Estimation of Discrete Games with Weak Assumptions on Information*

Lorenzo Magnolfi[†] and Camilla Roncoroni[‡]

March, 2022

Abstract

We propose a method to estimate static discrete games with weak assumptions on the information available to players. We do not fully specify the information structure of the game, but allow instead for all information structures consistent with players knowing their own payoffs. To make this approach tractable we adopt as a solution concept Bayes Correlated Equilibrium (BCE) (Bergemann and Morris, 2016). We characterize the sharp identified set under BCE and unrestricted equilibrium selection, and find that in simple games with limited variation in covariates identified sets are informative. In an application, we estimate a model of entry in the Italian supermarket industry and quantify the effect of large malls on local supermarkets. Estimates and predictions differ from those obtained under more restrictive assumptions.

Keywords: Estimation of games, Bayes Correlated Equilibrium, entry models, partial identification, supermarket industry

JEL Codes: C57, L10

*We are grateful to Phil Haile, Steve Berry, and Dirk Bergemann for their guidance and support. We thank the editor and three anonymous referees for very useful suggestions that substantially improved the paper. We thank Niccolò Lomys, Ken Hendricks, Tibor Heumann and Dan Quint for numerous suggestions. Many others, to whom we are grateful, provided excellent comments including seminar participants at John's Hopkins, Northwestern, NYU Stern, Penn State, University of Chicago IO lunch, UT-Austin, UW-Madison, Yale, Warwick and at the Eief-IGIER-Bologna IO Workshop, ES NASM 2016, ES ESM 2016, EARIE 2016, XXXI Jornadas de Economía Industrial and AEA 2017. We gratefully acknowledge financial support from the Cowles Foundation. All errors are our own.

[†] *Corresponding Author*: Department of Economics, University of Wisconsin-Madison. `magnolfi@wisc.edu`.

[‡] Department of Economics, University of Warwick. `c.roncoroni@warwick.ac.uk`

1 Introduction

Empirical models of static discrete games are important tools in industrial organization (IO), as they allow researchers to recover the determinants of firms' behavior while accounting for the strategic nature of firms' choices. Models in this class have been applied in contexts such as entry, product or location choice, advertising, and technology adoption.[1] Game-theoretic models' equilibrium predictions, and thus the map between the data and parameters of interest, depend crucially on the information that players have on each other's payoffs. However, the nature of firms' information about their competitors is often ambiguous in applications. Restrictive assumptions, when not satisfied in the application at hand, may result in inconsistent estimates of the payoff structure of the game.

We propose a new method to estimate the distribution of players' payoffs relying only on assumptions on the minimal information players have. In particular, we assume that players know at least their own payoffs, the distribution of opponents payoffs, parameters and observable covariates. We admit any information structure that satisfies these assumptions. Our model is thus incomplete, in the spirit of Manski (2003), Tamer (2003), and Haile and Tamer (2003). More precisely, our model may produce any prediction that results from a Bayes Nash Equilibrium (BNE) under an admissible information structure, without assumptions on equilibrium selection. Our object of interest is the set of parameters that are identified given this incomplete model.

Our method nests the two main approaches in the existing literature: complete information, adopted by the pioneering work in this area (Bjorn and Vuong, 1984; Jovanovic, 1989; Bresnahan and Reiss, 1991a; Berry, 1992), and private information (Seim, 2006; de Paula and Tang, 2012). It also nests the class of information structures considered by Grieco (2014). Moreover, our model is flexible in other dimensions: the information structure of the game may vary across markets and be asymmetric across agents.

To make this approach tractable, we rely on the connection between equilibrium behavior and information, and adopt *Bayes Correlated Equilibrium* (BCE) as solution concept. BCE, introduced by Bergemann and Morris (2013, 2016), has the property of describing BNE predictions for a range of information structures. We show that, for every vector of parameters in the identified set under BCE, there exists an admissible information structure and a BNE that deliver predictions compatible with the data. Exploiting the convexity of the set of equilibria, we also provide a tractable characterization of the sharp identified set of parameters. These results motivate the use of BCE to estimate the distribution of players' payoffs while being agnostic with respect to the information structure.

Weaker assumptions yield weaker identification. We investigate the identification power

[1] See e.g., Bresnahan and Reiss (1991b), Berry (1992), Jia (2008), Ciliberto and Tamer (2009) for entry, Seim (2006) for product choice, Sweeting (2009) for advertising, Ackerberg and Gowrisankaran (2006) for technology adoption.

of BCE in simple entry games with linear payoffs and find that the identified sets are informative about the model's primitives. Point identification is obtained under the assumption of full-support variation in excluded covariates, as in Tamer (2003). More generally, the payoff parameters and of the joint distribution of payoff types are set identified. We perform inference as in Chernozhukov, Hong, and Tamer (2007).

We apply our method to the investigation of the effect of large malls on supermarkets in Italy. The discussion on the impact of these big outlets echoes the US debate on "Wal-Mart effects." Advocates of stricter regulation of large retailers claim that the superstores in malls drive out existing supermarkets, leaving consumers without local stores. Economic theory (e.g., Zhu, Singh, and Dukes, 2011) and evidence from other markets suggest instead that supermarkets may benefit from the agglomeration economies created by the mall, or be differentiated enough not to compete with the mall's grocery anchor.

We estimate a static entry game using our method, and find mixed evidence on the effect of large malls on supermarkets. For all players in the industry the competition from a rival supermarket group seems to have a larger effect on profits than the competition from malls has. This is consistent with a substantial degree of differentiation between malls and local supermarkets, and thus a limited effect of malls on the availability of grocery stores. Our findings are in line with existing studies that have found a limited impact of supercenters on entry by small grocery retailers in the US (Ellickson and Grieco, 2013).

We compare these estimates with those obtained using a model of complete information. Results differ in important ways: high values (in absolute value) of competitive effects are rejected under strong assumptions on information, but not with our method. This is because the assumption of complete information imposes that players fully anticipate competitors' decisions. As a consequence, the more restrictive complete information model may lead to underestimate how much players' profits are affected by the presence of competitors in a market.

In a policy experiment, we evaluate the effect on market structure of removing large malls from markets that currently have no other supermarket. Under weak assumptions on information, we find that the absence of the mall may or may not foster the emergence of a market structure with at least two competing industry players. The model with complete information predicts instead that removing large malls results in a substantial increase in the average maximal probability[2] of observing at least two entrants. In this application, a model with restrictive assumptions on information leads us to strong conclusions, which are dispelled by more robust methods.

This article contributes to the literature on identification and estimation of static dis-

[2]As our model does not yield a unique prediction, we follow Ciliberto and Tamer (2009) in reporting the average across markets and the maximum over equilibrium selections of the probability of market structure outcomes.

crete games, surveyed by de Paula (2013). We follow Tamer (2003) and Berry and Tamer (2006) by not restricting equilibrium selection and allowing for set identification of parameters. We also rely on ideas in Beresteanu, Molchanov, and Molinari (2011), who provide a useful characterization of the sharp identified set for game-theoretic models.

Several existing articles relax the standard assumptions of either complete or perfectly private information. Aguirregabiria and Mira (2007) study a dynamic game which includes public, finite support state variables and private payoff types. Grieco (2014) defines a parametric class of flexible information structures for static games where players receive both public and private signals.[3] We adopt a complementary approach by imposing weaker assumptions on information, but without achieving in general point identification, nor estimating the information structure.

We build on the work of Bergemann and Morris (2013, 2016). They define the equilibrium concept used in this article and describe its robust prediction property.[4] Their theoretical characterization inspires our use of a similarly robust framework in empirical applications. More recently, Bergemann, Brooks, and Morris (2022) show how to perform counterfactual analysis under a *fixed latent information structure.* We use their method in our policy experiment, finding that it can help to obtain sharper predictions.

Aradillas-Lopez and Tamer (2008) study identification under rationalizability. Our approach is neither more general nor more restrictive than theirs, as they relax the equilibrium assumption, but impose restrictive assumptions on information. Yang (2009) estimates payoffs parameters in discrete games of complete information using the non-sharp restrictions imposed by correlated equilibrium to simplify computation. The assumption of correlated equilibrium under complete information is nested in our approach. Beyond discrete games, Bergemann, Brooks, and Morris (2017) characterize BCEs of first-price auctions, Syrgkanis, Tamer, and Ziani (2018) use BCE to perform inference in this class of models, and Gualdani and Sinha (2019) study identification and inference in single-agent discrete choice models using BCE.

Our application is related to studies that model market structure to examine the effect of entry of large store formats - especially Wal-Mart in the US - on other retailers, such as Jia (2008) and Arcidiacono, Bayer, Blevins, and Ellickson (2016). In a related article (Magnolfi and Roncoroni, 2016) we study the role of political connections in shaping market structure in the Italian supermarket industry.

The structure of the article is as follows. In Section 2 we define a class of a discrete games, and in Section 3 we discuss identification in this class of models. In Section 4

[3] Other studies that go beyond complete or perfectly private information include Bajari, Hahn, Hong, and Ridder (2011), Hu and Shum (2013), Igami and Yang (2016), Aguirregabiria and Mira (2019) and Marcoux (2020).

[4] Bergemann and Morris (2013) also discuss identification in the context of games with a continuum of players, symmetric quadratic payoff functions, and normally distributed uncertainty.

we describe how to compute and estimate the identified set. In Section 5 we illustrate the informativeness of our identified set. In Sections 6 and 7 we develop the empirical application and present our policy experiment. Section 8 concludes. All proofs are in Appendix B.

2 Model

We outline the general class of discrete games that we consider in this article, and then develop our leading example: a two-player entry game.

2.1 A General Empirical Discrete Game

We consider a class of static games, indexed by realizations of covariates $x \in X$. Let $N = \{1, ..., n\}$ be the finite set of players; each player $i \in N$ chooses an action y_i from the finite set Y_i. Both the actions' space $Y = \times_{i \in N} Y_i$ and N are the same across different games. We outline the other primitives of the game in the next subsections, describing separately the payoff structure and the information structure. The game is common knowledge among players.

2.1.1 Payoff Structure

Each player i is characterized by a payoff type $\varepsilon_i \in \mathcal{E}_i$.[5] Payoff types $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ are distributed according to the cdf $F(\cdot; \theta_\varepsilon)$, parametrized by the finite dimensional vector $\theta_\varepsilon \in \Theta_\varepsilon$. We also refer to F as players' prior. Payoffs to player i, denoted by π_i, depend on her payoff type and on action profiles. Observable covariates x and finite dimensional payoff parameters $\theta_\pi \in \Theta_\pi$ also affect payoffs. The payoff function π_i, for any pair (x, θ_π), is $\pi_i(\cdot; x, \theta_\pi) : Y \times \mathcal{E}_i \to \mathbb{R}$. A realization of x and a vector of parameters $\theta = (\theta_\pi, \theta_\varepsilon) \in \Theta$ pin down a payoff structure. Throughout the article we assume that ε is independent of x.

Example 1. Consider a game of oligopoly entry as in Bresnahan and Reiss (1991a). Players are firms that can either "Enter" or "Not enter" a market; these actions correspond to $y_i = 1$ and $y_i = 0$, respectively. Researchers observe firms making entry decisions in a cross-section of markets characterized by covariates x. Firms earn a profit of zero by not entering; when entering, firm i's profits are $\pi_i(y, \varepsilon_i; x, \theta_\pi) = \Pi_i(y_{-i}; x, \theta_\pi) + \varepsilon_i$. The additive ε_i represents factors that affect firms' profits or fixed costs and are unobservable to a researcher.

2.1.2 Information Structure

Every player i knows her payoff type ε_i, as well as parameters θ and covariates x, but may not observe ε_{-i}. To model this uncertainty, we assume that in each game with covariates x players receive a private signal $\tau_i^x = (\varepsilon_i, \tilde{\tau}_i^x)$. Hence players observe the realization of their

[5] The payoff type for player i need not be a scalar, although we mostly focus on examples with scalar payoff types.

own ε_i, and an additional random $\tilde{\tau}_i^x$ that may carry information on the opponents' payoff types ε_{-i}. An information structure S^x specifies the set of signals a player may receive and the probability of receiving them. Formally:

$$S^x = \left(T^x, \left\{P^x_{\tilde{\tau}|\varepsilon} :\ \varepsilon \in \mathcal{E}\right\}\right),$$

where $T^x = \mathcal{E} \times \tilde{T}^x$ and $\tilde{T}^x$ is a complete, separable metric space that represents the support of the vector of additional signals $\tilde{\tau}^x = (\tilde{\tau}_1^x, \ldots, \tilde{\tau}_n^x)$. The uncertainty on signal realizations is modeled by the probability kernel $\{P^x_{\tilde{\tau}|\varepsilon} :\ \varepsilon \in \mathcal{E}\}$, which contains the distributions of $\tilde{\tau}^x$ conditional on all realizations of ε. Signals need not have the same marginal distribution for each player, and may be correlated across players.

Sets and distributions of signals depend on x because the information structure may change with covariates. We denote as $S = (S^x)_{x \in X}$ the array that includes information structures for all realizations of x: a generic S provides to players, who know their own ε_i, additional signals $\tilde{\tau}^x$ for each x. The set $\mathcal{S}$ contains all such information structures:

$$\mathcal{S} = \left\{S :\ \forall x \in X,\ T^x{=}\mathcal{E} \times \tilde{T}^x \text{ for } \tilde{T}^x \text{ complete, separable metric space, } P^x_{\tilde{\tau}|\varepsilon} \in \mathcal{P}_{\tilde{T}^x}\right\},$$

where $\mathcal{P}_{\tilde{T}^x}$ is the set of all probability distributions on $\tilde{T}^x$.

Example 2. (Example 1 continued) In Bresnahan and Reiss (1991a) firm i not only observes its own payoff type ε_i, but also observes $\varepsilon_{-i} = (\varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_{i+1}, \ldots, \varepsilon_n)$, the payoff types of rival potential entrants. The information structure is hence *complete information*, denoted by $\overline{S}$. The additional signal space coincides with the opponents' type space, or $\tilde{T}_i^x = \mathcal{E}_{-i}$, and players observe perfectly informative signals: $P^x_{\tilde{\tau}_i|\varepsilon}([\tilde{\tau}_i = \varepsilon_{-i}]) = 1$ for all $x \in X$, $i \in N$.

2.1.3 Equilibrium

The parameter vector θ and the information structure S characterize a game $\Gamma^x(\theta, S)$ for every x. To specify the data-generating process (DGP), linking primitives of the game to outcomes y, we need an equilibrium notion. We describe strategies for player i as functions $\sigma_i : \mathcal{E}_i \times \tilde{T}_i^x \to \mathcal{P}_{Y_i}$, which map payoff types and signals into distributions over actions, and adopt as a solution concept the standard notion of Bayes Nash Equilibrium.

Definition 1. (Bayes Nash Equilibrium) A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ is a Bayes Nash Equilibrium (BNE) of the game $\Gamma^x(\theta, S)$ if for every $i \in N$, $\varepsilon_i \in \mathcal{E}_i$ and $\tilde{\tau}_i \in \tilde{T}_i^x$, whenever for some $y_i \in Y_i$ the corresponding $\sigma_i(y_i \mid \varepsilon_i, \tilde{\tau}_i) > 0$, then

$$E_{\sigma_{-i}}\left[\pi_i\left(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi\right) \mid \varepsilon_i, \tilde{\tau}_i\right] \geq E_{\sigma_{-i}}\left[\pi_i\left(y_i', y_{-i}, \varepsilon_i; x, \theta_\pi\right) \mid \varepsilon_i, \tilde{\tau}_i\right], \quad \forall\, y_i' \in Y_i,$$

where the expectation of y_{-i} is taken with respect to $\sigma_{-i}(y_{-i} \mid \varepsilon_j, \tilde{\tau}_j) = \Pi_{j \neq i} \sigma_j(y_j \mid \varepsilon_j, \tilde{\tau}_j)$.[6]

The information structure has important implications for Bayes Nash equilibrium. When players receive informative signals on their opponents' payoff types, their beliefs and hence their equilibrium actions reflect this information. The more informative the signals that player i receives about ε_{-i}, the more we expect player i's actions to vary with ε_{-i}. Conversely, players who receive uninformative signals only base their actions on their prior. We denote as $BNE^x(\theta, S)$ the set of all BNE strategy profiles for $\Gamma^x(\theta, S)$.

In addition to BNE, we also introduce the notion of Bayes Correlated Equilibrium (BCE), due to Bergemann and Morris (2013, 2016).

Definition 2. (BCE) A Bayes Correlated Equilibrium $\nu \in \mathcal{P}_{Y,\mathcal{E},\tilde{T}}$ for the game $\Gamma^x(\theta, S)$ is a probability measure ν over action profiles, payoff types, and signals that is:

1. *Consistent with the prior*: for all $\varepsilon \in \mathcal{E}$, $\tilde{\tau} \in \tilde{T}$,

$$\sum_{y \in Y} \int_{[t \leq \tilde{\tau}]} \int_{[e \leq \varepsilon]} \nu\left(y, e, t\right) \mathrm{d}t \mathrm{d}e = \int_{[t \leq \tilde{\tau}]} \int_{[e \leq \varepsilon]} P_{\tilde{\tau}|e}(t) \mathrm{d}F\left(e; \theta_\varepsilon\right) \mathrm{d}t;$$

2. *Incentive Compatible*: for all $i, \varepsilon_i, \tilde{\tau}_i, y_i$ such that $\nu\left(y_i \mid \varepsilon_i, \tilde{\tau}_i\right) > 0$,

$$E_\nu\left[\pi_i\left(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi\right) \mid y_i, \varepsilon_i, \tilde{\tau}_i\right] \geq E_\nu\left[\pi_i\left(y_i', y_{-i}, \varepsilon_i; x, \theta_\pi\right) \mid y_i, \varepsilon_i, \tilde{\tau}_i\right], \qquad \forall y_i' \in Y_i,$$

where the expectation $E_\nu[\cdot \mid y_i, \varepsilon_i, \tilde{\tau}_i]$ is taken with respect to $\nu(y_{-i}, \varepsilon_{-i}, \tilde{\tau}_{-i} \mid y_i, \varepsilon_i, \tilde{\tau}_i)$.

BCE is a generalization of Correlated Equilibrium (Aumann, 1974, 1987) to an incomplete information environment. A BCE is defined as a probability measure ν over outcomes, signals and payoff types.[7] This is in contrast to BNE, defined as a strategy profile. We denote as $BCE^x(\theta, S)$ the set of all BCE distributions for $\Gamma^x(\theta, S)$.

The consistency property of BCE requires ν (through its marginal over payoff types) to reflect common knowledge of the prior distribution of ε. The incentive compatibility property may be illustrated with the mediator metaphor: players receive payoff type- and signal-dependent recommendations from an omniscient mediator, and in equilibrium it is optimal for them to follow these recommendations. Whereas the product structure of BNE

[6]The conditional expectation $E_{\sigma_{-i}}$ with respect to the posterior distribution $\sigma_{-i}(y_{-i} \mid \varepsilon_j, \tilde{\tau}_j)$ is:

$$E_{\sigma_{-i}}\left[\pi_i\left(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi\right) \mid \varepsilon_i, \tilde{\tau}_i\right] =$$
$$\sum_{y_{-i} \in Y_{-i}} \int_{\mathcal{E}_{-i}} \int_{\tilde{T}^x_{-i}} \pi_i\left(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi\right) \left(\Pi_{j \neq i} \sigma_j\left(y_j | \varepsilon_j, \tilde{\tau}_j\right)\right) P^x_{\tilde{\tau}_{-i}|\varepsilon}\left(\mathrm{d}\tilde{\tau}_{-i} | \varepsilon_i, \varepsilon_{-i}\right) \mathrm{d}F\left(\varepsilon_{-i} | \varepsilon_i; \theta_\varepsilon\right).$$

[7]Throughout the article we also use the symbol ν for conditional distributions derived from the joint measure. For instance, in the incentive compatibility condition of Definition 2 we use the symbol ν to represent $\nu_{y_{-i}, \varepsilon_{-i}, \tilde{\tau}_{-i} | y_i, \varepsilon_i, \tilde{\tau}_i}$.

implies that correlation in players' actions may only stem from correlation in payoffs or signals, BCE may feature correlation in behavior due to mediator recommendations.

Example 3. (Example 1 continued) In the oligopoly entry game of complete information, if a BNE strategy σ prescribes that firm i enters a market, entry must be optimal given the firm's knowledge of ε and equilibrium expectations, or $E_\sigma[\Pi_i(y_{-i}; x, \theta_\pi) \mid \varepsilon] \geq -\varepsilon_i$. BCEs are instead distributions over $(Y \times \mathcal{E})$ whose marginal over payoff types coincides with the common prior, and such that whenever entry is recommended with a positive probability, or $\nu(y_i = 1 \mid \varepsilon) > 0$, then it must be $E_\nu[\Pi_i(y_{-i}, \varepsilon_i; x, \theta_\pi) \mid y_i = 1, \varepsilon] \geq -\varepsilon_i$ for all i, ε.

2.2 Illustration: the Two-player Entry Game

We introduce here our leading example, also related to the application in Section 6: a two-player entry game that specializes the model of Example 1 to the case of $N = \{1, 2\}$. Outcomes are either a duopoly when $(1, 1)$ is realized, or monopolies when either $(1, 0)$ or $(0, 1)$ are realized, or a market with no entrants with $(0, 0)$. In line with the literature (e.g., Bresnahan and Reiss, 1991a; Tamer, 2003) we let $\pi_i(y, \varepsilon_i; x, \theta_\pi) = y_i(x_i^T\beta_i + \Delta_{-i}y_{-i} + \varepsilon_i)$, so that payoff parameters are $\theta_\pi = (\beta_i, \Delta_i)_{i=1,2}$. The parameter Δ_i, called *competitive effect*, quantifies how entry by firm i affects firm $-i$'s payoffs. Payoffs are:

Player 1 / Player 2:	0	1
0	$(0,\ 0)$	$\left(0,\ x_2^T\beta_2 + \varepsilon_2\right)$
1	$\left(x_1^T\beta_1 + \varepsilon_1,\ 0\right)$	$\begin{pmatrix} x_1^T\beta_1 + \Delta_2 + \varepsilon_1, \\ x_2^T\beta_2 + \Delta_1 + \varepsilon_2 \end{pmatrix}$

A parametric assumption on the joint distribution of payoff types (e.g., iid uniform) completes the payoff structure. In what follows we also use a simplified version of this payoff structure: the *one-parameter entry game* where covariates are suppressed, competitive effects are equal across firms, or $\Delta_1 = \Delta_2 = \Delta$, and payoff types are iid uniform over $[-1, 1]$.

2.2.1 Information Structures and Bayes Nash Equilibrium

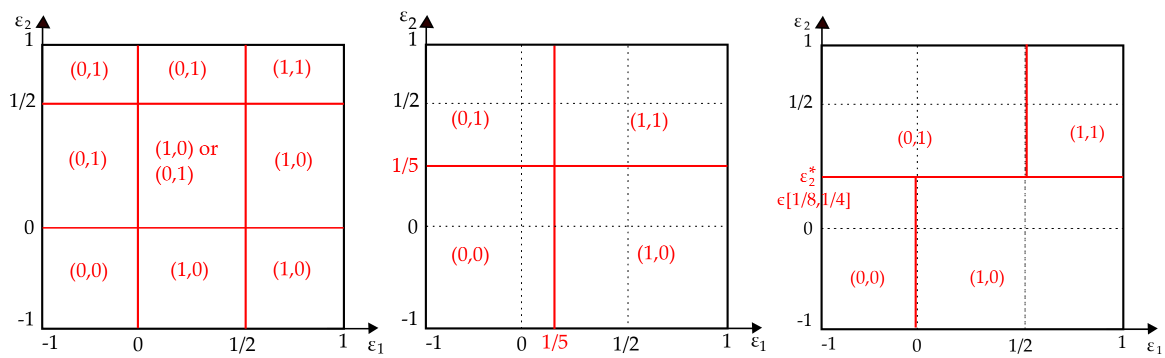
In addition to the *complete information* structure $\bar{S}$ of Example 2, other salient information structures fit our framework. For instance the *perfectly private information* structure, denoted by $\underline{S}$, does not provide to players any information on the realizations of others' payoff types so that players only know their own type. This is because players receive no additional signals: $\tilde{\tau}_i^x = \varepsilon_i$ for all $x \in X$, $i \in N$. This information structure - coupled with the additional assumption of conditional independence across players - is often called *incomplete information* or *independent private types* and widely adopted in the literature on empirical games and models of social interaction (e.g., Brock and Durlauf, 2001; Seim, 2006; Sweeting, 2009).

Both the environments of perfectly private information and complete information are symmetric across players. An alternative information structure still contained in the set $\mathcal{S}$ is *privileged information* S^P, in which only one player knows the type of her opponent. For the informed player i signals are $\tilde{\tau}_i^x = \varepsilon_{-i}$ for all $x \in X$. For the uninformed player j signals are $\tilde{\tau}_j^x = \varepsilon_j$ for all $x \in X$, so that she has no information beyond her payoff type ε_j.

Our model also nests the class of *flexible information* structures S^F. In this setting, the payoff type has two components, or $\varepsilon_i = (\eta_i, \epsilon_i)$, where η_i is public (observed by all players) and ϵ_i is private to player i. Hence additional private signals are $\tilde{\tau}_i = \eta_{-i}$. Examples of this class of information structures include Aguirregabiria and Mira (2007) and Grieco (2014).

Definition 1 characterizes BNE strategy profiles for a game with information structure S. As we vary the S, equilibria vary considerably. We illustrate this point in Figure 1, which depicts equilibrium outcomes in the space of payoff types for the one-parameter entry game with $\Delta = -1/2$. In Panel A we represent equilibrium outcomes for a game of complete information. For every realization of ε, common knowledge for players, there are one or two pure-strategy equilibria. In Panel B, equilibrium for the game of perfectly private information takes the form of threshold strategies: each player does $y_i = 1$ iff $\varepsilon_i \geq 1/5$. In Panel C the privileged information structure produces equilibria where player 1 knows - and can condition her action on - both ε_1 and ε_2. Player 2 only knows ε_2 and follows a threshold strategy. There is a continuum of such equilibria with thresholds $\varepsilon_2^* \in [1/8, 1/4]$.

Figure 1: Information and Equilibrium Outcomes



(A) Complete Information (B) Perfectly Private Information (C) Privileged Information

Note: We represent BNE outcomes in the space $(\varepsilon_1, \varepsilon_2)$ for the one-parameter entry game with payoffs $\pi_i(y, \varepsilon) = y_i(-\frac{1}{2}y_{-i} + \varepsilon_i)$ and $\varepsilon_i \sim U[-1, 1]$. A represents complete information pure-strategy Nash Equilibrium outcomes, B represents perfectly private information outcomes, C represents privileged information outcomes.

2.2.2 Bayes Correlated Equilibrium

We further illustrate the properties of BCE for the one-parameter entry game with $\Delta = -1/2$ and perfectly private information $\underline{S}$. In this case BCE distributions are in the set $\mathcal{P}_{Y\times\mathcal{E}} = \mathcal{P}_{\{0,1\}^2\times[-1,1]^2}$ since signals only contain information on players' own payoff type.

Because ε_i are iid uniform, for each $\varepsilon \in [-1,1]^2$ BCE distributions must satisfy

$$\sum_{y\in\{0,1\}^2}\int_{[e\leq\varepsilon]}\nu\left(y,e\right)\mathrm{d}e = \left(\frac{\varepsilon_1+1}{2}\right)\left(\frac{\varepsilon_2+1}{2}\right)$$

for consistency with the prior. Moreover, in any BCE ν if player i receives the recommendation to enter upon observing ε_i, then $\nu(y_{-i}=1 \mid y_i=1,\varepsilon_i) \leq 2\varepsilon_i$. Conversely player i will stay out if $\nu(y_{-i}=1 \mid y_i=0,\varepsilon_i) \geq 2\varepsilon_i$. Many BCEs satisfy these constraints; consider for instance the distribution ν':

$\varepsilon_1/\varepsilon_2$:	$\leq 1/5$	$> 1/5$
$\leq 1/5$	$\nu'(0,0,\varepsilon) = \frac{9}{25}$	$\nu'(0,1,\varepsilon) = \frac{6}{25}$
$> 1/5$	$\nu'(1,0,\varepsilon) = \frac{6}{25}$	$\nu'(1,1,\varepsilon) = \frac{4}{25}$

Checking consistency is immediate, as this BCE distribution prescribes pure strategies for every ε. Incentive compatibility is also satisfied since $\nu'(y_{-i}=1 \mid y_i=1,\varepsilon_i) = \nu'(y_{-i}=1 \mid y_i=0,\varepsilon_i) = 2/5$. As another example, consider ν'':

$\varepsilon_1/\varepsilon_2$:	$\leq 1/8$	$> 1/8$
≤ 0	$\nu''(0,0,\varepsilon) = \frac{9}{32}$	$\nu''(0,1,\varepsilon) = \frac{7}{32}$
$> 0,\ \leq 1/2$	$\nu''(1,0,\varepsilon) = \frac{9}{64}$	$\nu''(0,1,\varepsilon) = \frac{7}{64}$
$> 1/2$	$\nu''(1,0,\varepsilon) = \frac{9}{64}$	$\nu''(1,1,\varepsilon) = \frac{7}{64}$

The behavior induced by ν' and ν'' corresponds to the outcomes in Panel B and Panel C of Figure 1, respectively, which represent BNE play in the games of perfectly private information and privileged information. The correspondence between BCE distributions and BNEs for different information structures previews a result in the next section: the set of BNE outcomes for any information structure can be generated by BCE play in the perfectly private information game. Our identification strategy builds on this result.

3 Identification

3.1 BNE Predictions and Identified Set

We study in this section the identification of θ_π from cross-sectional data on outcomes y and covariates x. Payoff types ε are unobservable to the researcher. Thus, econometric unobservable and players' payoff types, two conceptually distinct objects, coincide. This is not without loss of generality - we discuss in Section 3.2.3 how this assumption can be relaxed by changing how $\mathcal{S}$ is formed. The setup is summarized in Assumption 1.

Assumption 1. (Observables) *The researcher observes the distribution $P_{x,y}$ of the random vector (x, y). This joint distribution induces a set of conditional probability measures $\{P_{y|x} \in \mathcal{P}_Y : x \in X\}$, where $\mathcal{P}_Y$ is the set of probability distributions over the finite set Y.*

For a game in the class of Section 2, equilibrium strategies $\sigma \in BNE^x(\theta, S)$ result in predictions on observable actions:

Definition 3. (BNE Prediction) A BNE prediction for an equilibrium σ of the game $\Gamma^x(\theta, S)$ is a distribution over outcomes q_σ such that

$$q_\sigma(y) = \int_{\mathcal{E}} \int_T \left(\prod_{i \in N} \sigma_i(\varepsilon_i, \tilde{\tau}_i)(y_i) \right) \mathrm{d}P_{\tau|\varepsilon} \mathrm{d}F, \ \forall y \in Y.$$

Since the game $\Gamma^x(\theta, S)$ may have multiple equilibria and we are agnostic about equilibrium selection, implications of equilibrium are summarized by sets of predictions. Moreover, when data are generated by an *arbitrary equilibrium selection mechanism*, defined as any probability distribution over the set of equilibria $BNE^x(\theta, S)$, the set of predictions is convexified as in Beresteanu et al. (2011).[8] The prediction correspondence $Q^{BNE}_{\theta,S} : X \rightrightarrows \mathcal{P}_Y$ that describes the set of BNE predictions in the game $\Gamma^x(\theta, S)$ is thus

$$Q^{BNE}_{\theta,S}(x) = \mathrm{co}\left[\{q \in \mathcal{P}_Y : \exists\, \sigma \in BNE^x(\theta, S) \text{ such that } q = q_\sigma\}\right],$$

where co[·] denotes the convex hull of a set.

Example 4. Consider the one-parameter entry game with $\Delta = -1/2$ and complete information $S = \overline{S}$ of Figure 1, Panel A. There are three BNE strategies for this game, namely two pure and one mixed-strategy Nash equilibria, and the corresponding set of predictions is

$$Q^{BNE}_{\Delta=-1/2,\overline{S}} = \mathrm{co}\left[\left\{\left(\frac{1}{4}, \frac{3}{8}, \frac{5}{16}, \frac{1}{16}\right), \left(\frac{1}{4}, \frac{5}{16}, \frac{3}{8}, \frac{1}{16}\right), \left(\frac{17}{64}, \frac{21}{64}, \frac{21}{64}, \frac{5}{64}\right)\right\}\right],$$

where vectors q_σ list the probabilities of outcomes $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$.

3.1.1 Data Generating Process and Identified Set

We assume that for the *true payoff and information structure* $(\theta_0, S_0) \in \Theta_0 \times \mathcal{S}$ at least one equilibrium in $BNE^x(\theta_0, S_0)$ exists for every $x \in X$, and that the data are generated by BNE play in games $\Gamma^x(\theta_0, S_0)$. As $S_0 = (S^x_0)_{x \in X}$, the true information structure may be different for each x.[9] The properties of the DGP are summarized by Assumption 2.

[8] Although the assumption that equilibrium selection mechanisms are representable as probability distributions is not fully general (Epstein, Kaido, and Seo, 2016), it is standard in the applied literature.

[9] Given the flexibility in the definition of information structure, the assumption of a unique S^x_0 for each x in the DGP is also hardly restrictive. For example, consider two information structures in $\mathcal{S}$: S_1 with signals $t^x_1 \in T^x_1$ and

Assumption 2. (DGP) *For all $x \in X$, the set $BNE^x(\theta_0, S_0)$ is non-empty and outcomes y are generated by BNE play of the game $\Gamma^x(\theta_0, S_0)$ and an arbitrary equilibrium selection mechanism, so that $P_{y|x} \in Q^{BNE}_{\theta_0,S_0}(x)$.*

Given this link between game-theoretic model and observables, we want to recover θ_0 but we do not know (nor attempt to recover) the true information structure S_0. Under Assumptions 1 and 2, we define the sharp identified set as:

$$\Theta_I^{BNE}(\mathcal{S}) = \left\{\theta \in \Theta : \exists\, S \in \mathcal{S} \text{ such that } P_{y|x} \in Q^{BNE}_{\theta,S}(x)\,,\ P_x - a.s.\right\}.$$

The set $\Theta_I^{BNE}(\mathcal{S})$ captures all the restrictions on parameters implied by assuming that players know *at least* their own payoff shock, or equivalently that S_0 belongs to the set $\mathcal{S}$. Because of how $\mathcal{S}$ is constructed, identification with $\Theta_I^{BNE}(\mathcal{S})$ is general in several dimensions. First, for a given realization of x, the spaces of signals $\tilde{T}^x$ may contain rich structures of non-payoff-relevant signals and generate correlation in players' actions. Second, the information structure S^x may vary in an unrestricted way across different realizations of x. In the rest of the article we refer to $\Theta_I^{BNE}(\mathcal{S})$ as the identified set under *weak assumptions on information.*

3.1.2 Assumptions on Information and Identification

Our strategy, centered on the identification of $\Theta_I^{BNE}(\mathcal{S})$, is in contrast with the prevalent approach in the literature, based instead on further restrictions on the information structure for the DGP. This is done by choosing $S' \in \mathcal{S}$ that yields a tractable set of equilibrium predictions $Q^{BNE}_{\theta,S'}$, and by focusing on

$$\Theta_I^{BNE}(S') = \left\{\theta \in \Theta :\ P_{y|x} \in Q^{BNE}_{\theta,S'}(x)\,,\ P_x - a.s.\right\}.$$

For example, seminal articles (e.g., Bresnahan and Reiss, 1991a; Berry, 1992; Tamer, 2003) assume complete information, or $S' = \overline{S}$. Conversely, other authors (e.g., Sweeting, 2009; de Paula and Tang, 2012) restrict S' to be the perfectly private information structure $\underline{S}$.

Example 5. (Example 4 continued) Suppose we observe data generated by the one-parameter entry game with $\Delta_0 = -1/2$ and $S_0 = \overline{S}$. If we perform identification of the competitive effect Δ under the *true* restriction $S' = \overline{S}$, the probability of duopolies $P_y(1,1) = 1/16$ identifies

$$\Theta_I^{BNE}\left(\overline{S}\right) = \left\{\Delta : \Pr\left\{\varepsilon_i > -\Delta\right\}^2 = P_y(1,1)\right\} = \{-1/2\}.$$

S_2 with signals $t_2^x \in T_2^x$ for all $x \in X$. The *mixed* information structure S_3 with signals $t_3^x = t_1^x$ with probability $p \in (0,1)$, and $t_3^x = t_2^x$ with probability $(1-p)$ is also in $\mathcal{S}$. As we sample from a DGP with information structure S_3, different draws may have signals t_1^x or t_2^x.

Restrictions on information have important consequences. Ideally, the restriction imposed on the information structure S' is *true*, that is $S' = S_0$ as in Example 5. In this case the identified set $\Theta_I^{BNE}(S')$ is nonempty and coincides with $\Theta_I^{BNE}(S_0)$. In typical applications there is, however, little evidence on the nature of S_0. If the restriction on information is such that $S' \neq S_0$, the model is *misspecified.* In this scenario the identified set $\Theta_I^{BNE}(S')$ may not contain the true θ_0, or it may be empty so that the model is *falsified.* Performing estimation in these circumstances leads to inconsistent estimates of θ_0.

3.2 BCE and Robust Identification

The main hurdle for performing identification under weak assumptions on information is to characterize the identified set $\Theta_I^{BNE}(\mathcal{S})$ in a tractable way. To accomplish this, we use the implications for payoff parameters of Bayes Correlated Equilibrium in games with the perfectly private information structure $\underline{S}$.[10] BCE distributions for games $\Gamma^x(\theta, \underline{S})$ are probability measures $\nu \in \mathcal{P}_{Y \times \mathcal{E}}$[11] as in Definition 2; to each ν corresponds a prediction on behavior, obtained as the marginal over players' actions.

Definition 4. (BCE Prediction) The BCE ν induces a prediction $q_\nu(y) = \int_{\mathcal{E}} \nu(y, \varepsilon) \mathrm{d}\varepsilon$.

The observable implications of BCE are thus described by the prediction correspondence

$$Q_\theta^{BCE}(x) = \{q \in \mathbb{P}_Y : \exists\, \nu \in BCE^x(\theta) \text{ such that } q = q_\nu\}.$$

Before proceeding with the identification results, we highlight the assumptions on equilibrium selection embedded in this approach. Assumption 2 restricts the data to be generated by BNE play and an arbitrary equilibrium selection. This assumption, allowing for all distributions over equilibria, results in the convexification of the set $Q_{\theta,S}^{BNE}(x)$. To model predictions corresponding to any equilibrium selection over BCE distributions, further convexification of the set of predictions is not needed. Because the set $BCE^x(\theta)$ is convex, any convex combination of BCE predictions is also a BCE prediction.[12] Hence $Q_\theta^{BCE}(x)$ captures predictions corresponding to any equilibrium selection over BCE distributions. In contrast, $Q_\theta^{BNE}(x)$ requires convexification to correspond to a BNE DGP with arbitrary equilibrium selection mechanism.

3.2.1 Robust Prediction

Bergemann and Morris (2013) establish the robust prediction property of BCE. In our

[10] This choice corresponds to our definition of the information structures that form $\mathcal{S}$; we discuss extensions in Section 3.2.3.

[11] Although in general Definition 2 specifies BCE distributions as $\nu \in \mathcal{P}_{Y,\mathcal{E},\tilde{T}}$, for the special case of perfectly private information where $\tilde{\tau}_i^x = \varepsilon_i$ considering BCE distributions $\nu \in \mathcal{P}_{Y \times \mathcal{E}}$ is without loss.

[12] See Remark 1 in Appendix B for a proof of these statements.

setup, this property translates into the equivalence, for any given θ, between the BCE predictions Q_{θ}^{BCE} and the union of BNE equilibrium predictions $Q_{\theta,S}^{BNE}$ over all $S \in \mathcal{S}$.

Lemma 1. *(Bergemann and Morris, 2013) For all $\theta \in \Theta$ and $x \in X$,*

1. *If $q \in Q_{\theta}^{BCE}(x)$, then $q \in Q_{\theta,S}^{BNE}(x)$ for some $S \in \mathcal{S}$.*
2. *Conversely, for all $S \in \mathcal{S}$, $Q_{\theta,S}^{BNE}(x) \subseteq Q_{\theta}^{BCE}(x)$.*

Part 1 of the lemma states that, given a BCE, we can generate corresponding BNE predictions. This is done by constructing an information structure where signals correspond to BCE mediator recommendations. Conversely, in a BNE players receive signals on their opponents' payoffs and an equilibrium is selected. Adopting the mediator metaphor of BCE, for every payoff type and signal realization, the mediator suggests play according to the BNE strategies selected by the equilibrium selection mechanism. Hence, we can construct a BCE where each player i receives the recommendation to play action y_i if and only if payoff types and equilibrium selection are such that (y_i, y_{-i}) is the outcome of BNE play. Thus BCE predictions for the game with information structure $\underline{S}$ encompass BNE predictions for all games with any information structure $S \in \mathcal{S}$.

Example 6. (Example 4 continued) Figure 2 depicts the set of BCE outcomes for $\Delta_0 = -1/2$. Panel A shows that BCE imposes weaker restrictions on equilibrium behavior than BNE does for a specific information structure: the sets of BNE predictions are subsets of the set of BCE predictions, as stated in Lemma 1. Panel B illustrates instead that BCE predictions are a relatively small subset of all possible outcomes (represented by the simplex).

3.2.2 Robust Identification

We are most interested in the implications of adopting BCE for identification. Under the assumption of BCE, the identified set of parameters in this class of games is

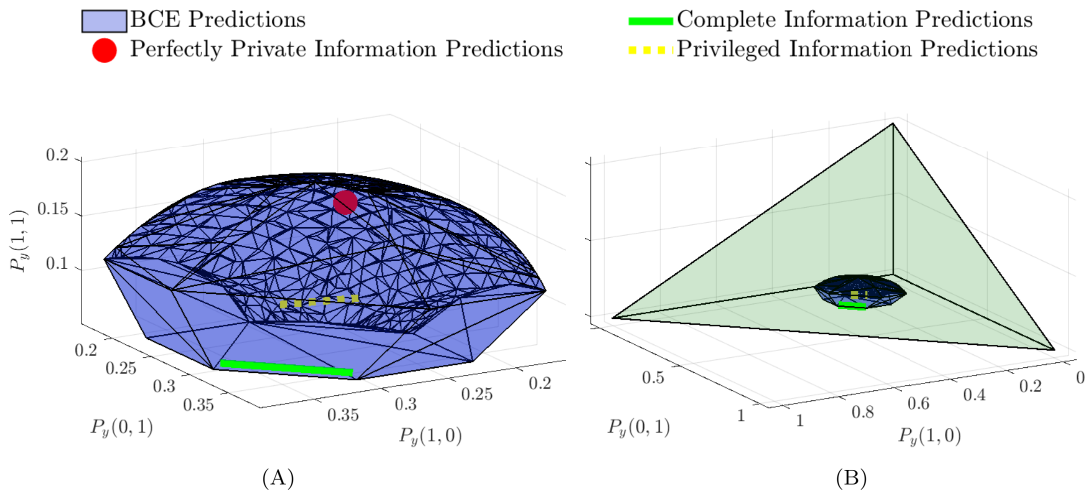
$$\Theta_I^{BCE} = \left\{\theta \in \Theta :\ P_{y|x} \in Q_{\theta}^{BCE}(x)\ P_x - a.s.\right\}. \tag{3.1}$$

Building on the robust prediction property of BCE, we establish the following proposition:

Proposition 1. (Robust identification) *Let Assumptions 1 - 2 hold. Then $\Theta_I^{BCE} = \Theta_I^{BNE}(\mathcal{S})$.*

Proposition 1 translates the robust prediction insight, due to Bergemann and Morris (2013, 2016) and summarized in Lemma 1, into a robust identification result and is the foundation for the use of BCE for identification. Adopting BCE provides an alternative

Figure 2: BCE Predictions



Note: We compare BCE predictions Q_θ^{BCE} with the BNE predictions $Q_{\theta,S}^{BNE}$ obtained under different information structures for the one-parameter entry game with payoffs $\pi_i(y,\varepsilon) = y_i(-1/2y_{-i} + \varepsilon_i)$ and $\varepsilon_i \sim U[-1,1]$. The axes represent probabilities of outcomes P_y. A shows that Q_θ^{BCE} contains the BNE predictions. B shows the set of BCE predictions inside the unit simplex. Computational details are in Appendix C in the Supplementary Materials.

(and, as we discuss in the next section, analytically convenient) characterization of identification under weak assumptions on information. We do not use BCE as an alternative equilibrium assumption on the DGP: our Assumption 2 maintains that data are generated by BNE play. Instead, in light of Proposition 1, we use the BCE identified set Θ_I^{BCE} to recover the identified set under weak assumptions on information $\Theta_I^{BNE}(\mathcal{S})$.

Note an important difference between Lemma 1 and Proposition 1. The set of BCE predictions contains predictions corresponding to *all* BNEs for games with information structures $S \in \mathcal{S}$. However, the identification perspective of Proposition 1 does not imply that for *all* $S \in \mathcal{S}$ there exists some $\theta \in \Theta_I^{BCE}$ that is compatible with the data and with S. Instead, the set Θ_I^{BCE} contains only those parameters for which there exists an information structure and a corresponding BNE that generate predictions matching the data.

3.2.3 Extensions: Alternative Baseline Information Structures

The robust identification result of Proposition 1 can be extended beyond the class of games described in Section 2. For instance, our description of payoff structures in Section 2.1 included restrictions such as independence of payoff types and covariates, and parametric assumptions on the distribution of payoff types. While these are useful to preserve the link with the applied literature, they are not necessary for Proposition 1 to hold.

More importantly, the definitions of a generic information structure S and of the set

$\mathcal{S}$ in Section 2.1 do not impose parametric restrictions on information, but assume that players know *at least* their own payoff type and all other elements of the game such as x. This restriction simplifies exposition, is plausible in our application, and strikes a balance between maintaining weak assumptions and providing identification power in practice. However, other definitions of the set of information structures are possible and may be more appropriate in applications. For instance, we could assume that players do not know elements of x, thus breaking the perfect overlap between the econometric unobservable and players' domain of uncertainty in the incomplete information game.

Proposition 1 extends to models where the set of information structures is defined by first choosing a *baseline* level of information S^*, and then constructing the set $\mathcal{S}(S^*)$ of all information structures such that players receive the signals in S^* and some *additional* signal. More precisely, the set $\Theta_I^{BNE}(\mathcal{S}(S^*))$, which corresponds to BNE play in games with any information structure in $\mathcal{S}(S^*)$, coincides with $\Theta_I^{BCE}(S^*)$, the BCE identified set for the game with information structure S^*. A formal statement of this result is in Appendix F in the Supplementary Materials. Although we use $\underline{S}$ as baseline in most of the article, we discuss in the example below and in Section 5 a model with a different baseline.

Example 7. Consider the one-parameter entry game with $\Delta = -1/2$ and additively separable payoff types $\varepsilon_i = \eta_i + \epsilon_i$. We set as baseline S^* the flexible information structure S^F where ϵ_i is private information and η_i is public information. Thus, player i receives signals $\tau_i = (\varepsilon_i, \eta_{-i}, \tilde{\tau}_i)$ where $\tilde{\tau}_i$ is an additional private signal, which we leave unrestricted. When data are generated by BNE play the identified set is $\Theta_I^{BCE}(S^F)$ as Proposition 1 extends to this setup - see Appendix F for details. This example speaks to many IO applications, where firms know more than the researcher about market conditions that affect their profit.

3.3 Illustration: Assumptions on Information and BCE Identification

We consider again identification of the competitive effect $\Delta \in [-1, 0]$ in the one-parameter entry game, where the data are generated by the payoff structure $\Delta_0 = -1/2$. Restrictive assumptions on information have a substantial impact on identification. For illustration, we consider the non-sharp identified set[13]

$$\tilde{\Theta}_I^{BNE}(S') = \left\{ \Delta \in \Theta \mid \exists\, q \in Q_{\Delta,S'}^{BNE} \text{ such that } q([y = (1,1)]) = P_y(1,1) \right\},$$

obtained by using only the probability of the outcome $(1, 1)$. Table 1 summarizes the identified set $\tilde{\Theta}_I^{BNE}(S')$ under several combinations of S' and S_0.

Table 1 shows that overstating the amount of information available to players leads to

[13]We use in this example the non-sharp identified set based on one moment of P_y – as opposed to the sharp identified set $\Theta_I^{BNE}(\mathcal{S}')$ – to build intuition on the direction of the misspecification bias.

Table 1: Information and Identification

True Information Structure:	$S_0 = \overline{S}$	$S_0 = S^P$	$S_0 = \underline{S}$
Projection of $\tilde{\Theta}_I^{BNE}(S')$:			
$S' = \overline{S}$	$\{-0.50\}$	$\{-0.34\}$	$\{-0.2\}$
$S' = S^P$	$[-0.72, -0.63]$	$[-0.5, -0.44]$	$[-0.29, -0.26]$
$S' = \underline{S}$	$\emptyset$	$\emptyset$	$\{-0.50\}$
Projection of Θ_I^{BCE}	$[-1, -0.39]$	$[-1, -0.26]$	$[-1, -0.28]$

Note: We report the identified sets for the one-parameter entry model with payoffs $\pi_i(y, \varepsilon_i; \Delta) = y_i(\Delta y_{-i} + \varepsilon_i)$ and $\varepsilon_i \sim U[-1, 1]$. The non-sharp identified sets $\tilde{\Theta}_I^{BNE}(S')$ are obtained under assumptions on information S' (corresponding to rows) and true S_0 (corresponding to columns). The true parameter value is $\Delta_0 = -1/2$. Details on the computation of Θ_I^{BCE} are in Appendix A.

an identified parameter that is smaller, in magnitude, than the true parameter value.[14] This is because the probability of a duopoly predicted by the model depends on Δ and on players' degree of certainty that their competitor also enters. With complete information players know that the equilibrium outcome is $(1, 1)$ whenever a duopoly is realized. Hence this model predicts, for a given Δ, the lowest $P_y(1, 1)$ across all information structures. Adopting complete information when data are generated by a model with some level of incomplete information leads thus to understate the magnitude of Δ.

The example illustrates how misspecification of the information structure may result in bias, and this intuition generalizes to more complex models. Estimation of Θ_I^{BCE} avoids bias as this set always contains the true parameter value (in the example, $\Delta_0 = -1/2$). However the example also shows that Θ_I^{BCE} may be large, casting doubt on the informativeness of our approach. We return to this issue in Section 5, after introducing in Section 4 computational and inferential tools that make our approach applicable.

4 Computation and Inference

4.1 Support Function Characterization of the Identified Set

Proposition 1 establishes that Θ_I^{BCE} contains all parameters compatible with the observables and with any information structure in $\mathcal{S}$. To estimate Θ_I^{BCE} we need however a more practical characterization, as it is not obvious how to compute the set from Equation (3.1).

We already noticed that the set $Q_\theta^{BCE}(x)$ is convex for every x and thus it can be represented by its support function as in Beresteanu et al. (2011).[15] Let B denote the closed unit ball in $\mathbb{R}^{|Y|}$ centered at zero and let $h(\cdot; Q_\theta^{BCE}(x)) : B \to \mathbb{R}$ denote the *support function* of $Q_\theta^{BCE}(x)$

$$h\left(b; Q_\theta^{BCE}(x)\right) = \sup_{q \in Q_\theta^{BCE}(x)} b^T q.$$

[14] This bias has already been noted by by Bergemann and Morris (2013) and Aguirregabiria and Magesan (2020).

[15] Because BCE yields a convex set of predictions, we do not need to use Aumann expectations as in Beresteanu et al. (2011). Appendix E describes how our characterization of the identified set maps into their framework.

The support function provides a representation of the set of predictions

$$q \in Q_\theta^{BCE}(x) \Longleftrightarrow b^T q \leq h\left(b; Q_\theta^{BCE}(x)\right) \;\forall\, b \in B.$$

We have then a new characterization of the identified set:

$$\begin{aligned}\Theta_I^{BCE} &= \left\{\theta \in \Theta :\; b^T P_{y|x} \leq h\left(b; Q_\theta^{BCE}(x)\right) \;\forall\, b \in B,\; P_x - a.s.\right\} \\ &= \left\{\theta \in \Theta :\; \max_{b \in B} \min_{q \in Q_\theta^{BCE}(x)} \left[b^T P_{y|x} - b^T q\right] = 0,\; P_x - a.s.\right\}.\end{aligned}$$

The computation of this object can be further simplified: because the inner program is a linear constrained minimization, we can consider its dual maximization program. This step allows us to check whether θ belongs to Θ_I^{BCE} by solving a single constrained maximization problem. Appendix A discusses computational details.

4.2 Inference

Suppose that the researcher observes an iid sample of actions and covariates $\{y_j, x_j\}_{j=1}^n$, where the set of covariates X is discrete and finite.[16] We adopt an extremum estimation approach to perform inference, after Chernozhukov et al. (2007). To this aim, we redefine the identified set as the set of minimizers of a non-negative criterion function G,[17] or $\Theta_I^{BCE} = \{\theta \in \Theta :\; G(\theta) = 0\}$, where $G(\theta) = \int_X \sup_{b \in B}[b^T P_{y|x} - h(b; Q_\theta^{BCE}(x))]\mathrm{d}P_x$. The sample analogue of the population criterion function is

$$G_n(\theta) = n^{-1} \sum_{j=1}^{n} \sup_{b \in B} \left[b^T \hat{P}_{y|x_j} - h\left(b; Q_\theta^{BCE}(\bar{x}_j)\right)\right],$$

where $\hat{P}_{y|x_j}$ is the empirical frequency of y in observations with covariates $x = x_j$.

As long as the payoff function is continuous in payoff parameters, the population criterion function inherits a smoothness property from the upper hemi-continuity of the equilibrium correspondence, and we obtain a consistent estimator of the identified set as in

[16] This assumption is needed to apply the method of Chernozhukov et al. (2007). Although alternative inferential methods do not require discrete covariates (Andrews and Shi, 2013), they are computationally intensive in the context of our model. Other methods (e.g., Andrews and Soares, 2010; Kaido, Molinari, and Stoye, 2019) are designed for models that generate a finite number of (conditional) moment inequalities, and do not immediately apply to our setup.

[17] Since the set of predictions $Q_\theta^{BCE}(x)$ is a subset of the $(|Y|-1)$-dimensional simplex, in our application it is sufficient to adopt the equivalent criterion function

$$\tilde{G}(\theta) = \int_X \sup_{\tilde{b} \in B^{|Y|-1}} \left[\tilde{b}^T \tilde{P}_{y|x} - h\left(\tilde{b}; \tilde{Q}_\theta^{BCE}(x)\right)\right] \mathrm{d}P_x,$$

where $B^{|Y|-1}$ is the $(|Y|-1)$-dimensional closed ball, $\tilde{P}_{y|x}$ is the first $|Y|-1$ elements of $P_{y|x}$ and $\tilde{Q}_\theta^{BCE}(x)$ is the set of the first $|Y|-1$ elements of BCE predictions. With an argument analogous to Theorem B.1 in the Supplemental Materials of Beresteanu et al. (2011) it is immediate to establish that $\{\theta \in \Theta : \tilde{G}(\theta) = 0\} = \{\theta \in \Theta : G(\theta) = 0\}$.

Chernozhukov et al. (2007). We state this result formally in Proposition 2 in Appendix B, where we establish that $\hat{\Theta}_I^{BCE} = \{\theta \in \Theta : nG_n(\theta) \leq a_n\}$ is a consistent estimator of Θ_I^{BCE} for $a_n \to \infty$ and $\frac{a_n}{n} \to \infty$. As in Ciliberto and Tamer (2009), we do not use this estimator directly to perform inference, but rather use it as a basis to construct a confidence set C_n for $\theta \in \Theta_I^{BCE}$. The set C_n, for a confidence level $\alpha = 0.05$, has the coverage property

$$\lim_{n\to\infty} \inf P\{\theta \in C_n\} \geq 1 - \alpha, \ \forall \theta \in \Theta_I^{BCE},$$

Appendix C.4 in Supplementary Materials describes how we compute C_n.

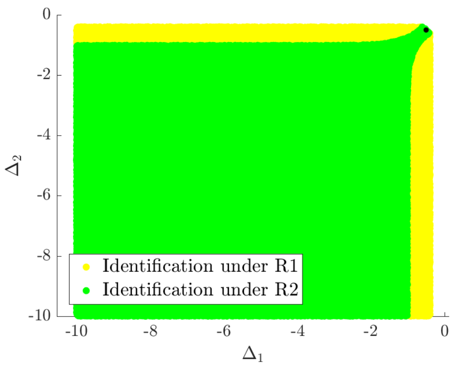
5 Identifying Power of BCE

We discuss in this section the informativeness of the identified set under BCE. When relaxing identifying restrictions there is a trade off between robustness and informativeness of identified sets. This is true not only for assumptions on information, the focus of this article, but more generally. For instance, the assumption of BNE or BCE play could be weakened to non-equilibrium concepts such as k-level rationality (Aradillas-Lopez and Tamer, 2008).

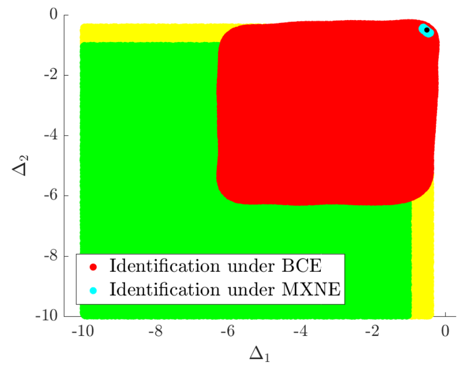
Consider the simple *two-parameter entry game*, a variant of the two-player entry game described in Section 2.2 with no covariates, firm-specific competitive effects Δ_i and iid standard normal payoff types. In Figure 3, Panels A and B we explore how assumptions on equilibrium and information affect the identification of competitive effects for data generated by equilibrium play under complete information. BCE emerges as a compromise between robustness and informativeness. On the one hand, the set Θ_I^{BCE} (in red, Panel B) is larger than the identified set under the (correct) assumption of Nash Equilibrium with complete information (in light blue, Panel B). On the other hand, identified sets obtained under level-1 and level-2 rationality (Panel A) are hardly informative.

Although the exposition in most of the article adopts $\underline{S}$ as baseline, some applications may warrant a different choice. To model this case, we specify payoff types as in Example 7: $\varepsilon_i = \eta_i + \epsilon_i$, where η_i is iid uniform in $\{-1/2, 1/2\}$, and $\epsilon_i \sim N(0,1)$. In addition to $\underline{S}$ we consider another assumption on the baseline information structure: S^F, where ϵ_i is private and η_i is public information. As we assume that the DGP is complete information, setting S^F as baseline is a *correct* restriction on information. In Figure 3, Panel C we represent BCE identified sets for the two-parameter entry game under the two baselines. The baseline S^F results in a smaller identified set (in blue) as compared to $\underline{S}$ (in purple). This suggests that - whenever compatible with the empirical environment - researchers should consider applying our method with baselines that are richer than $\underline{S}$.

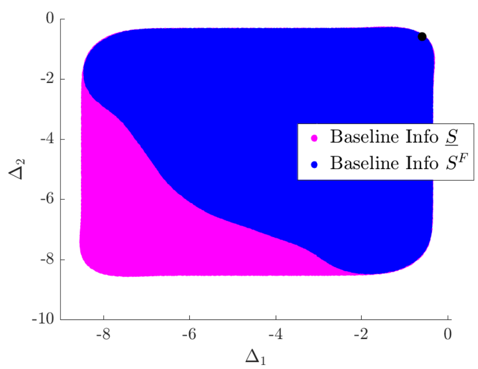
Figure 3: Equilibrium Assumptions, Information and Identification



(A) Identification under R1, R2



(B) Identification under BCE, NE



(C) BCE with $\underline{S}$, S^F baseline

Note: We represent the identified sets for (Δ_1, Δ_2) under different restrictions on behavior in a two-player game with payoffs $\pi_i = y_i(y_{-i}\Delta_i + \varepsilon_i)$. Data are generated by Nash Equilibrium play with complete information, and true parameters $\Delta_1 = \Delta_2 = -1/2$ are represented by black dots. In Panels A and B, $\varepsilon_i \sim N(0,1)$. In Panel C, $\varepsilon_i = \eta_i + \epsilon_i$, where η_i is iid uniform in $\{-1/2, 1/2\}$, and $\epsilon_i \sim N(0,1)$. See Appendices A, C and F for computational details.

5.1 The Role of Covariates: Point and Set Identification

Figure 3 shows that BCE generates tighter identification than non-equilibrium restrictions do, but also that Θ_I^{BCE} may be much larger than $\Theta_I^{BNE}(\overline{S})$. To shrink Θ_I^{BCE} we introduce a key source of identifying power: variation in x. In particular, full-support variation of player-specific covariates yields point identification of payoff parameters θ_π under BCE. The full-support assumption ensures that covariates take values such that players have a dominant strategy for almost all payoff types: identification of payoffs then proceeds as in single-agent discrete choice models. This identification strategy, first proposed for games of complete information under pure Nash Equilibrium play (Tamer, 2003), also applies with weak assumptions on information.[18] In fact, BCE does not allow agents to play dominated actions: this is sufficient for point identification of θ_π. A formal statement of this intuition in a simple setting (two-player entry games with linear index payoffs) is in Proposition 3 in Appendix B.

Although we do not expect the large support assumptions to always hold in applications, the identification at infinity argument points to a source of variation that aids identification even when covariates have finite support. To illustrate the identifying power of BCE in the latter case, we compute Θ_I^{BCE} for a two-player entry game with linear index payoffs. Table 2 reports projections of Θ_I^{BCE} for three DGPs characterized by information structures $S_0 = \bar{S}$, $S_0 = \underline{S}$, and $S_0 = S^P$ respectively,[19] and for two sets of uniformly distributed covariates with finite support, X' and X''. The set $X' = X'_1 \times X'_2 \times X'_C$ is characterized by $X'_i = X'_C = \{-1, 0, 1\}$; the set $X'' = X''_1 \times X''_2 \times X''_C$ is instead characterized by player-specific $X''_i = \{-3, 0, 3\}$ for $i = 1, 2$ and $X''_C = X'_C$.

Table 2: Identification with Finite Support

Panel (A): X'	β^C	β_i	Δ_1	Δ_2	ρ
θ_0	1	1	-1	-1	–
$S_0 = \bar{S}$	$[0.82, 1.17]$	$[0.82, 1.18]$	$[-2.41, -0.87]$	$[-2.37, -0.87]$	–
$S_0 = \underline{S}$	$[0.82, 1.17]$	$[0.82, 1.16]$	$[-1.79, -0.92]$	$[-2.43, -0.87]$	–
$S_0 = S^P$	$[0.83, 1.01]$	$[.83, 1.12]$	$[-2.35, -0.87]$	$[-2.33, -0.87]$	–
Panel (B): X''					
θ_0	1	1	-1	-1	–
$S_0 = \bar{S}$	$[0.96, 1.16]$	$[0.92, 1.03]$	$[-1.16, -0.99]$	$[-1.16, -0.99]$	–
$S_0 = \underline{S}$	$[0.96, 1.15]$	$[0.93, 1.03]$	$[-1.19, -0.98]$	$[-1.16, -0.99]$	–
$S_0 = S^P$	$[0.96, 1.15]$	$[0.92, 1.03]$	$[-1.19, -0.98]$	$[-1.16, -0.99]$	–
Panel (C): X'' and correlated payoff types					
θ_0	1	1	-1	-1	0.80
$S_0 = \bar{S}$	$[0.94, 1.15]$	$[0.82, 1.23]$	$[-1.17, -0.78]$	$[-1.16, -0.83]$	$[0.08, 0.80]$

Note: We report projections of Θ_I^{BCE} for the two-player game with payoffs $\pi_i(y, \varepsilon_i; x, \theta_\pi) = y_i(x_c^T \beta^C + x_i^T \beta_i^E + \Delta_{-i} y_{-i} + \varepsilon_i)$. Payoff types are $\varepsilon_i \sim N(0, 1)$ in A and B, and $\varepsilon \sim N(0, \Sigma)$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ in C. The first row in each panel reports the true θ_0; subsequent rows report projections of Θ_I^{BCE} for different S_0. Computational details are in Appendices A and C.

Results indicate that discrete covariates have identifying power in this model. The identified set, measured by projections for each parameter, shrinks considerably as we increase variation in covariates. In particular, the projection of the identified sets along Δ_1 and Δ_2 shrinks by a factor of about 5 to 7 (depending on the DGP) when the support of covariates is enlarged from X' to X''.

[18] Other articles (e.g., Bajari, Hong, and Ryan, 2010b; Grieco, 2014; Kline, 2015) establish point identification of players' utility functions under *at infinity* variation for games with different sets of assumptions on information, equilibrium selection and parametric restrictions on primitives.

[19] We select an equilibrium when multiple equilibria exist in the DGP. For $S_0 = \overline{S}$ we select with equal probability one of the two pure-strategy equilibria, and for $S_0 = S^P$ we select the equilibrium that maximizes player 2's probability of entry. Selecting different equilibria does not change qualitatively the informativeness of our identified sets.

Whereas full-support covariates generate point identification of θ_π, we do not have a corresponding point identification result for the payoff type parameters θ_ε. To understand intuitively the challenges posed by the identification of θ_ε under BCE, we discuss the equivalent problem of identification under BNE and weak assumptions on information. Some information structures in $\mathcal{S}$ feature correlated signals that may induce highly correlated BNE play despite low correlation in payoff types. If there exist information structures that, for a given correlation in payoff types, can generate any correlation in actions, then the bounds on θ_ε implied by our method would be trivial. Instead, the equilibrium assumption disciplines outcomes even under weak assumptions on information. For instance, in a two-player entry game, rationalizing high values of correlation in payoff types when this correlation is zero in the DGP requires the model to systematically produce duopolies that are not profitable, and thus cannot represent BNE outcomes.

To formalize this argument, we construct in Appendix G non-sharp bounds using a simple implication of equilibrium behavior: observed play cannot correspond to dominated actions. These bounds show that moments of the *joint* distribution of outcomes have significant identifying power with respect to the parameter θ_ε. Although the characterization of Θ_I^{BCE} in Section 4 uses an infinite number of moment inequalities to identify θ_ε and hence is hardly interpretable, these non-sharp bounds reveal that simple moments of the data provide identification power in an intuitive way.

6 Application: Large Malls and Supermarket Entry

In this section we apply our method to quantify the effects of large grocery-anchored malls on local supermarkets in Italy. The sign and intensity of this effect need to be determined empirically. Malls' "anchor" grocery stores may be strong competitors for nearby supermarkets, thus discouraging their presence. Alternatively, format differentiation may results in little competition between local supermarkets and malls' anchor stores, which instead may benefit from demand spillovers. If this is the case, restrictions to entry by malls may be harmful to consumers. To shed light on the research question, we estimate a game-theoretic model where supermarket chains choose whether to operate stores in geographic grocery markets, and malls may affect supermarkets' profits. We model the cross-section of market-structure outcomes as equilibrium of a simultaneous game, following a large literature (Bresnahan and Reiss, 1991b; Berry, 1992; Seim, 2006; Ciliberto and Tamer, 2009).[20]

The empirical methods developed in the previous sections of this article are well suited for this application. The institutional features of the industry offer limited guidance on the

[20] Although dynamic methods are appealing for applications where inter-temporal incentives are of first-order importance, most empirical models of dynamic games require strong assumptions on the nature of information and of unobserved heterogeneity that we want to avoid.

information available to firms, which base their entry decisions on both private and public information. For instance, local authorities impose entry costs that vary across stores and are mostly private information to firms. Moreover, industry players are heterogeneous in their ability to collect and process private information.[21] We also estimate the game under the assumptions of complete information and perfectly private information, and discuss the consequences of using methods that rely on these more restrictive assumptions.

Results from our method indicate a substantial degree of differentiation between the grocery stores in malls and local supermarkets. We do not reject a zero effect of malls on supermarkets, while we reject large negative effects. As a consequence, in the policy experiment of Section 7 we find that a market structure with at least two competing industry players may not be more likely without the mall. Adopting weak assumptions on information is key for this finding. Assuming that data is generated by BNE in the game of perfectly private information or by pure strategy Nash equilibrium in the game of complete information yields negative estimates of the effect of malls on supermarkets, thus predicting an increase of the probability of observing two or more local stores upon removing the mall from small markets.

6.1 Data and Institutional Details

We use data on the cross-section of all supermarkets in Northern and Central Italy in 2013 from the market research firm IRI. We complement these with hand-collected information on malls and mall size, obtained from public online directories. We focus on Northern and Central Italy because the structure of grocery markets in the South differs markedly, with traditional stores and open-air markets playing an important role and fewer large malls. We obtain data on population and demographics from the 2011 official census, and municipality level data on income for 2013 from the Italian Ministry of the Economy.

Defining the relevant markets for our study requires specifying which store formats are direct competitors and the geographical extent of grocery markets. The Italian antitrust authority distinguishes between stores with floor space up to 1,500 m^2 (16,146 ft^2) and stores above this threshold, pointing out that these two categories differ in location, assortments, and applicable regulation (Viviano, Aimone Gigio, Ciapanna, Coin, Colonna, Lagna, and Santioni, 2012). Larger stores have seen the fastest growth in the industry in the period of our study, suggesting that firms and consumers prefer these modern formats. As larger stores are most relevant to welfare outcomes and most likely to compete with the grocery anchors in malls, we consider stores with a floor space of at least 1,500 m^2 (16,146 ft^2)as the relevant market for our study.

No existing administrative unit provides a natural way of defining local grocery markets in Italy. Because commuting patterns capture consumers' daily movements better than

[21] Magnolfi and Roncoroni (2016) explore one of the sources of this heterogeneity: firms' political connections.

administrative units do, we delimit markets starting from the geographic commuting areas defined by ISTAT, the national statistical agency, and split commuting areas that are too large.[22] The geographic extension of these markets is consistent with industry sources and existing studies.[23] We drop from our sample large cities with population greater than 300,000, as the density of urban areas makes it hard to separate markets. This leaves us with 484 grocery markets. Markets with no large malls are systematically smaller, have a lower per capita income, and have on average one supermarket. Summary statistics are available in Table A2 of the Supplementary Materials.

Firms operating in the Italian supermarket industry are heterogeneous. Coop Italia and Conad, organized as cooperatives, have the largest market share. Despite their organizational form, we assume that their entry behavior is profit maximizing.[24] Auchan and Carrefour, two major French retail firms, entered the Italian market mostly in the early 2000s. Additionally, five other Italian firms (Esselunga, Bennet, PAM, Finiper and Selex) operate chains of large stores and have a market share greater than 2.5% in 2013. Given the similarities among supermarket groups with comparable organizational structures, we consider in our analysis three types of players: cooperatives, Italian supermarket groups, and French groups. The industry is subject to extensive regulation, and entry in local markets may be delayed significantly by zoning and other laws. We assume that all players that found profitable to enter a market were able, by 2013, to do so.

We define large malls as shopping centers with at least 50 shops, including a grocery anchor. These anchor supermarkets may receive rent subsidies from mall operators, as they attract consumers that shop at other stores in the mall. Malls' catchment areas are substantially larger than those of supermarkets, attracting shoppers who drive up to 30 minutes from a region that only partially coincides with the local grocery market.

To obtain preliminary evidence on the impact of large malls on grocery markets, we estimate linear and ordered probit regressions. The estimates, reported in Table A2 of the Supplementary Materials, indicate a small and negative covariation between presence of large malls and market structure outcomes such that as the number of supermarkets or the number of industry players operating in a market. However, the market structure that would emerge if malls were not present in a market also depends on the competitive effect that supermarket industry groups have on each other's entry decisions: credible predictions require estimates of these parameters.

[22] We split the commuting area along municipality borders if it contains more than two towns that have population greater than 15,000, and are in a radius of 20 minutes of driving distance.

[23] Evidence collected by various European antitrust authorities indicates that most consumers travel little to do their grocery shopping. For example, UK's Competition Commission considers all large stores in a radius of 10-15 minutes by car to belong to the same market. Pavan, Pozzi, and Rovigatti (2020) use the same Italian commuting areas we use as a basis for market definition in their study of gasoline markets.

[24] This is in line with our findings in Duarte, Magnolfi, and Roncoroni (2021), where we find that Coop's pricing behavior is profit maximizing.

6.2 Game-theoretic Model

To capture strategic interaction among players in the supermarket industry we estimate a static model of entry similar to the example in Section 2.2. Players choose whether to enter each market based on exogenous market characteristics, presence of other players, and firm-market specific characteristics unobserved by the researcher. Payoffs from entry for player i in market m are

$$\pi_i(\cdot; x_m, \theta_\pi) = x_{i,m}^T \beta_i + \sum_{j \neq i} y_{j,m} \Delta_j + \varepsilon_{i,m},$$

whereas payoffs from staying out of the market are normalized to zero.[25]

Market level covariates $x_{i,m}$ include market size (measured as the product of population and logarithm of income),[26] an indicator for the presence of large malls in the market, and a player-specific home-region indicator. The coefficient measuring the effect of market size on profits is constant across players. The coefficients that measure the effect of malls on supermarket players, the home-region indicator and competitive effects Δ_i are instead heterogeneous across players. The vector of payoff types $(\varepsilon_{i,m})_{i \in \mathcal{I}}$ is jointly distributed according to a distribution $F(\varepsilon; \rho)$. We assume that for every i, $\varepsilon_{i,m}$ has a logistic distribution with zero mean and unit scale parameter. The correlation of payoff types is modeled by a normal copula, with correlation ρ between any pair $(\varepsilon_{i,m}, \varepsilon_{j,m})$.

In principle, supermarket groups may enter a market with several stores, or choose different store formats for different markets. For simplicity actions y_i are binary in the model. Moreover, we consider a game with three players, lumping together cooperatives, independent Italian groups and French groups.[27] Hence player i (for example, independent Italian groups) can take a binary action $y_{im} \in \{0, 1\}$ in market m, so that $y_{im} = 1$ corresponds to entry by at least one Italian group with at least one supermarket in market m. These substantial simplifications respond to the need to limit the complexity of the model while maintaining the flexibility necessary to consider a policy experiment.[28]

We also assume that the presence of large malls is exogenous to outcomes in the supermarket industry.[29] This strong assumption is not unrealistic in our environment. Malls have a larger catchment area than supermarkets, as they attract consumers from a region that

[25] A structural interpretation of this payoff specification is discussed in Berry (1989).

[26] This measure of market size captures the decline of the share of income that consumers spend in groceries as their income level increases. Alternative formulations of market size generate conditional correlations between entry outcomes and presence of large malls, and are thus unlikely to drive our results.

[27] As in Ciliberto and Tamer (2009), this assumption is appropriate as long as the firms that are aggregated behave similarly in the markets in our sample.

[28] Existing studies of market structure in retail industries have explored aspects that are absent from our analysis, which provides instead greater flexibility with respect to the information structure. For instance, economies of density (Holmes, 2011) and chain-effects (Jia, 2008) have been found to be important in the US discount retail industry.

[29] Grieco (2014) and Ackerberg and Gowrisankaran (2006) similarly assume exogenous entry for a large player.

only partly overlaps with the local grocery market. Moreover, regulation and the scarcity of available land may push developers to locate malls far from their ideal location.

We estimate the model under weak assumptions on information: by Proposition 1, this is equivalent to estimating Θ_I^{BCE}. This approach nests all the information structures adopted thus far in the empirical literature, and allows for asymmetries in players' information that are relevant for this empirical setting but not compatible with existing models. To compare our method with standard techniques, we also obtain estimates under two more restrictive assumptions: complete information (Ciliberto and Tamer, 2009), and perfectly private information Su (2014). Beyond assumptions on information, these methods require additional restrictions on the payoff structure and on equilibrium selection that are not necessary with our method. In particular, we assume that payoff types are independent across players when performing estimation under perfectly private information.[30]

The discussion in Section 5 guides our intuition on what variation in the data identifies the parameters. Although our model includes a firm specific covariate, the home-region indicator variable, this variable does not have full support thus parameters are set identified. Bounds on β are identified by covariation of market characteristics and entry patterns. Identification of Δ_j stems from the comparison between the probability of entry for firms $-j$ in markets where firm j is unlikely to enter, and the corresponding probability in markets where firm j is likely to enter. The data have some identification power with respect to ρ, the correlation between unobservable payoff types. A high correlation between entry decisions across firms in markets that have different profitability across firms (based on data and other parameters) is particularly informative about the lower bound on ρ. Similarly, low correlation between entry decisions across firms in markets that have uniform observed profitability across firms establishes an upper bound on ρ.

6.3 Estimation Results

Column I in Table 3 presents projections of the estimated 95% confidence set for Θ_I^{BCE}. Constructing the joint confidence set, as opposed to estimating projections directly, results in more conservative inference and is computationally costlier (Bugni, Canay, and Shi, 2017; Kaido et al., 2019), but is necessary to perform policy experiments. We report, for each parameter, its lowest and highest values in the confidence set. We also report values of $\hat{\theta}_0$, the minimizer of G_n. Coefficients' magnitudes are not interpretable, but the policy experiment in the next section illustrates their implications for outcomes.

The evidence on the effect of the presence of large malls on supermarkets is mixed. The effect of malls is not significantly different from zero for any of the players, although the

[30]In line with Ciliberto and Tamer (2009) we restrict equilibrium selection assuming that data are generated by pure-strategy Nash equilibrium. Following Su (2014) we restrict in an important way the payoff structure (by assuming that payoff types are iid) and equilibrium selection (by assuming that a *unique* BNE is played in the data).

Table 3: Confidence Sets

Parameter	Weak Assumptions on Info - BCE (I)	Complete Info - Nash (II)	Perfectly Private Info - BNE (III)
Constant	$[-2.15, -0.21]$ -1.46	$[-3.26, -1.51]$ -2.08	$[-3.52, -3.17]$ -3.33
Market Size	$[3.00, 7.64]$ 3.66	$[3.67, 6.23]$ 4.28	$[2.60, 3.95]$ 3.07
Home-region:			
Cooperatives	$[-0.91, 1.95]$ 0.61	$[-0.21, 1.16]$ 0.64	$[1.37, 1.81]$ 1.60
Italian Groups	$[-0.39, 2.62]$ 1.34	$[-0.14, 1.66]$ 0.97	$[1.54, 2.00]$ 1.72
French Groups	$[-1.46, 1.96]$ 1.10	$[-0.50, 1.15]$ 0.62	$[1.11, 1.58]$ 1.32
Presence of Large Malls:			
Cooperatives	$[-3.26, 1.79]$ 1.35	$[-2.37, 0.45]$ -1.19	$[-2.03, -1.19]$ -1.62
Italian Groups	$[-3.77, 1.49]$ 0.87	$[-2.63, -0.53]$ -1.41	$[-1.47, -0.43]$ -1.05
French Groups	$[-2.94, 1.02]$ -1.04	$[-4.39, -0.19]$ -1.31	$[-1.31, -0.47]$ -0.80
Competitive Effects:			
Cooperatives	$[-5.30, -1.11]$ -2.70	$[-2.40, -0.73]$ -1.76	$[-0.56, 0.47]$ 0.02
Italian Groups	$[-6.11, -1.69]$ -2.46	$[-2.45, -1.34]$ -1.84	$[-1.09, -0.30]$ -0.67
French Groups	$[-7.12, -1.55]$ -5.61	$[-3.46, -0.39]$ -1.49	$[1.73, 3.52]$ 2.88
ρ - Correlation Of Unobservable Profitability	$[0.36, 0.96]$ 0.69	$[0.90, 0.99]$ 0.96	–

Note: We report estimates for the game-theoretic model of Section 6.2 under different assumptions on information. For the models in columns I and II, which are set identified, we report projections of C_n (on top) and $\hat{\theta}_0$, the minimizer of G_n. Estimates in column II are obtained under the additional assumption of pure strategy Nash equilibrium. The model in column III is point identified, and we report point estimates and 95% confidence intervals. Estimates for this model are obtained under the additional assumption that payoff types are iid and that the data are generated by a unique BNE. See Appendices A and C for computational details.

corresponding confidence sets lie mostly on the negative real line. Competitors' presence in a market makes entry less profitable: the confidence set includes large negative competitive effects. Projected confidence sets for ρ are firmly positive, pointing to a substantial correlation among unobserved determinants of supermarkets' profits.

In column II of Table 3 we report the projections of the 95% confidence intervals for the identified set under complete information. For the constant, market size, and home-region parameters the confidence sets corresponding to the weak assumptions on information and complete information models are largely similar. Assuming complete information makes a difference, however, for the effect of large malls and for competitive effects. Although the sign of the effect of malls is not identified under weak assumptions on information, with complete information this effect is negative for two out of three players in the industry.

The importance of assumptions on information is also apparent when we consider the estimates of competitive effects. Under complete information the competitive effects are milder than those obtained with weak assumptions on information. This finding is in line with the discussion in Section 3.3: by assuming complete information we impose that players who enter a market have correct expectations on competitors' presence. Instead, under BCE expectations may incorporate uncertainty about competitors' actions, so that more negative values for the competitive effects are not rejected. The interval for ρ is smaller under complete information and includes only values close to 1. Instead, lower values of ρ are not rejected under weak assumptions on information, as in this model correlation in players' actions can be rationalized as correlation of signals.

In column III of Table 3 we report parameter estimates obtained under the assumption of perfectly private information.[31] This method finds negative and precisely estimated effects of malls, and very weak or positive competitive effects. This latter finding is inconsistent with economic intuition, suggesting that the restriction of perfectly private information - coupled with the additional strong assumptions of a single equilibrium in the data and independent payoff types that are necessary for estimation with standard methods - does not fit the data well. This result echoes previous observations in the literature.[32] For this reason, we focus on the models with weak assumptions on information and with complete information in the next section.

The fact that the confidence set we estimate under restrictive assumption on information is not nested in the confidence set estimated under the weaker BCE assumption deserves some discussion. Proposition 1 establishes that the complete information identified set is a subset of the BCE identified set. However, when going from identification to finite-sample estimates, sampling variation may cause the sets to be non-nested. Additionally, misspecification due to restrictive assumption may lead the complete information model to be falsified, and hence have an empty identified set.[33]

7 Policy Experiment: Removing Malls from Small Markets

To quantify how market structure is affected by the presence of malls, we consider a policy experiment in which regulation prevents malls from locating in small markets. We examine eight small markets that have a mall but no supermarkets in the data, and compute predicted entry outcomes when the mall is removed.

Our policy experiment involves a change in x: we denote x^{pre} as the market-level covariates in the data, and x^{post} as the covariates in the scenario that removes large malls.[34] To

[31] We compute these estimates using the method of Su (2014); in Appendix C we discuss the details of the estimation procedure, and compare it the alternative method in Bajari, Hong, Krainer, and Nekipelov (2010a).

[32] Aguirregabiria and Mira (2007) and Igami and Yang (2016) observe that ignoring correlation among payoff types biases estimates of competitive effects. In Grieco (2014) the perfectly private information model is rejected.

[33] This discussion suggests a possible procedure for rejecting assumptions on information, although the implementation is not straightforward in our inferential setup, and we do not pursue formal testing in this article.

[34] Although our inferential methods require that we use a discretized set of covariates in estimation, in the policy experiment we use for each market m the actual value of x_m^{pre} and the value x_m^{post} where the mall is removed.

perform estimation under weak assumptions on information we assume that the data are generated by BNE in the game $\Gamma^{x^{pre}}(\theta_0, S_0)$ for some unspecified $S_0 \in \mathcal{S}$ (see Assumption 2). Consistent with this approach, the model's predictions are subject to two dimensions of uncertainty. First, as we recover a confidence set C_n for Θ_I^{BCE}, all parameters $\theta \in C_n$ are candidates for generating predictions. Second, for a given θ, predicted outcomes are generated by convex combinations of BNEs in games $\Gamma^{x^{post}}(\theta, S)$, where the information structure may be any $S \in \mathcal{S}$. To capture the multiplicity of information structures and equilibrium selection mechanisms, we rely on Lemma 1 and consider predictions generated by all BCE distributions $\nu \in BCE^{x^{post}}(\theta)$. As there are several ways of constructing predictions, we give formal definitions in the next subsection.

7.1 Model Predictions

We are interested in predicting the probabilities of specific market structures. We define the model's predictions as functions W of equilibrium distributions, parameters and covariates. In particular, the predicted probability of market structure outcomes $\hat{Y} \subset Y$ is

$$W_{\hat{Y}}(\nu, \theta, x) = \sum_{y \in Y} \int 1\left\{y \in \hat{Y}\right\} \nu(y, \varepsilon) \, \mathrm{d}\varepsilon,$$

where the dependence on (θ, x) arises because $\nu \in BCE^x(\theta)$.

We consider several approaches to summarize predictions across equilibrium distributions and parameter values in $\{(\nu, \theta) : \ \nu \in BCE^x(\theta), \ \theta \in C_n\}$. The most conservative prediction intervals have upper (lower) bounds constructed by maximization (minimization) over all equilibrium distributions and parameters in the confidence set:

$$I_W^x = \left[\min_{\{(\nu,\theta):\, \nu \in BCE^x(\theta),\, \theta \in C_n\}} W(\nu, \theta, x), \ \max_{\{(\nu,\theta):\, \nu \in BCE^x(\theta),\, \theta \in C_n\}} W(\nu, \theta, x) \right].$$

When x is set to x^{post}, we denote the corresponding I_W^x as I_W^{post} for ease of notation. To gain further insight into the properties of our method, it is useful to separate the uncertainty in prediction due to the multiplicity of parameters in C_n from the uncertainty due to equilibrium multiplicity. To do so, we fix the value $\hat{\theta}_0$ that minimizes G_n, and denote $I_W^{post}(\hat{\theta}_0)$ as the intervals that correspond to I_W^{post} when C_n is reduced to $\hat{\theta}_0$.[35]

We also define upper bound probabilities for a generic prediction W, denoted as $\overline{W}(\theta, x) = \max_{\nu \in BCE^x(\theta)} W(\nu, \theta, x)$. These are of particular interest when the predicted outcomes are desirable for regulators, who may nudge firms to select equilibria that maximize the proba-

[35] This is subject to an important caveat. In our set-identified model $\hat{\theta}_0$ is not more likely to be the true parameter value than any other $\theta \in C_n$. Hence this exercise is an illustration of the properties of the model, rather than an empirical evaluation of the policy experiment.

bility of such outcomes. Focusing on upper bounds is also in line with important articles in the literature (Ciliberto and Tamer, 2009). Hence we compute for each $\theta \in C_n$ the difference in $\overline{W}$, averaged across markets, as:

$$\Delta\overline{W}(\theta) = \left(\frac{1}{|\hat{X}|} \sum_{m \in \hat{X}} \overline{W}\left(\theta, x_m^{post}\right) - \frac{1}{|\hat{X}|} \sum_{m \in \hat{X}} \overline{W}\left(\theta, x_m^{pre}\right) \right),$$

where $\hat{X}$ denotes the set of markets affected by the policy experiment, and then report bounds across parameter values in the confidence set

$$I_{\overline{W}} = \left[\min_{\{\theta \in C_n\}} \Delta\overline{W}(\theta), \max_{\{\theta \in C_n\}} \Delta\overline{W}(\theta) \right].$$

7.2 Variable and Fixed Latent Information Structure

The predictions defined in the previous subsection do not restrict the latent information structure after the policy experiment to be equal to S_0, the information structure in the DGP. We refer to this approach as *variable latent information structure*. This approach may not fit all empirical settings. For instance, in our application removing malls may not affect the information structure of the game between supermarket chains.

Thus we also pursue a *fixed latent information structure* approach, following Bergemann et al. (2022). This involves computing predicted outcomes compatible with BNEs of $\Gamma^{x^{post}}(\theta_0, S_0)$, keeping S_0 fixed. Bergemann et al. (2022) show that this approach is feasible by considering BCEs of the *linked game*. These equilibrium distributions specify how players choose both y for the game in the data and predicted y', imposing consistency with common knowledge of the primitives and with incentives.

In our setting, the set of BCEs of the linked game $B\tilde{C}E^{x^{pre},x^{post}}(\theta)$ contains augmented equilibrium distributions $\tilde{\nu} \in \Delta(Y \times Y \times \mathcal{E})$. These need to be consistent with the prior, incentive compatible for both factual and predicted actions, and consistent with the observed outcomes. We modify functions W as $\tilde{W}(\tilde{\nu}, \theta, x^{pre}, x^{post})$, so that predictions depend on augmented equilibrium distributions $\tilde{\nu}$. Appendix D.2 in the Supplementary Materials contains formal definitions and an extended discussion of this approach.

For any generic value (x, x'), with latent information kept fixed as in the game with covariates x, predicted intervals for the game with covariates x' are thus

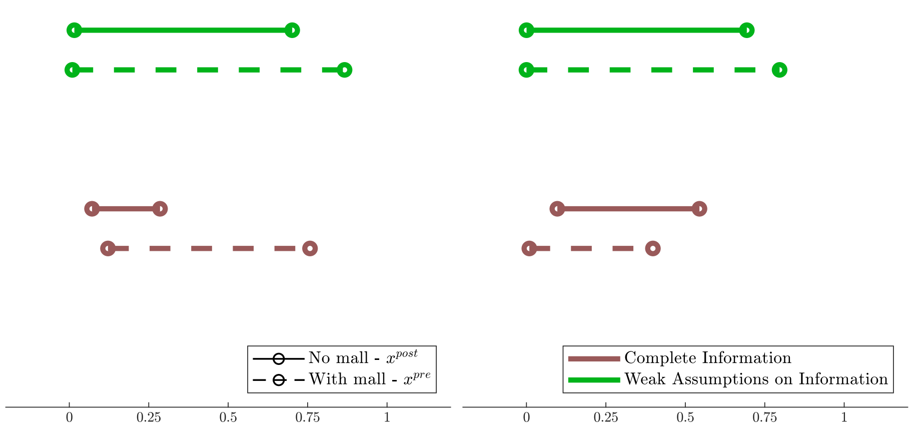
$$\tilde{I}_{\tilde{W}}^{x,x'} = \left[\min_{\left\{(\tilde{\nu},\theta):\, \tilde{\nu} \in B\tilde{C}E^{x,x'}(\theta),\, \theta \in C_n\right\}} \tilde{W}(\tilde{\nu}, \theta, x, x'), \max_{\left\{(\tilde{\nu},\theta):\, \tilde{\nu} \in B\tilde{C}E^{x,x'}(\theta),\, \theta \in C_n\right\}} \tilde{W}(\tilde{\nu}, \theta, x, x') \right].$$

When (x, x') are set to (x^{pre}, x^{post}), we denote the corresponding $\tilde{I}_{\tilde{W}}^{x,x'}$ as $\tilde{I}_{\tilde{W}}^{post}$ for ease of notation. Similarly, predicted intervals for the parameter $\hat{\theta}_0$ are $\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)$.

7.3 Policy Experiment Results

We report in Figure 4 the intervals obtained by averaging across markets the endpoints of I_W^x. Panel A represents probabilities of no entrants, and Panel B represents probabilities of at least two entrants. Intervals for x^{post} (without mall) are solid lines, whereas intervals for x^{pre} (with mall) are dashed lines. The two intervals at the top of the two panels refer to the model with weak assumptions on information, whereas the two intervals at the bottom refer to the complete information model.

Figure 4: Probabilities of Market Structure Outcomes



(A) Average Interval I_W^x for the Probability of No Entrants

(B) Average Interval I_W^x for the Probability of at Least Two Entrants

Note: We represent the intervals obtained by averaging across markets the endpoints of I_W^x for two different outcomes: the probability of no entry (A), and the probability of at least two entrants (B). The two green lines at the top refer to the model with weak assumptions on information. The two red lines at the bottom refer to the model with complete information. In each panel intervals I_W^x are represented as solid line segments for x^{post} and as dashed lines for x^{pre}.

In the model with weak assumptions on information prediction intervals are wide. The effect of removing malls from small markets is ambiguous: the average upper bound probability of observing no entrants decreases, but also the average upper bound probability of observing at least two entrants decreases slightly. Thus, examining intervals I_W^x does not give a definitive answer to our question of interest. [36] Although these results are obtained with the variable latent information approach, changes in intervals for fixed latent information are broadly similar, with an important difference further explored in Section 7.4:

[36]This result is not an artifact of averaging across markets, as it holds for most individual markets, and also holds for other outcomes of interest, such as the expected number of entrants. See Figures A2 and A3 in the Supplementary Materials for more details.

the fixed latent information intervals are smaller than those obtained with variable latent information. In contrast, the complete information model generates a sharper conclusion. The intervals I_W^x obtained under the assumption of complete information indicate that removing a mall from a small market may decrease the probability of no entrants, and may increase the probability of at least two entrants.

We turn in Table 4 to the analysis of the changes in upper bound probabilities $I_{\overline{W}}$. In the table we also report this object for the complete information model, for models with restrictive assumptions on information, and for simple reduced form models.

Table 4: Predicted Change in Probability of Outcomes

Outcome	Weak Assumptions on Info - BCE (I)	Complete Info - Nash (II)	Perf Private Info - BNE (III)	Reduced Form (IV)
No Entry	$[-0.27, 0.24]$	$[-0.56, 0.04]$	$[-0.35, -0.23]$ -0.30	$[-0.23, -0.01]$ -0.10
At least 2 Players	$[-0.24, 0.42]$	$[0.07, 0.40]$	$[0.20, 0.27]$ 0.23	$[0.00, 0.29]$ 0.15
Entry by:				
Cooperatives	$[-0.19, 0.40]$	$[0.04, 0.70]$	$[0.22, 0.33]$ 0.28	$[0.07, 0.37]$ 0.22
Italian Groups	$[-0.15, 0.60]$	$[-0.17, 0.70]$	$[0.15, 0.29]$ 0.23	$[-0.08, 0.28]$ 0.10
French Groups	$[-0.12, 0.56]$	$[-0.16, 0.51]$	$[0.05, 0.14]$ 0.09	$[-0.10, 0.12]$ 0.01

Note: We report in this table predicted changes in upper bound probabilities $I_{\overline{W}}$ for different models. Columns I and II correspond to $I_{\overline{W}}$ for the model with weak assumptions on information and with complete information, respectively. In column III we report the 95% confidence interval and point estimate for the model with perfectly private information. In column IV we report 95% confidence intervals and point estimates for changes in outcome probabilities obtained from simple parametric models. These are ordered probit regressions to predict probabilities of no entry and Entry by at least 2 Players, and a probit regression to predict the binary outcomes Entry by Cooperatives, Entry by Italian Groups, Entry by French Groups.

Predictions on the effect of removing malls as summarized by $I_{\overline{W}}$ are overall inconclusive for the model with weak assumptions on information, in line with the confidence sets and with the evidence in Figure 4.[37] Intervals for entry of at least two players and entry of individual supermarket groups include negative values. This means that the upper bound probabilities of these events may decrease upon removing malls.

Models with more restrictive assumptions on information yield a different conclusion. The model with complete information in column II predicts a decrease of the probability of no entry for most parameter values, and an increase in the probabilities of at least two players operating in a market, or of entry by specific players. The predictions of the game-theoretic model with perfectly private information (column III) and of the reduced form model (column IV) are similar to the predictions of the complete information model.[38] The

[37] Predictions for the fixed latent information structure approach are similar.

[38] The similarity between complete information predictions in column II and reduced form predictions in column IV is not surprising: the ordered probit model we use to predict the probability of no entrants or at least two entrants is analogous to a Bresnahan and Reiss (1991b) specification with homogeneous payoffs and complete information. More generally, reduced form predictions can be interpreted as counterfactual predictions from a game-theoretic model

models in columns III and IV are point identified, so that they yield point predictions of $I_{\overline{W}}$. These are quite precisely estimated and in most cases close to the midpoints of the intervals produced by the complete information model.

7.4 The Empirical Content of BCE: Informativeness of Predictions

A recurring theme in our policy experiments is that the model with weak assumptions on information yields wide prediction intervals. This uncertainty in the model's predictions deserves further investigation: intervals may be wide because of the the size of the confidence set of parameters, or because of equilibrium multiplicity. The former source of uncertainty may be addressed by better variation in the data or by adopting less conservative inferential methods. Equilibrium multiplicity instead is a fundamental feature of the model, and is linked to weak assumptions on information. In fact, in light of Lemma 1, the multiplicity of BCEs reflects the many information structures and equilibrium selection mechanisms that may have generated the data. Anchoring predictions to the *unobserved* information structure in the DGP, as in the fixed latent information approach, may help reduce the uncertainty in prediction stemming from multiple equilibria. Hence, in this subsection we investigate to what extent the uncertainty in prediction is due to equilibrium multiplicity. This is important to clarify the empirical content of BCE.

We first compare intervals I_W^{post} with $I_W^{post}(\hat{\theta}_0)$, the predictions that are obtained fixing the parameter value at $\hat{\theta}_0$, the parameter that minimizes the criterion function G_n. Fixing a parameter value in the identified set removes the uncertainty stemming from partial identification. Table 5, Panel A reports average ratios $|I_W^{post}(\hat{\theta}_0)|/|I_W^{post}|$ and $|\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)|/|\tilde{I}_{\tilde{W}}^{post}|$ that represent the relative width of prediction intervals computed at $\hat{\theta}_0$ with respect to general intervals (incorporating uncertainty due to $\mathcal{C}_n$) for the variable and fixed latent information approaches, respectively. The table shows that there is still considerable uncertainty in prediction in our model even when we fix payoff parameters at $\hat{\theta}_0$. Prediction intervals shrink by about 28 percent on average for the variable latent information approach, and by about 37 percent on average for the fixed latent information approach. Despite the potential of sharper inference and better data to reduce the width of prediction intervals, uncertainty in prediction is an unavoidable cost associated with our approach.

The fixed latent information approach may mitigate this cost. To better understand the role of fixing the latent information structure in our application, we compare prediction intervals $I_W^{post}(\hat{\theta}_0)$ and $\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)$ obtained by fixing the parameters at $\hat{\theta}_0$. We report in Table 5, Panel B the ratio $|\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)|/|I_W^{post}(\hat{\theta}_0)|$, which represents how much the fixed latent information approach shrinks prediction intervals relative to the variable latent information

under the assumption that the equilibrium selection rule in the counterfactual game stays "equivalent up to reduced forms" to that of the games in the data (Canen and Song, 2021).

approach.

Table 5: Relative Size of Prediction Intervals

Panel (A):	Relative Size of I_W^{post} and $I_W^{post}\left(\hat{\theta}_0\right)$			
Variable latent info ratio		Fixed latent info ratio		
No Entrants	At least Two Entrants	No Entrants	At least Two Entrants	
0.74	0.71	0.68	0.58	
Panel (B):	Relative Size of $I_{\tilde{W}}^{post}\left(\hat{\theta}_0\right)$ and $I_W^{post}\left(\hat{\theta}_0\right)$			
No Entrants	At least Two Entrants	Coop Entry	Ita Entry	Fr Entry
0.88	0.78	0.91	0.54	0.92

Note: We report the average (across markets) of ratios $|I_W^{post}(\hat{\theta}_0)|/|I_W^{post}|$ and $|\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)|/|\tilde{I}_{\tilde{W}}^{post}|$ in panel A, and of ratios $|\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)|/|I_W^{post}(\hat{\theta}_0)|$ in panel B. All intervals are computed for the value $\hat{\theta}_0$ which minimizes the empirical criterion function. See Tables A3 and A4 in Supplementary Materials for a market-by-market breakdown.

Overall, the fixed latent information method delivers an average reduction of about 20 percent in the width of prediction intervals. For some outcomes the effect is larger: for the probability of entry of Italian groups, intervals are shrunk by almost 50 percent. The fixed latent information approach of Bergemann et al. (2022) is thus an useful tool to deliver sharper predictions in this class of models.

Three main conclusions emerge from our policy experiment. The first concerns the answer to our empirical question in this application: what is the effect of large malls on supermarkets? Our data - read through the lens of a flexible model - do not dispel the uncertainty. In particular, our results do not rule out that the presence of malls in small markets may have positive spillovers for supermarkets, and caution against any policy that limits the presence of malls.

The second takeaway is that assumptions on information, maintained both in estimation and in prediction with empirical discrete games, have a large effect. The method developed in this article allows researchers to weaken assumptions on information, thus transparently showing what conclusions are robust and which ones are driven by assumptions. In our application, models with either complete information or perfectly private information suggest that removing malls may improve market outcomes in the supermarket industry. This conclusion does not stand when we remove strong restrictions on information.

Finally, we find that the robustness provided by our approach comes at the cost of uncertainty in prediction. This uncertainty, although partly due to the lack of variation in the data of our application, is mostly a by-product of the agnostic approach to information that we maintain in estimating payoff parameters and predicting counterfactual outcomes.

8 Conclusion

We present in this article a method to estimate empirical discrete games under weak assumptions on the information available to players about each other's payoffs. Assumptions on information matter, because the equilibrium predictions implied by different information structures result in parameter estimates that may be biased if the information structure is misspecified. We avoid strong assumptions on information by adopting Bayes Correlated Equilibrium (BCE), defined by Bergemann and Morris (2013, 2016), as solution concept. We argue that BCE is weak enough to make our method robust to assumptions on information, but informative enough to yield useful confidence sets for parameters. In an application, in which we study the effect of large malls on competition among supermarket groups, we show that assumptions on information may drive the results of policy experiments, whereas our method avoids restrictive assumptions.

We conclude with two avenues for future research left open by this article. First, we do not pursue in this article inference on information structures. Although trying to recover an information structure from data on binary outcomes may be too optimistic, richer data like those generated by play in games with continuous actions may allow researchers to identify the information structure of the game that generates the observable outcomes. Second, estimation of games under the BCE solution concept may be interesting beyond providing robustness to assumptions on information - the angle we explored in this article. Results in the theory of learning in games (e.g., Hart and Mas-Colell, 2013) establish regret-minimizing dynamics as a foundation for correlated equilibrium. Based on this insight, in ongoing work Lomys, Magnolfi, and Roncoroni (2021) suggest that BCE captures well outcomes of long-run interaction in incomplete information games, thus providing a connection between dynamic play and a static solution concept.

References

Ackerberg, D. A. and G. Gowrisankaran (2006): "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," *The RAND Journal of Economics*, 37, 738–761.

Aguirregabiria, V. and A. Magesan (2020): "Identification and estimation of dynamic games when players beliefs are not in equilibrium," *The Review of Economic Studies*, 87, 582–625.

Aguirregabiria, V. and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

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

Andrews, D. W. K. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

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

Arcidiacono, P., P. Bayer, J. R. Blevins, and P. B. Ellickson (2016): "Estimation of

Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," *The Review of Economic Studies*, 83, 889–931.

Aumann, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

——— (1987): "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, 1–18.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011): "A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models," *International Economic Review*, 52, 807–824.

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

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

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

Bergemann, D., B. Brooks, and S. Morris (2017): "First-Price Auctions With General Information Structures: Implications for Bidding and Revenue," *Econometrica*, 85, 107–143.

Bergemann, D., B. A. Brooks, and S. Morris (2022): "Counterfactuals with Latent Information," *American Economic Review*, 112, 343–368.

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.

Berry, S. T. (1989): "Entry in the Airline Industry," *Ph.D. Dissertation, Univ. of Wisconsin-Madison*.

——— (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 6, 889–917.

Berry, S. T. and E. Tamer (2006): "Identification in Models of Oligopoly Entry," *Advances in Economics and Econometrics: 9th World Congress of the Econometric Society*, 2, 46–85.

Bjorn, P. A. and Q. H. Vuong (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: a Game Theoretic Formulation with an Application to Labor Force Participation," *Working Paper*.

Bresnahan, T. F. and P. C. Reiss (1991a): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57–81.

——— (1991b): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 977–1009.

Brock, W. A. and S. N. Durlauf (2001): "Discrete choice with social interactions," *The Review of Economic Studies*, 68, 235–260.

Bugni, F. A., I. A. Canay, and X. Shi (2017): "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8, 1–38.

Canen, N. and K. Song (2021): "A Decomposition Approach to Counterfactual Analysis in Game-Theoretic Models," *Working Paper*.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for

Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

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.

de Paula, Á. and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80, 143–172.

Duarte, M., L. Magnolfi, and C. Roncoroni (2021): "The Competitive Conduct of Consumer Cooperatives," *Working Paper*.

Ellickson, P. B. and P. L. E. Grieco (2013): "Wal-Mart and the Geography of Grocery Retailing," *Journal of Urban Economics*, 75, 1–14.

Epstein, L. G., H. Kaido, and K. Seo (2016): "Robust Confidence Regions for Incomplete Models," *Econometrica*, 84, 1799–1838.

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 (2019): "Identification and inference in discrete choice models with imperfect information," *Working Paper*.

Haile, P. A. and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Hart, S. and A. Mas-Colell (2013): *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, EBL-Schweitzer, World Scientific.

Holmes, T. J. (2011): "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 79, 253–302.

Hu, Y. and M. Shum (2013): "Identifying Dynamic Games with Serially-Correlated Unobservables," *Advances in Econometrics*, 31, 97–113.

Igami, M. and N. Yang (2016): "Unobserved Heterogeneity in Dynamic Games: Cannibalization and Preemptive Entry of Hamburger Chains in Canada," *Quantitative Economics*, 7, 483–521.

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

Jovanovic, B. (1989): "Observable Implications of Models with Multiple Equilibria," *Econometrica*, 1431–1437.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence intervals for projections of partially identified parameters," *Econometrica*, 87, 1397–1432.

Kline, B. (2015): "Identification of Complete Information Games," *Journal of Econometrics*, 189, 117–131.

Lomys, N., L. Magnolfi, and C. Roncoroni (2021): "Estimation of Games under No Regret," *Working Paper*.

Magnolfi, L. and C. Roncoroni (2016): "Political Connections and Market Structure," *Working Paper*.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer.

Marcoux, M. (2020): "Games with Unobservable Heterogeneity and Multiple Equilibria: An Application to Mobile Telecommunications," *Working Paper*.

Pavan, G., A. Pozzi, and G. Rovigatti (2020): "Strategic entry and potential competition: Evidence from compressed gas fuel retail," *International Journal of Industrial Organization*, 69, 102566.

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *The RAND Journal of Economics*, 37, 619–640.

Su, C.-L. (2014): "Estimating discrete-choice games of incomplete information: Simple static examples," *Quantitative Marketing and Economics*, 12, 167–207.

Sweeting, A. (2009): "The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis using Multiple Equilibria," *The RAND Journal of Economics*, 40, 710–742.

Syrgkanis, V., E. Tamer, and J. Ziani (2018): "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.

Viviano, E., L. Aimone Gigio, E. Ciapanna, D. Coin, F. Colonna, F. Lagna, and R. Santioni (2012): "La Grande Distribuzione Organizzata e l'Industria Alimentare in Italia," *Occasional Papers, Bank of Italy.*

Yang, Z. (2009): "Correlated Equilibrium and the Estimation of Discrete Games of Complete Information," *Working Paper.*

Zhu, T., V. Singh, and A. Dukes (2011): "Local Competition, Entry, and Agglomeration," *Quantitative Marketing and Economics*, 9, 129–154.

Appendix

A Computation of G and G_n

To find the identified set and perform inference we need to compute the functions G and G_n. In this appendix we describe the steps necessary to compute these functions defined in Section 4.2. At the core of both G and G_n there is the maxmin program

$$\max_{b\in B}\min_{q\in Q_\theta^{BCE}(x)}\left[b^T P_{y|x} - b^T q\right], \qquad (P0)$$

which must be computed for every value of x.

Step 1 - Discretization: To make $(P0)$ feasible we approximate the infinite dimensional object ν by discretizing the set $\mathcal{E} = \times_i \mathcal{E}_i$. Let $\mathcal{E}^r \subset \mathcal{E}$ be the discretized set, with $|\mathcal{E}^r| = r$. We construct $\mathcal{E}^r$ as the product space of $\mathcal{E}_i^r \subset \mathcal{E}_i$, where every set $\mathcal{E}_i^r$ contains $r_i = \frac{r}{|N|}$ equally spaced quantiles of F_{ε_i}.[39] We also define $f^r(\cdot;\theta_\varepsilon)$ as the probability mass function over $\mathcal{E}^r$, where the mass of each $\varepsilon \in \mathcal{E}^r$ is generated by F_{ε_i} and a normal copula with correlation parameter $\rho = \theta_\varepsilon$. The program $(P0)$ is then approximated by the feasible program

$$\begin{array}{rcl}
\max\limits_{b\in\mathbb{R}^{|Y|}}\ \min\limits_{q\in\mathbb{R}^{|Y|},\ \nu\in\mathbb{R}^{|Y|\times r}} & b^T\left(P_{y|x} - q\right) & (P1)\\
s.t. & b^T b - 1 & \leq 0\\
\forall y\in Y & q(y) - \sum_\varepsilon \nu(y,\varepsilon) & = 0\\
\forall \varepsilon\in\mathcal{E}^r & \sum_y \nu(y,\varepsilon) - f^r(\varepsilon;\theta_\varepsilon) & = 0\\
 & \sum_{y,\varepsilon}\nu(y,\varepsilon) - 1 & = 0\\
\forall i, y_i, y_i', \varepsilon_i & \sum_{y_{-i}}\sum_{\varepsilon_{-i}}\nu(y,\varepsilon_i,\varepsilon_{-i})\left(\pi_i(y_i', y_{-i},\varepsilon_i;x,\theta) - \pi_i(y,\varepsilon_i;x,\theta)\right) & \leq 0.
\end{array}$$

Although in $(P0)$ the minimum is taken over $q \in Q_\theta^{BCE}(x)$ only, here we minimize over both a vector of predictions $q \in \mathcal{P}_Y$ and a distribution $\nu \in \mathcal{P}_{Y\times\mathcal{E}^r}$ whose marginal over Y corresponds to q. The restriction that q must be a BCE prediction is now incorporated by imposing that ν must satisfy the constraints of Definition 2.

Step 2 - Vectorization: The discretized ν is a matrix with dimensions $|Y|\times r$; we define $v = \text{vec}(\nu)$, the vectorized ν that stacks the columns of ν in a vector with $d_\nu = |Y|\cdot r$ rows. We then transform $(P1)$ by defining new variables $\tilde{p} = P_{y|x} - q$ and $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \tilde{p} \\ v \end{bmatrix}$. As the set of predictions is a subset of the $(|Y|-1)$-dimensional simplex, we modify the objective

[39] We have experimented with other discretization techniques (e.g., Halton sets, random draws) and have found negligible impact on our results as long as $\mathcal{E}^r$ includes at least some relatively extreme (both positive and negative) payoff types. Including such values of payoff types is important because for them the incentive compatibility constraint of BCE is more likely to be binding.

of the program to $\begin{bmatrix} \tilde{b} \\ 0 \end{bmatrix}^T \left(P_{y|x} - q\right)$, where $\tilde{b}$ is a vector in the $(|Y|-1)-$dimensional closed ball. As argued in footnote 17, this modified objective yields a value of zero if and only if the original program has a value of zero. The transformed program is

$$\begin{array}{rll} \max\limits_{\tilde{b}\in\mathbb{R}^{|Y|-1}} \min\limits_{z_1\in\mathbb{R}^{|Y|}, z_2\in\mathbb{R}_+^{d_\nu}} & \begin{bmatrix} \tilde{b} \\ 0_{d_\nu+1} \end{bmatrix}^T z, & (P2) \\ s.t. & \tilde{b}^T\tilde{b} & \leq 1 \\ & A_{eq}z & = a \\ & A_{ineq}z & \leq 0_{d_{ineq}}, \end{array}$$

where A_{eq} and A_{ineq} are matrices that stack, respectively, linear equality constraints and linear inequalities. These matrices have $d_{eq} = |Y| + r + 1$ and $d_{ineq} = \sum_{i\in N}(|Y_i| \cdot |Y_i - 1| \cdot r_i)$ rows, respectively. The object a is a vector of constants, and we use 0_d, 1_d and I_d to denote, respectively, the $d-$vector of zeros and ones, and the $d \times d$ identity matrix. To construct the matrix A_{eq}, notice that the equality constraints in $(P1)$ can be written as

$$I_{|Y|}\tilde{p} + A_{eq}^1 v = P_{y|x}, \qquad A_{eq}^2 v = f^r(\theta_\varepsilon), \qquad 1_{d_v}^T v = 1,$$

where A_{eq}^1 is a matrix of r copies of a $I_{|Y|}$, or $A_{eq}^1 = 1_r^T \otimes I_{|Y|}$, and A_{eq}^2 is a block-diagonal matrix with r rows and $1_{|Y|}^T$ on the diagonal, or $A_{eq}^2 = I_r \otimes 1_{|Y|}^T$. The $d_{eq} \times d_z$ matrix A_{eq} is then

$$A_{eq} = \begin{bmatrix} I_{|Y|} & A_{eq}^1 \\ 0_{(r\cdot|Y|)} & A_{eq}^2 \\ 0_{|Y|}^T & 1_{d_v}^T \end{bmatrix}$$

with $d_z = |Y| \cdot (r+1)$; a is a $d_{eq}-$vector defined as

$$a = \begin{bmatrix} P_{y|x} \\ f^r(\theta_\varepsilon) \\ 1 \end{bmatrix}.$$

The inequality constraints in $(P1)$ are also linear, so that A_{ineq} can be constructed in a similar way.

Step 3 - Duality and Maximization Program: Although $(P2)$ is in the form of a maxmin problem, it can be transformed into a maximization problem by considering the dual of the inner minimization:

$$\begin{array}{rcl} \max_{\tilde{b}\in\mathbb{R}^{|Y|-1},\lambda_{eq}\in\mathbb{R}^{d_{eq}},\lambda_{ineq}\in\mathbb{R}_{+}^{d_{ineq}}} & -\left[\begin{array}{c} a \\ 0_{d_{ineq}} \end{array}\right]^T \left[\begin{array}{c} \lambda_{eq} \\ \lambda_{ineq} \end{array}\right] & (P3) \\ s.t. & \tilde{b}^T\tilde{b} & \leq 1 \\ & \left(A^T\right)_{1:|Y|} \left[\begin{array}{c} \lambda_{eq} \\ \lambda_{ineq} \end{array}\right] & = -\left[\begin{array}{c} \tilde{b} \\ 0 \end{array}\right] \\ & \left(A^T\right)_{|Y|+1:d_z} \left[\begin{array}{c} \lambda_{eq} \\ \lambda_{ineq} \end{array}\right] & \geq 0_{d_\nu}, \end{array}$$

where A stacks A_{eq} and A_{ineq}, the vectors λ_{eq} and λ_{ineq} are the dual variables associated to the constraints of $(P2)$, $(A^T)_{1:|Y|}$ and $(A^T)_{|Y|+1:d_z}$ denote the first $|Y|$ and the last $r \cdot |Y|$ rows of the matrix A^T. By strong duality and existence of BCE, $(P3)$ has the same value than $(P2)$ and we compute it using the solver KNITRO in AMPL.

Computational Burden: - Due to the tractable nature of $(P3)$, the computational burden of mapping BCE identified sets and confidence intervals is manageable. For example, computation of $G(\theta)$ for the two-player game of Table 2 with $r = 50^2$ takes less than 30 seconds of CPU time on a 3.4Ghz processor. Computation times for the function $G_n(\theta)$ in our application, with $r = 10^3$, are similar. The total time necessary to map the identified set or confidence sets depends - for a fixed dimension of the problem - on the extent to which parallelization is implemented.[40]

Computing time also depends on the dimension of the game (i.e. number of players and number of strategies) and on the discretization adopted. The dimension of the program $(P3)$ - which needs to be solved for every value of x - is determined by the number of variables

$$\begin{aligned} |\left(\tilde{b}^T, \lambda_{eq}^T, \lambda_{ineq}^T\right)| &= |Y| - 1 + d_{eq} + d_{ineq} \\ &= 2|Y| + r\left(1 + \sum_{i\in N}(|Y_i| \cdot |Y_i - 1|)\right), \end{aligned}$$

and the number of equality constraints $|Y|$, and inequality constraints $r \cdot |Y|$. The exact relation between computing time and the scalars r, $|Y|$ and $|N|$ depends on the computing environment, but (as it is common in this class of models) the computational burden grows fast with the dimensions of the game. Further details on how to compute Θ_I and $\mathcal{C}_n$ are in Appendix C in the Supplementary Materials.

[40] Although parallel computation of $G(\theta)$ for different values of θ is not natively supported by AMPL, it can be implemented using the script `Parampl` (Olszak and Karbowski, 2018), available at www.parampl.com. We thank Arthur Olszak for kind and patient support with Parampl.

B Proofs

We first establish as Remark 1 the convexity property of the sets $BCE^x(\theta)$ and $Q_\theta^{BCE}(x)$.

Remark 1. The sets $BCE^x(\theta)$ and $Q_\theta^{BCE}(x)$ are convex for any x and θ.

Proof. Consider the equilibrium distributions ν_1 and ν_2 in $BCE^x(\theta)$. Let $\nu_\alpha = \alpha\nu_1 + (1-\alpha)\nu_2$ for $\alpha \in (0,1)$. Then, ν_α is consistent with the prior game since for all $\varepsilon \in \mathcal{E}$

$$\begin{aligned}\sum_{y\in Y}\int_{[e\leq\varepsilon]}\nu_\alpha(y,e,)\,\mathrm{d}e &= \sum_{y\in Y}\int_{[e\leq\varepsilon]}[\alpha\nu_1(y,e,)+(1-\alpha)\nu_2]\,\mathrm{d}e\\ &= \int_{[e\leq\varepsilon]}\mathrm{d}F(e;\theta_\varepsilon)\,t,\end{aligned}$$

where the second equality follows since $\nu_1, \nu_2 \in BCE^x(\theta)$. Moreover, ν_α is incentive compatible since for all i, ε_i, y such that $\nu(y_i \mid \varepsilon_i) > 0$ we have

$$\begin{aligned}&E_{\nu_\alpha}[\pi_i(y_i,y_{-i},\varepsilon_i;x,\theta_\pi)\mid y_i,\varepsilon_i]\\ &\quad=\alpha E_{\nu_1}[\pi_i(y_i,y_{-i},\varepsilon_i;x,\theta_\pi)\mid y_i,\varepsilon_i]+(1-\alpha)E_{\nu_2}[\pi_i(y_i,y_{-i},\varepsilon_i;x,\theta_\pi)\mid y_i,\varepsilon_i]\\ &\qquad\geq E_\nu[\pi_i(y_i',y_{-i},\varepsilon_i;x,\theta_\pi)\mid y_i,\varepsilon_i], \qquad \forall y_i'\in Y_i,\end{aligned}$$

where the equality follows from the linearity of the expectation operator, and the inequality follows from the incentive compatibility property of ν_1, ν_2. Hence, $BCE^x(\theta)$ is convex.

Consider instead $q_1, q_2 \in Q_\theta^{BCE}(x)$. By construction, there must exist distributions $\nu_1, \nu_2 \in BCE^x(\theta)$ that correspond to the predictions q_1, q_2. For any prediction $q_\alpha = \alpha q_1 + (1-\alpha)q_2$, $\alpha \in (0,1)$ we can then construct a corresponding distribution ν_α such that

$$\begin{aligned}q_\alpha &= \int_{\mathcal{E}}[\alpha\nu_1(y,\varepsilon)+(1-\alpha)\nu_2(y,\varepsilon)]\,\mathrm{d}\varepsilon,\\ &= \int_{\mathcal{E}}\nu_\alpha(y,\varepsilon)\,\mathrm{d}\varepsilon,\end{aligned}$$

and $\nu_\alpha \in BCE^x(\theta)$ since $BCE^x(\theta)$ is convex. Hence, $q_\alpha \in Q_\theta^{BCE}(x)$ and $Q_\theta^{BCE}(x)$ is convex. $\square$

Lemma 1 is a preliminary result needed to prove Proposition 1. In the lemma we restate and adapt to our context the robust prediction property of BCE, established as Theorem 1 in Bergemann and Morris (2016).

Lemma 1. *For all $\theta \in \Theta$ and $x \in X$,*

1. *If $q \in Q_\theta^{BCE}(x)$, then $q \in Q_{\theta,S}^{BNE}(x)$ for some $S \in \mathcal{S}$.*

2. *Conversely, for all $S \in \mathcal{S}$, $Q^{BNE}_{\theta,S}(x) \subseteq Q^{BCE}_{\theta}(x)$.*

Proof. Fix $\theta \in \Theta$ and $x \in X$ throughout.

1. Consider $q \in Q^{BCE}_{\theta}(x)$. Then there exists $\nu \in BCE^x(\theta)$ such that $q = q_\nu$. We need to show that there exists an information structure S and a strategy profile σ such that $q_\sigma = q_\nu$ and $q_\sigma \in Q^{BNE}_{\theta,S}(x)$. To this aim, let $T^x = Y$ and define a probability kernel $\left\{P^x_{\tau|\varepsilon} : \varepsilon \in \mathcal{E}\right\}$[41] such that

$$\int_E \left(P_{\tau|\varepsilon}([\tau = y])\right) \mathrm{d}F = \nu(y, E), \; \forall E \in \mathcal{B}(\mathcal{E}) : \int_E \mathrm{d}F > 0, \; y \in Y.$$

Also, for all ε_i, τ_i, let $\sigma_i(\varepsilon, \tau_i)(y_i) = 1$ if $y_i = \tau_i$, and $\sigma_i(\varepsilon_i, \tau_i)(y_i) = 0$ if $y_i \neq \tau_i$. Hence, the incentive compatibility conditions of BCE guarantee that σ is a BNE of the game $\Gamma^x(\theta, S)$.

2. Suppose that $q = \sum_{k=1}^K \alpha_k q_k \in Q^{BNE}_{\theta,S}(x)$ for $K < \infty$, positive weights α_k such that $\sum_{k=1}^K \alpha_k = 1$ and predictions q_k each corresponding to a $\sigma_k \in BNE^x(\theta, S)$ for all $k = 1, ..., K$. Then, for each σ_k we can obtain $\nu_k \in BCE^x(\theta)$ as

$$\nu_k(y, E) = \int_E \int_T \left(\prod_{i \in N} \sigma_i(\varepsilon_i, \tau_i)(y_i)\right) \mathrm{d}P_{\tau|\varepsilon} \mathrm{d}F,$$

for all $y \in Y$ and $E \in \mathcal{B}(\mathcal{E})$. Hence, $\sum_k \alpha_k \nu_k = \nu \in BCE^x(\theta)$, and the corresponding $q_\nu = q \in Q^{BCE}_{\theta}(x)$. ☐

Proposition 1. (Robust identification) *Let Assumptions 1 - 2 hold. Then*

$$\Theta^{BCE}_I = \Theta^{BNE}_I(\mathcal{S}).$$

Proof. Let $\theta \in \Theta^{BNE}_I(\mathcal{S})$. Then $\exists\; S \in \mathcal{S}$ such that $P_{y|x} \in Q^{BNE}_{\theta,S}(x)$ $P_x - a.s.$ Since, by Lemma 1 again, we have $Q^{BNE}_{\theta,S}(x) \subseteq Q^{BCE}_{\theta}(x)$, $\theta \in \Theta^{BCE}_I$ and $\Theta^{BNE}_I(\mathcal{S}) \subseteq \Theta^{BCE}_I$. Consider instead $\theta \in \Theta^{BCE}_I$; by definition of Θ^{BCE}_I, there must be a collection of $(\nu^x)_{x \in X:}$ such that $p_{\nu^x} \in Q^{BCE}_{\theta}(x)$. It follows that, by Lemma 1, $p_{\nu^x} \in Q^{BNE}_{\theta,S}(x)$ $P_x - a.s.$ for some $S \in \mathcal{S}$. Hence, $\Theta^{BCE}_I \subseteq \Theta^{BNE}_I(\mathcal{S})$. Moreover, by Assumption 2, $P_{y|x} \in Q^{BNE}_{\theta_0,S_0}(x)$ almost surely with respect to P_x. Also, by Lemma 1, $Q^{BNE}_{\theta_0,S_0}(x) \subseteq Q^{BCE}_{\theta_0}(x)$. ☐

Proposition 2. *Assume that:*

1. *The function $\theta_\pi \to \pi_i(y, \varepsilon_i; x, \theta_\pi)$ is continuous for all i, x, y and ε_i, the quantity*

$$|\pi_i(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi) - \pi_i(y'_i, y_{-i}, \varepsilon_i; x, \theta_\pi)|$$

is bounded above, and the function $\theta_\varepsilon \to F(\cdot; \theta_\varepsilon)$ is continuous for all ε;

[41] For the existence of such a kernel, see Chang and Pollard (1997).

2. *The parameter space* Θ *is compact;*

3. *The following uniform convergence condition holds:* $\sup_{\theta\in\Theta}\sqrt{n}|G_n(\theta)-G(\theta)| = O_p(1)$;

4. *For all* $\theta\in\Theta_I$ *we have* $nG_n = O_p(1)$.

Then, the set $\hat{\Theta}_I^{BCE}=\{\theta\in\Theta:\ nG_n(\theta)\le a_n\}$ *is a consistent estimator of* Θ_I^{BCE} *for* $a_n\to\infty$ *and* $\frac{a_n}{n}\to\infty$.

Proof. We want to show that our setup satisfies the condition C.1 in Chernozhukov et al. (2007); the consistency of $\hat{\Theta}_I$ follows by their Theorem 3.1. To this aim, we need to establish that the function $G(\theta)$ is lower semicontinuous.

We start by showing that $\theta\rightrightarrows Q_\theta^{BCE}(x)$ is upper hemi-continuous for all $x\in X$. This correspondence is a compound correspondence between the BCE equilibrium correspondence $\theta\rightrightarrows BCE^x(\theta)$ and the marginal operator $\nu\to\int_{\mathcal{E}}\nu(y,\mathrm{d}\varepsilon)$. The latter is a continuous function mapping into a compact set. For the the equilibrium correspondence: consider a sequence $\theta^k\to\bar{\theta}\in\Theta$, for $\left\{\theta^k\right\}_{k=1}^{\infty}\in\Theta$, and a corresponding sequence $\{\nu_k\}_{k=1}^{\infty}$ such that $\nu_k\in BCE^x(\theta^k)$ for all k, and ν_k converges to $\overline{\nu}$. To see that $\overline{\nu}\in BCE^x(\bar{\theta})$, notice that (i) consistency of $\overline{\nu}$ follows for the continuity of the function $\theta_\varepsilon\to F(\cdot;\theta_\varepsilon)$ and absolute continuity of the marginal operator, and (ii) incentive compatibility of $\overline{\nu}$ results from the continuity of $\theta_\pi\to\pi_i(\cdot;x,\theta_\pi)$ (this can be shown by contradiction, as in Milgrom and Weber, 1985). Therefore the correspondence Q_θ^{BCE} is upper hemi-continuous.

Then, the function

$$\tilde{h}:\theta\to h\left(b;Q_\theta^{BCE}(x)\right)=\sup_{q\in Q_\theta^{BCE}(x)} b^Tq$$

is upper semicontinuous (Lemma 17.30 in Aliprantis and Border, 1994), for all values of x,b. It follows that the function $\theta\to -h\left(b;Q_\theta^{BCE}(x)\right)$ is lower semicontinuous, and so is $\theta\to\sup_{b\in B}\left(b^TP_{y|x}-h\left(b;Q_\theta^{BCE}(x)\right)\right)$, point-wise supremum of a family of lower semicontinuous functions (Proposition 2.41 in Aliprantis and Border, 1994). Hence, the function $G(\theta)$ is lower semicontinuous: for a sequence $\theta_n\to\theta$ in Θ

$$\begin{aligned}
\liminf_{n\to\infty}G(\theta_n) &= \liminf_{n\to\infty}\int\sup_{b\in B}\left[b^TP_{y|x}-h\left(b;Q_{\theta_n}^{BCE}(x)\right)\right]\mathrm{d}P_x\\
&\ge \int\liminf_{n\to\infty}\sup_{b\in B}\left[b^TP_{y|x}-h\left(b;Q_{\theta_n}^{BCE}(x)\right)\right]\mathrm{d}P_x\\
&\ge \int\sup_{b\in B}\left[b^TP_{y|x}-h\left(b;Q_{\theta_n}^{BCE}(x)\right)\right]\mathrm{d}P_x=G(\theta)
\end{aligned}$$

where the first inequality holds by Fatou's Lemma, and the second inequality holds for the lower semi continuity of $\theta \to \sup_{b \in B} \left(b^T P_{y|x} - h\left(b; Q_\theta^{BCE}(x)\right)\right)$. $\square$

Assumption 3. (Two-player entry game with linear payoffs) *Let* $|N| = 2$ *and* $Y = \{0,1\}^2$; *let payoffs be*

$$\pi_i(y, \varepsilon_i; x, \theta_\pi) = y_i \left(x_c^T \beta^C + x_i^T \beta_i^E + \Delta_{-i} y_{-i} + \varepsilon_i\right).$$

Assume moreover:

1. *Vectors of covariates are partitioned as* $x = (x_1, x_2, x_c) \in X_1 \times X_2 \times X_C = X$, *and the distribution* P_x *is such that* x_i *has everywhere positive Lebesgue density conditional on* x_c, x_{-i}, *for* $i = 1, 2$, *and there exists no linear subspace* E *of* $X_i \times X_C$ *such that* $P_x(E) = 1$.

2. *Payoff types* $(\varepsilon_1, \varepsilon_2)$ *are independent of covariates* x, *and distributed according to an absolutely continuous cdf* $F(\cdot; \theta_\varepsilon)$, *defined on* $\mathcal{E} = \mathbb{R}^2$.

Proposition 3. *Suppose the researcher observes the distribution of the data* $\left\{P_{y|x} : x \in X\right\}$, *generated by BCE play of a game. Then, under Assumption 3,*

1. *Payoff parameters* β^C, β^E *and* Δ *are point identified as in single-agent threshold crossing models; and*

2. *The structure implies bounds on the payoff type parameter* θ_ε.

Proof. 1. Consider first the identification of β^C, β_2^E. We want to show that, for appropriate values of x, we have

$$P_{y_2=1|x} = \int_{\left\{\varepsilon_2 : \varepsilon_2 \geq -x_c^T \beta^C - x_2^T \beta_2^E\right\}} \mathrm{d}F_2(\cdot; \theta_\varepsilon), \tag{B.1}$$

where $F_i(\cdot; \theta_\varepsilon)$ is the marginal over ε_i of $F(\cdot; \theta_\varepsilon)$. The model implies the following link between the observables and the structure, for all $x \in X$ and $\nu^x \in BCE^x(\theta)$

$$\begin{aligned} P_{y_2=1|x} &= \nu^x([y_1 = 1, y_2 = 1]) + \nu^x\left(\left[y_1 = 0, y_2 = 1, \varepsilon : \varepsilon_2 < -x_c^T \beta^C - x_2^T \beta_2^E\right]\right) + \\ &+ \nu^x\left(\left[y_1 = 0, y_2 = 1, \varepsilon : \varepsilon_2 \geq -x_c^T \beta^C - x_2^T \beta_2^E\right]\right) \end{aligned}$$

Assume $\beta_{1k}^E > 0$ without loss of generality, and let $x_{1k} \to -\infty$. Conditional on such values of x, $\pi_1(1, y_2, \varepsilon_1; x, \theta_\pi) < 0$ for all values of y_2 $\varepsilon_1 - a.s.$ By the equilibrium optimality condition, $\nu^x(y_1 = 1 \mid y_2, \varepsilon_1) = 0$ whenever $\pi_1(1, y_2, \varepsilon_1; x, \theta_\pi) < 0$. It follows that

$$\lim_{x_{1k} \to -\infty} \nu^x([y_1 = 1, y_2 = 1]) \leq \lim_{x_{1k} \to -\infty} \int_{\mathcal{E}_1} \nu^x([y_1 = 1] | \varepsilon_1) \, \mathrm{d}F_1(\cdot; \theta_\varepsilon) = 0.$$

Moreover, $\lim_{x_{1k}\to-\infty} \nu^x\left(\left[y_1=0, y_2=1, \varepsilon : \varepsilon_2 < -x_c^T\beta^C - x_2^T\beta_2^E\right]\right) = 0$, as in the limit $\varepsilon_2 < -x_c^T\beta^C - x_2^T\beta_2^E$ implies $y_2 = 0$. For a similar application of the incentive compatibility property of BCE,

$$\nu^x\left(\left[y_1=0, y_2=1, \varepsilon : \varepsilon_2 \geq -x_c^T\beta^C - x_2^T\beta_2^E\right]\right) = \int_{\left\{\varepsilon_2 : \varepsilon_2 \geq -x_c^T\beta^C - x_2^T\beta_2^E\right\}} \mathrm{d}F_2\left(\cdot;\theta_\varepsilon\right).$$

The result in Equation (B.1) follows; this equation describes a single-agent threshold crossing model: under Assumption 3, $\left(\beta^C, \beta_2^E\right)$ and F_i are point identified (Manski, 1988). Player 1's parameter β_1 is identified by a symmetric argument. To prove identification of Δ parameters, consider instead $x_{1k}\to\infty$; the same steps lead to

$$\lim_{x_{1k}\to\infty} P_{y_2=1|x} = \int_{\left\{\varepsilon_2 : \varepsilon_2 \geq -x_c^T\beta^C - x_2^T\beta_2^E - \Delta_1\right\}} \mathrm{d}F_2\left(\cdot;\theta_\varepsilon\right).$$

2. Let $\theta_\pi = (\beta, \Delta)$ be identified. We can derive (non-sharp) bounds on the distribution of observable outcomes, and thus on the joint distribution of payoff types $F\left(\varepsilon;\theta_\varepsilon\right)$. To construct lower bounds on the probabilities of outcomes, we need to define regions where such outcomes are the product of dominant strategies. For instance, let

$$\underline{\mathcal{E}}^{(1,1)}\left(x,\theta\right) = \left\{\varepsilon_1 \geq -x_c^T\beta^C - x_1^T\beta_1^E - \Delta_2, \varepsilon_2 \geq -x_c^T\beta^C - x_2^T\beta_2^E - \Delta_1\right\}.$$

For any $x \in X$, Definition 2 implies that for $\varepsilon \in \underline{\mathcal{E}}^{(1,1)}$ we have $\nu^x([y=(1,1)] \mid \varepsilon) = 1$ for every $\nu^x \in BCE^x(\theta)$. We can similarly define a region $\underline{\mathcal{E}}^y(x,\theta)$ for any action profile y.

For each y, we can also construct upper bounds by defining regions where for any i, no y_i is dominated. Hence, let

$$\overline{\mathcal{E}}^y\left(x,\theta\right) = \left\{\varepsilon : \max_{\nu^x\in BCE^x(\theta)} \nu^x\left(y \mid \varepsilon\right) > 0\right\};$$

for the outcome $y = (1,1)$ for instance,

$$\overline{\mathcal{E}}^{(1,1)}\left(x,\theta\right) = \left\{\varepsilon_1 \geq -x_c^T\beta^C - x_1^T\beta_1^E, \varepsilon_2 \geq -x_c^T\beta^C - x_2^T\beta_2^E\right\}.$$

We can then construct the bounds

$$LB_y\left(\theta_\varepsilon;x\right) = \int_{\underline{\mathcal{E}}^y} \mathrm{d}F\left(\cdot;\theta_\varepsilon\right) \leq P_{y|x} \leq \int_{\overline{\mathcal{E}}^y(x,\theta)} \mathrm{d}F\left(\cdot;\theta_\varepsilon\right) = UB_y\left(\theta_\varepsilon;x\right),$$

and define the set $BD(y) = \{\theta_\varepsilon : LB_y(\theta_\varepsilon;x) \leq P_{y|x} \leq UB_y(\theta_\varepsilon;x),\ a.e.-x\}$. Variation in x shifts the regions $\underline{\mathcal{E}}$ and $\overline{\mathcal{E}}$, thus providing useful restrictions on θ_ε and shrinking the set of parameters $\cap_{y\in Y} BD(y)$ that are compatible with the bounds. $\square$

Appendix References

Aliprantis, C. D. and K. Border (1994): *Infinite Dimensional Analysis*, Springer-Verlag.

Bergemann, D. and S. Morris (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

Chang, J. T. and D. Pollard (1997): "Conditioning as Disintegration," *Statistica Neerlandica*, 51, 287–317.

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

Manski, C. F. (1988): "Identification of Binary Response Models," *Journal of the American statistical Association*, 83, 729–738.

Milgrom, P. R. and R. J. Weber (1985): "Distributional strategies for games with incomplete information," *Mathematics of operations research*, 10, 619–632.

Olszak, A. and A. Karbowski (2018): "Parampl: A Simple Tool for Parallel and Distributed Execution of AMPL Programs," *IEEE Access*, 6, 49282–49291.

Supplementary Materials - For Online Publication

C Further Computational Details: Identification and Estimation

C.1 Computation of the Set of BCE Predictions

We describe here the computation of the set of BCE predictions shown in Figure 2 in the article. For the simple one-parameter game in Example 6, the figure represents the simplex and sets of BCE and BNE predictions for different assumptions on information. Since the computation of BNE outcomes in discrete games is standard, we focus on how to compute and draw the set:

$$Q_{\Delta}^{BCE} = \left\{ q \in \mathbb{P}_Y : \exists\, \nu \in BCE(\Delta) \text{ such that } q = \int_{\mathcal{E}} \nu\left(y, \varepsilon\right) \mathrm{d}\varepsilon \right\},$$

for $\Delta = -1/2$.[1]

First, notice that - since the set of BCE distributions is convex - this set can also be described as the convex hull of infinitely many *extreme points.* Such extreme points, for any direction $b \in B$, where B denotes the closed unit ball centered at zero in $\mathbb{R}^{|Y|}$, are

$$q\left(b\right) = \arg \sup_{q \in Q_{\theta}^{BCE}(x)} b^T q,$$

so that - by using the property of the support function already used in Section 4.1:

$$q \in Q_{\theta}^{BCE}\left(x\right) \Longleftrightarrow b^T q \leq b^T q\left(b\right)\ \forall\, b \in B$$

we have that:

$$Q_{\Delta}^{BCE} = \mathrm{co}\left\{ q\left(b\right) \mid b \in B \right\}.$$

To feasibly compute Q_{Δ}^{BCE} we proceed in two steps. First, we approximate it as a convex polyhedron by discretizing the set B. Formally, we consider $B^s \subset B$ as the discretized set, with $|B^s| = s$. We set $s = 200$ to draw Figure 2. Second, we compute for each $b \in B^s$ the

[1]The discussion here, although specialized to the game of Example 6 in the article, is more general and can be applied to drawing any set Q_{θ}^{BCE} corresponding to a known payoff structure.

value $q(b)$ as the arg max of the following program:

$$\begin{array}{rcl}
\max_{q\in\mathbb{R}^{|Y|},\ \nu\in\mathbb{R}^{|Y|\times r}} & b^T q & \quad (PPoly)\\
s.t. & & \\
\forall y\in Y & q(y)-\sum_{\varepsilon}\nu(y,\varepsilon) & =0\\
\forall \varepsilon\in\mathcal{E}^r & \sum_y \nu(y,\varepsilon)-f^r(\varepsilon;\theta_\varepsilon) & =0\\
 & \sum_{y,\varepsilon}\nu(y,\varepsilon)-1 & =0\\
\forall i, y_i, y_i', \varepsilon_i & \sum_{y_{-i}}\sum_{\varepsilon_{-i}}\nu(y,\varepsilon_i,\varepsilon_{-i})\left(\pi_i(y_i',y_{-i},\varepsilon_i;x,\theta)-\pi_i(y,\varepsilon_i;x,\theta)\right) & \leq 0.
\end{array}$$

This is a linear program, which can be solved with any LP solver. We use the solver KNITRO in AMPL, but found similar results with CPLEX. Finally, we draw Figure 2 by representing the set

$$\check{Q}_{\Delta}^{BCE}=\mathrm{co}\left\{q(b)\mid b\in B^s\right\}$$

using Matlab's MPT Toolbox (Kvasnica, Grieder, Baotić, and Morari, 2004).

C.2 Computation of Identified Sets Θ_I^{BCE}

We describe in this appendix how to compute Θ_I^{BCE} to construct Figure 3 and Table 2 in the article. The identified set is defined in Section 3 as:

$$\Theta_I^{BCE}=\left\{\theta\in\Theta:\ G(\theta)=0\right\},$$

where $G(\theta)=\int_X \sup_{b\in B}\left[b^T P_{y|x}-h\left(b;Q_\theta^{BCE}(x)\right)\right]\mathrm{d}P_x$. Appendix A outlines how to compute $G(\cdot)$, and the choice of discretization for $\mathcal{E}$; we denote with $\check{G}(\cdot)$ the computed $G(\cdot)$.

As a high-dimensional search over the whole set Θ is infeasible, we conduct a search over a subset $\check{\Theta}$. Moreover, since by construction $\check{G}(\cdot)>0$, we specify a threshold and report the computed analog of the identified set:

$$\check{\Theta}_I^{BCE}=\left\{\theta\in\check{\Theta}:\ \check{G}(\theta)\leq c_I\right\}.$$

There is no general rule to construct an upper bound for this discretization error that is valid for every game and data generating process. However, for the two-player binary game with independent payoff types considered in Table 2, r^{-1} (where r is the dimension of the discrete grid of ε_i that we use to compute $\check{G}(\cdot)$) is an upper bound of the discretization error if we restrict $Q_\theta^{BCE}(x)$ to $Q_\theta^{PSNE}(x)$. Since r^{-1} is representative of the order of magnitude of the discretization error, we use $c_I=r^{-1}$. Our findings on the informativeness of identified sets are similar if we use higher values for c_I.

To construct $\check{\Theta}$, we proceed sequentially. We first specify $\check{\Theta}_1$ as a large Halton set of points around θ_0, then find:

$$Bds = \left[\left(\min_{\theta^k} \left\{ \theta \in \check{\Theta}_1 : \check{G}(\theta) \leq c_I \right\} \right)_{k=1,\ldots,d_\Theta}, \left(\max_{\theta^k} \left\{ \theta \in \check{\Theta}_1 : \check{G}(\theta) \leq c_I \right\} \right)_{k=1,\ldots,d_\Theta} \right]$$

and construct $\check{\Theta}_2$ as another Halton set within $Bds \times 1.2$. This procedure is aimed at constructing more precise boundaries for the identified set. Increasing the umber of points in $\check{\Theta}_1$ and $\check{\Theta}_2$ increases the precision in the computation of the identified set, at the cost of computing time. For Table 2, we use $|\check{\Theta}_1| = 20,000$ and $|\check{\Theta}_2| = 5,000$.

C.3 Computation of Identified Sets $\Theta_I^{BNE}\left(\overline{S}\right)$

In Figure 3 in the main text we compute the sharp identified set under the assumption of complete information and Nash equilibrium behavior, allowing for mixed strategies. The sharp identified set for this case can be obtained by first defining the criterion function:

$$G^{MXNE}(\theta) = \sup_{b \in Dir} \left[b^T P_{y|x} - \sup_{p \in Q_\theta^{MXNE}(x)} b^T p \right]_+ \tag{A1}$$

where Dir denotes the core-determining class (Galichon and Henry, 2011) and $Q_\theta^{MXNE}(x_j)$ contains the Nash equilibrium predictions for a game with covariates x and parameters θ. Since Dir is a discrete set, the computation of G^{MXNE} is simple for games with a small number of players and actions. Then, we have:

$$\Theta_I^{BNE}\left(\overline{S}\right) = \left\{ \theta \in \check{\Theta} : \ G^{MXNE}(\theta) = 0 \right\}.$$

Figure 3 also shows the the identified sets under different behavioral assumptions, R1 and R2. The computation of the corresponding identified sets is analogous to our description of the construction of $\Theta_I^{BNE}\left(\overline{S}\right)$. Under the assumptions of R1 and R2, respectively, we obtain the functions G^{R1} and G^{R2} by substituting $Q_\theta^{R1}(x)$ and $Q_\theta^{R2}(x)$ for Q_θ^{MXNE} into the function G^{MXNE}. Notice that, as the set of predictions is relatively simple, the computation of Q_θ^{MXNE} (as well as of $Q_\theta^{R1}(x)$ and of $Q_\theta^{R2}(x)$) does not involve numerical simulation of the values of ε.

C.4 Computation of Confidence Sets C_n for Θ_I^{BCE}

We begin by discretizing the space of covariates in three steps. First, we compute the median of market size and code a binary variable $D_m = 1\{\text{market size}_m \geq Median\}$. Second, we consider the set $\tilde{M}$ of all combinations of D_m and of the other four discrete regressors in our model (home-region dummies for each player, and presence of large malls), and classify each market m as one such combination $\tilde{m}$. Out of 2^5 such bins $\tilde{m}$, 20 contain a positive number of markets m. Finally, for each $\tilde{m}$ we compute discretized values of market size as

$$\text{market size}_{\tilde{m}} = \frac{1}{|\tilde{m}|} \sum_{m \in \tilde{m}} \text{market size}_m.$$

We end up with a discretization of the market size variable with 20 distinct values. This procedure preserves the correlations of entry patterns with the exogenous variables in the data.

To construct a confidence set s C_n for parameters in the identified sets Θ_I^{BCE} we follow the procedure outlined in Ciliberto and Tamer (2009). The procedure is based on the values of the empirical criterion G_n, whose computation is described in Appendix A. We compute the confidence set via the following steps:

1. We construct deterministic parameter grids using Halton sets around the parameter values of probit regressions, and select among these 40 starting points for a simulated annealing routine, which runs for 10,000 iterations.

2. We collect all the parameters visited by simulated annealing, and consider the corresponding set $\check{\Theta}$ as an approximation of Θ. We define as $g_n = \min_{\theta' \in \check{\Theta}} G_n(\theta')$, and can then obtain for all $\theta \in \check{\Theta}$:

$$\tilde{G}_n(\theta) = G_n(\theta) - g_n.$$

3. We extract $T = 100$ subsamples of size $n_t = n/4$. Subsample size can be an important tuning parameter in this class of models, as argued by Bugni (2016). We follow Ciliberto and Tamer (2009) in the choice of this parameter. For each subsample s, we compute the criterion function using the subsampled observations, so that:

$$G_n^s(\theta) = \frac{1}{n_t} \sum_{j=1}^{n_t} \sup_{b \in B} \left[b^T \hat{P}_{y|x_j}^s - h\left(b; Q_\theta^{BCE}(x_j)\right) \right],$$

and then we find $g_n^s = \min_{\theta \in \Theta} G_n^s(\theta)$ running a Nelder-Mead algorithm.

4. We choose the cutoff value $\hat{c}_0 = n g_n \times 1.25$, and define the set:

$$\hat{\Theta}_I(\hat{c}_0) = \left\{ \theta \in \check{\Theta} : \; n\tilde{G}_n(\theta) \leq \hat{c}_0 \right\}.$$

5. For all $\theta \in \hat{\Theta}_I(\hat{c}_0)$, we obtain then $\tilde{G}_n^s(\theta) = G_n^s(\theta) - g_n^s$ and the threshold $\hat{c}_1(\theta)$ as the 95th percentile of the distribution across subsamples of the statistic $n_t \tilde{G}_n^s$. We

compute then

$$\hat{c}_1 = \sup_{\theta \in \hat{\Theta}_I(\hat{c}_0)} \hat{c}_1(\theta),$$

and

$$\hat{\Theta}_I(\hat{c}_1) = \left\{\theta \in \check{\Theta} : \, n\tilde{G}_n(\theta) \leq \hat{c}_1\right\}.$$

6. Iterating steps 4,5 we obtain $\hat{c}_2$ and report the confidence set:

$$C_n = \left\{\theta \in \check{\Theta} : \, n\tilde{G}_n(\theta) \leq \min \hat{c}_2\right\}.$$

   Further iterations of this procedure do not alter significantly our results.

We report results for confidence sets for parameters in the identified sets. For both Θ_I^{BCE} and $\Theta_I^{BNE}\left(\overline{S}\right)$, constructing confidence sets for the identified set, as opposed to constructing confidence sets for all points in the identified set, yields similar results (as in Ciliberto and Tamer, 2009).

C.5 Computation of Confidence Sets for $\Theta_I^{BNE}\left(\overline{S}\right)$

The construction of the confidence set for parameters in $\Theta_I^{BNE}\left(\overline{S}\right)$ is analogous to the procedure followed to compute the confidence set under the assumption of BCE behavior, except that it is based on the empirical criterion function:

$$G_n^{PSNE}(\theta) = \frac{1}{n}\sum_{j=1}^{n} \sup_{b \in Dir}\left[b^T \hat{P}_{y|x_j} - \sup_{q \in Q_\theta^{PSNE}(x_j)} b^T q\right]_+,$$

where Dir contains vectors corresponding to core-determining class (Galichon and Henry, 2011) and $Q_\theta^{PSNE}(x_j)$ contains the pure-strategy Nash equilibrium predictions for a game with covariates x_j and parameters θ. We limit Nash equilibria to pure-strategy to maintain the parallel with Ciliberto and Tamer (2009), but the extension to mixed strategy is immediate and can be done by considering the empirical analogue of (A1). The confidence set for parameters identified under the assumption of pure-strategy Nash equilibrium and complete information is obtained going through the same steps 1.-6. described for the computation of C_n, where G_n is substituted with G_n^{PSNE}.

C.6 Computation of Perfectly Private Information Estimates

In the context of the model of our application in Section 6, we compute parameter estimates for a perfectly private information model $\hat{\theta}\left(\underline{S}\right)$. To apply standard methods in the literature we maintain the following assumptions:

- The information structure is $S = \underline{S}$,
- Payoff types are iid type-1 EV,
- The data are generated by an unique equilibrium σ^x for each $x \in X$.

The two latter assumptions impose strong restrictions on the payoff structure and on equilibrium selection that we do not maintain in either the general model of Section 2, or the application of Section 6 of the paper.

Under these assumptions, suppose that behavior is defined by the BNE strategy profile σ^x; for each player i, let $\tilde{\sigma}_i^x$ be the equilibrium probability of entry derived from σ_i^x. Define moreover the deterministic part of expected payoffs as:

$$\begin{aligned}
\Pi_i^E\left(x, \theta_\pi, \sigma\right) &= E_\sigma \Pi_i\left(y_{-i}; x, \theta_\pi\right) \\
&= \sum_{y_{-i}} \Pi_i\left(y_{-i}; x, \theta_\pi\right) \tilde{\sigma}_{-i}^x\left(y_{-i}\right) \\
&= x_{im}^T \beta_i + \sum_{j \neq i} \Delta_j \tilde{\sigma}_j^x.
\end{aligned}$$

The definition of BNE implies that, for all y_i such that $\tilde{\sigma}_i^x\left(y_i\right) > 0$ we have that

$$\begin{aligned}
\tilde{\sigma}_i^x &= \int_{\left\{\varepsilon_i \mid \Pi_i^E\left(x, \theta_\pi, \sigma\right) + \varepsilon_i > 0\right\}} 1\left\{\varepsilon_i = e_i\right\} dF\left(e_i\right), \\
&= \Phi\left(x_{im}, \tilde{\sigma}_{-i}^x; \beta, \Delta\right).
\end{aligned}$$

and using the EV distributional assumption, this becomes:

$$\tilde{\sigma}_i^x = \frac{\exp\left\{x_{im}^T \beta_i + \sum_{j \neq i} \Delta_j \tilde{\sigma}_j^x\right\}}{1 + \exp\left\{x_{im}^T \beta_i + \sum_{j \neq i} \Delta_j \tilde{\sigma}_j^x\right\}}. \tag{A2}$$

This expression motivates two estimation strategies: the first adopts a Maximum Likelihood approach (Su, 2014); the second adopts instead a two-step approach (Bajari, Hong, Krainer, and Nekipelov, 2010). We use the former to produce the estimates in the paper.

Maximum Likelihood Estimation - Su (2014) - We can reinterpret Equation (A2) as the equilibrium map $\sigma^x = \Phi\left(\sigma^x; x, \theta_\pi\right)$, and can thus form the likelihood function of the data:

$$L\left(\sigma; x, y\right) = \sum_{i,m} \left\{y_{im} \times \log\left(\sigma_i^{x_m}\right) + \left(1 - y_{im}\right) \times \log\left(1 - \sigma_i^{x_m}\right)\right\}.$$

To perform estimation we adopt an MPEC approach by recovering:

$$\hat{\theta}_\pi = \arg\max_{\theta_\pi,\sigma} L(\sigma; x, y)$$
$$s.t. \qquad \sigma^x = \Phi(\sigma^x; x, \theta_\pi) \ \forall x.$$

Standard errors can be derived analytically or obtained via bootstrap. Notice that for this method we need a discrete set of covariates X; we proceed to discretize the set as we do for the estimation of confidence sets under weak assumption on information.

Two Step Estimation - Bajari et al. (2010) - Assume that there are firm-specific covariates: hence, for each player i, we have that $\Pi_i(y_{-i}; x_m, \theta_\pi) = \Pi_i(y_{-i}; x_{im}, \theta_\pi)$. Then, we can first recover estimates of (marginals of) equilibrium strategies; in our context this is equivalent to recovering how entry probabilities vary as a function of x. This can be done by estimating for each player a function $\hat{\sigma}_i(x) = \Pr(y_i = 1|x)$, for instance by fitting a linear model with OLS. In a second step, we plug the first-step estimates into Equation (A2) and obtain:

$$\Pr\{y_i = 1\} = \frac{\exp\left\{x_{im}^T\beta_i + \sum_{j\neq i}\Delta_j\hat{\sigma}_j^x\right\}}{1 + \exp\left\{x_{im}^T\beta_i + \sum_{j\neq i}\Delta_j\hat{\sigma}_j^x\right\}}. \tag{A3}$$

This equation can then be estimated as a logit model. Standard errors need to be recovered by bootstrap or with a two-step correction. As opposed to the Maximum Likelihood method, this method does not require discretization of the covariates.

Comparison of Results - In Table A1 we report estimation results for both methods. The estimates are qualitatively similar, although (bootstrap) standard errors are systematically smaller for the Maximum Likelihood method, reflecting its greater efficiency.

D Further Computational Details: Policy Experiment

D.1 Computation of Predictions for the model with Weak Assumptions on Information: the Variable Latent Information Approach

All of the predicted objects described in Section 7.1 - I_W^x, $I_{\overline{W}}$ and $I_W^x\left(\hat{\theta}_0\right)$ - can be easily obtained from the computation of

$$\overline{W}(\theta, x) = \max_{\nu\in BCE^x(\theta)} W(\nu, \theta, x) \qquad (PC0)$$
$$\underline{W}(\theta, x) = -\max_{\nu\in BCE^x(\theta)} -W(\nu, \theta, x),$$

Table A1: Perfectly Private Information Estimation - Two Methods

Parameter	PP Info ML - Su (2014)	PP Info Two Step - Bajari et al. (2010)
Constant	-3.32	-2.89
	$[-3.51, -3.16]$	$[-3.30, -2.49]$
Market Size	3.06	2.62
	$[2.59, 3.94]$	$[1.63, 3.62]$
Home-region:		
Cooperatives	1.60	1.95
	$[1.36, 1.81]$	$[1.48, 2.42]$
Italian Groups	1.72	1.60
	$[1.54, 1.99]$	$[1.03, 2.18]$
French Groups	1.32	1.66
	$[1.10, 1.58]$	$[1.13, 2.19]$
Presence of Large Malls:		
Cooperatives	-1.61	-0.96
	$[-2.03, -1.19]$	$[-1.93, 0.011]$
Italian Groups	-1.04	-0.50
	$[-1.46, -0.42]$	$[-1.54, 0.53]$
French Groups	-0.80	-0.28
	$[-1.30, -0.46]$	$[-1.15, 0.59]$
Competitive Effects:		
Cooperatives	0.02	-0.069
	$[-0.55, 0.47]$	$[-1.22, 1.08]$
Italian Groups	-0.66	-1.57
	$[-1.08, -0.29]$	$[-2.62, -0.52]$
French Groups	2.88	2.93
	$[1.73, 3.51]$	$[0.69, 5.17]$

Note: We report estimates for the game-theoretic model of Section 6 in the article, obtained using the two methods for the estimation of perfectly private information games described in this appendix. Bootstrap standard errors are in parenthesis, and are calculated from 200 bootstrap samples.

for all values of $\theta \in C_n$, where W is a function such as $W_{\hat{Y}}(\nu, \theta, x)$ or $W_N(\nu, \theta, x)$ which is *linear in* ν. For simplicity, we focus on the computation of $\overline{W}(\theta, x)$ since $\underline{W}(\theta, x)$ can be obtained with minimal changes.

With the same discretization applied in Appendix A, the program $(PC0)$ can be approximated by the feasible program:

$$
\begin{array}{rcr}
\max\limits_{\nu \in \mathbb{R}^{|Y| \times r}} & W(\nu, \theta, x) & (PC1) \\
s.t. & \sum_{y,\varepsilon} \nu(y, \varepsilon) - 1 & = 0 \\
\forall \varepsilon \in \mathcal{E}^r & \sum_{y} \nu(y, \varepsilon) - f^r(\varepsilon; \theta_\varepsilon) & = 0 \\
\forall i, y_i, y_i', \varepsilon_i & \sum_{y_{-i}} \sum_{\varepsilon_{-i}} \nu(y, \varepsilon_i, \varepsilon_{-i}) \left(\pi_i(y_i', y_{-i}, \varepsilon_i; x, \theta) - \pi_i(y, \varepsilon_i; x, \theta)\right) & \leq 0.
\end{array}
$$

This is a linear program which can be solved with standard solvers; we compute it using the solver KNITRO in AMPL. Alternative solvers (e.g., CPLEX) gave us similar results.

D.2 Computation of Predictions for the model with Weak Assumptions on Information: the Fixed Latent Information Approach

We first introduce the notion of the BCE for the linked game, encompassing both the factual game with covariates x, and the predicted game with covariates x':

Definition 5. (BCE of the Linked Game) A Bayes Correlated Equilibrium $\tilde{\nu} \in \mathcal{P}_{Y,Y,\mathcal{E},\tilde{T}}$ for the linked game $\Gamma^{x,x'}(\theta, S)$ is a probability measure $\tilde{\nu}$ over *factual* and *counterfactual* actions profiles, payoff types, and signals that is:

1. *Consistent with the prior*: for all $\varepsilon \in \mathcal{E}$, $\tilde{\tau} \in \tilde{T}$,

$$\sum_{y,y' \in Y} \int_{[t \leq \tilde{\tau}]} \int_{[e \leq \varepsilon]} \tilde{\nu}(y, y', e, t)\, \mathrm{d}t\mathrm{d}e = \int_{[t \leq \tilde{\tau}]} \int_{[e \leq \varepsilon]} P_{\tilde{\tau}|e}(t) \mathrm{d}F(e; \theta_\varepsilon)\, \mathrm{d}t;$$

2. *Incentive Compatible*: for all $i, \varepsilon_i, \tilde{\tau}_i, y_i, y'_i$ such that $\tilde{\nu}(y_i \mid \varepsilon_i, \tilde{\tau}_i, y'_i) > 0$,

$$E_{\tilde{\nu}}\left[\pi_i(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi) \mid y_i, \varepsilon_i, \tilde{\tau}_i\right] \geq E_{\tilde{\nu}}\left[\pi_i(\tilde{y}_i, y_{-i}, \varepsilon_i; x, \theta_\pi) \mid y_i, \varepsilon_i, \tilde{\tau}_i\right], \qquad \forall \tilde{y}_i \in Y_i,$$

   and for all $i, \varepsilon_i, \tilde{\tau}_i, y_i, y'_i$ such that $\tilde{\nu}(y'_i \mid \varepsilon_i, \tilde{\tau}_i, y_i) > 0$,

$$E_{\tilde{\nu}}\left[\pi_i(y'_i, y'_{-i}, \varepsilon_i; x', \theta_\pi) \mid y'_i, \varepsilon_i, \tilde{\tau}_i\right] \geq E_{\tilde{\nu}}\left[\pi_i(\tilde{y}_i, y'_{-i}, \varepsilon_i; x', \theta_\pi) \mid y'_i, \varepsilon_i, \tilde{\tau}_i\right], \qquad \forall \tilde{y}_i \in Y_i,$$

   where the expectation operators $E_{\tilde{\nu}}[\cdot \mid y_i, \varepsilon_i, \tilde{\tau}_i]$ are taken with respect to the conditional equilibrium distributions $\tilde{\nu}(y_{-i}, \varepsilon_{-i}, \tilde{\tau}_{-i} \mid y_i, y'_i, \varepsilon_i, \tilde{\tau}_i)$ and $\tilde{\nu}(y'_{-i}, \varepsilon_{-i}, \tilde{\tau}_{-i} \mid y_i, y'_i, \varepsilon_i, \tilde{\tau}_i)$, respectively in the two inequalities.

3. *Consistent with factual outcomes*: equilibrium behavior in the factual game must be consistent with factual outcomes, so that

$$\sum_{y'} \int_{\tilde{T}} \int_{\mathcal{E}} \tilde{\nu}(y, y', e, t)\, \mathrm{d}t\mathrm{d}e = P_{y|x}(y), \qquad \forall y \in Y.$$

This definition is closely related to the one in Bergemann, Brooks, and Morris (2022). Let $\widetilde{BCE}^{x,x'}(\theta)$ be the set of all BCEs of the linked game $\Gamma^{x,x'}(\theta, \underline{S})$. The choice of the perfectly private information structure $\underline{S}$ corresponds to our choice of baseline information structure in the paper, and allows us to omit the presence of additional signals $\tilde{\tau}_i$.

Our use of fixed latent information predictions is based on the following lemma, which restates and adapts Theorem 1 in Bergemann et al. (2022).

Lemma 2. *(Bergemann et al., 2022) Suppose that observed play in the factual game is $P_{y|x} \in Q^{BNE}_{\theta_0,S_0}(x)$, and play in the counterfactual game is $q' \in Q^{BNE}_{\theta_0,S_0}(x')$. Suppose the re-*

searcher only assumes that $S_0 \in \mathcal{S}$, but has identified θ_0. Then, we can characterize all counterfactual outcomes q' consistent with this setup as those for which there exists a BCE of the linked game $\tilde{\nu} \in \widetilde{BCE}^{x,x'}(\theta_0)$ such that (i) the marginal of $\tilde{\nu}$ on the space of factual actions is $P_{y|x}$, and (ii) the marginal of $\tilde{\nu}$ on the space of counterfactual actions is q'.

The lemma establishes that BCEs of the linked game enable a tractable characterization of predicted or counterfactual outcomes generated by BNE play under fixed latent information. The perspective adopted in the Lemma is consistent with the rest of our article, since the researcher does not know the true information structure that generated the data and only maintains weak assumptions on information in the sense that $S_0 \in \mathcal{S}$. Different Bayes Nash equilibria and equilibrium selection mechanisms may be generating the data in the factual and counterfactual games.

Although we state the lemma for a known $\theta_{0,}$ in our application we only know that $\theta \in C_n$, so that we have to account for the extra uncertainty due to set identification and estimation of the game's structure. To compute predicted objects $\tilde{I}^{post}_{\tilde{W}}, \tilde{I}_{\overline{W}}$ and $\tilde{I}^{post}_{\tilde{W}}\left(\hat{\theta}_0\right)$ and implement the fixed latent information approach we need to compute

$$\overline{\tilde{W}}(\theta) = \max_{\tilde{\nu} \in \widetilde{BCE}^{x,x'}(\theta)} \tilde{W}(\tilde{\nu}, \theta, x, x'), \qquad (PC2)$$

where the function $\tilde{W}$ adapts the corresponding W in a natural way, that is:

$$\begin{array}{rcr}
\max\limits_{\tilde{\nu} \in \mathbb{R}^{|Y| \times |Y| \times r}} & \tilde{W}(\tilde{\nu}, \theta, x, x') & (PC3) \\
s.t. & \sum_{y,y',\varepsilon} \tilde{\nu}(y, y', \varepsilon) - 1 & = 0 \\
\forall y \in Y & P_{y|x}(y) - \sum_{\varepsilon, y' \in Y} \tilde{\nu}(y, y', \varepsilon) & = 0 \\
\forall \varepsilon \in \mathcal{E}^r & \sum_{y,y'} \tilde{\nu}(y, y', \varepsilon) - f^r(\varepsilon; \theta_\varepsilon) & = 0 \\
\forall i, y_i, y_i', \tilde{y}_i, \varepsilon_i & \sum_{y_{-i}, y'_{-i}} \sum_{\varepsilon_{-i}} \tilde{\nu}(y, y', \varepsilon_i, \varepsilon_{-i}) \left(\pi_i(\tilde{y}_i, y_{-i}, \varepsilon_i; x, \theta) - \pi_i(y, \varepsilon_i; x, \theta)\right) & \leq 0 \\
\forall i, y_i, y_i', \tilde{y}_i, \varepsilon_i & \sum_{y_{-i}, y'_{-i}} \sum_{\varepsilon_{-i}} \tilde{\nu}(y, y', \varepsilon_i, \varepsilon_{-i}) \left(\pi_i(\tilde{y}_i, y'_{-i}, \varepsilon_i; x', \theta) - \pi_i(y', \varepsilon_i; x', \theta)\right) & \leq 0.
\end{array}$$

To compute predictions we need thus to operationalize $(PC3)$, which we do with the usual discretization of $\mathcal{E}$.

There is another consideration when implementing this method. At the identification level, if $\theta \in \Theta_I^{BCE}$ there exists a latent information structure such that BCE predictions can match $P_{y|x}$. However, the inferential procedure that we employ implies only that for $\theta \in C_n$ BCE predictions need to match $P_{y|x}$ *approximately.* In fact, $\theta \in C_n$ only implies that θ is *close* to the minimizer of the empirical criterion function G_n built using a finite sample $\{x_i, y_i\}_{i=1}^{\infty}$.

Moreover, a feasible implementation of our empirical strategy involves several approximations: we estimate the set C_n by relying on a discretized set of covariates X, and we discretize the support of $\mathcal{E}$ in order to compute ν. Hence, finite-sample and computational error may make the restriction $marg_{\tilde{\nu}}(y) = P_{y|x}$ for some parameters $\theta \in C_n$ impossible to satisfy exactly, thus rendering $(PC3)$ unfeasible. To address this problem, we compute $\tilde{P}_{y|x}(\theta)$, the distribution of the observables that best fits the data for a given parameter θ. More formally, $\tilde{P}_{y|x}(\theta)$ is equal to q that solves the program:

$$\max_{b \in B} \min_{q \in Q_\theta^{BCE}(x)} \left[b^T P_{y|x} - b^T q \right].$$

Intuitively, $\tilde{P}_{y|x}(\theta)$ is the distribution of the observables corresponding to the BCE that *best fits the data* $P_{y|x}$ and the parameter value θ. In our experience, these $\tilde{P}_{y|x}(\theta)$ are reasonably close to the data for parameters θ in the confidence set. We can thus compute the predicted quantity of interest $\overline{\tilde{W}}(\theta)$ as the solution to the program:

$$\begin{array}{rcr}
\max\limits_{\tilde{\nu} \in \mathbb{R}^{|Y|\times|Y|\times r}} & \tilde{W}(\tilde{\nu}, \theta, x, x') & (PC4) \\
s.t. & \sum_{y,y',\varepsilon} \tilde{\nu}(y, y', \varepsilon) - 1 & = 0 \\
\forall \varepsilon \in \mathcal{E}^r & \sum_{y,y'} \tilde{\nu}(y, y', \varepsilon) - f^r(\varepsilon; \theta_\varepsilon) & = 0 \\
\forall y \in Y & |\tilde{P}_{y|x}(y; \theta) - \sum_{\varepsilon, y' \in Y} \tilde{\nu}(y, y', \varepsilon)| & \leq \epsilon \\
\forall i, y_i, y_i', \tilde{y}_i, \varepsilon_i & \sum_{y_{-i}, y_{-i}'} \sum_{\varepsilon_{-i}} \tilde{\nu}(y, y', \varepsilon_i, \varepsilon_{-i}) \left(\pi_i(\tilde{y}_i, y_{-i}, \varepsilon_i; x, \theta) - \pi_i(y, \varepsilon_i; x, \theta)\right) & \leq 0 \\
\forall i, y_i, y_i', \tilde{y}_i, \varepsilon_i & \sum_{y_{-i}, y_{-i}'} \sum_{\varepsilon_{-i}} \tilde{\nu}(y, y', \varepsilon_i, \varepsilon_{-i}) \left(\pi_i(\tilde{y}_i, y_{-i}', \varepsilon_i; x', \theta) - \pi_i(y', \varepsilon_i; x', \theta)\right) & \leq 0.
\end{array}$$

Notice that the constraint of consistency with the (pseudo-) factual outcomes is enforced with some slack, to maintain feasibility in light of the multiple approximations involved in the program. We set $\epsilon = 0.001$ in our computation; experimenting with tighter and looser tolerances did not alter the results substantially.

D.3 Computation of Predictions for Models with more Restrictive Assumptions on Information

Complete Information - Under the assumption of complete information and Nash equilibrium in pure strategies, the lower and upper bound probabilities of market structure outcomes have analytical expressions for our three-player entry game (Tamer, 2003; Ciliberto and Tamer, 2009). We obtain in this way $\overline{W}^{\overline{S}}(\theta, x)$ and $\underline{W}^{\overline{S}}(\theta, x)$, and we can compute from these the complete information intervals in the bottom part of Figure 3 and the results in column III of Table 3 in the article.

Perfectly Private Information - Market structure outcomes $\hat{Y}$ can be expressed as functions of a vector of strategy profiles σ as $W^{\underline{S}}_{\hat{Y}}(\sigma)$. For instance, the probability that there are no entrants is

$$W^{\underline{S}}_{\{(0,0,0)\}}(\sigma) = \Pi_{i\in I}(1-\sigma_i).$$

Hence, upper bound probabilities for a market structure outcome under the perfectly private information model can be obtained as

$$\begin{aligned} \overline{W}^{\underline{S}}_{\hat{Y}}(x) &= \arg\max_{\sigma} W^{\underline{S}}_{\hat{Y}}(\sigma) \\ s.t. \qquad \sigma &= \Phi\left(\sigma; x, \hat{\theta}^{\underline{S}}_{\pi}\right), \end{aligned}$$

where $\hat{\theta}^{\underline{S}}_{\pi}$ is the parameter estimate obtained under the assumption of perfectly private information (see C.6 above) and Φ represents the equilibrium mapping in Equation (A3). The average changes in upper bounds probabilities for the perfectly private information model reported in column IV of Table 3 are computed from $\overline{W}^{\underline{S}}_{\hat{Y}}(x)$. Confidence intervals for the prediction are based on 200 bootstrap samples, and account for uncertainty in $\hat{\theta}^{\underline{S}}_{\pi}$.

E BMM Representation of the Identified Set

Beresteanu, Molchanov, and Molinari (2011), henceforth BMM, provide a computable characterization of the identified set of partially identified models making use of random set theory. In this appendix, we show how our characterization of the identified set maps into their framework.

Let $z=(x,y)$ and ε be respectively the vector of observable outcomes and covariates, and the vector of payoff types. The random vectors are defined on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{G}$ be the sigma algebra generated by the random vector x. We also adopt the assumptions 3.1(i),(iii) and 3.2 in BMM, and substitute 3.1(ii) with the assumption of BCE behavior. We restate these assumptions below for ease of reference:

Assumption 4. *Assume that:*

1. *The discrete set of strategy profiles of the game, Y, is finite.*

2. *Payoffs $\pi_i(y, \varepsilon_i; x, \theta_\pi)$ have a known parametric form, and are continuous in x and ε_j.*

3. *The observed outcome y of the game is the result of BCE behavior in the game of perfectly private information $\underline{S}$.*

4. *The conditional distribution of outcomes $P_{y|x}$ is identified by the data, and ε has a continuous distribution function.*

Let us adapt our notation and denote the set of BCE equilibrium distributions ν with $BCE_\theta(x)$, for any given realization of x. Considering $x(\omega)$ as a random vector, $BCE_\theta(x(\omega)) = BCE_\theta(\omega)$ is a random set. Let $\mathrm{Sel}(BCE_\theta)$ denote the set of all $\nu(\omega)$, measurable selections of $BCE_\theta(\omega)$. In order to characterize the identified set, we need to map these equilibria into observable outcomes of the game for each $\omega \in \Omega$. A realization of ω implies both a realization of $(x(\omega), \varepsilon(\omega))$, and also a BCE distribution $\nu(\omega)$, which in turn determine the following probability distribution over outcomes:

$$q(\nu(\omega)) = \nu(\cdot \mid \varepsilon(\omega)) \in \mathcal{P}_Y,$$

where $\nu(\cdot|\varepsilon(\omega))$ is the conditional distribution implied by the joint distribution $\nu(\omega) \in \mathcal{P}_{Y,\mathcal{E}}$, and the realization $\varepsilon(\omega)$. $\tilde{Q}_\theta$ is the set of all equilibrium predictions:

$$\tilde{Q}_\theta = \{q(\nu) : \nu \in \mathrm{Sel}(BCE_\theta)\}.$$

Then the conditional Aumann expectation of this random set is:

$$\mathbb{E}\left(\tilde{Q}_\theta \mid x\right) = \{E(q(\nu) \mid x) : \nu \in \mathrm{Sel}(BCE_\theta)\}.$$

Notice however that:

$$\begin{aligned} E(q(\nu) \mid x) &= E[\nu(\cdot \mid \varepsilon(\omega)) \mid x] \\ &= \int_{\mathcal{E}} \nu(y \mid \varepsilon) \, \mathrm{d}F \\ &= \int_{\mathcal{E}} \nu(y, \mathrm{d}\varepsilon), \end{aligned}$$

so that $\mathbb{E}\left(\tilde{Q}_\theta|x\right) = Q_\theta^{BCE}(x)$. Hence, our characterization of the identified set is equivalent to the one proposed in BMM.

F A More General Model

The model in Section 2 of the paper embeds an important restriction: our definition of the class of information structures $\mathcal{S}$ maintains the assumption that players know the realization of market-level covariates x and of their own payoff type ε_i. This restriction in turn is important for the definition of $\Theta_I^{BNE}(\mathcal{S})$ and the equivalence result in Proposition 1. In this appendix we discuss identification under more general assumptions.

As in the article, we assume that players receive signals with two components, but now $\tau_i^x = (\tau_i^{*x}, \tilde{\tau}_i^x)$. The first component of τ_i^x is a private random signal τ_i^{*x} that is part of their *baseline information structure* S^* defined as:

$$S^{*x} = \left(T^{*x}, \left\{P^x_{\tau^*|\varepsilon} : \varepsilon \in \mathcal{E}\right\}\right).$$

This definition of *baseline* information structure allows for both non-informative signals and perfectly informative signals on ε. The model in Section 2 specifies this baseline to be $\tau_i^{*x} = \varepsilon_i$; we consider here cases where the baseline could be either more or less informative.

In addition to the baseline signal, every player receives an *extra* private random signal $\tilde{\tau}_i^x$, which may also be informative about the full vector of baseline types τ^* and the full vector of ε. An information structure $\tilde{S}^x$ specifies, for a game with covariates x, the set of extra signals a player may receive and the probability of receiving them, given the realization of the vector of payoff types and baseline signals. Formally:

$$\tilde{S}^x = \left(\tilde{T}^x, \left\{\tilde{P}^x_{\tilde{\tau}|\tau^*,\varepsilon,} : (\tau^*, \varepsilon) \in T^{*x} \times \mathcal{E}\right\}\right),$$

Whereas S^* describes the baseline information, the information structure $\tilde{S}$ denotes the extra information players might receive. We use $\mathcal{S}(S^*)$ to denote the class of information structures that are compatible with the baseline information structure S^*,[2] so that a generic $S \in \mathcal{S}(S^*)$ is the information structure where players receive signals $\tau_i^x = (\tau_i^{*x}, \tilde{\tau}_i^x)$ for a fixed baseline S^* and some generic $\tilde{S}$. The game $\Gamma^x(\theta, S)$ for $S \in \mathcal{S}(S^*)$ is then analogous to the game as defined in the article, except that players now observe baseline signals according to S^* and extra signals according to $\tilde{S}$.

We also redefine the BNE concept used in the paper:

Definition 6. (Bayes Nash Equilibrium) A strategy profile $\sigma = \times_{i\in N}\sigma_i,\ \sigma_i : T_i^{*x} \times \tilde{T}_i^x \to \mathcal{P}_{Y_i}$ is a Bayes Nash Equilibrium (BNE) of the game $\Gamma^x(\theta, S)$ if for every $i \in N$, $\tau_i^* \in T_i^{*x}$ and $\tilde{\tau}_i \in \tilde{T}_i^x$ we have that, for every $y_i \in Y_i$ such that $\sigma_i(y_i \mid \tau_i^*, \tilde{\tau}_i) > 0$:

$$E_\sigma\left[\pi_i\left(y_i, y_{-i}, \varepsilon_i; x, \theta_\pi\right)|\tau_i^*, \tilde{\tau}_i\right] \geq E_\sigma\left[\pi_i\left(y_i', y_{-i}, \varepsilon_i; x, \theta_\pi\right)|\tau_i^*, \tilde{\tau}_i\right], \quad \forall\, y_i' \in Y_i.$$

Based on this modified notion of BNE, which defines equilibrium strategies as functions of both baseline signals τ^* and extra signals $\tilde{\tau}$, it's immediate to redefine the set of BNE predictions $Q^{BNE}_{\theta,S}(x)$ for the game $\Gamma^x(\theta, S)$ and the BNE identified set $\Theta_I^{BNE}(\mathcal{S}(S^*))$.

Our main result relies on the following Lemma, a re-statement of Lemma 1 in the article:

Lemma 3. *For all $\theta \in \Theta$ and $x \in X$,*

1. *If $q \in Q^{BCE}_{\theta,S^*}(x)$, then $q \in Q^{BNE}_{\theta,S}(x)$ for some $S \in \mathcal{S}(S^*)$.*

[2] In the language of Bergemann and Morris (2016), $\mathcal{S}(S^*)$ contains all *expansions* of S^*.

2. Conversely, for all $S \in \mathcal{S}(S^*)$, $Q_{\theta,S}^{BNE}(x) \subseteq Q_{\theta,S^*}^{BCE}(x)$.

We may now extend our Proposition 1 to this environment:

Proposition 4. (Robust identification) *Let Assumptions 1 and 2, aptly modified for the game $\Gamma^x(\theta, S)$, hold. Then*

$$\Theta_I^{BCE}(S^*) = \Theta_I^{BNE}(\mathcal{S}(S^*)).$$

For $S^* = \underline{S}$, this Proposition coincides with our Proposition 1 in Section 3 of the article. However, the explicit reference to the baseline information structure allows us to generalize the result to environments where players do not observe their own ε_i, or observe not only ε_i but also other components of the vector of payoff types ε.

It is interesting to investigate how identified sets vary for different assumptions on $\tilde{S}$. The set $\Theta_I^{BCE}(\mathcal{S}(S^*))$ is certainly not invariant to S^*; intuitively, as information structures get more informative, the set of BCE predictions gets smaller, and hence fewer parameters are compatible with the observables, so that the identified set shrinks. This notion of "more informative" can be made precise (as in Bergemann and Morris, 2016).

Having established that $\Theta_I^{BCE}(S^*)$ is not invariant to S^*, we still maintain that using $S^* = \underline{S}$ best suits the goal of the paper. Baseline information structures S^* that are less informative than $\underline{S}$ are likely to result in limited identifying power, whereas baselines that are more informative than S^* may not be justified in applications.

Computational Burden: The change in baseline affects computation of the functions G and G_n that define identified set and confidence set of parameters. With respect to the steps described in Appendix A, a baseline $S^* \neq \underline{S}$ changes the number of inequalities in the program ($P0$). In fact, now each player i has an incentive constraint for all values of $(y_i, y_i', \varepsilon_i, \tau_i^*)$. This implies that - if the space of signals for player i has dimension $|T_i^{*x}| = s_i$ - the number of inequalities is now $d_{ineq,S^*} = \sum_{i \in N}(|Y_i| \cdot |Y_i - 1| \cdot r_i \cdot s_i)$ as opposed to the original $d_{ineq} = \sum_{i \in N}(|Y_i| \cdot |Y_i - 1| \cdot r_i)$ for the perfectly private baseline. This results in a corresponding increase in the number of variables for the computable program ($P3$). In practice, simple increases in the baseline information structure remain feasible computationally. For example, for the game of Figure 3, Panel C in the article going from a perfectly private baseline to the flexible information baseline S^F where the public type η_i has a discrete support taking two different values generates a number of inequalities $d_{ineq,S^*} = 2 \cdot d_{ineq}$, thus resulting in an increase of 400 variables in the maximization problem to compute $G(\theta)$ when the continuously distributed private type ϵ_i is discretized with a support taking 50 different values.

G Identification of Correlation among Payoff Types

In this subsection we further investigate the identifying power of BCE with respect to θ_ε. The characterization of the sharp identified set Θ_I^{BCE} in Section 4.2 of the article uses an infinite number of moment inequalities to set identify θ_ε so that the resulting mapping between data and parameters is not transparent. To gain intuition on how the parameter θ_ε is identified, we assume that θ_π is point identified and we construct non-sharp bounds using a simple implication of equilibrium behavior. In any BCE distribution dominated actions must occur with zero probability, since otherwise incentive compatibility constraints would be violated.

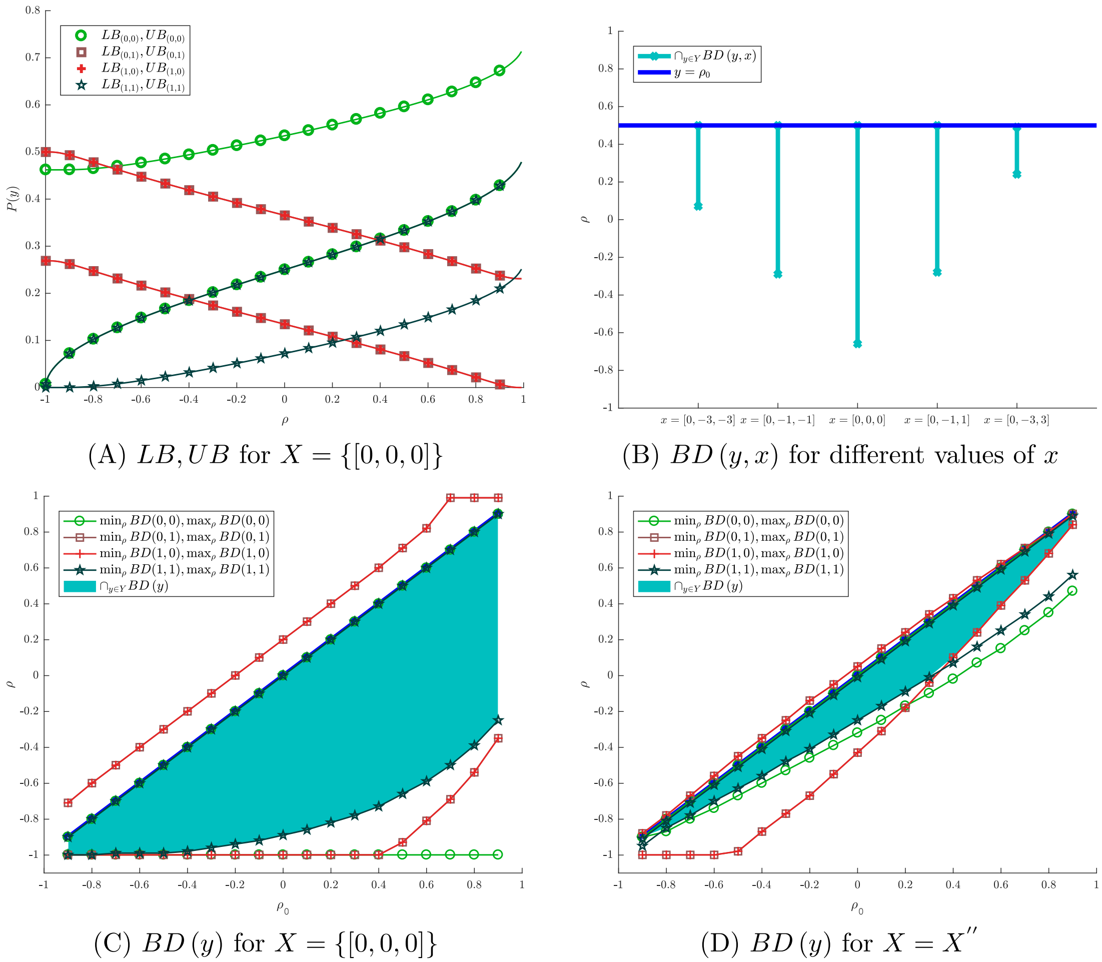
For each value of θ_ε and x this observation implies thus a lower bound $LB_y(\theta_\varepsilon; x)$ on the probability of observing any outcome $y \in Y$, constructed as the integral of the cdf of ε over the region of $\mathcal{E}$ where y is a *dominant outcome* (i.e. y_i is dominant for every player i). Similarly we construct an upper bound $UB_y(\theta_\varepsilon; x)$ for the probability of each outcome y by integrating over all areas of $\mathcal{E}$ where y is a *non dominated outcome,* that is for all players i no other $y_i' \neq y_i$ is dominant. We can then construct bounds for θ_ε. For each outcome $y \in Y$ we define a set $BD(y, x)$ which includes values of θ_ε such that $P_{y|x}$ falls within the bounds $LB_y(\theta_\varepsilon; x)$ and $UB_y(\theta_\varepsilon; x)$. Finally, the sets $BD(y) = \cap_{x \in X} BD(y, x)$ summarize the identification power of the bounds constructed using outcome y. Variation in x shrinks the sets $BD(y)$. The construction of the bounds outlined above is described more formally in Proposition 3 in Appendix B for a two-player entry game with point identified θ_π.

In Figure A1 we show the bounds on outcome probabilities LB and UB, and the sets BD of parameters θ_ε compatible with these bounds for a two-player entry game with point identified payoffs. For this figure we assume that payoff types are jointly normal so that we can focus on the identification of the parameter $\theta_\varepsilon = \rho$ that represents the correlation of players' payoff types. Panel A depicts bounds on outcome probabilities when all covariates are zero. Although the bounds are wide, encompassing a range of realizations of $P_{y|x}$, they are non-trivial. As the correlation in ρ increases, players are more likely to choose the same action: bounds on the probabilities of outcomes $(0, 0)$ and $(1, 1)$ increase with ρ, whereas bounds on outcomes $(0, 1)$ and $(1, 0)$ decrease with ρ.

Panel B depicts sets $\cap_{y \in Y} BD(y, x)$ of parameters ρ that generate bounds compatible with the data for a given value of x. To understand what variation in covariates is most helpful in identifying ρ we plot these sets as vertical segments for different values of x. In this example, where the upper bound of BD is sharp,[3] values of covariates that generate the largest dispersion in payoffs across players are the most informative about the lower

[3] This is a because in the DGP we chose (complete information, $S_0 = \overline{S}$) the probability of observing firms doing the same action (hence, selecting either $(0, 0)$ or $(1, 1)$) is the lowest across all possible S_0: every other information structure implies a higher probability of observing $(1, 1)$ or $(0, 0)$ for any given ρ. In turn, this means that there is no level of correlation in unobservables $\rho > \rho_0$ that is compatible with the data, since for such ρ the value of LB_y for $y \in \{(0, 0), (1, 1)\}$ would exceed the corresponding $P_{y|x}$.

Figure A1: Identification of Correlation among Payoff Types



(A) LB, UB for $X = \{[0,0,0]\}$

(B) $BD(y,x)$ for different values of x

(C) $BD(y)$ for $X = \{[0,0,0]\}$

(D) $BD(y)$ for $X = X''$

Note: We represent bounds on probabilities of outcomes LB_y and UB_y (Panel A) and bounds $BD(y)$ on the parameter ρ (Panels B-D) for the two-player game with payoffs $\pi_i(y, \varepsilon_i; x, \theta_\pi) = y_i \left(x_c^T \beta^C + x_i^T \beta_i^E + \Delta_{-i} y_{-i} + \varepsilon_i \right)$ for $i = 1, 2$. Payoff types are distributed $\varepsilon \sim N(0, \Sigma)$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and payoff parameters θ_π are the same as those in Table 2 in the article. The vector x takes values in $\{[0,0,0]\}$ for Panels A and C, and in X'' in Panel D. See Section 5.1 in the article for the definition of X''. Values of x are indicated on the horizontal axis in Panel B.

bound of BD. This is because if the observed level of correlation in actions is high even if the deterministic part of players' payoffs is very different, then the value of correlation in payoff types cannot be too low. Symmetrically, values of covariates that generate identical payoffs are most informative about the upper bound on the correlation parameter ρ.

Panel C shows bounds on parameters implied by the inequalities LB and UB for different values of correlation ρ_0 in the data generating process. The upper bound on ρ in $\cap_{y \in Y} BD(y)$ is sharp (it coincides with ρ_0), but only the moment $P_{(1,1)}$ generates a non-trivial lower bound for most values of ρ_0. This is not surprising: if we observe a certain frequency of duopolies, it must be the case that correlation in payoff types is not too low. Panel D exemplifies how the bounds on parameters implied by the inequalities shrink as the amount of variation in x increases. As x takes values on a wider support, the moments $P_{(0,1)}$ and $P_{(1,0)}$ start being informative on the lower bound for ρ, and the set of values compatible with the inequalities becomes reasonably small.

We remark that the set $\cap_y BD\,(y)$ need not be a subset of the projection of Θ_I^{BCE} onto the direction of ρ: to obtain $\cap_y BD\,(y)$ we have assumed a point identified θ_π and we only use part of the information contained in the model, whereas Θ_I^{BCE} considers joint sharp identification of the full vector of parameters. However, the figure provides reassurance that the structure of the model - together with moments of the *joint* distribution of outcomes - have significant identifying power with respect to the parameter θ_ε that summarizes the distribution of payoff types.

H Additional Tables and Figures

We report in Table A2, Panel A summary statistics for the data used in our application. In Panel B we report coefficient estimates for linear regressions and ordered probit regressions of market structure outcomes on market-level covariates. See Section 6.1 in the article for more discussion.

Table A2: Descriptive Statistics and Regressions

Panel (A): Demographics of Local Grocery Markets

Variable	Mean	Std Dev.	Median	Max	Min
Large Mall in Market	0.130	0.337	0	1	0
421 *Markets with no Large Malls*:					
Population	$44,629.22$	$40,341.88$	$31,730$	$297,510$	$3,276$
Surface, in km^2	329.90	242.72	275.72	$1,969.64$	25.19
Tax Income Per Capita, in EUR	$13,223.8$	$1,730.34$	$13,204.92$	$18,288.90$	$8,020.68$
# of Supermarkets	1.46	1.95	1	16	0
# of Players in Market	0.85	0.93	1	3	0
63 *Markets with Large Malls*:					
Population	$117,614.10$	$56,195.42$	$103,925$	$249,852$	$35,768$
Surface, in km^2	447.84	377.92	359.95	$2,243.54$	95.33
Tax Income Per Capita, in EUR	$14,411.47$	$1,650.48$	$14,475.88$	$18,627.36$	$10,333.89$
# of Supermarkets	3.77	2.89	3	13	0
# of Players in Market	1.58	0.87	2	3	0

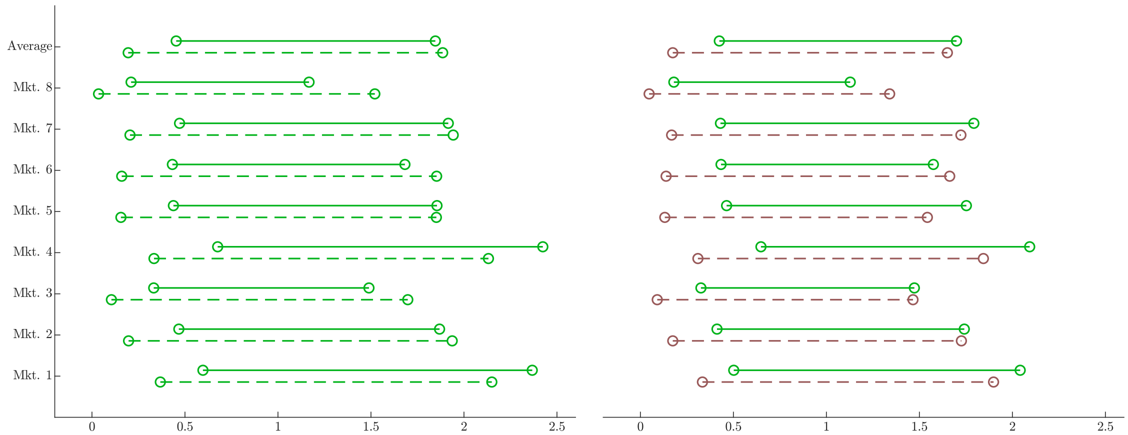
Panel (B): Regressions of Market Structure on Presence of Large Malls

Model	Linear Regression	Ordered probit	Linear Regression	Ordered probit
Variable	# of Supermarkets		# of Players in Market	
	(I)	(II)	(III)	(IV)
Large Mall in Market	-0.437	-0.222	-0.150	-0.242
	(0.278)	(0.165)	(0.145)	(0.175)
Market Size	3.764	2.658	1.213	1.766
	(0.236)	(0.158)	(0.109)	(0.143)
Constant	0.167		0.022	
	(0.378)		(0.230)	
N	484	484	484	484
R^2	0.677	0.255	0.434	0.225

Note: Panel A reports market-level descriptive statistics for the 484 markets used for the application in Section 6 of the article. Panel B reports coefficient estimates and standard errors (in parenthesis) from linear regressions (columns I and III) and ordered probit models (columns II and IV). The dependent variable is the number of supermarkets of at least 1500 m^2 in column I and II, or the number of supermarket players in column III and IV. Market size is the product of population and log of tax income per capita. All regressions include fixed effects for 13 administrative regions. Values of R^2 refer to McFadden's pseudo-R^2 for the ordered probit regressions.

Figure A2 represents intervals I_W^x for the expected number of entrant, considering both a variable and a fixed latent information approach

Figure A2: Expected Number of Entrants

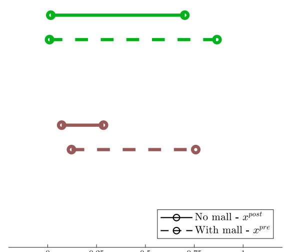


(A) *Variable* Latent Information (B) *Fixed* Latent Information

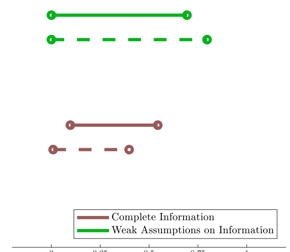
Note: We represent predicted intervals I_W^x (on the left) and $\tilde{I}_{\tilde{W}}^{x,x'}$ (on the right) for the expected number of entrants. The interval I_W^x is computed with variable latent information, while $\tilde{I}_{\tilde{W}}^{x,x'}$ is computed with fixed latent information. Each figure represents intervals I_W^x and $\tilde{I}_{\tilde{W}}^{x,x'}$ as solid line segments for x^{post} and as dashed lines for x^{pre}. Segments for different markets and average values are stacked vertically.

Figure A3 includes market-by-market details for Figure 4 of the main text.

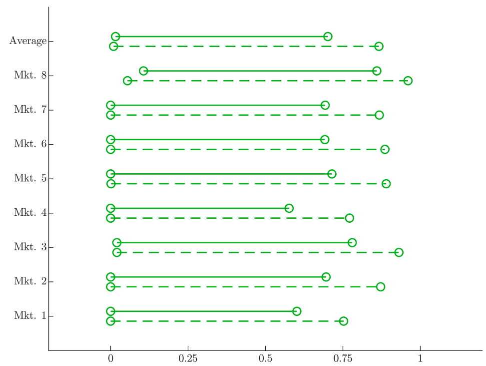
Figure A3: Probabilities of Market Structure Outcomes With and Without Malls



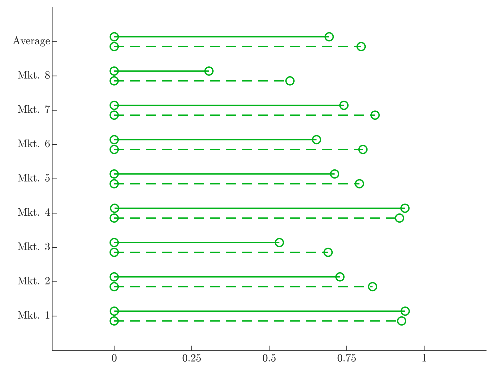
(A) Average Interval I_W^x for the Probability of No Entrants



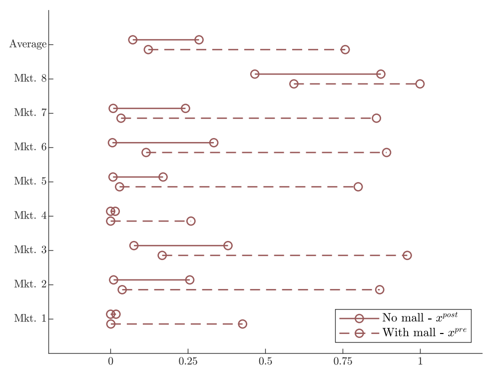
(B) Average Interval I_W^x for the Probability of At Least Two Entrants



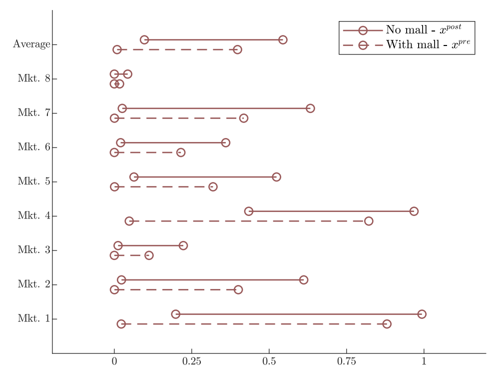
(C) Weak Assumpt. on Information



(D) Weak Assumpt. on Information



(E) Complete Information



(F) Complete Information

Note: The figure reproduces Panels A and B of Figure 4 in the article, and adds a market-by-market breakdown of the intervals I_W^x under two different assumptions on information. The green lines depict intervals obtained for the model with weak assumptions on information; the red lines at refer to the model with complete information. Intervals are represented as solid line segments for x^{post} and as dashed lines for x^{pre}. Figures A, C and E represent intervals for the outcome: "No entrants"; figures B, D and F represent intervals for the outcome: "At least two entrants."

Tables A3 and A4 include a detailed breakdown across markets and across values of covariates x^{pre} and x^{post} of the results summarized in Table 4 in the article.

Table A3: Relative Size of I_W^x and $I_W^x(\hat{\theta}_0)$

	Variable latent info ratio		Fixed latent info ratio	
	No Entrants	Two+ Entrants	No Entrants	Two+ Entrants
Average	**0.74**	**0.71**	**0.68**	**0.58**
Mkt. 1	0.67	0.68	0.65	0.62
Mkt. 2	0.80	0.67	0.75	0.52
Mkt. 3	0.78	0.76	0.74	0.53
Mkt. 4	0.78	0.70	0.71	0.60
Mkt. 5	0.79	0.76	0.67	0.65
Mkt. 6	0.73	0.69	0.71	0.64
Mkt. 7	0.81	0.68	0.82	0.50
Mkt. 8	0.56	0.74	0.37	0.55

Note: We report ratios $|I_W^{post}(\hat{\theta}_0)|/|I_W^{post}|$ and $|\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)|/|\tilde{I}_{\tilde{W}}^{post}|$ for two outcomes of interest: observing no entrants, and observing at least two entrants. Intervals $I_W^{post}(\hat{\theta}_0)$ and $\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)$ are computed for the value $\hat{\theta}_0$ which minimizes the empirical criterion function.

Table A4: Relative Size of Variable and Fixed Latent Information Intervals

	Fixed to Variable latent info ratio $\lvert\tilde{I}_{\tilde{W}}^{post}(\hat{\theta}_0)\rvert/\lvert I_W^{post}(\hat{\theta}_0)\rvert$				
	No Entrants	Two+ Entrants	Coop Entry	Ita Entry	N of entrants
Average	**0.88**	**0.78**	**0.91**	**0.54**	**0.82**
Mkt. 1	0.96	0.86	0.98	0.72	0.91
Mkt. 2	0.88	0.75	0.86	0.47	0.82
Mkt. 3	0.93	0.66	0.96	0.52	0.70
Mkt. 4	0.92	0.84	0.98	0.54	0.70
Mkt. 5	0.77	0.79	0.92	0.48	0.83
Mkt. 6	0.95	0.91	0.96	0.77	0.88
Mkt. 7	0.98	0.73	0.92	0.38	0.86
Mkt. 8	0.61	0.73	0.70	0.46	0.69

Note: We report ratios of intervals $|\tilde{I}_{\tilde{W}}^{x,x'}(\hat{\theta}_0)|/|I_W^x(\hat{\theta}_0)|$ for each market and on average. All intervals are computed for the value $\hat{\theta}_0$ which minimizes the empirical criterion function.

References

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

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

Bergemann, D., B. A. Brooks, and S. Morris (2022): "Counterfactuals with Latent Information," *American Economic Review*, 112, 343–368.

Bergemann, D. and S. Morris (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

Bugni, F. A. (2016): "Comparison of inferential methods in partially identified models in terms of error in coverage probability," *Econometric Theory*, 32, 187–242.

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

Galichon, A. and M. Henry (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, 78, 1264–1298.

Kvasnica, M., P. Grieder, M. Baotić, and M. Morari (2004): "Multi-parametric toolbox (MPT)," in *International workshop on hybrid systems: Computation and control*, Springer, 448–462.

Su, C.-L. (2014): "Estimating discrete-choice games of incomplete information: Simple static examples," *Quantitative Marketing and Economics*, 12, 167–207.

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