Stable Outcomes and Information in Games: An Empirical Framework[*]

Paul S. Koh[†]

November 18, 2021

Job Market Paper

Download the latest version [here]

Abstract

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 in the presence of incomplete information, we build an empirical framework by introducing a novel solution concept that we call *Bayes stable equilibrium.* Our framework allows the researcher to be agnostic about players' information and the equilibrium selection rule. The Bayes stable equilibrium identified set collapses to the complete information pure strategy Nash equilibrium identified set under strong assumptions on players' information. Furthermore, all else equal, it is weakly tighter than the Bayes correlated equilibrium identified set. We also propose computationally tractable approaches for estimation and inference. In an application, we study the strategic entry decisions of McDonald's and Burger King in the US. Our results highlight the identifying power of informational assumptions and show that the Bayes stable equilibrium identified set can be substantially tighter than the Bayes correlated equilibrium identified set. In a counterfactual experiment, we examine the impact of increasing access to healthy food on the market structures in Mississippi food deserts.

Keywords: Estimation of games, stability, Bayes stable equilibrium, informational robustness, partial identification, entry models, burger industry, food deserts

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 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'

---

[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).

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

3

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 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

---

[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; (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.

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 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

---

[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).

[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.

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 (2020)). Bergemann and Morris (2016) established the informational robustness property of Bayes correlated equilibrium by showing that Bayes correlated equilibrium captures the 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 information robustness property when the underlying solution concept is rational expectations equilibrium.

Finally, our empirical application contributes to the literature on entry competition in

---

[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.

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 2 introduces the notion of Bayes stable equilibrium in a general finite game of incomplete information and studies its property. Section 3 sets up the econometric model and provides identification results. Section 4 provides econometric strategies for computationally tractable estimation and inference. Section 5 applies our framework to the entry game played by McDonald's and Burger King in the US. Section 6 concludes. All proofs are in Appendix A.

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 $\Delta_{y|x}$ 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 $a = (a_1, ..., a_I) = (a_i, a_{-i})$.

---

[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.

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 the 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. For empirical work, 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 2.1, we lay out the game environment. In Section 2.2, we formalize the notion of stable outcomes and motivate our solution concept. In Section 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 Section 2.4, we introduce Bayes stable equilibrium. Then, in Section 2.5, 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 2.6, we compare the proposed solution concept to pure strategy Nash equilibrium and Bayes correlated equilibrium.

---

[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.

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 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 model for our empirical application.

Example 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, $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$,

---

[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.

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 2.5. ∎

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 $t = (t_1, ..., t_I)$,

---

[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

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 (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 $i = 1, ..., I$,

$$\mathbb{E}_{\varepsilon \sim \mu^i}[u_i(a, \varepsilon)] \geq \mathbb{E}_{\varepsilon \sim \mu^i}[u_i(a_i', a_{-i}, \varepsilon)] \tag{1}$$

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 depend on ad hoc assumptions. In the following sections, we propose a simple and pragmatic approach to the problem.

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

---

"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.

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 and the signal distribution, 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 $\mathbb{E}_\varepsilon^\delta[u_i(a_i', a_{-i}, \varepsilon)|t_i, a_i, a_{-i}]$ denote the expected payoff to player i from choosing a_i' conditional on observing private signal t_i and action profile (a_i, a_{-i}).

Definition 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

---

[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.

$\Pr^{\delta}(t_i, a_i, a_{-i}) > 0$, we have

$$\mathbb{E}^{\delta}_{\varepsilon}[u_i(a_i, a_{-i}, \varepsilon)|t_i, a_i, a_{-i}] \geq \mathbb{E}^{\delta}_{\varepsilon}[u_i(a_i', a_{-i}, \varepsilon)|t_i, a_i, a_{-i}] \qquad (2)$$

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) 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.

14

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 ψ 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 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 the players always have no incentives to deviate from the realized action profiles, we call σ a Bayes stable equilibrium.

Definition 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

$$\mathbb{E}_{\varepsilon}^{\sigma}\left[u_i(a_i, a_{-i}, \varepsilon) | t_i, a_i, a_{-i}\right] \geq \mathbb{E}_{\varepsilon}^{\sigma}\left[u_i(a_i', a_{-i}, \varepsilon) | t_i, a_i, a_{-i}\right] \tag{3}$$

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 (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 (3) can be rewritten as follows:[11]

$$
\sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i \left(a, \varepsilon \right) \geq \sum_{\varepsilon, t_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon,t} u_i \left(a_i', a_{-i}, \varepsilon \right), \quad \forall i \in \mathcal{I}, t_i \in \mathcal{T}_i, a \in \mathcal{A}, a_i' \in \mathcal{A}_i. \quad (4)
$$

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

2.5 Informational Robustness of Bayes Stable Equilibrium

In Section 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 generating process in the real world. Clearly, attempts to characterize all feasible predictions

---

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

(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 equilibria predictions that can arise when the players might observe more information than specified by the analyst.

We formalize the idea as follows. First, to formalize 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 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 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 (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 σ.

Theorem 1 can be framed in terms of marginal distributions on the action profiles. This characterization is more relevant to 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 (Observational Equivalence). *For any basic game G and information structure S, $\mathcal{P}^{BSE}_a(G, S) = \bigcup_{S^* \succsim_E S} \mathcal{P}^{REE}_a(G, S^*)$.*

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 D 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.

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 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 2.1 states that in any game with complete information structure, assuming rational expectations equilibrium is 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 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 2.2 implies that in a class of games where the state of the world ε is simply a vector 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.

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

$\sigma : \mathcal{E} \times \mathcal{T} \rightarrow \Delta(\mathcal{A})$ in (G, S) and announces it 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 $\sigma : \mathcal{E} \times \mathcal{T} \rightarrow \Delta(\mathcal{A})$ 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,

$$\sum_{\varepsilon, t_{-i}, a_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon, t} u_i(a_i, a_{-i}, \varepsilon) \geq \sum_{\varepsilon, t_{-i}, a_{-i}} \psi_\varepsilon \pi_{t|\varepsilon} \sigma_{a|\varepsilon, t} u_i(a'_i, a_{-i}, \varepsilon), \quad \forall i, t_i, a_i, a'_i. \tag{5}$$

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 (5) which integrate out opponents' actions a_{-i} since each player i needs to anticipate a_{-i}, Bayes stable equilibrium conditions (4) condition on a_{-i} because it is assumed that all actions are publicly observed at the equilibrium situation. The following is immediate.

Theorem 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 2. When there is a single player, the two solution concepts are identical because there is no informational feedback from observing opponents' actions.

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 its properties.

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 $S^x = \langle (\mathcal{T}_i)_{i=1}^I, \pi^x \rangle$.[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 $\psi^{x,\theta} \in \Delta(\mathcal{E})$ 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 burden. See Appendix B for the details on how we make discrete approximations to continuous distributions.

$u_i^{x,\theta} : \mathcal{A} \times \mathcal{E} \to \mathbb{R}$, 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 (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 $\phi = (\phi_{(0,0)}, \phi_{(0,1)}, \phi_{(1,0)}, \phi_{(1,1)})$ 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, the identified set of parameters can be defined when the solution concept and the information structure are specified. For any game $(G^{x,\theta}, S^x)$, let $\mathcal{P}_a^{SC}(G^{x,\theta}, S^x)$ 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 5 (Identified set of parameters). Given Assumption 1, 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.

3.2 Identification and Informational Robustness

We translate the observational equivalence between rational expectations equilibrium and Bayes stable equilibrium (Corollary 1) in terms of identified sets. Consider the following assumption.

Assumption 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 $\tilde{S}^{x,0}$ which is an expansion of S^x ($\tilde{S}^{x,0} \succsim_E S^x$).*

Assumption 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 and 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 2 is computationally infeasible because it requires searching over the set of information structures which is large. We show that Assumption 2 can be replaced with the following assumption, which does not rely on unknown information structures.

Assumption 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; Assumption 2 and Assumption 3 are observationally equivalent.

Theorem 4 (Equivalence of identified sets). *The identified set under Assumptions 1 and 2 is equal to the identified set under Assumptions 1 and 3.*

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.

Our identification results make no assumptions on the equilibrium selection rule. The Bayes stable equilibrium identified set under Assumptions 1 and 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 (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]

---

[14] Also see Syrgkanis, Tamer, and Ziani (2021) Lemma 2 for a general argument on why it is without loss

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). 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 5 (Relationship between identified sets). *Suppose Assumption 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 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 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 5.2 (which follows from Theorem 2) says that Bayes stable equilibrium and pure strategy Nash equilibrium are observationally equivalent when $S^{complete}$ is assumed.[15] Furthermore, due to Assumption 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

---

to assume that the data are generated by a single equilibrium if the set of predictions is convex.

[15]When Assumption 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.

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 5.3 (which follows from Theorem 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.

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.
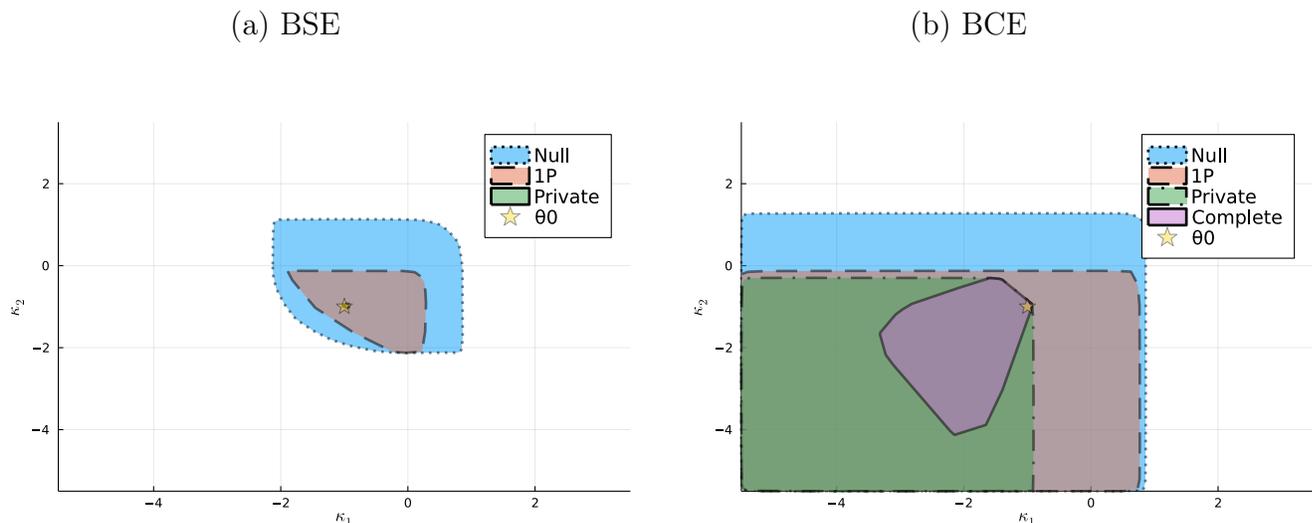
Figure 1-(a) shows the BSE identified sets obtained under different baseline information structures. The identified sets shrink as the informational assumptions get stronger. We omit the complete information case since $\Theta_I^{BSE}(S^{private}) = \Theta_I^{BSE}(S^{complete})$. Setting the baseline 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

---

[16]Computational details can be found in Appendix B.

[17]Specifically, we generate population choice probability by finding a feasible $\sigma : \mathcal{E} \to \Delta(\mathcal{A})$ which satisfies the inequalities in (8) as described in Section 4.1.

Figure 1: Convex Hulls of Identified Sets

(a) BSE   (b) BCE

(which corresponds to the PSNE identified set). Similarly, Figure 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.

Comparing Figure 1-(a) and 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, BSE identified set can be much tighter than the BCE identified set. Therefore, using restrictions that incorporate observability of opponents' actions can add substantial identifying power.

4   Estimation and Inference

We propose a computationally attractive approach for estimation and inference. In Section 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 4.2, we show that this property

can be combined with the insights from Horowitz and Lee (2021) to make construction of 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 4.3, we provide some practical suggestions for computational implementations.

4.1 A Linear Programming Characterization

The following proposition provides a computationally attractive characterization of the identified set $\Theta_I \equiv \Theta_I^{BSE}(S)$. 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}). Recall our notation: $\sigma^x \in \Delta_{a|\varepsilon,t}$ if and only if $\sigma_{a|\varepsilon,t}^x \geq 0$ for all a, ε, t and $\sum_a \sigma_{a|\varepsilon,t}^x = 1$.

Theorem 6 (Linear programming characterization). *Under Assumptions 1 and 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 $i \in \mathcal{I}$, $t_i \in \mathcal{T}_i$, $a \in \mathcal{A}$, $a_i' \in \mathcal{A}_i$,*

$$\sum_{\varepsilon \in \mathcal{E}, t_{-i} \in \mathcal{T}_{-i}} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x \partial u_i^{x,\theta}(a_i', a, \varepsilon_i) \leq 0. \tag{6}$$

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

$$\phi_a^x = \sum_{\varepsilon \in \mathcal{E}, t \in \mathcal{T}} \psi_\varepsilon^{x,\theta} \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x. \tag{7}$$

Theorem 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 states that the nuisance parameter σ^x should be a decision rule that satisfies the Bayes stable equilibrium conditions. The second condition states that the observed conditional choice probabilities must be consistent with those induced by the equilibrium decision rule. Given a candidate θ, $\psi_\varepsilon^{x,\theta}$, $\pi_{t|\varepsilon}^x$, $\partial u_i^{x,\theta}$, and ϕ_a^x are known objects. Then, since the variables of optimization σ^x enter the constraints linearly, the program is linear.

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 5 and Theorem 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 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 $i \in \mathcal{I}$, $\varepsilon_i \in \mathcal{E}_i$, $a \in \mathcal{A}$, $a_i' \in \mathcal{A}_i$,

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

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

$$\sum_{\varepsilon_{-i}} \psi_\varepsilon \sigma_{a|\varepsilon} \left((a_i' - a_i)(\kappa_i a_{-i} + \varepsilon_i) \right) \leq 0, \quad \forall i, \varepsilon_i, a_i, a_{-i}, a_i' \tag{8}$$

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

which is a linear feasibility program. ∎

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 6. The statistical sampling uncertainty arises only from the estimation of 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 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

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

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

---

[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 can be translated to posterior statements on θ using a known mapping between them.

with high probability asymptotically.

Assumption 4 (Convex confidence set for CCP)**.** *Let $\alpha \in (0,1)$. A set Φ_n^α such that*

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

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

$$\widehat{\Theta}_I^\alpha \equiv \bigcup_{\tilde{\phi}\in\Phi_n^\alpha} \Theta_I\left(\tilde{\phi}\right). \tag{9}$$

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

Theorem 7. *Suppose Φ_n^α satisfies Assumption 4 and $\widehat{\Theta}_I^\alpha$ is constructed as (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 7.1 follows directly from (9) and the assumption on Φ_n^α. To understand Theorem 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 (6), (7), and $\phi \in \Phi_n^\alpha$ are satisfied. Compared to the population program described in Theorem 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 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]

4.3 Implementation

We propose a practical routine for obtaining the confidence set $\widehat{\Theta}_I^\alpha$. Theorem 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 previous 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).

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

---

[20]See Appendix B.2 for details. We also provide Monte Carlo evidence that the proposed method has desirable coverage probabilities even when $\mathcal{X}$ has many elements.

following convex program.

$$\min_{q,\sigma,\phi} \sum_{x \in \mathcal{X}} w^x q^x \quad \text{subject to} \tag{10}$$

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

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

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

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

Intuitively, $q^x \geq 0$ 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 (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 8 (Implementation). *1. For any $\theta \in \Theta$, program (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 (10) via the envelope theorem.

In particular, Theorem 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

---

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

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 8.3, one can use gradient-based optimization algorithms to identify a minimizer of the criterion function.[22] The ability to quickly identify $\arg\min_{\tilde{\theta}} \widehat{Q}_n^\alpha\left(\tilde{\theta}\right)$ 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 B.3 for details.

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 a rich dataset. 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 and 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 corresponds to our definition of markets) in the contiguous US over 1997-2019. For instance, the probability that McDonald's has an outlet in operation in a local market conditional on it having an outlet in operation three years ago is 0.95. Together with the assumption that the costs of revising decisions are sufficiently

---

[22]When program (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.

low, the evidence supports the claim that firms' decisions are best-responses to opponents' decisions that are readily observed.[23]

Table 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 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. 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

---

[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)

policy exercise that studies how the market structures in Mississippi food deserts respond after increasing access to healthy food.

5.1 Data Description

We combine multiple datasets to construct the final dataset for structural estimation of the entry game. 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 in 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 which decennial census data were available. We describe the main features of our dataset below. Further details on data construction are provided in Appendix C.

---

[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 C 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.

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 3. Continuous variables are discretized to binary variables by using cutoffs around their medians. Clearly, the entry probability of McDonald's 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.

---

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

Table 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 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.

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 4 reports the average marginal effects computed from the regression results.

Table 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 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 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 when studying a policy experiment. In the next section, we estimate the entry game using Bayes stable equilibrium as a solution concept.

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: $S^{null}, S^{1P}, S^{private}$. 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 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

---

[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.

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.

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 B 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 used simultaneous confidence bands based on the method described in Fitzpatrick and Scott (1987); using simultaneous confidence bands makes the evaluation of

---

[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.

the criterion function a linear program.

5.4.1 The Role of Informational Assumptions on Identification

Table 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 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

---

[31]Specifically, we consistently find that the minimum of the criterion function under $S^{private}$ is strictly

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 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 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.

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 6. We can readily see that the Bayes correlated equilib-

---

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.

Table 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.

rium assumption produces a much larger set under the same 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 5 and 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 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.

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 B.4 for computational details.

We report the results of the counterfactual analysis in Table 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.

---

[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). Some studies point out that consumers' eating habits may not change even after healthy food options increase (e.g., Allcott et al. (2019)). Our model assumes that consumption responds to supply-side factors.

[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 the responses of healthy food providers.

Table 7: The Impact of Increasing Access to Healthy Food in Mississippi Food Deserts

| | Data | $BSE(S^{1P})$ | | $BCE(S^{1P})$ | |
		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.

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 direction. 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 on equilibrium selection and admit all information structures that are expansions of the baseline information structure. 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 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.

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 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.

References

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.

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.

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.

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.

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

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.

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. 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.

Doval, L. and J. C. Ely (2020): "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.

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.

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

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.

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.

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

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.

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.

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.

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.

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.

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

A   Proofs

A.1   Proof of Theorem 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

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

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

Lemma 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 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 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.[36] 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 (G, S). That is, we have

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

---

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

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). $\qquad\square$

The statement of the theorem then follows directly from Lemma 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). $\square$

A.2 Proof of Corollary 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, 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, 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 $\phi \in \mathcal{P}_a^{BSE}(G,S)$. $\square$

A.3 Proof of Theorem 2

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

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

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^*.

($\Longleftarrow$) Suppose that $\delta : \mathcal{E} \to \Delta(\mathcal{A})$ 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. $\square$

2. ($\Longleftarrow$) Let $\delta : \mathcal{E} \to \Delta(\mathcal{A})$ 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 $\left(G, S^{private} \right)$.

($\Longrightarrow$) Conversely, let $\delta : \mathcal{E} \to \Delta(\mathcal{A})$ be a REE of $(G, S^{private})$. To show that δ is a REE of $\left(G, S^{complete} \right)$, by Theorem 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

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

$$\Leftrightarrow \varphi \left(a, \varepsilon_i \right) u_i \left(a, \varepsilon_i \right) \geq \varphi \left(a, \varepsilon_i \right) u_i \left(a_i', a_{-i}, \varepsilon_i \right), \quad \forall i, \varepsilon_i, a, a_i'$$

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.4   Proof of Theorem 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

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

Note that

$$\Theta_I^{BSE}(S) \equiv \left\{\theta \in \Theta : \forall x \in \mathcal{X}, \ \phi^x \in \mathcal{P}_a^{BSE}\left(G^{x,\theta}, S^x\right)\right\} \tag{11}$$

and

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

$$= \left\{\theta \in \Theta : \forall x \in \mathcal{X}, \ \phi^x \in \bigcup_{\tilde{S}^x \succsim_E S^x} \mathcal{P}_a^{REE}\left(G^{x,\theta}, \tilde{S}^x\right)\right\}. \tag{12}$$

By Corollary 1, for any given $\theta \in \Theta$ and $x \in \mathcal{X}$, we have

$$\mathcal{P}_a^{BSE}\left(G^{x,\theta}, S^x\right) = \bigcup_{\tilde{S}^x \succsim_E S^x} \mathcal{P}_a^{REE}\left(G^{x,\theta}, \tilde{S}^x\right). \tag{13}$$

That (11) and (12) are equal follows from (13), which is what we wanted. $\square$

A.5 Proof of Theorem 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 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 assump-

tion is placed on the equilibrium selection rule. Since the set of probability distribu-
tions 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.
□

3. The statement follows from Theorem 3. Theorem 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. □

A.6 Proof of Theorem 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 $\Theta_I(\phi) \subseteq \Theta_I(\bar{\phi})$.) Taking the limits on both sides gives the desired result.
□

2. The second statement follows from the fact that ϕ enters the population program (see
Theorem 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

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

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

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

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.7 Proof of Theorem 8

1. First, let use show that (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 (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. $\square$

B Computational Details

B.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 2, 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 2: True Correlation vs. Estimated Correlation ($N_E = 10$)
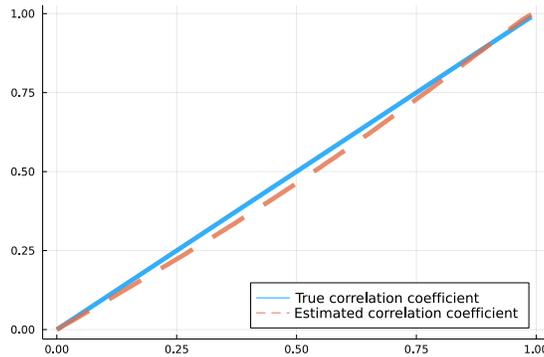
Note that whereas Kennan (2006) shows an "optimal" way of discretizing the support of a univariate random variable, we do not have such optimality result for a multivariate case. Thus, our approach should be understood as being heuristic.

B.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.)
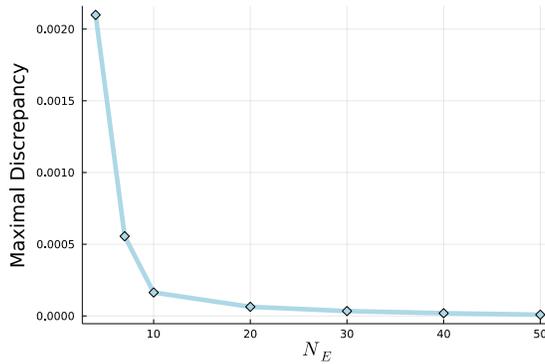
Figure 3: Discrete approximation error

Figure 3 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.

B.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.[37]

Let $\mathcal{X}$ be a finite set of covariates and $|\mathcal{X}|$ its cardinality. Let $\phi_a^x \in \mathbb{R}$ be the population 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

---

[37]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.

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:

$$\Phi_{n^x}^{x,\beta_\alpha} \equiv \left\{ \phi^x : \ \phi_a^x \in \hat{\phi}_a^x \pm \frac{z\left(\beta_\alpha/4\right)}{2\sqrt{n^x}}, \quad \forall a \in \mathcal{A} \right\}, \tag{14}$$

where $z\left(\tau\right) \in \mathbb{R}$ denotes the upper $100\left(1 - \tau\right)\%$ quantile of the standard normal distribution.[38] Note that $\Phi_{n^x}^{x,\beta_\alpha}$ consists of $|\mathcal{A}|$ number of confidence intervals.

Finally, we define a confidence region for ϕ as:

$$\Phi_n^\alpha \equiv \left\{ \phi : \ \phi^x \in \Phi_{n^x}^{x,\beta_\alpha}, \quad \forall x \in \mathcal{X} \right\}. \tag{15}$$

The following proposition states that, under regular conditions, Φ_n^α constructed as (15) has the desired asymptotic coverage probabilities for the population conditional choice probabilities ϕ.

Proposition 1. *Let Φ_n^α be defined as (15). Suppose that samples are independent across*

---

[38]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.

$x \in \mathcal{X}$, 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 $\beta_\alpha \le 0.032$, we have

$$\lim_{n \to \infty} Pr(\phi \in \Phi_n^\alpha) \ge 1 - \alpha.$$

To prove Proposition 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 (14) are used.

Lemma 2 (Fitzpatrick and Scott (1987) Theorem 1). *Let $\Phi_{n^x}^{x,\beta_\alpha}$ be defined as (14). Then*

$$\lim_{n^x \to \infty} Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) \ge \mathcal{L}\left(\beta_\alpha\right)$$

where

$$\mathcal{L}\left(\beta_\alpha\right) = \begin{cases} 1 - \beta_\alpha, & \text{if } \beta_\alpha \le 0.032 \\ 6\Phi\left(\frac{3z(\beta_\alpha/4)}{\sqrt{8}}\right) - 5, & \text{if } 0.032 \le \beta_\alpha \le 0.3 \end{cases}.$$

Now let us prove Proposition 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 (Šidák correction).

Proof. We have

$$\Pr\left(\phi \in \Phi_n^\alpha\right) = \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}, \quad \forall x \in \mathcal{X}\right)$$
$$= \prod_{x \in \mathcal{X}} \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) \tag{16}$$

where (16) follows from the independence across $x \in \mathcal{X}$. Given that β_α is sufficiently small,

taking the limit gives

$$\lim_{n \to \infty} \prod_{x \in \mathcal{X}} \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) = \prod_{x \in \mathcal{X}} \lim_{n^x \to \infty} \Pr\left(\phi^x \in \Phi_{n^x}^{x,\beta_\alpha}\right) \tag{17}$$

$$\geq \prod_{x \in \mathcal{X}} (1 - \beta_\alpha) \tag{18}$$

$$= (1 - \beta_\alpha)^{|\mathcal{X}|}$$

$$= \left(1 - \left\{1 - (1-\alpha)^{1/|\mathcal{X}|}\right\}\right)^{|\mathcal{X}|} \tag{19}$$

$$= 1 - \alpha.$$

where (17) follows from the product rule of limits, (18) follows from Fitzpatrick and Scott (1987) Theorem 1, and (19) follows from the definition of β_α. $\qquad\square$

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 3. 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 $z(\beta_\alpha/4)/(2\sqrt{400}) = 0.0623$ and $z(\beta_\alpha/4)/(2\sqrt{600}) = 0.0509$, our Φ_n^α is defined by the following inequalities:

$$\hat{\phi}_a^l - 0.0623 \leq \phi_a^l \leq \hat{\phi}_a^l + 0.0623, \quad \forall a \in \mathcal{A}$$

$$\hat{\phi}_a^h - 0.0509 \leq \phi_a^h \leq \hat{\phi}_a^h + 0.0509, \quad \forall a \in \mathcal{A}.$$

■

B.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 8: Coverage Probability of Simultaneous Confidence Bands from Simulation

$N_{\mathcal{X}} \backslash n^x$	(A) $\alpha = 0.05$					(B) $\alpha = 0.01$				
	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 8 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.

B.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 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 $\eta \sim N(0, \sigma_\eta^2)$ is a vector of random shocks. We then evaluate $Q(\tilde{\theta}^1)$ 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 $\theta^0, \theta^1, \theta^2, \dots$ 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).

B.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 $h : \mathcal{A} \times \mathcal{E} \to \mathbb{R}$ 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:

$$\min / \max_{\sigma^x \in \Delta_{a|\varepsilon,t}} \sum_{\varepsilon,t,a} \psi_\varepsilon^x \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x h(a, \varepsilon) \quad \text{subject to} \tag{20}$$

$$\sum_{\varepsilon, t_{-i}} \psi_\varepsilon^x \pi_{t|\varepsilon}^x \sigma_{a|\varepsilon,t}^x \partial u_i^{x,\theta}(\tilde{a}_i, a, \varepsilon) \le 0, \quad \forall i, t_i, a, \tilde{a}_i.$$

Note that (20) 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	$h(a, \varepsilon)$
Number of entrants	$1 \times (\mathbb{I}\{a = (0,1)\} + \mathbb{I}\{a = (1,0)\}) + 2 \times \mathbb{I}\{a = (1,1)\}$
McDonald's entry	$\mathbb{I}\{a = (1,0)\} + \mathbb{I}\{a = (1,1)\}$
Burger King entry	$\mathbb{I}\{a = (0,1)\} + \mathbb{I}\{a = (1,1)\}$
No entry	$\mathbb{I}\{a = (0,0)\}$

Suppose θ is given. At a fixed covariate x, we can obtain bounds on the expected value of h by solving (20). 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).

B.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 B.1. We then declare optimization program using `JuMP` interface (Dunning et al., 2017).[39] We construct the simultaneous confidence sets for the conditional choice probabilities using the approach described in B.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

---

[39]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.

scanning algorithm described in Section B.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 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 (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.

Supplementary Materials

C   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.[40] 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.

---

[40]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.[41] The criteria for identifying a census tract as low-income are from the Department of Treasury's New Markets Tax Credit (NMTC) program.

C.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.

---

[41] 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.[42]

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

---

[42]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.

this with the help of `Stata`'s geocoding function, which helps identify census tract id's corresponding 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.[43]

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.

D   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. 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

---

[43]Accessible from https://s4.ad.brown.edu/Projects/Diversity/Researcher/Pooling.htm.

(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 actions frictionlessly, the researcher may want to consider alternative solution concepts as identifying restrictions. We illuminate this point in the examples below.

D.1 Instability of Nash Equilibrium Outcomes

Figure 4: Nash equilibrium in entry game
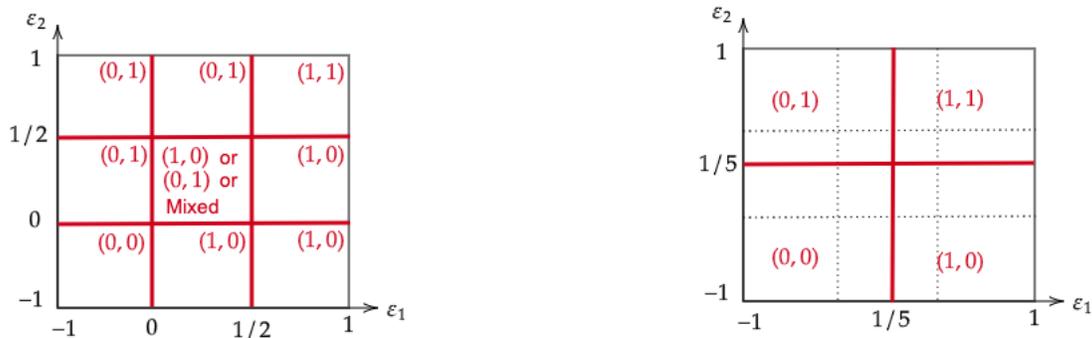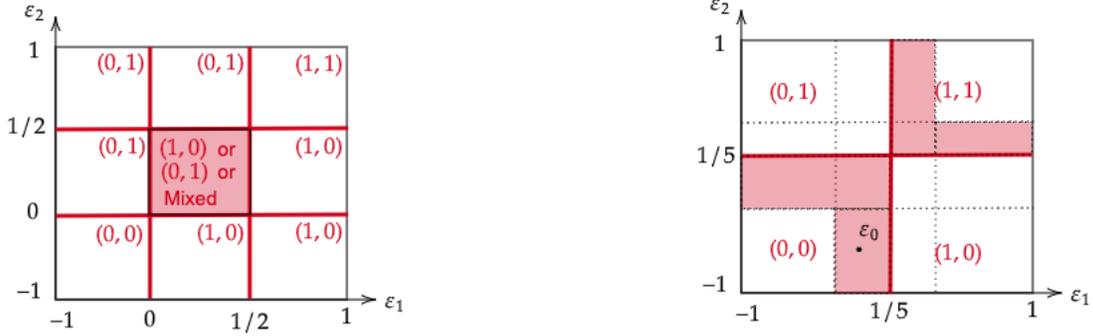

(a) Complete information ($S^{complete}$)   (b) Private information ($S^{private}$)

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 4 summarizes the Nash equilibrium predictions under each informational assumption. In Figure 4-(a), as is well-known, the center region admits multiple equilibria, including a (totally) mixed strategy equilibrium. Figure 4-(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

Figure 5: Unstable region under Nash equilibria

(a) Complete information ($S^{complete}$)

(b) Private information ($S^{private}$)

actions after observing the opponents' realized actions. Figure 5 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 5-(a), regret may arise when either $a = (0,0)$ 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 5-(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.

D.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 6: Bayes stable equilibrium in a simple two-player entry game

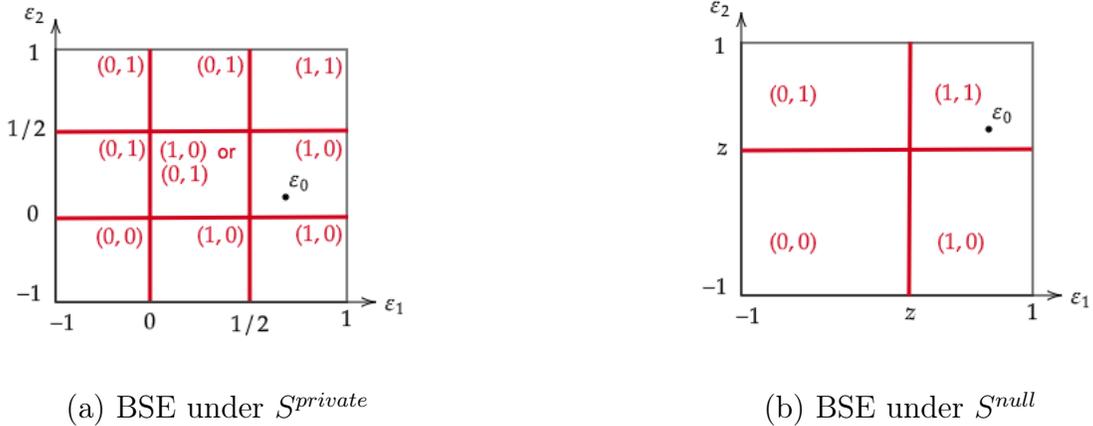(a) BSE under $S^{private}$ 　　　　　　　　　(b) BSE under S^{null}

Figure 6-(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 $a = (1, 0)$.[44]

---

[44]Note the similarity between Figure 6-(a) and 5-(a). Figure 6-(a) also illustrates Theorem 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 $S^{complete}$.

Figure 6-(b) shows an example of a Bayes stable equilibrium under S^{null} in which 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 6-(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 6-(b), any σ with $z \in [0, 1]$ constitutes a BSE under S^{null}.