	$[-1.32, 0.00]$	$[-1.13, 0.00]$

Notes: The table reports the average projection intervals of the confidence set for each sample size; each end point is the average of the end points over 1,000 simulations.

where $\hat{\mu}(y|x) \equiv \log(\phi(y|x))$. (Note that we have used the approximation $e^x \approx e^{\hat{x}} + e^{\hat{x}}(x - \hat{x})$.)

In sum, the projection of $\widehat{\Theta}_I^\alpha$ in direction p can be obtained by solving

$$\min_{\theta,\mu} p^T\theta \quad \text{subject to (3.23), (3.24), and (3.26).}$$

In addition, the criterion function approach described in Section 3.3.2 can be used in the same manner.

3.4.3 Monte Carlo Simulation

To examine the tightness of the proposed confidence set, we conduct a simulation study using Example 3.1. As before, we set $(\beta_1, \beta_2, \Delta_1, \Delta_2) = (0.0, 0.0, -0.5, -0.5)$, and simulate data assuming that the equilibrium selection rule is symmetric. Given each simulated data, we estimate the conditional choice probability using the nonparametric frequency estimator and find the projections of the confidence set by following the procedure described in Section 3.4.2.

Table 3.3 reports the average intervals over 1,000 simulations given each sample size. Clearly, the width of the interval is decreasing in the number of observations. Note that each estimated projection interval, obtained as the projection of the confidence set for the identified set, is conservative in the sense that they may cover the true projection interval with unnecessarily high probability (see, e.g., Kaido, Molinari, and Stoye (2019)).[17]

3.5 Empirical Applications

We illustrate the usefulness of our methodology using real-world datasets. The first example uses the model and the data in Kline and Tamer (2016), and the second example uses those in Ellickson and Misra (2011); these papers have made the data publicly available. In both cases, our methodology is very easy to implement, produces estimates that are (qualitatively) similar to the original papers, and takes *less than one CPU second* to obtain the projections of the confidence sets.

3.5.1 Empirical Application to Kline and Tamer (2016) Dataset

In the two-player entry game model of Kline and Tamer (2016), player 1 is LCC (low cost carriers) and player 2 is OA (other airlines). The payoff of player i in market m is

$$u_{im}\left(y_{im}, y_{jm}, x_m, \xi_{im}; \theta\right) = y_{im}\left(\beta_i^{cons} + \beta_i^{size} x_{m,size} + \beta_i^{pres} x_{im,pres} + \Delta_i y_{jm} + \xi_{im}\right).$$

The explanatory variables $x_{m,size} \in \mathbb{R}$, $x_{LCCm,pres} \in \mathbb{R}$, and $x_{OAm,pres} \in \mathbb{R}$ are binary.

We assume that each $\xi_{im} \in \mathbb{R}$ is independent and follows the standard logistic distribution.[18] The partially identified parameter is $\theta = (\beta_i^{cons}, \beta_i^{size}, \beta_i^{pres}, \Delta_i)_{i=1,2}$ is 8-dimensional.

Table 3.4 reports the projection intervals of the 99% confidence set.[19] The results are qualitatively similar to the estimates reported in Kline and Tamer (2016).[20] It took 0.102 CPU second to obtain the table.

[17] To obtain confidence intervals for projections of partially identified parameters that are not conservative, Kline and Tamer (2016)'s approach is also attractive in the sense that their approach can be easily applied together with our proposed strategy.

[18] Kline and Tamer (2016) assumes that ξ_{LCCm} and ξ_{OAm} are jointly normally distributed and can be correlated. We assume that the correlation is zero.

[19] As in Kline and Tamer (2016), we have allowed the criterion function to be slightly above zero to admit small numerical/misspecification error.

[20] See Figure 2 of Kline and Tamer (2016) for their posterior probability estimates.

Table 3.4: Projections of the Confidence Set ($\alpha = 0.01$)

Parameter	Projection Interval
β_{LCC}^{cons}	$[-4.38, -2.69]$
β_{OA}^{cons}	$[0.42, 0.77]$
β_{LCC}^{size}	$[0.25, 0.88]$
β_{OA}^{size}	$[0.32, 0.89]$
β_{LCC}^{pres}	$[2.61, 4.19]$
β_{OA}^{pres}	$[0.94, 1.33]$
Δ_{LCC}	$[-0.98, -0.08]$
Δ_{OA}	$[-1.04, -0.32]$

3.5.2 Empirical Application to Jia (2008) Dataset

Ellickson and Misra (2011) (henceforth EM) uses Jia (2008) dataset to illustrates different methodologies for estimating static games.[21] EM uses assumptions that render the traditional likelihood approaches applicable; this includes assuming that the players are symmetric (Bresnahan and Reiss (1991)) or assuming a particular form of equilibrium selection rule (Berry (1992)). On the other hand, our approach allows the players to be heterogeneous and is agnostic about the underlying equilibrium selection rule, so the traditional likelihood approach is not applicable.

The dataset includes observations on entry decisions of Walmart and Kmart in $n = 2,065$ local markets. To find the conditional choice probabilities, we discretized the continuous covariates by running k-means clustering algorithm. The set of continuous covariates includes *population*, *retail sales per capita*, *urban*, and *distance to Bentonville*, but not *South* and *Midwest* which are indicator variables. We group the continuous variables in to 20 clusters and find the within-cluster means of each variables. Crossed with the other binary variables, our discretization procedure yields a total of 48 covariate bins.

Table 3.5 compares the estimation results reported in EM and ours. The columns under

[21] Ellickson and Misra (2011) make simplifying assumptions to make the estimation problem easier. Specifically, while the original model in Jia (2008) assumes that the players are playing a single network game across a large number of markets, Ellickson and Misra (2011) assumes that the games were played independently across markets.

Table 3.5: Comparison of Estimation Results

	Ellickson and Misra (2011)				This paper
		Berry	Berry	Berry	Projection Intervals
Variable	BR	(Profit)	(Walmart)	(Kmart)	($\alpha = 0.01$)
Common effects					
Population	1.32	1.69	1.67	1.69	$[2.07, 2.42]$
Retail sales per capita	1.13	1.54	1.52	1.54	$[1.22, 1.55]$
Urban	1.03	1.20	1.19	1.20	$[2.31, 4.13]$
δ	0.65	0.39	0.40	0.38	$[0.12, 0.42]$
Walmart-specific effects					
Intercept (Walmart)	−14.03	−11.87	−11.76	−11.90	$[-12.45, -9.36]$
Distance to Bentonville, AK		−1.07	−1.06	−1.07	$[-1.58, -1.34]$
South		0.72	0.72	0.71	$[1.02, 1.40]$
Kmart-specific effects					
Intercept (Kmart)	−14.03	−19.76	−19.56	−19.57	$[-23.29, -20.15]$
Midwest		0.37	0.37	0.37	$[0.36, 0.75]$

Notes: The numbers under the column Ellickson and Misra (2011) are from Table 1 of their paper. The last column reports the projection intervals of the confidence set with $\alpha = 0.01$.

Ellickson and Misra (2011) are copied from Table 1 of their paper. They report the parameter estimates obtained from using approaches designed to estimate models under complete information assumption. The column denoted “BR” is obtained using the methodology of Bresnahan and Reiss (1991) which assumes that the players are symmetric. The column denoted “Berry” uses the methodology in Berry (1992) which relies on assumptions on the order of moves: the specification “Profit” assumes that a more profitable firm moves first; “Walmart” assumes that Walmart is the first mover; “Kmart” assumes that Kmart is the first mover. Finally, the last column reports the projection intervals of the confidence set obtained by applying our methodology. Clearly, our estimates are very tight and qualitatively similar to the numbers reported in EM, indicating that our approach yields reasonable results in practice.[22] It took 0.728 CPU second to obtain the projections of the confidence set reported in Table 3.5.

[22] As before, we allow for the criterion function to be slightly above zero to account for numerical/misspecification error.

3.6 Conclusion

This paper proposes a novel approach to estimating static discrete games of complete information with finite actions and finite players and no assumptions on the equilibrium selection rule. We propose an (non-sharp) identified set that is very easy to compute under standard assumptions. We also propose a simple approach to constructing confidence sets for the identified set. Numerical examples show that our method performs well in terms of speed and tightness.

Our identified set is non-sharp. At the minimum, however, our approach can be used to rule out parameters that are rejected by the model. Then, given the outer set, the researcher can use existing approaches for obtaining the sharp set (e.g., Beresteanu et al. (2011), Galichon and Henry (2011), and Henry, Méango, and Queyranne (2015)).

We conclude with some suggestions on future directions. First, it will be important to develop a computationally tractable approach while allowing the latent variables to be correlated. While we have worked with a restrictive assumption that the latent variables are independently distributed, researchers are often interested in cases where there might be correlation across alternatives or players. In these cases, correlation structure can be captured using Archimedian copula that permit closed-form expressions. Second, it will be interesting to explore how our approach can be extended to allow for rich unobserved heterogeneity.

References

AGUIRREGABIRIA, V., A. COLLARD-WEXLER, AND S. P. RYAN (2021): "Dynamic games in empirical industrial organization," in *Handbook of Industrial Organization*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier, vol. 4 of *Handbook of Industrial Organization, Volume 4*, 225–343.

AGUIRREGABIRIA, V. AND J. JEON (2020): "Firms' Beliefs and Learning: Models, Identification, and Empirical Evidence," *Review of Industrial Organization*, 56, 203–235.

AGUIRREGABIRIA, V. AND A. MAGESAN (2020): "Identification and Estimation of Dynamic Games When Players' Beliefs Are Not in Equilibrium," *The Review of Economic Studies*, 87, 582–625.

AGUIRREGABIRIA, V. AND P. MIRA (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

——— (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156, 38–67.

AIZAWA, N. AND H. FANG (2020): "Equilibrium Labor Market Search and Health Insurance Reform," *Journal of Political Economy*, 128, 4258–4336.

ALLCOTT, H., R. DIAMOND, J.-P. DUBÉ, J. HANDBURY, I. RAHKOVSKY, AND M. SCHNELL (2019): "Food Deserts and the Causes of Nutritional Inequality," *The Quarterly Journal of Economics*, 134, 1793–1844.

ANDREWS, D. W. K., S. T. BERRY, AND P. J. BARWICK (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," *SSRN Electronic Journal*.

Andrews, D. W. K. and S. Kwon (2019): "Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification," Tech. Rep. 2184, Cowles Foundation Discussion Paper.

Aradillas-López, A. (2020): "The Econometrics of Static Games," *Annual Review of Economics*, 12, 135–165.

Aradillas-Lopez, A. and E. Tamer (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business & Economic Statistics*, 26, 261–283.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010a): "Estimating Static Models of Strategic Interactions," *Journal of Business & Economic Statistics*, 28, 469–482.

Bajari, P., H. Hong, and S. P. Ryan (2010b): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78, 1529–1568.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models With Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

Bergemann, D. and S. Morris (2013): "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11, 487–522.

——— (2017): "Belief-free rationalizability and informational robustness," *Games and Economic Behavior*, 104, 744–759.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889–917.

Boyd, S. and L. Vandenberghe (2004): *Convex optimization*, Cambridge: Cambridge university press.

Bresnahan, T. F. and P. C. Reiss (1991): "Empirical models of discrete games," *Journal of Econometrics*, 48, 57–81.

Canay, I. A. and A. M. Shaikh (2017): "Practical and Theoretical Advances in Inference for Partially Identified Models," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge: Cambridge University Press, vol. 2 of *Econometric Society Monographs*, chap. 9, 271–306.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

Chiappori, P.-A. and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54, 832–861.

Ciliberto, F. and I. C. Jäkel (2021): "Superstar exporters: An empirical investigation of strategic interactions in Danish export markets," *Journal of International Economics*, 129, 103405.

Ciliberto, F., C. Murry, and E. Tamer (2021): "Market Structure and Competition in Airline Markets," *Journal of Political Economy*, 129, 2995–3038.

Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

de Paula, Á. (2013): "Econometric Analysis of Games with Multiple Equilibria," *Annual Review of Economics*, 5, 107–131.

Doraszelski, U. and J. F. Escobar (2010): "A theory of regular Markov perfect equilibria in dynamic stochastic games: Genericity, stability, and purification," *Theoretical Economics*, 5, 369–402.

Doval, L. and J. C. Ely (2020a): "Sequential Information Design," *Econometrica*, 88, 2575–2608.

——— (2020b): "Sequential Information Design," *Econometrica*, 88, 2575–2608.

Draganska, M., S. Misra, V. Aguirregabiria, P. Bajari, L. Einav, P. Ellickson, D. Horsky, S. Narayanan, Y. Orhun, P. Reiss, K. Seim, V. Singh, R. Thomadsen, and T. Zhu (2008): "Discrete choice models of firms' strategic decisions," *Marketing Letters*, 19, 399–416.

Dunning, I., J. Huchette, and M. Lubin (2017): "JuMP: A modeling language for mathematical optimization," *SIAM Review*, 59, 295–320.

Egesdal, M., Z. Lai, and C.-L. Su (2015a): "Estimating dynamic discrete-choice games of incomplete information," *Quantitative Economics*, 6, 567–597.

——— (2015b): "Estimating dynamic discrete-choice games of incomplete information," *Quantitative Economics*, 6, 567–597.

Einav, L. (2010): "Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game," *Economic Inquiry*, 48, 369–390.

Ellickson, P. B. and S. Misra (2011): "Structural workshop paper—Estimating discrete games," *Marketing Science*, 30, 997–1010.

Ericson, R. and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62, 53–82.

Fershtman, C. and A. Pakes (2012): "Dynamic games with asymmetric information: A framework for empirical work," *The Quarterly Journal of Economics*, 127, 1611–1661.

Fitzpatrick, S. and A. Scott (1987): "Quick Simultaneous Confidence Intervals for Multinomial Proportions," *Journal of the American Statistical Association*, 82, 875–878.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, vol. 1, The MIT Press.

Galichon, A. and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *The Review of Economic Studies*, 78, 1264–1298.

Gayle, P. G. and Z. Luo (2015): "Choosing between Order-of-Entry Assumptions in Empirical Entry Models: Evidence from Competition between Burger King and McDonald's Restaurant Outlets," *The Journal of Industrial Economics*, 63, 129–151.

Goll, D. (2000): "Seattle's 'other' coffee giant moves in," *The Business Journals*.

Green, J. R. and J.-J. Laffont (1987): "Posterior Implementability in a Two-Person Decision Problem," *Econometrica*, 55, 69–94.

Grieco, P. L. E. (2014): "Discrete games with flexible information structures: an application to local grocery markets," *The RAND Journal of Economics*, 45, 303–340.

Gualdani, C. and S. Sinha (2020): "Identification and inference in discrete choice models with imperfect information," *arXiv:1911.04529 [econ]*, arXiv: 1911.04529.

Henry, M., R. Méango, and M. Queyranne (2015): "Combinatorial approach to inference in partially identified incomplete structural models," *Quantitative Economics*, 6, 499–529.

Ho, K. and A. M. Rosen (2017): "Partial Identification in Applied Research: Benefits and Challenges," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge: Cambridge University Press, vol. 2 of *Econometric Society Monographs*, chap. 10, 307–359.

Horowitz, J. L. and S. Lee (2021): "Inference in a class of optimization problems: Confidence regions and finite sample bounds on errors in coverage probabilities," ArXiv: 1905.06491.

Igami, M. and N. Yang (2016): "Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada," *Quantitative Economics*, 7, 483–521.

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

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87, 1397–1432, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA14075.

Kalai, E. (2004): "Large Robust Games," *Econometrica*, 72, 1631–1665.

Kasahara, H. and K. Shimotsu (2012): "Sequential Estimation of Structural Models With a Fixed Point Constraint," *Econometrica*, 80, 2303–2319.

Kédagni, D., L. Li, and I. Mourifié (2021): "Discordant Relaxations of Misspecified Models," *arXiv:2012.11679 [econ]*, arXiv: 2012.11679.

Kennan, J. (2006): "A Note on Discrete Approximations of Continuous Distributions," Unpublished Manuscript. URL: https://www.ssc.wisc.edu/~jkennan/research/DiscreteApprox.pdf (Last accessed: October 2021).

Kennan, J. and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79, 211–251.

Kline, B. and E. Tamer (2016): "Bayesian inference in a class of partially identified models," *Quantitative Economics*, 7, 329–366.

Koh, P. (2022): "Stable Outcomes and Information in Games: An Empirical Framework," Working Paper.

Lee, S. Y. T. and A. Seshadri (2019): "On the Intergenerational Transmission of Economic Status," *Journal of Political Economy*, 127, 855–921.

Leung, S. (2003): "A Glutted Market Is Leaving Food Chains Hungry for Sites: Finding Spots for New Outlets Takes Heaps of Research, an Eye for Details," *The Wall Street Journal*.

Liu, Q. (2020): "Stability and Bayesian Consistency in Two-Sided Markets," *American Economic Review*, 110, 2625–2666.

Magnolfi, L. and C. Roncoroni (2021): "Estimation of Discrete Games with Weak Assumptions on Information," Working Paper.

Mailath, G. J. and L. Samuelson (2006): *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press, USA.

Makris, M. and L. Renou (2021): "Information Design in Multi-stage Games," *arXiv:2102.13482 [econ]*, arXiv: 2102.13482.

Minehart, D. and S. Scotchmer (1999): "Ex Post Regret and the Decentralized Sharing of Information," *Games and Economic Behavior*, 27, 114–131.

Minelli, E. and H. Polemarchakis (2003): "Information at equilibrium," *Economic Theory*, 21, 573–584.

Molinari, F. (2020): "Microeconometrics with partial identification," in *Hankbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, vol. 7A of *Handbook of Econometrics*, chap. 5, 355–486.

Olea, J. L. M. and M. Plagborg-Møller (2019): "Simultaneous confidence bands: Theory, implementation, and an application to SVARs," *Journal of Applied Econometrics*, 34, 1–17.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *The RAND Journal of Economics*, 38, 373–399.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334.

Pesendorfer, M. and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *The Review of Economic Studies*, 75, 901–928.

Radner, R. (1979): "Rational expectations equilibrium: Generic existence and the information revealed by prices," *Econometrica*, 655–678.

Ridley, D. B. (2008): "Herding versus Hotelling: Market Entry with Costly Information," *Journal of Economics & Management Strategy*, 17, 607–631.

Rust, J. (1994): "Estimation of dynamic structural models, problems and prospects: discrete decision processes," in *Advances in Econometrics: Sixth World Congress*, ed. by C. A. Sims, Cambridge: Cambridge University Press, vol. 2 of *Econometric Society Monographs*, 119–170.

Su, C.-L. and K. L. Judd (2012): "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 80, 2213–2230.

Syrgkanis, V., E. Tamer, and J. Ziani (2021): "Inference on Auctions with Weak Assumptions on Information," Working Paper.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

Thomadsen, R. (2007): "Product Positioning and Competition: The Role of Location in the Fast Food Industry," *Marketing Science*, 26, 792–804.

Toivanen, O. and M. Waterson (2005): "Market Structure and Entry: Where's the Beef?" *The RAND Journal of Economics*, 36, 680–699.

Train, K. (2009): *Discrete choice methods with simulation*, Cambridge: Cambridge University Press, 2nd ed.

Yang, N. (2012): "Burger King and McDonald's: Where's the Spillover?" *International Journal of the Economics of Business*, 19, 255–281.

——— (2020): "Learning in retail entry," *International Journal of Research in Marketing*, 37, 336–355.

Zhu, T. and V. Singh (2009): "Spatial competition with endogenous location choices: An application to discount retailing," *Quantitative Marketing and Economics*, 7, 1–35.

Zhu, T., V. Singh, and M. D. Manuszak (2009): "Market Structure and Competition in the Retail Discount Industry," *Journal of Marketing Research*, 46, 453–466.

Appendix for Chapter 1

A.1 Proofs

A.1.1 Proof of Theorem 1.1

Let S^* be an expansion of S. Let $\delta : \mathcal{T} \times \tilde{\mathcal{T}} \to \Delta(\mathcal{A})$ be an outcome function in (G, S^*). We say that an outcome function δ in (G, S^*) *induces* a decision rule $\sigma : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ in (G, S) if

$$\sigma(a|\varepsilon, t) = \sum_{\tilde{t}} \lambda\left(\tilde{t}|\varepsilon, t\right) \delta\left(a|t, \tilde{t}\right)$$

for each a whenever $\Pr(\varepsilon, t) > 0$.

Lemma A.1. *A decision rule σ is a Bayes stable equilibrium of (G, S) if and only if, for some expansion S^* of S, there is a rational expectations equilibrium of (G, S^*) that induces σ.*

The proof of Lemma A.1 closely follows the proof in Theorem 1 of Bergemann and Morris (2016). The only if ($\Rightarrow$) direction is established by (i) letting the Bayes stable equilibrium decision rule σ a signal function that generates public signals (recommendations of outcomes) for every given (ε, t), and (ii) constructing an outcome function δ as a degenerate self-map that places unit mass on a whenever a is drawn from $\sigma(\cdot|\varepsilon, t)$. Conversely, the if ($\Leftarrow$) direction is established by constructing a decision rule by integrating out the players' signals from a given outcome function.

Proof of Lemma A.1. ($\Rightarrow$) Suppose σ is a Bayes stable equilibrium of (G, S). That is,

$$\sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i(a, \varepsilon_i) \geq \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i(a_i', a_{-i}, \varepsilon_i), \quad \forall i, t_i, a, a_i'.$$

We want to find an expansion S^* of S and a rational expectations equilibrium outcome function δ in (G, S^*) that induces σ. Construct an expansion S^* of S as follows. With some abuse in notation, let λ be a signal distribution that generates a *public* signal such that

$$\lambda\left(\tilde{t}^p = a|\varepsilon, t\right) = \sigma\left(a|\varepsilon, t\right).$$

where $\tilde{t}^p$ denotes a public signal.[1] Let an outcome function be degenerate as follows:

$$\delta\left(\tilde{a}|t, \tilde{t}^p = a\right) = \begin{cases} 1 & \text{if } \tilde{a} = a \\ 0 & \text{if } \tilde{a} \neq a \end{cases}.$$

That is, when the players observe $\tilde{t}^p = a$ as a public signal, the outcome function dictates that a be played as an outcome of the game. It remains to show that every outcome a generated by the outcome function δ is optimal to the players. The rational expectations equilibrium condition is

$$\sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \lambda_{\tilde{t}^p|\varepsilon, t} \delta_{\tilde{a}|t, \tilde{t}^p} u_i\left(\tilde{a}, \varepsilon\right) \geq \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \lambda_{\tilde{t}^p|\varepsilon, t} \delta_{\tilde{a}|t, \tilde{t}^p} u_i\left(\tilde{a}'_i, \tilde{a}_{-i}, \varepsilon\right), \quad \forall i, t_i, \tilde{t}^p, \tilde{a}, \tilde{a}'_i$$

But since $\lambda\left(\tilde{t}^p = a|\varepsilon, t\right) = \sigma\left(a|\varepsilon, t\right)$ and the inequality is trivially satisfied when $\tilde{t}^p \neq \tilde{a}$ (both sides become zero), the rational expectations equilibrium condition reduces to

$$\sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon, t} u_i\left(a, \varepsilon\right) \geq \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon, t} u_i\left(\tilde{a}'_i, a_{-i}, \varepsilon\right), \quad \forall i, t_i, a, \tilde{a}'_i$$

which holds by the assumption that σ is a Bayes stable equilibrium of (G, S).

($\Leftarrow$) Suppose that δ is a rational expectations equilibrium of (G, S^*) and δ induces σ in

[1] More formally, the agents receive signals that are perfectly correlated, i.e., $\lambda\left(\tilde{t}_1 = a, ..., \tilde{t}_I = a|\varepsilon, t\right) = \sigma\left(a|\varepsilon, t\right)$.

(G, S). That is, we have

$$\sum_{\varepsilon,t_{-i},\tilde{t}_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \lambda_{\tilde{t}|\varepsilon,t} \delta_{a|t,\tilde{t}} u_i(a, \varepsilon) \geq \sum_{\varepsilon,t_{-i},\tilde{t}_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \lambda_{\tilde{t}|\varepsilon,t} \delta_{a|t,\tilde{t}} u_i(a_i', a_{-i}, \varepsilon), \quad \forall i, t_i, \tilde{t}_i, a, a_i'$$

Integrating out $\tilde{t}_i$ from both sides gives

$$\sum_{\varepsilon,t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \left(\sum_{\tilde{t}} \lambda_{\tilde{t}|\varepsilon,t} \delta_{a|t,\tilde{t}} \right) u_i(a, \varepsilon) \geq \sum_{\varepsilon,t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \left(\sum_{\tilde{t}} \lambda_{\tilde{t}|\varepsilon,t} \delta_{a|t,\tilde{t}} \right) u_i(a_i', a_{-i}, \varepsilon), \quad \forall i, t_i, a, a_i'$$

$$\Leftrightarrow \sum_{\varepsilon,t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i(a, \varepsilon) \geq \sum_{\varepsilon,t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i(a_i', a_{-i}, \varepsilon), \quad \forall i, t_i, a, a_i'$$

which is the Bayes stable equilibrium condition for σ in (G, S). □

The statement of the theorem then follows directly from Lemma A.1 because any decision rule $\sigma : \mathcal{E} \times \mathcal{T} \to \Delta(\mathcal{A})$ in (G, S) pins down the joint distribution on $\mathcal{E} \times \mathcal{T} \times \mathcal{A}$ (the prior distribution ψ on $\mathcal{E}$ is fixed by G and the signal distribution $\pi : \mathcal{E} \to \Delta(\mathcal{T})$ is fixed by S). □

A.1.2 Proof of Corollary 1.1

($\subseteq$) Take any $\phi \in \mathcal{P}_a^{BSE}(G, S)$. By definition, there is a BSE σ in (G, S) that induces ϕ. By Theorem 1.1, there exists an expansion S^* of S and a REE δ of (G, S^*) that induces σ. Since δ induces σ and ϕ induces ϕ, δ induces ϕ. It follows that $\phi \in \bigcup_{S^* \succsim_E S} \mathcal{P}_a^{REE}(G, S^*)$.

($\supseteq$) Take any $\phi \in \bigcup_{S^* \succsim_E S} \mathcal{P}_a^{REE}(G, S^*)$. By definition, there exists some $S^* \succsim_E S$ and a REE δ of (G, S^*) such that δ induces ϕ, (i.e., $\phi_a = \sum_{\varepsilon,t,\tilde{t}} \psi_\varepsilon \pi_{t|\varepsilon} \lambda_{\tilde{t}|\varepsilon,t} \delta_{a|t,\tilde{t}}$ for all $a \in \mathcal{A}$). Since $S^* \succsim_E S$ and δ is a REE of (G, S^*), by Theorem 1.1, δ induces a decision rule σ in (G, S) that is a BSE of (G, S). Since δ induces σ, it follows that σ induces ϕ. Therefore, we have $\phi \in \mathcal{P}_a^{BSE}(G, S)$. □

A.1.3 Proof of Theorem 1.2

1. ($\Rightarrow$) Since δ is a REE of $\left(G, S^{complete}\right)$, it satisfies

$$\psi_{\varepsilon}\delta_{a|\varepsilon}u_i\left(a,\varepsilon\right) \geq \psi_{\varepsilon}\delta_{a|\varepsilon}u_i\left(a_i',a_{-i},\varepsilon\right), \quad \forall i,\varepsilon,a,a_i'.$$

   Fix any $\varepsilon^* \in \mathcal{E}$ such that $\psi_{\varepsilon^*} > 0$ (with the full support assumption, we have $\psi_{\varepsilon} > 0$ for all ε). Consider any $a^* \in \mathcal{A}$ such that δ places a positive mass at ε^*, i.e., $\delta_{a^*|\varepsilon^*} > 0$. Since $\psi_{\varepsilon^*}\delta_{a^*|\varepsilon^*} > 0$, the REE condition reduces to

$$u_i\left(a^*,\varepsilon^*\right) \geq u_i\left(a_i',a_{-i}^*,\varepsilon^*\right), \quad \forall i,a_i'$$

   which is exactly the PSNE condition of a^* at state ε^*.

   ($\Leftarrow$) Suppose that $\delta : \mathcal{E} \to \Delta\left(\mathcal{A}\right)$ is constructed in a way such that $\delta_{a|\varepsilon} > 0$ implies that a is a PSNE outcome at ε. Since any on-path outcome a at ε is a PSNE at ε, it immediately follows that the outcome is optimal to each player who observes (a_i, a_{-i}) and ε, satisfying the REE condition. □

2. ($\Leftarrow$) Let $\delta : \mathcal{E} \to \Delta\left(\mathcal{A}\right)$ be a REE of $\left(G, S^{complete}\right)$. By definition, we have

$$\psi_{\varepsilon}\delta_{a|\varepsilon}u_i\left(a,\varepsilon_i\right) \geq \psi_{\varepsilon}\delta_{a|\varepsilon}u_i\left(a_i',a_{-i},\varepsilon_i\right), \quad \forall i,\varepsilon,a,a_i'$$

   Integrating both sides with respect to ε_{-i} gives

$$\sum_{\varepsilon_{-i}}\psi_{\varepsilon}\delta_{a|\varepsilon}u_i\left(a,\varepsilon_i\right) \geq \sum_{\varepsilon_{-i}}\psi_{\varepsilon}\delta_{a|\varepsilon}u_i\left(a_i',a_{-i},\varepsilon_i\right), \quad \forall i,\varepsilon_i,a,a_i'$$

   which is exactly the REE condition for $(G, S^{private})$.

   ($\Rightarrow$) Conversely, let $\delta : \mathcal{E} \to \Delta\left(\mathcal{A}\right)$ be a REE of $(G, S^{private})$. To show that δ is a REE of $\left(G, S^{complete}\right)$, by Theorem 1.2.1, it is enough to show that for each ε, $\delta_{a|\varepsilon} > 0$ implies that a is a PSNE of Γ_ε.

Since δ is a REE of $(G, S^{private})$, by definition, we have

$$\sum_{\varepsilon_{-i}} \psi_\varepsilon \delta_{a|\varepsilon} u_i(a, \varepsilon_i) \geq \sum_{\varepsilon_{-i}} \psi_\varepsilon \delta_{a|\varepsilon} u_i(a'_i, a_{-i}, \varepsilon_i), \quad \forall i, \varepsilon_i, a, a'_i$$
$$\Leftrightarrow \varphi(a, \varepsilon_i) u_i(a, \varepsilon_i) \geq \varphi(a, \varepsilon_i) u_i(a'_i, a_{-i}, \varepsilon_i), \quad \forall i, \varepsilon_i, a, a'_i$$

where $\varphi(a, \varepsilon_i) := \sum_{\varepsilon_{-i}} \psi_\varepsilon \delta_{a|\varepsilon}$.

Now fix ε and consider any a such that $\delta_{a|\varepsilon} > 0$. But $\delta_{a|\varepsilon} > 0$ implies $\varphi(a, \varepsilon_i) > 0$ which in turn implies that

$$u_i(a, \varepsilon_i) \geq u_i(a'_i, a_{-i}, \varepsilon_i), \quad \forall i, a'_i$$

which is exactly the PSNE condition of a at ε. $\square$

A.1.4 Proof of Theorem 1.4

Let $S \equiv (S^x)_{x\in\mathcal{X}}$ and $\tilde{S} \equiv \left(\tilde{S}^x\right)_{x\in\mathcal{X}}$. Let $\tilde{S} \succsim_E S$ if and only if $\tilde{S}^x \succsim_E S^x$ for each $x \in \mathcal{X}$. We want to show

$$\Theta_I^{BSE}(S) = \bigcup_{\tilde{S}\succsim_E S} \Theta_I^{REE}\left(\tilde{S}\right).$$

Note that

$$\Theta_I^{BSE}(S) \equiv \left\{\theta \in \Theta : \ \forall x \in \mathcal{X}, \ \phi^x \in \mathcal{P}_a^{BSE}\left(G^{x,\theta}, S^x\right)\right\} \tag{A.1}$$

and

$$\bigcup_{\tilde{S}\succsim_E S} \Theta_I^{REE}\left(\tilde{S}\right) \equiv \bigcup_{\tilde{S}\succsim_E S} \left\{\theta \in \Theta : \ \forall x \in \mathcal{X}, \ \phi^x \in \mathcal{P}_a^{REE}\left(G^{x,\theta}, \tilde{S}^x\right)\right\}$$
$$= \left\{\theta \in \Theta : \ \forall x \in \mathcal{X}, \ \phi^x \in \bigcup_{\tilde{S}^x\succsim_E S^x} \mathcal{P}_a^{REE}\left(G^{x,\theta}, \tilde{S}^x\right)\right\}. \tag{A.2}$$

By Corollary 1.1, for any given $\theta \in \Theta$ and $x \in \mathcal{X}$, we have

$$\mathcal{P}_a^{BSE}\left(G^{x,\theta}, S^x\right) = \bigcup_{\tilde{S}^x \succsim_E S^x} \mathcal{P}_a^{REE}\left(G^{x,\theta}, \tilde{S}^x\right). \tag{A.3}$$

That (A.1) and (A.2) are equal follows from (A.3), which is what we wanted. □

A.1.5 Proof of Theorem 1.5

1. Let G be an arbitrary basic game. We suppress the covariates x since they do not play a role. Let S^1 and S^2 be arbitrary information structures such that $S^1 \succsim_E S^2$. It is enough to show that a BSE in (G, S^1) always induces a BSE in (G, S^2) because it will imply that the set of feasible CCPs in (G, S^1) is a subset of the feasible CCPs in (G, S^2).

   Since S^1 is an expansion of S^2, we can express the signal function in S^1 as $\pi^1\left(t, \tilde{t}|\varepsilon\right) = \pi^2\left(t|\varepsilon\right)\lambda\left(\tilde{t}|\varepsilon, t\right)$ where $\tilde{t}$ denotes the extra signals available in S^1. We show that if $\sigma^1 : \mathcal{E} \times \mathcal{T} \times \tilde{\mathcal{T}} \to \Delta\left(\mathcal{A}\right)$ is a BSE in (G, S^1), then σ^1 induces a decision rule $\sigma^2 : \mathcal{E} \times \mathcal{T} \to \Delta\left(\mathcal{A}\right)$ in (G, S^2) that is a BSE of (G, S^2). Since σ^1 is a BSE of (G, S^1), we have

$$\sum_{\varepsilon, t_{-i}, \tilde{t}_{-i}} \psi_\varepsilon \pi^1_{t,\tilde{t}|\varepsilon} \sigma^1_{a|\varepsilon,t,\tilde{t}} u_i^\theta\left(a, \varepsilon_i\right) \geq \sum_{\varepsilon, t_{-i}, \tilde{t}_{-i}} \psi_\varepsilon \pi^1_{t,\tilde{t}|\varepsilon} \sigma^1_{a|\varepsilon,t,\tilde{t}} u_i^\theta\left(a_i', a_{-i}, \varepsilon_i\right), \quad \forall i, t_i, \tilde{t}_i, a, a_i'.$$

   Integrating out $\tilde{t}_i$, and defining σ^2 such that $\pi^2_{t|\varepsilon}\sigma^2_{a|\varepsilon,t} \equiv \sum_{\tilde{t}} \pi^1_{t,\tilde{t}|\varepsilon}\sigma^1_{a|\varepsilon,t,\tilde{t}} = \pi^2_{t|\varepsilon}\left(\sum_{\tilde{t}} \lambda_{\tilde{t}|\varepsilon,t}\sigma^1_{a|\varepsilon,t,\tilde{t}}\right)$ for each a, ε, t, we get

$$\sum_{\varepsilon, t_{-i}} \psi_\varepsilon \left(\sum_{\tilde{t}} \pi^1_{t,\tilde{t}|\varepsilon}\sigma^1_{a|\varepsilon,t,\tilde{t}}\right) u_i^\theta\left(a, \varepsilon_i\right) \geq \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \left(\sum_{\tilde{t}} \pi^1_{t,\tilde{t}|\varepsilon}\sigma^1_{a|\varepsilon,t,\tilde{t}}\right) u_i^\theta\left(a_i', a_{-i}, \varepsilon_i\right), \quad \forall i, t_i, a, a_i'.$$

$$\Leftrightarrow \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi^2_{t|\varepsilon}\sigma^2_{a|\varepsilon,t} u_i^\theta\left(a, \varepsilon_i\right) \geq \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi^2_{t|\varepsilon}\sigma^2_{a|\varepsilon,t} u_i^\theta\left(a_i', a_{-i}, \varepsilon_i\right), \quad \forall i, t_i, a, a_i'$$

which is the BSE condition for σ^2 in (G, S^2). It follows that any CCP that can be induced by a BSE in (G, S^1) can be induced by a BSE in (G, S^2), which is what we wanted to show. $\square$

2. The statement follows from Theorem 1.2. In particular, note that when pure strategy Nash equilibrium is the relevant solution concept, the decision rule (or the outcome function) simply represents an arbitrary equilibrium selection mechanism; no assumption is placed on the equilibrium selection rule. Since the set of probability distributions over $\mathcal{A}$ on each realization of ε is the same across Bayes stable equilibria and pure strategy Nash equilibria, the resulting identified set of parameters must be identical. $\square$

3. The statement follows from Theorem 1.3. Theorem 1.3 says that for any (G, S), if a decision rule σ in (G, S) is a Bayes stable equilibrium of (G, S), then it is a Bayes correlated equilibrium of (G, S). This implies that we will have $\mathcal{P}_a^{BSE}(G, S) \subseteq \mathcal{P}_a^{BCE}(G, S)$ for any (G, S) which leads to the statement. $\square$

A.1.6 Proof of Theorem 1.7

1. The first statement follows directly from construction:

$$\Pr\left(\Theta_I \subseteq \widehat{\Theta}_I^{\alpha}\right) = \Pr\left(\Theta_I(\phi) \subseteq \bigcup_{\tilde{\phi} \in \Phi_n^{\alpha}} \Theta_I\left(\tilde{\phi}\right)\right) \geq \Pr\left(\phi \in \Phi_n^{\alpha}\right)$$

(The inequality follows from the possibility that there may exist $\bar{\phi} \neq \phi$ such that $\bar{\phi} \in \Phi_n^{\alpha}$ but $\Theta_I(\phi) \subseteq \Theta_I\left(\bar{\phi}\right)$.) Taking the limits on both sides gives the desired result. $\square$

2. The second statement follows from the fact that ϕ enters the population program (see Theorem 1.6) in an additively separable manner, and that $\phi \in \Phi_n^{\alpha}$ represents a set of convex constraints. To see this, note that $\theta \in \widehat{\Theta}_I^{\alpha}$ if and only if the following program

is feasible: For each $x \in \mathcal{X}$, find $\sigma^x \in \Delta_{a|\varepsilon,t}$ and $\phi^x \in \Delta_a$ such that

$$\sum_{\varepsilon,t_{-i}} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x \partial u_i^{x,\theta}(a_i', a, \varepsilon_i) \le 0, \quad \forall i, t_i, a, a_i'$$

$$\phi_a^x = \sum_{\varepsilon,t} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x, \quad \forall a, x$$

$$\phi \in \Phi_n^\alpha.$$

That is, compared to the population program which treats ϕ as known, we let ϕ be a variable of optimization and add convex constraints $\phi \in \Phi_n^\alpha$. Under the assumption that $\phi \in \Phi_n^\alpha$ represents convex constraints, the above program is convex. $\square$

A.1.7 Proof of Theorem 1.8

1. First, let use show that (1.10) is always feasible for any θ. Pick any $\bar{\phi} \in \Phi_n^\alpha$. For any $\bar{\phi}$, we can find a $\bar{\sigma}$ satisfying $\bar{\phi}_a^x = \sum_{\varepsilon,t} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x$ for all a, x. Finally, there exists a non-negative vector of $\{q_x\}_{x \in \mathcal{X}}$ such that $\sum_{\varepsilon,t_{-i}} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x \partial u_i^{x,\theta}(\tilde{a}_i, a, \varepsilon_i) \le q_x$ for all $i, x, t_i, a, \tilde{a}_i$. Therefore, the feasible set of (q, σ, ϕ) is always non-empty. Second, convexity of program (1.10) follows from the fact that all the constraints are linear in (q, σ, ϕ) and that $\phi \in \Phi_n^\alpha$ represents a set of convex constraints. $\square$

2. It is straightforward to show that $\hat{Q}_n^\alpha(\theta) = 0$ if and only if $\theta \in \widehat{\Theta}_I^\alpha$. If $\hat{Q}_n^\alpha(\theta) = 0$, then it must be that $q_x^* = 0$ for all $x \in \mathcal{X}$, implying that $\theta \in \widehat{\Theta}_I^\alpha$. Conversely, if $\theta \in \widehat{\Theta}_I^\alpha$, then we can get $\hat{Q}_n^\alpha(\theta) = 0$ by plugging in $q_x = 0$ for all $x \in \mathcal{X}$. $\square$

3. Finally, we can obtain $\nabla \widehat{Q}_n^\alpha(\theta)$ as a byproduct to the convex program using the envelope theorem. $\square$

A.2 Computational Details

A.2.1 Discretization of Unobservables

Our approach to econometric analysis requires a discrete approximation to the distribution of payoff shocks which are often assumed to be continuous. We follow a discretization approach similar to that taken in Magnolfi and Roncoroni (2021), which requires finding a finite set of representative points on the support, and assigning appropriate probability mass on each point of the discretized support. The only difference is that Magnolfi and Roncoroni (2021) uses equally spaced quantiles of the distribution of ε_i's to find the discretized support whereas we use the approach introduced in Kennan (2006) to find the discretized support.

First, to discretize the space of each $\varepsilon_i \in \mathbb{R}$, we adopt the recommendations by Kennan (2006), which have been used in several works, e.g., Kennan and Walker (2011), Lee and Seshadri (2019), and Aizawa and Fang (2020). Let us briefly describe the procedure. Let F_0 be the true continuous distribution of a scalar random variable ε_i with support $\mathcal{E}_0$. Suppose we want to find an N-point discrete approximation to F_0. Specifically, we want to find a pair $(\mathcal{E}, F)$ where $\mathcal{E}$ contains N points and F describes the probability mass on each of the n points. How should we choose $\mathcal{E}$ and F?

Kennan (2006) characterizes the "best" discrete approximation $(\mathcal{E}, F)$ to $(\mathcal{E}_0, F_0)$, measured in L^p norm (for any $p > 0$) when the researcher can choose N points. We restate the proposition introduced in Kennan (2006).

Proposition (Kennan 2006)**.** *The best N-point approximation F to a given distribution F_0 has equally-weighted support points $\mathcal{E} \equiv \left\{x_j^*\right\}_{j=1}^N$ given by*

$$F\left(x_j^*\right) = \frac{2j-1}{2N}$$

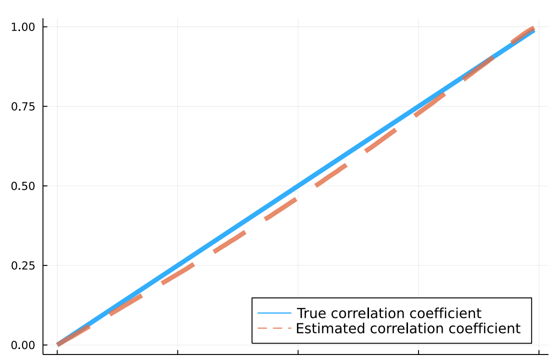
for $j = 1, ..., N$.

Following the proposition, we discretize unobservables as follows. In a two-player game with binary actions, we take the benchmark distribution of firm i's random shock ε_i to be the standard normal distribution. We fix the number of grid points N (we use $N = 10$ for empirical application) and find $\mathcal{E}_i \equiv \left\{x_j^*\right\}_{j=1}^N$ as described above. Then we take the Cartesian product of $\mathcal{E}_1$ and $\mathcal{E}_2$ to set the discrete support of $(\varepsilon_1, \varepsilon_2)$. In the baseline case where ε_1 is uncorrelated with ε_2, we construct the discretized prior distribution ψ as an $N \times N$ matrix whose entries are constant at $\frac{1}{N \times N}$. Thus, $\psi(\varepsilon_1, \varepsilon_2) = \frac{1}{N \times N}$ for any $(\varepsilon_1, \varepsilon_2) \in \mathcal{E} \equiv \mathcal{E}_1 \times \mathcal{E}_2$. For example, when each ε_i is approximated with $N = 20$ points, we have $20^2 = 400$ points in $\mathcal{E}$ with ψ assigning mass $1/400$ to each point in $\mathcal{E}$.

Second, to capture correlated unobservables, we apply weights to each point in $\mathcal{E}$ where the weights are generated using the density of the Gaussian copula. Specifically, we find the weight at each point $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathcal{E}$ to be proportional to the density of bivariate Gaussian copula evaluated at the point with correlation matrix $R = \begin{bmatrix} 1 & \rho \ ; \ \rho & 1 \end{bmatrix}$. In the special case $\rho = 0$, the approach applies uniform weights to each point on $\mathcal{E}$, and we return to the case where ψ has constant mass on every point on $\mathcal{E}$. Extension to the case with more than two players is straightforward.

In Figure A.1, we plot the true correlation coefficient against the estimated correlation coefficient obtained using the discretization approach with $N_E = 10$. The figure shows that discretized distribution has estimated correlation coefficient slightly smaller than the true (input) correlation coefficient ρ.

Figure A.1: True Correlation vs. Estimated Correlation ($N_E = 10$)



Note that whereas Kennan (2006) shows an "optimal" way of discretizing the support of a univariate random variable, we do not have such optimality result for a multivariate case. Thus, our approach should be understood as being heuristic.

A.2.1.1 Maximal Error from Discrete Approximation

Given that our approach relies on discrete approximations (as done in Syrgkanis, Tamer, and Ziani (2021) and Magnolfi and Roncoroni (2021)), a natural question is how accurate the approximation is. We provide a simple numerical evidence which supports the claim that the approximation error is at most mild.

Consider a two-player entry game with payoff $u_i(a_i, a_j, \varepsilon_i) = a_i(\kappa_i a_j + \varepsilon_i)$. We generate observed choice probability data at $(\kappa_1, \kappa_2) = (-0.5, -0.5)$ using a continuous distribution $\varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$, and symmetric equilibrium selection probability. The population choice probability is $(\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}) \approx (0.25, 0.3274, 0.3274, 0.0952)$.

If we use the discrete approximation procedure described above, how much error can there be? Our measure of discrepancy is the solution to

$$\min_{t\in\mathbb{R}, \sigma\in\Delta_{a|\varepsilon}} t \quad \text{subject to}$$

$$\sum_{\varepsilon_{-i}} \psi_\varepsilon \sigma_{a|\varepsilon} \partial u_i(\tilde{a}_i, a, \varepsilon_i) \leq t, \quad \forall i, \varepsilon_i, a, \tilde{a}_i$$

$$\sum_{\varepsilon} \psi_\varepsilon \sigma_{a|\varepsilon} - \phi_a \leq t, \quad \forall a$$

$$\phi_a - \sum_{\varepsilon} \psi_\varepsilon \sigma_{a|\varepsilon} \leq t, \quad \forall a$$

The solution t^* measures the maximal relaxation required for the equilibrium conditions and the consistency conditions. If $t^* = 0$, there is no approximation error. In general, we can expect $t^* > 0$. Let N_E be the number of grid points used for approximating $N(0,1)$. (We use $N_E = 10$ for ε_1 and ε_2 in our empirical application which produces $10^2 = 100$ points for the support of ψ.)

Figure A.2: Discrete approximation error

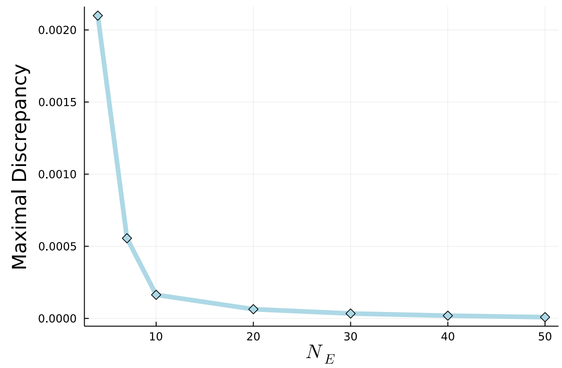


Figure A.2 plots t^* ("maximal discrepancy") against N_E. The figure shows that the discrepancy is decreasing in N_E and at most modest after $N_E = 10$. Since we construct confidence sets for the conditional choice probabilities when we do inference, it is likely that the approximation error will be controlled together. For this reason, it seems quite unlikely that discretization error will contaminate the estimation results.

A.2.2 Construction of Convex Confidence Sets for Conditional Choice Probabilities

In this section, we describe a simple approach to constructing confidence sets for the conditional choice probabilities, which we use for the empirical application. We construct simultaneous confidence intervals based on Fitzpatrick and Scott (1987). The basic idea is to construct confidence intervals for each multinomial proportion parameter so that the confidence set for the conditional choice probabilities can be characterized as a set of constraints that are *linear* in the population conditional choice probabilities. While there are many ways of constructing simultaneous confidence bands for a vector of means (e.g., see Olea and Plagborg-Møller (2019) and the references therein), we follow Fitzpatrick and Scott (1987) because it provides a very simple approach to constructing simultaneous confidence intervals for multinomial proportion parameters.[2]

Let $\mathcal{X}$ be a finite set of covariates and $|\mathcal{X}|$ its cardinality. Let $\phi_a^x \in \mathbb{R}$ be the population

[2] Specifically, Fitzpatrick and Scott (1987) shows a particular simultaneous confidence intervals for multinomial proportion parameters that are extremely easy to construct and characterizes the asymptotic coverage probabilities.

choice probability of outcome $a \in \mathcal{A}$ at bin $x \in \mathcal{X}$. At each bin x, the conditional choice probabilities $\phi^x \equiv (\phi_a^x)_{a\in\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$ represent the proportion parameters of a multinomial distribution. The entire vector of conditional choice probabilities is denoted $\phi \equiv (\phi^x)_{x\in\mathcal{X}} \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{X}|}$. Let $n^x \in \mathbb{Z}$ be the number of observations at each bin x, and let $n \equiv \sum_{x\in\mathcal{X}} n^x$ be the total number of observations in the data.

Our strategy is described as follows. Our objective is to construct a confidence set Φ_n^α that covers ϕ with probability at least $1-\alpha$ asymptotically where $\alpha \in (0,1)$. To do so, we will construct a confidence set $\Phi_{n^x}^{x,\beta_\alpha}$ at each bin x that covers ϕ^x with probability at least $1-\beta_\alpha$ asymptotically where $\beta_\alpha = 1-(1-\alpha)^{1/|\mathcal{X}|}$ (β_α arises from applying the Šidák correction for testing $|\mathcal{X}|$ number of independent hypotheses with family-wise error rate α; note that the samples in each bin x are independent from each other when the data are generated from independent markets). Next, we will construct Φ_n^α by taking intersections of $\Phi_{n^x}^{x,\beta_\alpha}$ across x; making the coverage probability for ϕ^x at each x be no less than $1-\beta_\alpha$ ensures that the overall coverage probability for ϕ is no less than $1-\alpha$. Moreover, if, for each x, $\Phi_{n^x}^{x,\beta_\alpha}$ can be represented by a set of constraints linear in ϕ^x, then Φ_n^α will be represented by a set of constraints linear in ϕ by construction.

At each $x \in \mathcal{X}$, we define the confidence set for ϕ^x as follows. Let $\hat{\phi}_a^x \equiv n_a^x/n^x \in \mathbb{R}$ be the nonparametric frequency estimator of ϕ_a^x where $n_a^x \in \mathbb{Z}$ is the number of observations with outcome a at bin x. Then construct $\Phi_{n^x}^{x,\beta_\alpha}$ as:

$$\Phi_{n^x}^{x,\beta_\alpha} \equiv \left\{\phi^x : \ \phi_a^x \in \hat{\phi}_a^x \pm \frac{z\left(\beta_\alpha/4\right)}{2\sqrt{n^x}}, \quad \forall a \in \mathcal{A}\right\}, \tag{A.4}$$

where $z(\tau) \in \mathbb{R}$ denotes the upper $100(1-\tau)\,\%$ quantile of the standard normal distribution.[3] Note that $\Phi_{n^x}^{x,\beta_\alpha}$ consists of $|\mathcal{A}|$ number of confidence intervals.

[3]Although the intervals may include values lower than 0 or higher than 1, we impose the condition that $\phi_a^x \in [0,1]$ for each a, x and $\sum_a \phi_a^x = 1$ for each x in the optimization problem.

Finally, we define a confidence region for ϕ as:

$$\Phi_n^\alpha \equiv \left\{\phi: \ \phi^x \in \Phi_{n^x}^{x,\beta_\alpha}, \quad \forall x \in \mathcal{X}\right\}. \tag{A.5}$$

The following proposition states that, under regular conditions, Φ_n^α constructed as (A.5) has the desired asymptotic coverage probabilities for the population conditional choice probabilities ϕ.

Proposition A.1. *Let Φ_n^α be defined as (A.5). Suppose that samples are independent across $x \in \mathcal{X}$, and $n^x \to \infty$ for each $x \in \mathcal{X}$ as $n \to \infty$. If α is sufficiently low or $|\mathcal{X}|$ is sufficiently large so that $\beta_\alpha \leq 0.032$, we have*

$$\lim_{n\to\infty} Pr(\phi \in \Phi_n^\alpha) \geq 1-\alpha.$$

To prove Proposition A.1, we use Theorem 1 of Fitzpatrick and Scott (1987) as a lemma. The lemma characterizes the asymptotic lower bounds on the coverage probabilities of $\Phi_{n^x}^{x,\beta_\alpha}$ for ϕ^x when the intervals of form (A.4) are used.

Lemma A.2 (Fitzpatrick and Scott (1987) Theorem 1)**.** *Let $\Phi_{n^x}^{x,\beta_\alpha}$ be defined as (A.4). Then*

$$\lim_{n^x\to\infty} Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) \geq \mathcal{L}\left(\beta_\alpha\right)$$

where

$$\mathcal{L}\left(\beta_\alpha\right) = \begin{cases} 1-\beta_\alpha, & \text{if } \beta_\alpha \leq 0.032 \\ 6\Phi\left(\frac{3z(\beta_\alpha/4)}{\sqrt{8}}\right) - 5, & \text{if } 0.032 \leq \beta_\alpha \leq 0.3 \end{cases}.$$

Now let us prove Proposition A.1. The proof uses the fact that (i) the samples are independent across $x \in \mathcal{X}$, (ii) $\Phi_{n^x}^{x,\beta_\alpha}$ covers ϕ^x with probability no less than β_α asymptotically, and (iii) β_α is chosen in a way that ensures the overall coverage probability for ϕ becomes no less than $1-\alpha$ asymptotically (Šidak correction).

Proof. We have

$$\Pr\left(\phi \in \Phi_n^\alpha\right) = \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}, \quad \forall x \in \mathcal{X}\right)$$
$$= \prod_{x\in\mathcal{X}} \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) \tag{A.6}$$

where (A.6) follows from the independence across $x \in \mathcal{X}$. Given that β_α is sufficiently small, taking the limit gives

$$\lim_{n\to\infty} \prod_{x\in\mathcal{X}} \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) = \prod_{x\in\mathcal{X}} \lim_{n^x\to\infty} \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) \tag{A.7}$$
$$\geq \prod_{x\in\mathcal{X}} (1-\beta_\alpha) \tag{A.8}$$
$$= (1-\beta_\alpha)^{|\mathcal{X}|}$$
$$= \left(1 - \left\{1-(1-\alpha)^{1/|\mathcal{X}|}\right\}\right)^{|\mathcal{X}|} \tag{A.9}$$
$$= 1-\alpha.$$

where (A.7) follows from the product rule of limits, (A.8) follows from Fitzpatrick and Scott (1987) Theorem 1, and (A.9) follows from the definition of β_α. □

The main advantage of using Fitzpatrick and Scott (1987) is its simplicity. The method is easily applicable even when there are zero count cells, i.e., $n_a^x = 0$ for some $a \in \mathcal{A}$ and $x \in \mathcal{X}$. Zero count cells often occur when the sample size is small and may require some correction if other popular approaches (e.g., normal approximation for each ϕ_a^x taken as a Bernoulli parameter) were used. The simultaneous confidence bands can be conservative, but retains a linear structure, which is computationally attractive.

Example A.1. Suppose there are two bins $\mathcal{X} = \{l, h\}$, and that the number of observations at each bin is $n^l = 400$ and $n^h = 600$. Suppose that $\mathcal{A} = \{00, 01, 10, 11\}$ so that $\phi^x = (\phi_{00}^x, \phi_{01}^x, \phi_{10}^x, \phi_{11}^x)$ and that we obtained $\hat{\phi}^l = (0.1, 0.1, 0.4, 0.4)$ and $\hat{\phi}^h = (0.2, 0.3, 0.3, 0.2)$ using nonparametric frequency estimators at each bin. If $\alpha = 0.05$, then $\beta_\alpha = 1-(1-\alpha)^{1/2} =$

0.0253. Then $z(\beta_\alpha/4) = z(1-0.0253/4) = 2.4931$. Finally, since $z(\beta_\alpha/4)/\left(2\sqrt{400}\right) = 0.0623$ and $z(\beta_\alpha/4)/\left(2\sqrt{600}\right) = 0.0509$, our Φ_n^α is defined by the following inequalities:

$$\hat{\phi}_a^l - 0.0623 \leq \phi_a^l \leq \hat{\phi}_a^l + 0.0623, \quad \forall a \in \mathcal{A}$$
$$\hat{\phi}_a^h - 0.0509 \leq \phi_a^h \leq \hat{\phi}_a^h + 0.0509, \quad \forall a \in \mathcal{A}.$$

■

A.2.2.1 Monte Carlo Experiment

We conduct Monte Carlo experiments to examine whether the simultaneous confidence bands have correct coverage probabilities and confirm that the approach works well. Let $\mathcal{X} = \{1, 2, ..., N_{\mathcal{X}}\}$ be a finite set indices (covariates). The following constitutes a single trial. We randomly generate a probability vector $\phi^x \in \mathbb{R}^4$ for $x = 1, ..., N_{\mathcal{X}}$ by taking a 4-dimensional uniform vector and normalize the vector so that it sums to one. Then, at each $x \in \mathcal{X}$, we generate a random sample by taking a draw from a multinomial distribution with parameter (n^x, ϕ^x) where n^x is the number trials. Finally, we test whether the simultaneous confidence bands, constructed as described above, covers ϕ^x. We repeat this procedure for $100,000$ times and find the coverage probability.

Table A.1: Coverage Probability of Simultaneous Confidence Bands from Simulation

	(A) $\alpha = 0.05$					(B) $\alpha = 0.01$				
$N_{\mathcal{X}} \backslash n^x$	100	200	500	1000	10000	100	200	500	1000	10000
4	0.9697	0.9707	0.9713	0.9744	0.9837	0.9950	0.9948	0.9957	0.9956	0.9975
10	0.9735	0.9731	0.9748	0.9754	0.9854	0.9955	0.9954	0.9957	0.9960	0.9978
50	0.9760	0.9760	0.9777	0.9797	0.9885	0.9958	0.9962	0.9962	0.9968	0.9981
100	0.9779	0.9788	0.9791	0.9811	0.9886	0.9959	0.9961	0.9964	0.9969	0.9982
200	0.9776	0.9783	0.9794	0.9816	0.9902	0.9964	0.9962	0.9966	0.9971	0.9984

Table A.1 reports the results of the Monte Carlo experiment. It shows that the confidence sets obtain desired coverage probabilities although they can be conservative. We conclude that the proposed approach works well.

A.2.3 Random Walk Surface Scanning Algorithm

Let Θ_I be the identified set of parameters. The identified set is defined as the level set

$$\Theta_I \equiv \{\theta \in \Theta : Q(\theta) \leq 0\}$$

where $Q(\theta)$ is a non-negative valued criterion function. (To obtain the confidence set, simply replace $Q(\theta)$ with $\widehat{Q}_n^{\alpha}(\theta)$.) Except for special cases (e.g., when Θ_I is convex), we need to approximate Θ_I by collecting a large number of points in Θ_I. A naive approach is to conduct an extensive grid search: draw a fine grid on the parameter space Θ (e.g., by taking quasi-Monte Carlo draws) and evaluate the criterion function at all point on the grid. However, a naive grid search can be computationally burdensome especially when the dimension of θ is large.

In our setup, Theorem 1.8 says that we can get the gradient information for free due to the envelope theorem. That is, once we evaluate $Q(\theta)$ at any θ, we can get $\nabla Q(\theta)$ as well. Exploiting the gradient information allows us to find a minimizer of $Q(\theta)$ far more efficiently because we can use gradient-based optimization algorithms (e.g., gradient descent or (L-)BFGS) as opposed to gradient-free algorithms. However, since we need to find *all* minimizers of $Q(\theta)$, solving $\min_\theta Q(\theta)$ is insufficient.

We propose a heuristic approach. First, we identify $\theta^0 = \arg\min_\theta Q(\theta)$ by using a gradient-based optimization algorithm. Second, we iteratively explore the neighbors of the identified set by running a random walk process from θ^0 and accepting points at which the criterion function is zero-valued. Being able to quickly identify a point in the identified set gives a considerable advantage over grid search algorithms because we do not have to explore points that are "far" from the identified set. The required assumption is that Θ_I is a connected set.

We use a random walk surface scanning algorithm described as follows. Let $\theta^0 = \arg\min_\theta Q(\theta)$ be the identified parameter and assume that $Q(\theta^0) = 0$ (otherwise the identified set is empty). From θ^0, we take a random candidate

$$\tilde{\theta}^1 \leftarrow \theta^0 + \eta$$

where $\eta \sim N\left(0, \sigma_\eta^2\right)$ is a vector of random shocks. We then evaluate $Q\left(\tilde{\theta}^1\right)$ and check whether the value is equal to zero. If $Q\left(\tilde{\theta}^1\right) = 0$, we accept the candidate $\tilde{\theta}^1$ and let $\theta^1 \leftarrow \tilde{\theta}^1$. If $Q\left(\tilde{\theta}^1\right) > 0$, then we draw a new $\tilde{\theta}^1$ until we find a point that is accepted. Iterating this process generates a random sequence of points $\theta^0, \theta^1, \theta^2, ...$ that "bounces" inside the level set Θ_I. We iterate this process until we find a large number of points in Θ_I.

To control the step size, we let σ_η adjust adaptively. Specifically, if a candidate point is accepted, we increase σ_η before a new draw is taken to make the search more aggressive. If a candidate point is rejected, we decrease σ_η to make the search more conservative (a lower bound can be placed to prevent excessively small step size).

A.2.4 Counterfactual Analysis

In this section, we explain the implementation details for counterfactual analysis. Let us first lay out the counterfactual prediction problem. Let us call the game before and after the counterfactual policy pre-game and post-game respectively. Suppose we have a counterfactual policy that changes the pre-game (G^{pre}, S) to post-game (G^{post}, S) (we assume that a counterfactual policy only changes the payoff-relevant primitives, but not the information structure). In our application, we assume the counterfactual policy changes the covariates from x^{pre} to x^{post} so that the payoff function changes from $u_i^{pre}(a, \varepsilon_i; \theta) \equiv u_i^{x^{pre},\theta}(a, \varepsilon_i)$ to $u_i^{post}(a, \varepsilon_i; \theta) \equiv u_i^{x^{post},\theta}(a, \varepsilon_i)$. We assume that the prior distribution ψ and the baseline information structure S do not change.

Let $h : \mathcal{A} \times \mathcal{E} \rightarrow \mathbb{R}$ be the counterfactual objective of interest, which is a function of realized state of the world and action profiles (see examples provided below). At a fixed $x \in \mathcal{X}$, we can obtain the lower/upper bounds on the expected value of h by finding the equilibria that will minimize/maximize the expected value of h:

$$\min / \max_{\sigma^x \in \Delta_{a|\varepsilon,t}} \sum_{\varepsilon,t,a} \psi^x_\varepsilon \pi^x_{t|\varepsilon} \sigma^x_{a|\varepsilon,t} h\left(a, \varepsilon\right) \quad \text{subject to} \tag{A.10}$$

$$\sum_{\varepsilon,t_{-i}} \psi^x_\varepsilon \pi^x_{t|\varepsilon} \sigma^x_{a|\varepsilon,t} \partial u^{x,\theta}_i \left(\tilde{a}_i, a, \varepsilon\right) \leq 0, \quad \forall i, t_i, a, \tilde{a}_i.$$

Note that (A.10) is a linear program.

We now connect the characterizations to the empirical application. Let $\mathcal{X}^{pre}$ be the set of covariates corresponding to the food deserts in Mississippi; there can be multiple values of $x^{pre} \in \mathcal{X}^{pre}$ because there are multiple markets with different observable covariates. By the definition of food deserts, all Mississippi food deserts have covariates with the low access to healthy food indicator equal to 1. For each market m, we define the counterfactual covariates as the vector obtained by changing the low access indicator from 1 to 0. For example, if $x^{pre} = \left(x^{highipc}, x^{lowaccess}\right) = (1, 1)$ for a particular market, we set $x^{post} = (1, 0)$. This changes the game since the players' payoff functions are changed. Then the set of covariates for the post-regime $\mathcal{X}^{post}$ is constructed by taking each $x^{pre} \in \mathcal{X}^{pre}$ and changing the low access indicator from 1 to 0.

We use four measures of market structure:

Counterfactual objective	$h\left(a, \varepsilon\right)$
Number of entrants	$1 \times \left(\mathbb{I}\left\{a = (0,1)\right\} + \mathbb{I}\left\{a = (1,0)\right\}\right) + 2 \times \mathbb{I}\left\{a = (1,1)\right\}$
McDonald's entry	$\mathbb{I}\left\{a = (1,0)\right\} + \mathbb{I}\left\{a = (1,1)\right\}$
Burger King entry	$\mathbb{I}\left\{a = (0,1)\right\} + \mathbb{I}\left\{a = (1,1)\right\}$
No entry	$\mathbb{I}\left\{a = (0,0)\right\}$

Suppose θ is given. At a fixed covariate x, we can obtain bounds on the expected value of h by solving (A.10). However, since $\mathcal{X}^{pre}$ is non-singleton, we find the weighted average of the bounds. Let $\{w^x\}_{x \in \mathcal{X}^{pre}}$ be the weights at each covariate vector where w^x is proportional to the number of markets corresponding to Mississippi food deserts in covariate bin $x \in \mathcal{X}^{pre}$; we scale the weights so that $\sum_{x \in \mathcal{X}^{pre}} w^x = 1$. The weighted average on h can be found by solving:

$$\min / \max_{\sigma} \sum_{x \in \mathcal{X}^{post}} w^x \sum_{\varepsilon,t,a} \psi_\varepsilon^x \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x h(a,x) \quad \text{subject to}$$

$$\sum_{\varepsilon,t_{-i}} \psi_\varepsilon^x \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x \partial u_i^{x,\theta}(\tilde{a}_i, a, \varepsilon) \leq 0, \quad \forall x \in \mathcal{X}^{post}, i, t_i, a, \tilde{a}_i$$

$$\sigma^x \in \Delta_{a|\varepsilon,t}, \quad \forall x \in \mathcal{X}^{post}$$

The bounds for the pre-counterfactual regime can be found by replacing $\mathcal{X}^{post}$ with $\mathcal{X}^{pre}$.

Finally, since Θ_I is set-valued, we repeat the above process for each θ in Θ_I and take the union of the bounds. Since there is a large number of points in Θ_I, to save computation time, we use k-means clustering on Θ_I to find a set of points that approximate Θ_I (we choose k equal to 2000 or larger and compare the projection of the original set to the projection of the approximating set to see if the approximation is accurate).

A.2.5 Overview of the Implementation

We provide a brief overview of how we obtain the confidence sets in the empirical application section. To prepare data for structural estimation, we use `Stata` to obtain discretized bins of covariates. We use estimate the conditional choice probabilities via nonparametric frequency estimator. We also compute the number of observations in each bin $x \in \mathcal{X}$ (which are inputs to constructing simultaneous confidence intervals for the CCPs) and define weights at each x (which are inputs to criterion function) as being proportional to the number of observations. The final dataset has $|\mathcal{X}|$ rows, where each row contains vector of covariate values corresponding to bin x, CCP estimates $\hat{\phi}_a^x$ for each outcome $a \in \mathcal{A}$, and the number of observations at the covariate bin. We then export the data to `Julia` where all computations for structural estimation are done.

To prepare feasible optimization programs, we discretize the space of shocks using the approach described in Section A.2.1. We then declare optimization program using `JuMP` interface (Dunning et al., 2017).[4] We construct the simultaneous confidence sets for the conditional choice probabilities using the approach described in A.2.2. This makes evaluation of the criterion functions $\widehat{Q}_n^\alpha(\theta)$ for each point $\theta \in \Theta$ a linear program. We use `Gurobi` to solve linear programs.

To approximate the confidence set $\widehat{\Theta}_I^\alpha$, we need to collect many points in Θ that satisfy the condition $\hat{Q}_n^\alpha(\theta) = 0$. Collecting these points are done by the random walk surface scanning algorithm described in Section A.2.3. To use this approach, it is important to quickly identify an initial point θ^0 such that $\hat{Q}_n^\alpha(\theta^0) = 0$ by solving $\min_\theta \hat{Q}_n^\alpha(\theta)$. This can be done efficiently by using gradients of $\hat{Q}_n^\alpha(\theta)$ obtained by the envelope theorem (see Theorem 1.8). We recommend using many initial points to increase the chance of convergence, and decreasing the tolerance for optimality conditions ($\|\nabla\hat{Q}_n^\alpha(\theta)\| < \varepsilon^{tol}$) for higher accuracy. We use `Knitro` to solve nonlinear programs.

Specifically, we identify $\arg\min_\theta \hat{Q}_n^\alpha(\theta)$ by solving the minimization problem in two steps:

$$\min_\theta \hat{Q}_n^\alpha(\theta) = \min_{\theta^\rho}\left(\min_{\theta^u} \hat{Q}_n^\alpha(\theta^u;\theta^\rho)\right)$$

where θ^u represent the parameters associated with the payoff functions and θ^ρ represent the correlation coefficient parameter for the distribution of payoff shocks. In the outer loop, we search for the minimum over a grid of θ^ρ on $[0,1]$. In the inner loop, taking θ^ρ as fixed, we solve $\min_{\theta^u} \hat{Q}_n^\alpha(\theta^u;\theta^\rho)$ by minimizing (1.10) *jointly* with θ^u. Solving the nested optimization program (the outer loop minimizes over θ^u and the inner loop minimizes over q, σ, ϕ) as a single optimization program is faster when the number of variables is manageable; this is similar to the key idea of Su and Judd (2012). Although we can obtain ψ^{x,θ^ρ} in closed form so that the minimization problem can be solved jointly in (θ^u, θ^ρ), we chose to divide the minimization problem as above because ψ^{x,θ^ρ} can be highly non-linear in θ^ρ.

[4] The main advantages of `JuMP` are its ease of use and its automatic differentiation feature which does not require the researcher to provide first- and second-order derivatives.

A.3 Data Appendix

This section describes the datasets used for our empirical application, which studies the entry game between McDonald's and Burger King in the US. The following table provides an overview of the datasets used in this paper.

Dataset Name	**Description**
Data Axle (Infogroup) Historical Business Database	Proprietary; accessed via Wharton Research Data Services https://wrds-www.wharton.upenn.edu/ using institutional subscription.[5] Data Axle (formerly known as Infogroup) is a data analytics marketing firm that provides digital and traditional marketing data on millions of consumers and businesses. Address-level records on business entities operating in the US are available for 1997-2019 at the annual level. We obtain the addresses of burger outlets in operation, which in turn are translated into tract-level entry decisions for each calendar year using the census shapefiles.
US Census Shapefiles	Accessible from https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html. Used to get 2010 census tract boundaries. Shapefiles are needed to find tract IDs corresponding to each physical store given their location coordinates.
Longitudinal Tract Data Base (LTDB)	Accessible from https://s4.ad.brown.edu/projects/diversity/researcher/bridging.htm. LTDB provides tract-level demographic information (from the census) for 1970-2010 harmonized to 2010 tract boundaries. We obtain population and income per capita for year 2000 and 2010 from here.
National Neighborhood Data Archive (NaNDA)	Accessible from https://www.openicpsr.org/openicpsr/nanda. NaNDA provides measures of business activities at each tract. We obtain the number of eating and drinking places for year 2010 at the tract level. Other variable such as the number of grocery stores (per square miles), the number of super-centers, and the number of retail stores are available.

[5] Wharton Research Data Services (WRDS) was used in preparing part of the data set used in the research reported in this paper. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

Food Access Research Atlas	Accessible from https://www.ers.usda.gov/data-products/food-access-research-atlas/. Food Access Research Atlas provides information on whether a census tract has limited access to supermarkets, super-centers, grocery stores, or other sources of healthy and affordable food. We obtain indicators for "low access to healthy food" and "food deserts" at the tract level for year 2010. A census tract is classified as a food desert if it is identified as having low access to healthy food and low income. A census tract is classified as low-access tract if at least 500 people or at least 33 percent of the population is greater than 1/2 mile from the nearest supermarket, supercenter, or large grocery store for an urban area or greater than 10 miles for a rural area.[6] The criteria for identifying a census tract as low-income are from the Department of Treasury's New Markets Tax Credit (NMTC) program.

A.3.1 Data Construction

We merge multiple datasets to construct the final sample used for structural estimation. The details are described as follows.

Panel data at tract-year level

Although we use 2010 cross-section for estimation of the structural model, we construct a panel dataset at a tract-year level to track the openings and closings of fast-food outlets in the US. We make the sample period run from 1997 to 2019, corresponding to the period for which business location data from Data Axle Historical Business Database are available.

We define the units for *markets* as 2010 census tracts designated by the US Census Bureau. (We define potential markets as 2010 urban tracts. See below for the definition of urban tracts.) Year 2010 was selected since it was the latest year for which the decennial census data was available when we started the empirical analysis. For all years in the sample period, we fix markets as 2010 census tracts; although census tract boundaries change slightly every decade, we fixed the boundaries for consistency across time.

[6] An alternative measure uses 1 mile radius for urban area. Using the 1 mile radius measure does not change the qualitative conclusion of our empirical analysis.

To construct tract-level data, we first download the 2010 census shapefiles from the US Census to obtain the list of all 2010 census tracts (there are 74,134 tracts defined for the 2010 decennial census in the US and its territories). Next, we exclude all tracts outside the contiguous US: Alaska, Hawaii, American Samoa, Guam, Northern Mariana Islands, Puerto Rico, and the Virgin Islands. We drop these regions since the data generating process (specifically how the game depends on observable market characteristics) is likely to differ from the rest.

Using the market-year panel data as a "blank sheet", we append relevant variables that include the firms' entry decisions in each tract for a given year and observable tract characteristics such as population. At this stage, we can create a variable distance to headquarter by measuring the distance between the location of a firm's headquarter and the centroid of a tract (McDonald's and Burger King have their headquarters in Chicago and Florida respectively).

In the final dataset used for the empirical application, we restrict attention to 2010 urban census tracts (i.e., we drop all rural tracts). A census tract is defined as urban if its population-weighted centroid is in an "urban area" as defined in the Census Bureau's *urbanized area* definition; a census tract is rural if not urban. We obtain the urban tract indicator from the Food Access Research Atlas.

Coding Entry Decisions

The primary source of data for our empirical application is Data Axle's *Historical Business Data*base. The dataset contains the list of local business establishments operating in the US over 1997-2019 at an annual level. Each establishment is assigned a unique identification number which can be used to construct establishment-level panel data. In addition, the dataset contains information such as company name, parent company, location of the establishment in coordinates, number of employees, industry codes.

We first need to download the entire list of burger outlets that were in operation. We download raw data from Wharton Research Data Services (WRDS) using the qualifier "SIC code=58" (retail eating places). We then identify relevant burger chains using company (brand) names and their parent number. In principle, each burger chain should have a unique parent number by the data provider. For example, all McDonald's outlets have parent number "001682400". Ideally, one can identify all burger chains that belong to a brand using their names and parent numbers. However, there are some errors due to misclassifications, which makes identifying all relevant burger chains more difficult. For example, McDonald's outlets will have different company names such as "MC DONALD'S", "MCDONALDS", and "MC DONALD". In addition, some McDonald's outlets have parent numbers missing for some subset of years, or some establishments have duplicate observations.[7]

To overcome this issue, we rely on the coordinates information to identify unique establishments. Since the same establishment can have different coordinates assigned over time depending on which point of place is used to measure the coordinates, we put each establishment in blocks approximately 250 meters in height and width. The idea is to put all observations whose coordinates are very close to each other in a single bin. Then we assign a unique establishment id to the stores in each block, i.e., we treat them as corresponding to a single store. We find that while it is challenging to avoid minor classification errors, the total number of burger chains outlets identified by our procedure closely is very close to the total number of outlets reported by other sources (e.g., reports in Statista `https://www.statista.com/`). Identifying unique establishments allows the construction of establishment-level panel data, which can be used to track firm entries and exits in each market.

The final step is to reshape the establishment-level panel data to market-level data to tabulate the number of burger chains operating in each market-year pair. We accomplish this with the help of `Stata`'s geocoding function, which helps identify census tract id's cor-

[7]The main hurdle in constructing establishment-level panel data is the following. Each establishment is assigned a unique "ABI number" which allows the analyst to track how the establishment operates over time. However, we found that some establishments had their ABIs changing over time or one establishment had duplicate observations with different ABI numbers assigned. When we inquired the original data provider support team about why this issue might be arising, they responded that it seems to be errors generated in the data recording stage.

responding to each coordinate (location of establishments). We then tabulate the number of outlets by each brand at a year-tract level.

In each market, we code entry decisions as binary variables. There were very few cases of a firm having more than one outlet in a single tract (approximately 1.5% of markets for McDonald's and 0.3% for Burger King). We also construct a firm-specific variable *own outlets in nearby markets*. This variable records the number of own-brand outlets operating in adjacent markets (neighboring markets that share the same borders). For example, if for market m, McDonald's nearby outlets are 2, it means that there were a total of 2 outlets operating in markets adjacent to market m. We constructed this variable with the help of a dataset downloaded from *Diversity and Disparities project* website that provides the list of 2010 census tracts and adjacent tracts.[8]

Market Characteristics

We obtain tract-level characteristics from multiple sources described in the table above. All of these datasets provide variables at tract-level for the year 2010. We append tract-level characteristics to the main dataset that has entry decisions and firm-specific variables at tract-level.

A.4 Information and Stability in a Two-player Entry Game Example

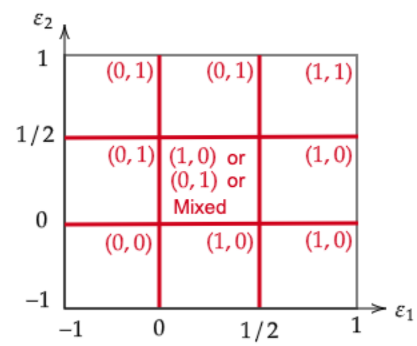
In this section, we compare Bayes stable equilibrium and static Nash equilibrium using a two-player entry game similar to Example 1.1. While the static Nash equilibrium framework has been a dominant approach for estimating games with cross-sectional data, we claim that it may not be applicable when the researcher is interested in analyzing stable outcomes (i.e., the researcher observes firms' decisions that are assumed to be stable). Static Nash equilibrium assumes that players' decisions are irreversible, i.e., players cannot revise their actions after observing opponents' actions. Thus, if the empirical setting allows players to revise their

[8]Accessible from https://s4.ad.brown.edu/Projects/Diversity/Researcher/Pooling.htm.

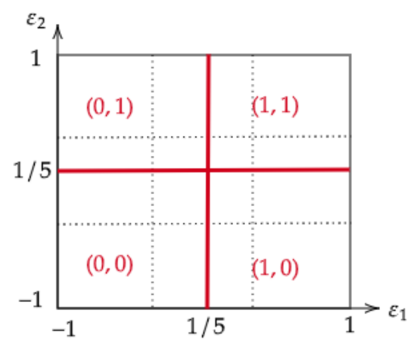
actions frictionlessly, the researcher may want to consider alternative solution concepts as identifying restrictions. We illuminate this point in the examples below.

A.4.1 Instability of Nash Equilibrium Outcomes

Figure A.3: Nash equilibrium in entry game



(a) Complete information ($S^{complete}$)



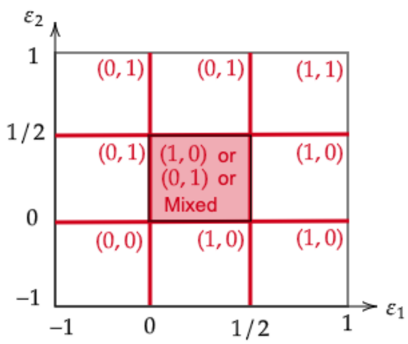
(b) Private information ($S^{private}$)

Consider a two-player entry game with payoffs $u_i(a_i, a_j, \varepsilon_i) = a_i(\kappa_i a_j + \varepsilon_i)$ for $i = 1, 2$ and assume $\varepsilon_i \overset{iid}{\sim} U[-1, 1]$. We set the true parameters at $\kappa_1 = \kappa_2 = -\frac{1}{2}$.

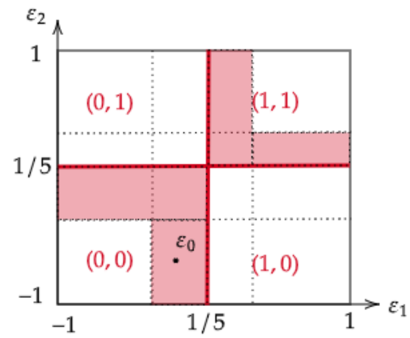
Consider $S^{complete}$ and $S^{private}$ which are the two information structures most commonly used in the empirical literature. Figure A.3 summarizes the Nash equilibrium predictions under each informational assumption. In Figure A.3-(a), as is well-known, the center region admits multiple equilibria, including a (totally) mixed strategy equilibrium. Figure A.3-(b) plots the predictions of Bayes Nash equilibrium in which outcomes are determined by a profile of threshold strategies such that $a_i = 1$ if and only if $\varepsilon_i \geq \frac{1}{5}$. Note that in both cases, the behavioral assumption underlying the static Nash equilibrium is that the players are trying to *predict* opponents' actions.

We claim that some outcomes might be unstable: players might want to revise their actions after observing the opponents' realized actions. Figure A.4 illustrates the instability of the Nash equilibrium outcomes. The shaded regions represent the set of ε's whose associated equilibrium outcomes may be unstable. In Figure A.4-(a), regret may arise when either

Figure A.4: Unstable region under Nash equilibria



(a) Complete information ($S^{complete}$)



(b) Private information ($S^{private}$)

$a = (0, 0)$ or $(1, 1)$ occurs "accidentally" due to totally mixed strategies; deviation incentives exist after observing the realized outcome. In this case, regrets occur with positive probability since players are mixing over their actions. It is also easy to see that if only pure strategies are allowed, then ex post regret problem does not arise because knowing others' pure strategies and the realization of $\varepsilon = (\varepsilon_1, \varepsilon_2)$ makes opponents' actions "known".

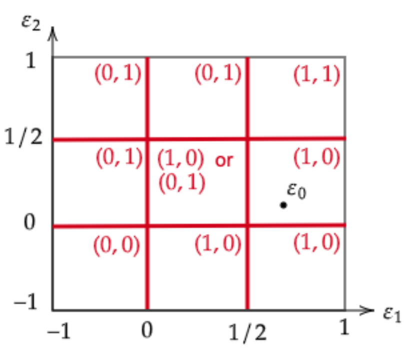
Figure A.4-(b) is more interesting. Outcomes in the shaded region are unstable because the revelation of opponents' actions determined by the Nash play may create incentives to revise the original actions. For instance, suppose ε_0 in the figure is realized so that the Bayes Nash strategy profile results in outcome $a = (0, 0)$. Then player 1 will find deviation to $a'_1 = 1$ strictly profitable because the profit from operating as a monopolist is strictly positive. Also note that rational agents can update their beliefs using information from the actions. For example, when the same ε_0 leads to $a = (0, 0)$, player 1 can infer that $\varepsilon_{0,2} \in [-1, 1/5]$ (although $\varepsilon_{0,2}$ is payoff-irrelevant to player 1 in this example). The refinement of information via observation of endogenous outcomes is exactly the idea of "rational expectations." The lesson is that endogenous actions reveal information and hence leads the posterior beliefs to become systematically different from the prior belief.

A.4.2 Examples of Bayes Stable Equilibria

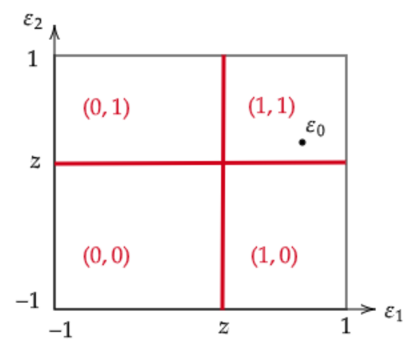
We provide simple examples of Bayes stable equilibria under $S^{private}$ and S^{null}. Consider a game $(G, S^{private})$ in which player i observes ε_i, but not ε_j, $j \neq i$. A decision rule $\sigma : \mathcal{E}_1 \times \mathcal{E}_2 \to \Delta(\mathcal{A}_1 \times \mathcal{A}_2)$ is a Bayes stable equilibrium of $(G, S^{private})$ if

$$\sum_{\varepsilon_{-i}} \psi_\varepsilon \sigma_{a|\varepsilon} u_i(a, \varepsilon_i) \geq \sum_{\varepsilon_{-i}} \psi_\varepsilon \sigma_{a|\varepsilon} u_i(a_i', a_{-i}, \varepsilon_i), \quad \forall i, \varepsilon_i, a, a_i'.$$

Figure A.5: Bayes stable equilibrium in a simple two-player entry game



(a) BSE under $S^{private}$



(b) BSE under S^{null}

Figure A.5-(a) shows the structure of the BSE under $S^{private}$. The listed outcomes represent the support of $\sigma(\cdot|\varepsilon)$ for each $\varepsilon \in [-1, 1]^2$. For example, if $\varepsilon_0 = (\varepsilon_{0,1}, \varepsilon_{0,2})$ in the figure is realized, the mediator publicly recommends $a = (1, 0)$ with probability one since $\sigma((1, 0)|\varepsilon_0) = 1$. When $a = (1, 0)$ is realized, each player i can partially infer ε_{-i} (e.g., player 1 infers that $\varepsilon_{0,2} \in [-1, 1/2))$ although ε_{-i} is payoff-irrelevant. One can readily check that each player has no incentive to deviate from $a = (1, 0)$.[9]

Figure A.5-(b) shows an example of a Bayes stable equilibrium under S^{null} in which

[9]Note the similarity between Figure A.5-(a) and A.4-(a). Figure A.5-(a) also illustrates Theorem 1.2 that says the predictions of Bayes stable equilibria under $S^{private}$ is the same and those of pure strategy Nash equilibria under $S^{complete}$. In a Bayes stable equilibrium under $S^{private}$, whenever an outcome (a_i, a_{-i}) is recommended, player i knows all of her payoff-relevant variables $(a_i, a_{-i}, \varepsilon_i)$; knowledge of ε_{-i} is irrelevant. Then, at an equilibrium situation, each player acts as if she is in a pure strategy Nash equilibrium of $S^{complete}$.

players' signals are null. The equilibrium condition is

$$\sum_{\varepsilon} \psi_{\varepsilon} \sigma_{a|\varepsilon} u_i (a, \varepsilon_i) \geq \sum_{\varepsilon} \psi_{\varepsilon} \sigma_{a|\varepsilon} u_i (a_i', a_{-i}, \varepsilon_i), \quad \forall i, a, a_i'.$$

In this case, the players have no private signals and the players' posteriors are derived solely from observing a. For example, suppose that the outcome function σ is given as shown in Figure A.5-(b), where z is an arbitrary threshold parameter. If ε_0 in the figure is realized, then σ dictates that outcome $a = (1, 1)$ is realized. Upon observing $a = (1, 1)$, Player 1 learns that $\varepsilon_{0,1} \in (z, 1)$ (the case for Player 2 is symmetric). For the recommended outcome to be incentive compatible, we need that

$$\delta_2 + \mathbb{E}^{\sigma} [\varepsilon_{0,1} | \varepsilon_{0,1} \in (z, 1)] \geq 0 \Leftrightarrow z \geq 0.$$

One can further verify that the obedience condition requires that $z \geq -1$, $z \leq 2$, and $z \leq 1$ for outcome $a = (1, 0)$, $(0, 1)$, and $(0, 0)$ to be incentive compatible to player 1. In sum, when the outcome function is assumed to have a parametric structure as in Figure A.5-(b), any σ with $z \in [0, 1]$ constitutes a BSE under S^{null}.

Appendix for Chapter 2

B.1 Proofs

B.1.1 Proof of Lemma 2.1

($\Rightarrow$) Let β be a MPE of (G, S). β induces $\left(G^{\beta}, S\right)$. The BNE best-response condition is directly implied by the MPE condition of (G, S).

($\Leftarrow$) Let β be a BNE of $\left(G^{\beta}, S\right)$. The BNE best-response condition implies that there is no profitable one-shot deviation given that each player expects all players to follow the prescription of β in the future. The absence of profitable one-shot deviation implies that β is a subgame perfect equilibrium of (G, S) by the one-shot deviation principle. Therefore, β is a MPE of (G, S). $\square$

B.1.2 Proof of Lemma 2.2

Omitted.

B.1.3 Proof of Lemma 2.3

It is enough to show that $V_i^{\beta}(x) = V_i^{\sigma}(x)$ for all $i \in \mathcal{I}$ and $x \in \mathcal{X}$ because then it follows that $v_i^{\beta}(a, x, \varepsilon) = v_i^{\sigma}(a, x, \varepsilon)$ for all $i \in \mathcal{I}$, $a \in \mathcal{A}$, $x \in \mathcal{X}$, and $\varepsilon \in \mathcal{E}$.

Recall that V_i^{β} and V_i^{σ} satisfies

$$V_i^{\beta}(x) = \sum_{\varepsilon,\tau,\tilde{\tau},a} \psi_{\varepsilon|x} \pi_{\tau|x,\varepsilon} \lambda_{\tilde{\tau}|x,\varepsilon,\tau} \beta_{a|x,\tau,\tilde{\tau}} \left(u_i(a, x, \varepsilon) + \delta \sum_{x'} V_i^{\beta}(x') f_{x'|a,x,\varepsilon} \right)$$

$$V_i^{\sigma}(x) = \sum_{\varepsilon,\tau,a} \psi_{\varepsilon|x} \pi_{\tau|x,\varepsilon} \sigma_{a|x,\varepsilon,\tau} \left(u_i(a, x, \varepsilon) + \delta \sum_{x'} V_i^{\sigma}(x') f_{x'|a,x,\varepsilon} \right)$$

for all $x \in \mathcal{X}$, where $\beta_{a|x,\tau,\tilde{\tau}} \equiv \prod_{j=1}^{I} \beta_j (a_j|x, \tau_j, \tilde{\tau}_j)$. But since β induces σ,

$$\begin{aligned} V_i^{\beta}(x) &= \sum_{\varepsilon,\tau,a} \psi_{\varepsilon|x} \pi_{\tau|x,\varepsilon} \left(\sum_{\tilde{\tau}} \lambda_{\tilde{\tau}|x,\varepsilon,\tau} \beta_{a|x,\tau,\tilde{\tau}} \right) \left(u_i(a,x,\varepsilon) + \delta \sum_{x'} V_i^{\beta}(x') f_{x'|a,x,\varepsilon} \right) \\ &= \sum_{\varepsilon,\tau,a} \psi_{\varepsilon|x} \pi_{\tau|x,\varepsilon} \sigma_{a|x,\varepsilon,\tau} \left(u_i(a,x,\varepsilon) + \delta \sum_{x'} V_i^{\beta}(x') f_{x'|a,x,\varepsilon} \right). \end{aligned}$$

Comparison of the equations defining V_i^{β} and V_i^{σ} implies that $V_i^{\beta} = V_i^{\sigma}$, which is what we wanted to show. □

B.1.4 Proof of Theorem 2.1

To prove the statement of the theorem, we make use of Theorem 1 of BM, which we restate below.

Lemma B.1 (BM Theorem 1)**.** *A decision rule σ is a Bayes correlated equilibrium of (G, S) if and only if, for some expansion S^* of S, there is a Bayes Nash equilibrium of (G, S^*) that induces σ.*

($\subseteq$) Suppose σ is a MCE of (G, S). We want to show that there exists an expansion S^* and a strategy profile β in (G, S^*) such that β is a MPE of (G, S^*) and β induces σ. Since σ is a MCE of (G, S), it is a BCE of (G^{σ}, S). By BM Theorem 1, there exists an expansion S^* of S and a strategy profile β of (G^{σ}, S^*) such that β is a BNE of (G^{σ}, S^*) and β induces σ. But since $G^{\sigma} = G^{\beta}$, β is also a BNE of $\left(G^{\beta}, S^*\right)$, which in turn implies that β is a MPE of (G, S^*).

($\supseteq$) Suppose β is a MPE of (G, S^*). We want to show that if β induces σ in (G, S), then σ is a MCE of (G, S). Since β is a MPE of (G, S^*), it is a BNE of $\left(G^{\beta}, S^*\right)$. By BM Theorem 1, if σ is induced by β, then σ is a BCE of $\left(G^{\beta}, S\right)$. But since $G^{\beta} = G^{\sigma}$, σ is also a BCE of (G^{σ}, S), which in turn implies that σ is a MCE of (G, S). □

B.1.5 Proof of Corollary 2.1

($\subseteq$) Take $\phi \in \mathcal{P}^{MCE}_{a|x}(G,S)$. By definition, there exists $q \in \mathcal{P}^{MCE}_{a|x,\varepsilon,\tau}(G,S)$ such that q induces ϕ, i.e.,

$$\phi_{a|x} = \sum_{\varepsilon\in\mathcal{E},\tau\in\mathcal{T}} \psi_{\varepsilon|x}\pi_{\tau|x,\varepsilon}q_{a|x,\varepsilon,\tau}, \quad \forall a \in \mathcal{A}, x \in \mathcal{X}.$$

Then there exists a S^* such that $S^* \succsim_E S$ and $q \in \mathcal{P}^{MPE}_{a|x,\varepsilon,\tau}(G,S^*)$, implying that $\phi \in \mathcal{P}^{MPE}_{a|x}(G,S^*)$.

($\supseteq$) Take $\phi \in \bigcup_{S^*\succsim_E S}\mathcal{P}^{MPE}_{a|x}(G,S^*)$. Then there exists S^* such that $S^* \succsim_E S$ and $\phi \in \mathcal{P}^{MPE}_{a|x}(G,S^*)$ which implies that there exists $q \in \mathcal{P}^{MPE}_{a|x,\varepsilon,\tau}(G,S^*)$ such that q induces ϕ. Then since $q \in \mathcal{P}^{MCE}_{a|x,\varepsilon,\tau}(G,S)$, we have $q \in \mathcal{P}^{MCE}_{a|x}(G,S)$. $\square$

B.1.6 Proof of Theorem 2.2

We want to show

$$\Theta^{MCE}_I(S) = \bigcup_{\tilde{S}\succsim_E S}\Theta^{MPE}_I\left(\tilde{S}\right).$$

To prove the equality, note that

$$\Theta^{MCE}_I(S) \equiv \left\{\theta\in\Theta:\ \phi\in\mathcal{P}^{MCE}_{a|x}\left(G^\theta,S\right)\right\}$$

and

$$\bigcup_{\tilde{S}\succsim_E S}\Theta^{MPE}_I\left(\tilde{S}\right) \equiv \bigcup_{\tilde{S}\succsim_E S}\left\{\theta\in\Theta:\ \phi\in\mathcal{P}^{MPE}_{a|x}\left(G^\theta,\tilde{S}\right)\right\}$$
$$= \bigcup_{\tilde{S}\succsim_E S}\left\{\theta\in\Theta:\ \phi\in\bigcup_{\tilde{S}\succsim_E S}\mathcal{P}^{MPE}_{a|x}\left(G^\theta,\tilde{S}\right)\right\}.$$

But by Corollary 2.1, we have

$$\mathcal{P}^{MCE}_{a|x}(G,S) = \bigcup_{S^*\succsim_E S}\mathcal{P}^{MPE}_{a|x}(G,S^*).$$

Thus, we must have $\Theta_I^{MCE}(S) = \bigcup_{\tilde{S} \succsim_E S} \Theta_I^{MPE}\left(\tilde{S}\right)$.

B.1.7 Proof of Theorem 2.4

To see the feasibility of the program, take any σ such that $\sigma \in \Delta_{a|x,\varepsilon,\tau}$ and $\phi = \phi(\sigma)$. Given σ, V is determined. Finally, given σ and V, there must exist some t that satisfies all the inequality conditions.

Next, that $Q(\theta) = 0$ if and only if $\theta \in \Theta_I$ is trivial. □

Appendix for Chapter 3

C.1 Proofs

C.1.1 Proof of Theorem 3.2

Before proceeding with the proof, it is helpful to review the logit probability formula (see Train (2009) Chapter 3). Let there be a decision maker who chooses from a finite set of alternatives. Let $U_j = V_j + \eta_j$ for alternatives $j = 1, ..., J$ where U_j is the payoff from choosing alternative j, V_j is the deterministic part of the payoff, and η_j is the stochastic part. If each η_k is independently and identically distributed as type 1 extreme value, then the probability of choosing alternative j is

$$P_j = \Pr\left(\{\eta : \; V_j + \eta_j > V_k + \eta_k, \; \forall k \neq j\}\right) = \frac{e^{V_j}}{\sum_l e^{V_l}} \tag{C.1}$$

which is the logit choice probability.

Now let us prove the statement of the theorem. By definition, $\mathcal{L}(y|x;\theta) = \nu\left(G^{-1}(y|x;\theta)\,|x;\theta\right)$. Under Assumption 3.1, $\nu(\cdot|x;\theta)$ describes the joint distribution of ε_i's, $i = 1, ..., I$ whose individual components $\varepsilon_i(y_i)$'s are i.i.d. type 1 extreme value distributed. Here, $G^{-1}(y|x;\theta) = \{\varepsilon \in \mathcal{E} : \; y \in G(\varepsilon|x;\theta)\}$ which describes the set of ε's that supports y as a pure-strategy Nash equilibrium outcome given x and θ. Thus, $\nu\left(G^{-1}(y|x;\theta)\,|x;\theta\right)$ represents the *maximal* probability of observing outcome y when the covariates and the parameter are fixed at x and θ respectively.

Let us show that $\mathcal{L}(y|x;\theta)$ admits a closed-form expression. For each given y, x, θ, we have

$$
\begin{aligned}
& G^{-1}(y|x;\theta) \\
= & \{\varepsilon \in \mathcal{E}: \; y \in G(\varepsilon|x;\theta)\} \\
= & \{\varepsilon \in \mathcal{E}: \; u_i(y,x,\varepsilon_i;\theta) \geq u_i(y_i', y_{-i}, x, \varepsilon_i;\theta), \quad \forall i, y_i' \neq y_i\} && \text{(C.2)} \\
= & \{\varepsilon \in \mathcal{E}: \; v_i(y,x;\theta) + \varepsilon_i(y_i) \geq v_i(y_i', y_{-i}, x;\theta) + \varepsilon_i(y_i'), \quad \forall i, y_i' \neq y_i\} && \text{(C.3)} \\
= & \bigcap_{i=1}^{I} \{\varepsilon_i \in \mathcal{E}_i: \; v_i(y,x;\theta) + \varepsilon_i(y_i) \geq v_i(y_i', y_{-i}, x;\theta) + \varepsilon_i(y_i'), \quad \forall y_i' \neq y_i\} && \text{(C.4)} \\
\equiv & \bigcap_{i=1}^{I} A_i(y_i|y_{-i},x;\theta) && \text{(C.5)}
\end{aligned}
$$

where

$$
A_i(y_i|y_{-i},x;\theta) := \{\varepsilon_i \in \mathcal{E}_i: \; v_i(y,x;\theta) + \varepsilon_i(y_i) \geq v_i(y_i', y_{-i}, x;\theta) + \varepsilon_i(y_i'), \quad \forall y_i' \neq y_i\}.
$$

In the above, (C.2) follows from the definition of pure strategy Nash equilibrium, (C.3) follows from the additive separability of the payoff functions (Assumption 3.1), (C.4) follows from the assumption that ε_i only enters player i's payoff, but not the payoff of $j \neq i$, and (C.5) follows from the definition of $A_i(y_i|y_{-i},x;\theta)$.

Now the generalized likelihood function for each outcome y is

$$
\begin{aligned}
& \mathcal{L}(y|x;\theta) \\
= & \nu(\{\varepsilon \in \mathcal{E}: \; y \in G(\varepsilon|x;\theta)\}|x;\theta) \\
= & \Pr\left(\bigcap_{i=1}^{I} A_i(y_i|y_{-i},x;\theta)\right) \\
= & \prod_{i=1}^{I} \Pr(A_i(y_i|y_{-i},x;\theta)) && \text{(C.6)} \\
= & \prod_{i=1}^{I} \frac{\exp(v_i(y,x;\theta))}{\sum_{y_i' \in \mathcal{Y}_i} \exp(v_i(y_i', y_{-i}, x;\theta))} && \text{(C.7)}
\end{aligned}
$$

where (C.6) follows from the independence of ε_i's across the players $i = 1, ..., I$, and (C.7) follows from the familiar logit choice probability formula (C.1). $\square$

C.1.2 Proof of Lemma 3.1

We show that $g_{y,x}(\theta)$ is convex in θ by using the fact that the sum of convex functions is convex (see p.79 of Boyd and Vandenberghe (2004)).

- Step 1: The first term $\log(\phi(y|x))$ is a constant.
  - This follows from the assumption that the conditional choice probabilities $\phi(y|x)$ are identified from the data.
- Step 2: $\log\left(\sum_{y_i' \in \mathcal{Y}_i} \exp\left(v_i\left(y_i', y_{-i}, x; \theta\right)\right)\right)$ is convex in θ.
  - We use two facts.
    - Fact 1: The log-sum-exp function $f(z) = \log\left(\sum_{l=1}^{L} \exp(z_l)\right)$ is convex in z (see p.72 of Boyd and Vandenberghe (2004)).
    - Fact 2: The composition of a convex function with affine function is convex, i.e., $g(z) \equiv f(Az + b)$ is convex in z if f is convex in z (see p.79 of Boyd and Vandenberghe (2004)).

  Since $v_i(y_i', y_{-i}, x; \theta) \equiv w_i(y_i', y_{-i}, x)^T \theta$ for some $w_i(y_i', y_{-i}, x)$ by Assumption 3.2, it follows that $\log\left(\sum_{y_i' \in \mathcal{Y}_i} \exp\left(v_i\left(y_i', y_{-i}, x; \theta\right)\right)\right)$ is convex in θ.
- Step 3: $\sum_{i=1}^{I} \left\{\log\left(\sum_{y_i' \in \mathcal{Y}_i} \exp\left(v_i\left(y_i', y_{-i}, x; \theta\right)\right)\right) - v_i(y, x; \theta)\right\}$ is convex in θ.
  - This follows from the convexity of the affine function $-v_i(y, x; \theta)$ and the fact that the sum of convex functions is convex.

From the above, it follows that each $g_{y,x}(\theta)$ is convex in θ, so the feasible set of θ's satisfying $g_{y,x}(\theta) \leq 0$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$ is convex. $\square$

C.1.3 Proof of Theorem 3.3

Since $g_{y,x}(\theta)$ is a convex function, the feasible set of θ's satisfying $g_{y,x}(\theta) \leq 0$ for a given y, x is convex. Then, since $\Theta_I \equiv \{\theta \in \Theta : \; g_{y,x}(\theta) \leq 0, \quad \forall y \in \mathcal{Y}, x \in \mathcal{X}\}$ is simply the intersection of the feasible sets at each y and x, it follows that Θ_I is convex. □

C.1.4 Proof of Theorem 3.4

1. First, note that (3.9) is a convex program because $g_{y,x}(\theta)$ is convex in θ (Lemma 3.1). Second, the equivalence between the non-emptiness of Θ_I and (3.9) follows directly from the definition of Θ_I. □

2. First, (3.10) is a convex program because the objective function is linear in θ and the constraints are convex in θ. Second, it is easy to see that the solution to (3.8) is equal to the solution to (3.10) because $\theta \in \Theta_I$ if and only if $g_{y,x}(\theta) \leq 0$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. □

C.1.5 Proof of Lemma 3.2

1. To show that the program (3.11)–(3.13) is convex, it is enough to show that (3.12) represents convex constraints. However, this is immediate given that $\log(1 + t_{y,x})$ is concave in t. The feasibility of the program is also easy to see because $g_{y,x}(\theta) < \infty$ for all y, x, θ. □

2. The criterion function $Q(\theta)$ is non-negative for all θ because the objective function is a weighted average of non-negative numbers. □

3. If $\theta \in \Theta_I$, then $g_{y,x}(\theta) \leq 0$ for all y, x indicating that we can plug in $t_{y,x} = 0$ for all y, x to (3.11)–(3.13). Conversely, if $Q(\theta) = 0$, then it must be that $t_{y,x} = 0$ for all y, x so that we have $g_{y,x}(\theta) \leq 0$ for all y, x. □

C.1.6 Proof of Theorem 3.5

The following lemma will be useful for proving the theorem.

Lemma C.1. *For any $c \geq 0$, the set of (θ, t) satisfying (3.12), (3.13), and $h^w(t) \leq c$ is convex.*

Proof. The statement follows from observing that each (3.12), (3.13), and $h^w(t) \leq c$ are constraints that are convex in (θ, t) respectively. □

We are ready to prove Theorem 3.5.

1. Let $c \geq 0$. Let us write

$$\begin{aligned} \Theta_I(c) &\equiv \{\theta \in \Theta : \; Q(\theta) \leq c\} \\ &= \{\theta \in \Theta : \; \exists t \text{ such that (3.12), (3.13), and } h^w(t) \leq c\}. \end{aligned} \tag{C.8}$$

   Take two $\theta^1, \theta^2 \in \Theta_I(c)$, and find the corresponding t^1, t^2 according to (C.8). We want to show that $\theta^\lambda := \lambda\theta^1 + (1-\lambda)\theta^2 \in \Theta_I(c)$ for any $\lambda \in [0,1]$. Let $t^\lambda := \lambda t^1 + (1-\lambda)t^2$. By Lemma C.1, $\left(\theta^\lambda, t^\lambda\right)$ satisfies (3.12), (3.13), and $h^w\left(t^\lambda\right) \leq c$, so $\theta^\lambda \in \Theta_I(c)$. □

2. First, note that

$$\min_\theta Q(\theta) \equiv \min_\theta \left(\min_t h^w(t) \quad \text{s.t. (3.12), (3.13)}\right)$$

   Thus, the reformulation of $\min_\theta Q(\theta)$ to (3.15) simply follows from solving the nested minimization problem as a single minimization problem. Second, the convexity of program (3.15) follows from the linearity of the objective function and the convexity of the constraints. □

3. The projection problem is

$$\min_\theta p^T\theta \quad \text{subject to } \theta \in \Theta_I(c) \tag{C.9}$$

The equivalence between (C.9) and (3.16) follows from (C.8). The convexity of program (3.16) follows from the linearity of the objective function and the convexity of the constraints. □

C.1.7 Proof of Theorem 3.6

Observe that

$$\begin{aligned}\Pr\left(\Theta_I \subseteq \widehat{\Theta}_I^{\alpha}\right) &= \Pr\left(\Theta_I\left(\phi\right) \subseteq \bigcup_{\tilde{\phi}\in\Phi_n^{\alpha}} \Theta_I\left(\tilde{\phi}\right)\right)\\ &\geq \Pr\left(\tilde{\phi} \in \Phi_n^{\alpha}\right).\end{aligned}$$

Taking the lim inf on both sides yields the desired result. □