Partial Identification Under Iterated Strict Dominance[*]

Francisco Garrido[†]

☠☣☢

PRELIMINARY AND INCOMPLETE

☠☣☢

Use this link to leave anonymous feedback on this or any other of my projects.

2024-02-27

Abstract

I use an Iterated Strict Dominance (ISD) argument to build bounds on the distribution of outcomes of games and use them to pin down an identified set for the parameters of interest. These bounds (ISD Bounds) are robust to equilibrium multiplicity, pure and mixed, and to any non-equilibrium play as long as it is consistent with ISD. Furthermore, ISD Bounds apply to games of complete or incomplete information, with discrete or continuous actions of any dimensionality, as well as games with unobserved heterogeneity. To maximize the "bite" of the ISD Bounds, I introduce Strategically Monotonic Supermodular Games, i.e., games of strategic complements/substitutes where players' payoffs are supermodular in their actions. I show that ISD rules out large swaths of the strategy set for this type of game via an easy-to-compute sequence of best-response iterations. Moreover, I show that for binary games the resulting identified set is sharp. Finally, I apply ISD Bounds to an entry game for the airline industry and use the estimates to evaluate the impact of the proposed merger between JetBlue and Spirit. This exercise proves that ISD Bounds are informative about the parameters of interest in practical applications.

Keywords: Partial identification, probability bounds, ISD, supermodularity, strategic substitutes, strategic complements.

JEL codes: L13, D43, C72.

---

[*]For useful comments, suggestions and discussions, I am indebted to Felipe Brugués, Federico Ciliberto, Andrew Kearns (2023 IIOC discussant), Nathan Miller, Adrian Rubli, Cristian Sánchez, José Tudón, Frank Verboven, as well as seminar participants at UT Austin, OSU, KU Leuven, the Latin American Meeting of the Econometric Society (Lima, 2022), EEA Conference (New York, 2023), MEA Conference (Cleveland, 2023), IIOC (Washington, 2023), IAAE (Oslo, 2023), Econometric Society - North American Winter Meeting (San Antonio, 2024). I gratefully acknowledge financial support from Asociación Mexicana de Cultura, A.C.

[†]Department of Business Administration, ITAM, Río Hondo 1, Ciudad de México 01080, México. Email: fagarrido@gmail.com. Phone: + 52 55 5628 4000

1 Introduction

In this paper I use an Iterated Strict Dominance (ISD) argument to build bounds on the (distribution of) outcomes of games and use them to pin down an identified set of the parameters of interest. These bounds (hereinafter, ISD Bounds) are extremely general in that they allow discrete, continuous, and discrete-continuous strategies of any dimensionality, and can account for arbitrary informational structures (i.e., complete and incomplete information). Furthermore, because they are built on the concept of ISD, which is weaker than (Bayes) Nash equilibrium, ISD Bounds are robust to equilibrium multiplicity, both in pure and mixed strategies, as well as various forms of non-equilibrium play. In particular, ISD Bounds are valid as long as players choose serially undominated strategies.

To maximize the bite of ISD Bounds I introduce a class of games which I dub Strategically Monotonic Supermodular Games (SMSGs). These are games where player's payoff are supermodular on their own actions, and best responses exhibit *strategic monotonicity*, i.e., for any pair of firms, actions are either strategic complements or strategic substitutes.[1] As I argue in the paper, SMSGs are a natural match for ISD Bounds as in these games ISD is informative, i.e., it eliminates large swaths of the strategy set, and it is easy to compute through a best response iteration.

To show that ISD has bite on SMSGs, I generalize a classic result from the literature of supermodular games due to Milgrom and Roberts (1990), and generalized to incomplete information environments by Van Zandt and Vives (2007). Assuming strategic complementarity, these papers show that all strategy profiles that survive ISD lie between a lower and an upper bound. I generalize this result to allow for any form of strategic monotonicity. In consequence, ISD Bounds apply to many important strategic environments that are not covered by the standard theory of supermodular games, such as Cournot games, entry games, and capacity investment games, all of which exhibit strategic substitutability.[2]

To understand how ISD Bounds in SMSGs work, consider an incomplete information entry game indexed by a parameter θ, where each player, f, receives an i.i.d. shock, ϵ_f, and independently chooses a strategy $\sigma_f(\epsilon_f) \in \{0, 1\}$, where 0 implies *no entry* and 1 implies *entry.* If the game is an SMSG for every θ, I show that there are strategy profiles σ^L and σ^H such that $\sigma^L(\epsilon; \theta) \leq \sigma^H(\epsilon; \theta)$ for all ϵ, and such that all strategies that survive ISD lie between σ^L and σ^H. ISD Bounds, then, result from the observation that for any strategy profile σ that survives ISD, and for any outcome y:

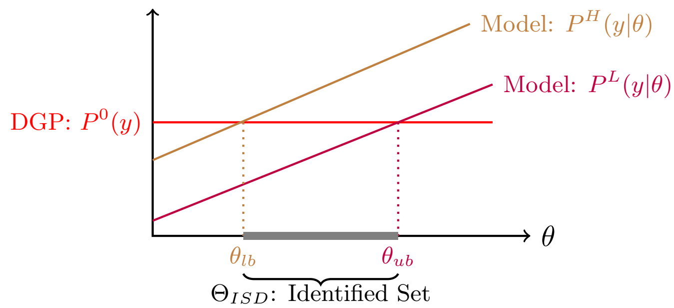
$$\underbrace{Pr(y \leq \sigma^L(\epsilon; \theta))}_{\equiv P^L(y|\theta)} \leq Pr(y \leq \sigma(\epsilon; \theta)) \leq \underbrace{Pr(y \leq \sigma^H(\epsilon; \theta))}_{\equiv P^H(y|\theta)}$$

The terms on the left and the right of these inequalities represent the lower and upper ISD Bounds,

[1] More precisely, a firm's payoffs need to exhibit supermodularity in own actions, and either increasing differences or decreasing differences with respect to its action and that of each its competitors.

[2] I should note that the comparative statics results from Milgrom and Roberts (1990) and Van Zandt and Vives (2007) do not generalize to the case of strategic monotonicity. These results, however, are irrelevant to the problem at hand.

Figure 1: Identified Set Example



Note: $P^H(y|\theta)$ and $P^L(y|\theta)$ are, respectively, are the upper and lower ISD bounds as functions of θ, and for a given outcome y. If $P^L(y|\theta) > P^0(y)$ of $P^L(y|\theta) < P^0(y)$, θ violates the ISD Bounds and therefore cannot be the "real" θ.

respectively. These terms bound the term in the middle which represents any distribution over outcomes consistent with players choosing serially undominated strategies.

Having built the ISD Bounds, the intuition behind the identified set lies in the idea that there is a data generating process (DGP), P^0, which corresponds to the "real world" distribution over outcomes.[3] If firms only play serially undominated strategies, any θ such that $P^L(y|\theta) > P^0(y)$ or $P^H(y|\theta) < P^0_f(y)$ cannot be part of the DGP. Hence, the ISD identified set is the set of θ's such that P^0 lies between P^L and P^H. Figure 1, depicts this intuition.

A natural question when dealing with set identification is how informative are the bounds. To address this issue, I provide two results for the case of binary games (i.e., games with binary actions). First, I show that under mild restrictions the ISD identified set is sharp relative to a DGP where players choose serially undominated strategies. Roughly speaking, this means that a parameter θ is in the ISD identified set *iff* there are strategies that survive ISD given θ that can match the DGP. Sharpness of the ISD identified set implies that the ISD Bounds are using all the information available in the ISD assumption, hence to get a smaller identified set one needs to assume a stronger solution concept.

Second, I find necessary and sufficient conditions that guarantee that no vector of parameters generates trivial ISD Bounds (i.e., $P^L(y|\theta) = 0$ and $P^H(y|\theta) = 1$ for all y).[4] These conditions emphasize the role that variables that uniformly shift payoffs across games play in narrowing down the identified set. For example, a large variation in *market size* in an entry game guarantees that whenever ISD Bounds are trivial in, say, a small market, ISD Bounds will be non-trivial in a large market.

To assess the performance of ISD Bounds I perform several Monte Carlo experiments on a standard entry game. First I focus on the case of incomplete information and I find that ISD Bounds do a good job of narrowing down the set of possible parameters. Second, I focus on the complete information case and compare ISD Bounds to the bounds in Ciliberto and Tamer (2009) (CT) and the bounds in Fan and Yang (2022) (FY). In an analytical example, I show that CT

[3] More precisely, $P^0(y) = Pr(\sigma^0(\epsilon) \geq y)$, where σ^0 is the "real world" strategy profile.

[4] Note that such θ lies in the identified set not because it fits the data, but because it does not rule anything out.

Bounds are tighter than ISD Bounds, which are tighter than FY Bounds. This should not be surprising since CT Bounds are based on a Nash equilibrium solution concept which is stronger than ISD, while FY Bounds are built using one strict dominance round, which is weaker than full blown ISD. In simulations, however, I find that the performance of ISD Bounds is comparable to CT bounds, and better than FY Bounds.

Finally, to put ISD Bounds to test, I propose and estimate an entry model with price competition for the airline industry, and use it to evaluate the proposed merger between JetBlue and Spirit which was blocked by a federal judge on January 2024. This exercise shows the practicality of ISD Bounds for applied research.

The rest of the paper is organized as follows. Section 1.1 reviews the relevant literature and places the current paper within it. Section 2 presents the model, introduces SMSGs, and shows that ISD has bite in SMSGs. Section 3 derives ISD Bounds the ISD identified set, while Section 4 explores the identifying power of ISD Bounds and pins down a source of cross-sectional variation that guarantees that the ISD Identified Set is bounded. Section 5 analyses the performance of ISD Bounds through Monte Carlo exercises, comparing it to other bounds proposed in the literature. Section 6 provides some background on the industry and details on the proposed merger, while Section 7 describes the data I use in this empirical exercise. Section 8 presents the estimation results and counterfactual simulations. Finally, Section 9 provides some concluding remarks.

1.1 Literature Review

This paper is related to two strands that have roughly run on independent lanes. The literature on estimation of discrete games of *complete information*, and the literature focusing on estimation of discrete games of *incomplete information*.

1.1.1 (Mostly) Complete Information Games

The problem of model incompleteness as described by Tamer (2003), has been a common thread throughout the literature studying estimation of discrete games of complete information. The early examples in this literature, such as Bresnahan and Reiss (1991), Berry (1992), and Mazzeo (2002), bypassed the problem of equilibrium multiplicity by making strong homogeneity assumptions on firms' payoffs that guaranteed that all equilibria could be mapped into a single outcome (e.g., number of firms), for which the model makes a unique prediction.

Later papers dealt with this issue using two broad approaches. The first approach consists of completing the model with an equilibrium selection mechanism and either assuming that it is known (e.g., Jia (2008), Li et al. (2018)) or estimating it from the data (e.g., Bajari et al. (2010)). This strategy is attractive because it brings us back to the world where standard estimation techniques work and point identification holds. The problem, however, is that economic theory provides little guidance when it comes to equilibrium selection, making any assumption related to the equilibrium selection mechanism hard to justify.

The second approach, the one that this paper takes, gives up on point identification and rather focuses on identifying a set for the parameters of interest. This approach was pioneered by Tamer (2003) and Ciliberto and Tamer (2009) (CT), who build the identified set by putting bounds on the probability of observing an outcome. In particular, the probability of observing an outcome y must be higher than the probability that said outcome is the *unique* Nash equilibria, and lower than the probability that it is *a* Nash equilibria. In this strand, Fan and Yang (2020) (FY) propose building the identified set using one round ISD, and Aradillas-Lopez and Tamer (2008) study the identification power of rationalizability as a solution concept.

Aradillas-Lopez and Tamer (2008) is perhaps the closest paper to the present one. That paper studies identification of k-level rationality in 2×2 games of complete and incomplete information, while imposing no assumptions on player's beliefs (beyond common priors and what is implied by k-level rationality). The present paper can be seen as a generalization of these ideas to much more flexible settings.

Relative to CT and FY bounds, ISD Bounds are neither more nor less general. CT bounds, on the one hand, are tighter but harder to compute, specially for large games, and require the data to be generated by equilibrium play. FY bounds, on the other hand, are wider but can be applied to a more general class of games and are easier to compute. Relative to CT Bounds, the main advantage of ISD Bounds is their tractability, their robustness to non-equilibrium play, and the fact that they can be applied to imperfect information environments. Relative to FY Bounds, the main advantages of ISD Bounds is that they are more informative while still being tractable, and that (at least in binary games) they produce a sharp identified set.

The present paper is also similar to Aradillas-Lopez (2011) and Aradillas-López and Rosen (2022) in using shape restrictions on payoffs, and restrictions on the action set to pin down an identified set of the parameters of interest. By restricting their attention to ordered actions and making appropriate concavity and increasingness assumptions, those papers are able to pin down an identified set based on Nash equilibrium conditions. The present paper, in contrast, allows for non-equilibrium play while making much weaker assumptions on the game's structure that allows me to estimate a more general class of games than the ones considered in these two papers. This generality, however, comes at the cost of pinning down a wider identified set.

As mentioned above, I argue that ISD Bounds are particularly useful in estimating SMSGs. Many of the static games estimated in the empirical literature are instance of SMSGs, and therefore can be estimated using the method I advance here. The models in Bresnahan and Reiss (1990), who estimate entry game for isolated retail and professional markets, Berry (1992), Tamer (2003), Ciliberto and Tamer (2009) all of whom estimate entry games for the airline industry, are all instances of SMSGs with strategic substitutes, and Ciliberto and Jäkel (2021) who estimate a game in which firms decide whether or not become exporters, are all examples of SMSGs. More recently, Wollmann (2018) estimates a two-stage model for the truck industry in which players can choose which truck varieties to offer and compete in prices. Although this model cannot be shown to be supermodular, as the payoffs depend on the reduced form variable profits in the pricing stage, economic intuition strongly suggest that strategic substitution should hold (i.e., the profit gain from introducing a variety is decreasing on the varieties of my competitors). Furthermore, supermodularity can be verified numerically from the pricing stage estimates.

A number of empirical papers explicitly exploit the theory of supermodular game to solve (and estimate) models with large strategy sets that would be computationally infeasible otherwise. Most prominently, Jia (2008) estimates an entry model for Wal-Mart and Kmart with spill over effects across markets. To solve this model, she shows that the duopolistic game can be written as a supermodular game, and proceeds with estimation assuming a known equilibrium selection mechanism. This trick, however, applies only to two-player games, so her methodology does not generalize to games with three or more players. Other empirical papers that exploit supermodularity are Uetake and Watanabe (2020) who study entry and merger decisions in a supermodular matching model, and Ackerberg and Gowrisankaran (2006) who study study technology adoption with network externalities. Both in the banking industry.

In all these papers the underlying model can be thought of as an SMSG, and therefore can be estimated using the approach I outline here. Furthermore, the approach I outline makes it feasible to relax the strong assumptions these papers made on equilibrium selection, information structure, equilibrium play.

The idea of exploiting supermodularity to estimate empirical models is not new to this paper. Molinari and Rosen (2008) and Uetake and Watanabe (2013) both proposed using the theory of supermodular games for set identification. However, due to its focus on SMSGs rather than only supermodular games, this paper is able to consider the much broader class of games characterized by strategic monotonicity. This generalization is particularly important in IO since strategic substitutability is likely more common than complementarity in empirical research.

1.1.2 Incomplete Information Games

As opposed to the complete information case, the literature on estimation of discrete games of incomplete information has, until recently, largely ignored the problem of model incompleteness in estimation. The reason for this asymmetry is that, under incomplete information, the econometrician and the players face the same uncertainty about other players' actions. As a result, by independently estimating the conditional choice probabilities of each player, the econometrician learns the distribution over actions that each player is facing, and can use this to estimate each player's payoffs as a single agent problem using the methods developed by Hotz and Miller (1993) and Aguirregabiria and Mira (2002) for single agent dynamic settings.

This approach, which is widely used in the literature (e.g. Seim (2006), Draganska et al. (2009), Atal et al. (2022)), rests on the assumption that all the data available comes from the same equilibrium, and that there is no unobserved heterogeneity. However, de Paula and Tang (2012), for static games, and Otsu et al. (2016) and Otsu and Pesendorfer (2022), for dynamic environments, propose statistical tests for these assumptions and find that, in commonly used datasets, the assumptions are violated.

The problem of model incompleteness in games of incomplete information is an area of active research. Two prominent efforts to deal with this issue are Aguirregabiria and Mira (2019), who study the problem of (point) identification in games with incomplete information and unobserved heterogeneity while estimating an equilibrium selection mechanism, and Otsu and Pesendorfer (2022) who treat equilibrium multiplicity as a market specific correlated latent variable. As in this paper, they provide results for set identification. As compared to these papers, the current work deals with the problem of equilibrium multiplicity in a more tractable way, by imposing bounds on (the distribution of) outcomes, and making fewer assumptions on the distribution of private shocks.

This paper is also related to work by Grieco (2014) and Magnolfi and Roncoroni (2020) both of whom study identification under weak informational assumptions. In particular, Grieco (2014) derives an exclusion restriction that allows him to set identify the parameters that control the informational structure behind two-player binary games. Magnolfi and Roncoroni (2020), on the other hand, build upon the partial identification results of Beresteanu et al. (2011) for models with convex predictions, and derive an identified set that is robust to any informational assumption. The ISD Bounds I propose in this paper are closer to those in Grieco (2014) in that they hold conditional on an informational structure. Nevertheless, ISD Bounds are more general in that they apply to games with any number of players and more flexible action sets.

1.1.3 Moment Inequalities/Revealed Preferences

A third popular route to estimation in discrete games is due Pakes et al. (2015), and has become known as the moment inequality approach.[5] Their approach is based on the idea that, if the data are generated by a Nash Equilibrium, then unilateral deviations from the observed actions should be unprofitable for the deviating firm. This reasoning generates profit inequalities that lend themselves for set identification, as any parameter vector that violates these inequalities cannot have generated the data.

The moment inequality approach has gained traction in the empirical literature due to its relative simplicity and tractability (e.g., Berry et al. (2016), Eizenberg (2014), Ellickson et al. (2013), Bontemps et al. (2023), and Wollmann (2018)). Nevertheless, the probability bounds approach has some advantages that can make it preferable. For example, as opposed to the moment inequalities approach, identified sets based on ISD Bounds need not assume that the observed data is generated by equilibrium play. Additionally, the revealed preference approach leaves parts of the game largely unspecified, which makes it map the estimates back to an structural model to perform counterfactuals with.

[5]This is a somewhat unfortunate name, as probability bounds like the ones proposed in this paper are basically moment inequalities too.

2 The Model, SMSGs, and ISD

In this section, I provide the building blocks of the model and introduce Strategically Monotonic Supermodular Games (SMSGs). Then I show that ISD has a bite in SMSGs.

2.1 Model Set-Up

Consider a finite set of players (firms), $\mathcal{F}$, indexed by f, who simultaneously choose a vector y_f from a compact action set, $\mathcal{Y}_f \subseteq \mathbb{R}^{\dim(\mathcal{Y}_f)}$, after receiving a private signal/shock, $\epsilon_f \in \mathcal{E}_f \subseteq \mathbb{R}^{\dim(\mathcal{E}_f)}$. Ex-post profits are given by:

$$\pi_f(y_f, y_{-f}, \epsilon_f; x, \theta, \xi)$$

where, as is standard, $y_{-f} = (y_g)_{g \neq f}$ is a vector containing f's competitors' actions, and where the vector of private shocks, $\epsilon = (\epsilon_f)_{f \in \mathcal{F}}$, follows a (possibly degenerate) joint distribution $G(\epsilon|x, \theta, \xi)$, which is common knowledge.

Each tuple (x, θ, ξ) indexes a different realization of the game, which I refer to as the (x, θ, ξ)-game. Here, $x \in \mathcal{X} \subseteq \mathbb{R}^{\dim(\mathcal{X})}$ represents a vector of observables, $\theta \in \Theta \subseteq \mathbb{R}^{\dim(\Theta)}$ is the vector of parameters of interest, and $\xi \in \Xi \subseteq \mathbb{R}^{\dim(\Xi)}$ is a vector of *common-knowledge unobservables*—variables known to firms but unobservable to the econometrician. Throughout the paper, I assume that π_f is continuous in ξ. Also, for brevity, hereafter I omit dependence to (x, θ, ξ) unless it may cause confusion.

A strategy for player f, denoted as σ_f, is measurable function mapping f's private information, ϵ_f, to a distribution over actions, i.e., $\sigma_f : \mathcal{E}_f \to \Delta(\mathcal{Y}_f)$ where $\Delta(\mathcal{Y}_f)$ denotes the set of distributions over $\mathcal{Y}_f$. If there is complete information, i.e., $\epsilon = \varnothing$, then σ_f is simply an element of $\Delta(\mathcal{Y}_f)$.[6] Let Σ_f represent the set of strategies of f. A strategy profile is a collection of strategies, one for each player: $\sigma = (\sigma_f)_f \in \Sigma \equiv \times_f \Sigma_f$.

For any strategy profile adopted by f's competitors, represented as $\sigma_{-f} = (\sigma_g)_{g \neq f}$, f's interim payoff is given by:

$$\Pi_f(y_f, \sigma_{-f}, \epsilon_f) = \int_{\mathcal{E}_{-f}} \pi_f(y_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) dG(\epsilon_{-f}|\epsilon_f) \tag{1}$$

where $G(\epsilon_{-f}|\epsilon_f)$ is the conditional distribution of $\epsilon_{-f} = (\epsilon_g)_{g \neq f}$.

Following the approach by Van Zandt and Vives (2007), I employ interim payoffs, rather than ex-ante payoffs, to define best responses. Specifically, σ_f is a best response to σ_{-f} if:

$$\Pi_f(\sigma_f(\epsilon_f), \sigma_{-f}, \epsilon_f) \geq \Pi_f(\sigma'_f(\epsilon_f), \sigma_{-f}, \epsilon_f), \forall \epsilon_f \in \mathcal{E}_f, \forall \sigma'_f \in \Sigma_f \tag{2}$$

Similarly, I define strict dominance in terms of interim payoffs rather than ex-ante ones. This has

[6] Throughout the paper, I use y_f, and variations therein, to denote an arbitrary non-random element of $\mathcal{Y}_f$, and treat $\sigma_f(\epsilon_f)$ as a random variable.

the advantage that it allows us to distinguish between strategies that are ex-ante equally attractive. For instance, consider two strategies, σ_f and σ'_f, equal everywhere except for a zero-measure subset of $\mathcal{E}_f$, in which σ_f is preferred to σ'_f. Ex-ante, these two strategies would be deemed equally good. However, in an interim evaluation, σ_f would be preferred to σ'_f because there are values of ϵ_f for which σ_f fares strictly better, even if this contingencies have zero probability.

Definition 1 (Strict Dominance and Strictly Dominated Strategies)**.** *Let $\tilde{\Sigma} \subseteq \Sigma$. Strategy $\sigma_f \in \tilde{\Sigma}_f$ strictly dominates strategy $\sigma'_f \in \tilde{\Sigma}_f$ relative to $\tilde{\Sigma}$ if:*

$$\Pi_f(\sigma_f(\epsilon_f), \sigma_{-f}, \epsilon_f) \geq \Pi_f(\sigma'_f(\epsilon_f), \sigma_{-f}, \epsilon_f), \forall \epsilon_f \in \mathcal{E}_f, \forall \sigma_{-f} \in \tilde{\Sigma}_{-f}$$

with strict inequality for at least one $\epsilon_f \in \mathcal{E}_f$.

Strategy $\sigma'_f \in \tilde{\Sigma}_f$ is strictly dominated relative to $\tilde{\Sigma}$ if there exists another strategy $\sigma_f \in \tilde{\Sigma}_f$ that strictly dominates it.

To close the model, rather than introducing a solution concept, I assign beliefs to each player and assume that they behave optimally given their beliefs. This allows me to represent multiple solution concepts—like (Bayes) Nash Equilibrium or ISD— within a unified framework.

Formally, I denote f's belief about g's behavior by a strategy $\rho^f_g \in \Sigma_g$.[7] and collect f's beliefs in $\rho^f = (\rho^f_g)_{g \neq f}$. Firm f's behavior is the result of implementing σ^ρ_f, which is a best response to the belief ρ^f as per (2). If the best response to ρ^f is not unique, I assume that f selects σ^ρ from the set of maximizers through some undetermined mechanism.

Before introducing the SMSGs, a quick comment regarding the informational structure of the model is in order. The model can accommodate arbitrary informational structures through the private shocks/signals, ϵ, and their distribution, G The *complete information* case, for example, can be represented by a degenerate distribution G, i.e., $\epsilon = \varnothing$. In this case, the randomness of the outcomes is driven by the randomness (from the perspective of the econometrician) of the common-knowledge unobservables, ξ.

Other informational structures can be represented by letting $\epsilon_f = (\tilde{\epsilon}_f, \tau_f)$ where $\tilde{\epsilon}_f$ is the payoff relevant shock and τ_f is a, payoff irrelevant, signal about other player's private information, as in Magnolfi and Roncoroni (2020). For example, the *independent private information* case corresponds to $\tilde{\epsilon}_f \perp \tilde{\epsilon}_{-f}$ and $\tau_f = \varnothing$ for all f. The *privileged information* case, where one player is perfectly informed and the rest only observe their private shocks, can be represented by $\tilde{\epsilon}_f \perp \tilde{\epsilon}_{-f}$ and $\tau_f = \varnothing$ for all f except the privileged party whose signal is $\tau_f = \epsilon_{-f}$. Similarly, the case with *independent partially observed information* corresponds to the case where $\epsilon_f \perp \epsilon_{-f}$ and $\tau_f = \epsilon_{-f} + \varsigma_{-f}$, where ς_{-f} is noise. In this dimension, the present paper can easily accommodate many more informational structures than previous research has allowed for.[8]

[7]We could allow f's belief to be a distribution over Σ_g. All results of the paper follow through.

[8]Other papers that allows for flexible information structures are Magnolfi and Roncoroni (2020), Aradillas-Lopez (2010), and Grieco (2014).

2.2 Strategically Monotonic Supermodular Games and ISD

Here, I introduce a class of games termed *Strategically Monotonic Supermodular Games* (SMSGs) and show that for this type of game ISD is informative, in that it rules out large swaths of the strategy set, and practical, in that it is easy to compute. Consequently, an estimation approach based on ISD is particularly promising for SMSGs.

The main result of this section, Theorem 1, says that in SMSGs there exist strategy profiles, σ^L and σ^H such that any strategy σ that survives ISD lies between σ^L and σ^H in the sense that $\sigma^L(\epsilon) \leq \sigma(\epsilon) \leq \sigma^H(\epsilon)$ for all ϵ, where "$\leq$" represents the standard vector inequality.[9] This result is the main building block for the ISD Bounds I derive in Section 3. To move in this direction, let us begin by investing in some definitions.

Definition 2 (Increasing Differences and Decreasing Difference)**.** *Let $h(z_1, z_2)$ be a function mapping from $\mathcal{Z}_1 \times \mathcal{Z}_2$ to $\mathbb{R}$, where $\mathcal{Z}_j \subseteq \mathbb{R}^{dim(\mathcal{Z}_j)}$ for $j = 1, 2$.*

2.a. ***Increasing Differences (ID):*** *h has increasing differences in (z_1, z_2) if for any distinct $z_1' \geq z_1$, and distinct $z_2' \geq z_2$:*

$$h(z_1', z_2') - h(z_1, z_2') > h(z_1', z_2) - h(z_1, z_2)$$

2.b. ***Decreasing Differences (DD):*** *h has decreasing differences in (z_1, z_2) if for any distinct $z_1' \geq z_1$, and distinct $z_2' \geq z_2$:*

$$h(z_1', z_2') - h(z_1, z_2') < h(z_1', z_2) - h(z_1, z_2)$$

Definition 3 (Complements and Substitutes)**.** *Let $y_{-\{f,g\}} = (y_{f'})_{f' \in \mathcal{F} \setminus \{f,g\}}$.*

3.a. ***Complements:*** *g is f's complement if $\pi_f(y_f, y_g, y_{-\{f,g\}}, \epsilon_f)$ has ID in (y_f, y_g) for all $(y_{-\{f,g\}}, \epsilon_f)$. The set of f's complements is denoted by $C(f) \subseteq \mathcal{F}$.*

3.b. ***Substitutes:*** *g is f's substitute if $\pi_f(y_f, y_g, y_{-\{f,g\}}, \epsilon_f)$ has DD in (y_f, y_g) for all $(y_{-\{f,g\}}, \epsilon_f)$. The set of f's substitutes is denoted by $S(f) \subseteq \mathcal{F}$.*

In Definition 2, ID and DD are notions of complementarity and substitutability, respectively. Intuitively, ID implies that the marginal return of y_f is increasing in y_g, hence the optimal y_f is increasing in y_g. Many games exhibit ID, such as games with complementary investments. Similarly, DD implies that the marginal return of y_f is decreasing in y_g, so the optimal y_f is decreasing in y_g. In IO settings, DD is more common than ID. Games of entry, capacity investment, and Cournot competition, for example, typically exhibit DD.

In Definition 3, a complement (substitute) of firm f is a firm, g, whose actions are strategic complements (substitutes) to f's actions. Note that if g is f's complement, this does not imply

[9] For $v, w \in \mathbb{R}^n$, $v \leq w$ if $v_i \leq w_i$ for all $i = 1, \ldots, n$.

that π_f is in increasing in y_g,[10] nor does it imply that f is g's complement (i.e., the complement relation is not necessarily symmetric). Similarly, if g is f's substitute, this does not imply that π_f is decreasing in y_g,[11] nor does it imply that f is g's substitute (i.e., the substitute relation is not necessarily symmetric).

Before moving to the definition of SMSGs, let us define the concept of a lattice, which is central to the theory of supermodular games which I exploit in this paper.

Definition 4 ((Complete) Lattice). *A set $\mathcal{Z}$ together with a partial order, $\leq$, constitute a lattice if for any $z, z' \in \mathcal{Z}$, $\sup\{z, z'\} \in \mathcal{Z}$ and $\inf\{z, z'\} \in \mathcal{Z}$. Furthermore, the tuple $(\mathcal{Z}, \leq)$ is a* complete lattice *if for every $Z \subseteq \mathcal{Z}$, $\inf\{Z\} \in \mathcal{Z}$ and $\sup\{Z\} \in \mathcal{Z}$.*

Definition 5 (SMSG). *A game is a Strategically Monotonic Supermodular Game if:*

5.a. **Complete Lattice Action Set:** *The action set, $\mathcal{Y}_f \subseteq \mathbb{R}^{\dim(\mathcal{Y}_f)}$, together with the standard vector inequality, "$\leq$", conform a complete lattice for all $f \in \mathcal{F}$.*[12] *Furthermore, $\mathcal{Y}_f$ is compact for all $f \in \mathcal{F}$.*

5.b. **Order Upper Semi-Continuity** *The profit function, π_f, is order upper semi-continuous in y_f. Formally, for any totally ordered set $O \subseteq \mathcal{Y}_f$:*[13]

$$\limsup_{y_f \in O, y_f \downarrow \inf(O)} \pi_f(y_f, y_{-f}, \epsilon_f,) \leq \pi_f(\inf(O), y_{-f}, \epsilon_f,)$$
$$\limsup_{y_f \in O, y_f \uparrow \sup(O)} \pi_f(y_f, y_{-f}, \epsilon_f,) \leq \pi_f(\sup(O), y_{-f}, \epsilon_f,)$$

for all $y_{-f} \in \mathcal{Y}_{-f}$, all $\epsilon_f \in \mathcal{E}_f$, and all $f \in \mathcal{F}$.

5.c. **Supermodularity:** *The profit function, π_f, is supermodular in y_f, i.e., for any $y_f, y'_f \in \mathcal{Y}_f$:*

$$\pi_f(\sup\{y_f, y'_f\}, y_{-f}, \epsilon_f) + \pi_f(\inf\{y_f, y'_f\}, y_{-f}, \epsilon_f) \geq \pi_f(y_f, y_{-f}, \epsilon_f) + \pi_f(y'_f, y_{-f}, \epsilon_f)$$

for all $y_{-f} \in \mathcal{Y}_{-f}$, all $\epsilon_f \in \mathcal{E}_f$, and all $f \in \mathcal{F}$.

5.d. **Strategic Monotonicity:** *For all $f, g \in \mathcal{F}$, either g is f's complement, i.e., $f \in C(f)$, or g is f's substitute, i.e., $g \in S(f)$.*

Point 5.a. of the definition is necessary to exploit the supermodular games infrastructure advanced by Milgrom and Roberts (1990) for games of complete information, and Van Zandt

[10] Say f and g produce differentiated goods, engage in Bertrand competition, and have to decide whether to adopt a cost-saving technology or not. If g adopts the technology it makes f worse off (f is harmed by the lower cost of g). Nevertheless, g adopting the technology may increase f's incentive to adopt, so that adoption decisions are strategic complements.

[11] For example, in a public good financing game, π_f might be increasing in y_g, i.e., the more g invests in the public good the higher the benefit for f, and π_f may have DD in (y_f, y_g), i.e., the more g invests in the public good the lower the marginal return for f to do so.

[12] Note that this definition allows $\mathcal{Y}_f$ to include $\{-\infty, +\infty\}$. Naturally, for this to work payoffs need to be well defined at infinity.

[13] A totally ordered (sub)set $C \subseteq \mathcal{Y}_f$ is a subset of $\mathcal{Y}_f$ such that for any $y_f, y'_f \in \mathcal{C}$ either $y_f \geq y'_f$ or $y_f \leq y'_f$.

and Vives (2007) for games of incomplete information. Although most empirical studies satisfy this assumption, it is easy to construct games in which it is violated. For example, consider an entry game with location choice as in Seim (2006). Firms have to choose between *no entry*, *entry in location* A, or *entry in location* B. Letting 1 (0) represent the case where f does (does not) enter a given location the action set is $\mathcal{Y}_f = \{(0,0),(0,1),(1,0)\}$, and it is easy to see that $\sup\{(0,1),(1,0)\} = (1,1) \notin \mathcal{Y}_f$. Point 5.b. is a technical condition necessary to guarantee that f's profit maximization problem has a solution. Order upper semi-continuity is satisfied if π_f is continuous or if the strategy set is discrete.

In point 5.c. of the definition, supermodularity of π_f represents a notion complementarity between the elements of y_f. If y_f is univariate then this condition is trivially satisfied. Otherwise, supermodularity is likely satisfied in cases where there are positive spillover effects between the different elements of y_f. Jia (2008) provides a prominent example of an empirical game exhibiting supermodularity. In her model, opening a Wal-Mart store in any location increases the profitability of opening a store in neighboring locations due to economies of scope in inventory management. Finally, point 5.d. says that there is *Strategic Monotonicity* meaning that each of f's competitors is either f's substitute or f's complement. For $g \in C(f)$ this implies that f's optimal behavior is increasing in y_g, whereas for $g \in S(f)$ this implies that f's optimal behavior is decreasing in y_g. Either way, the pairwise strategic relation is monotonic.

Theorem 1. [ISD in SMSGs] *Let the (x,θ,ξ)-game be an SMSG and say $\sigma \leq \sigma'$ if and only if $\sigma(\epsilon) \leq \sigma'(\epsilon)$ for all ϵ. Consider the following sequence:*

$$
\begin{array}{rcl}
\textbf{Set up:} & & \\
\Sigma^0_{ISD} & = & \Sigma \\
\sigma_f^{H,0} & = & \{\sup\{\mathcal{Y}_f\} : \epsilon_f \in \mathcal{E}_f\} \\
\sigma_f^{L,0} & = & \{\inf\{\mathcal{Y}_f\} : \epsilon_f \in \mathcal{E}_f\} \\
\Sigma^k_{ISD} & = & \{\sigma \in \Sigma : \sigma^{L,k} \leq \sigma \leq \sigma^{H,k}\} \\
\mathcal{Y}^k_f(\epsilon_f) & = & \{y_f \in \mathcal{Y}_f : \sigma_f^{L,k}(\epsilon_f) \leq y_f \leq \sigma_f^{H,k}(\epsilon_f)\} \\
\textbf{Best/Worst Case:} & & \\
\sigma_{-f}^{B,k} & = & (\sigma^{H,k}_{C(f)}, \sigma^{L,k}_{S(f)}) \\
\sigma_{-f}^{W,k} & = & (\sigma^{L,k}_{C(f)}, \sigma^{L,k}_{S(f)}) \\
\textbf{Update:} & & \\
\sigma_f^{H,k} & = & \sup\left\{ \underset{y_f \in \mathcal{Y}_f^{k-1}(\epsilon_f)}{\mathit{argmax}} \quad \Pi_f(y_f, \sigma_{-f}^{B,k-1}, \epsilon_f) : \epsilon_f \in \mathcal{E}_f \right\} \\
\sigma_f^{L,k} & = & \inf\left\{ \underset{y_f \in \mathcal{Y}_f^{k-1}(\epsilon_f)}{\mathit{argmax}} \quad \Pi_f(y_f, \sigma_{-f}^{W,k-1}, \epsilon_f) : \epsilon_f \in \mathcal{E}_f \right\}
\end{array}
\tag{3}
$$

The following holds:

1.a. For each $k = 1, 2, \ldots$, all $\sigma_f \ngeq \sigma_f^{k,L}$ and all $\sigma_f \nleq \sigma_f^{k,H}$ are dominated relative to Σ^{k-1}_{ISD}.

1.b. The set Σ^k_{ISD} contains all strategies that survives k ISD rounds.

1.c. For $k \to \infty$, $(\sigma^{k,L}, \sigma^{k,H}) \to (\sigma^L, \sigma^H)$ with $\sigma^L \leq \sigma^H$. Furthermore, the set:

$$\Sigma_{ISD} = \{\sigma \in \Sigma : \sigma^L \leq \sigma \leq \sigma^H\}$$

contains all strategies that survive ISD.

1.d. If $\sigma^L = \sigma^H$ then the game is dominance solvable, and this strategy profile is the unique (Bayes) Nash Equilibrium.

Proof. See Appendix A. □

Theorem 1 is a generalization of Theorem 5 in Milgrom and Roberts (1990) which focuses only on increasing differences, i.e., strategic complementarity. The present generalization from strategic complementarity to strategic monotonicity is crucial for the practical relevance of the approach to estimation I propose in this paper, as it implies that ISD Bounds can be applied to a much broader class of games than the classic theory of supermodular games considers. Namely, ISD Bounds can be applied to games of strategic substitutability which are likely the norm in industrial organization.

The sequence defined in equation (3) describes a best response iteration that results in the deletion of dominated strategies. Intuitively, the "best case" (resp. "worst case") strategies represent the strategies of f's competitors for which f's best response is maximal, $\sigma^{k,H}_f$ (resp. minimal, $\sigma^{k,L}_f$). Strategic monotonicity and supermodularity guarantee that any strategy that does not lie between $\sigma^{k,L}_f$ and $\sigma^{k,H}_f$ is strictly dominated. In the next subsection I give an intuition for how the proof operates using two entry game examples.[14]

2.3 Two Entry-Game Examples

Here I show the implications of Theorem 1 for two archetypal entry games. The one with independent private information, and the one with complete information.

2.3.1 Independent Private Information Entry Game

Two firms, $f = 1, 2$, simultaneously choose whether to enter a market ($y_f = 1$) or not ($y_f = 0$). Firm f's profit is:

$$\pi_f(y_f, y_g, \epsilon_f) = y_f\left(R_f(y_g) + \epsilon_f\right)$$

where $R_f(\cdot)$ is the non-stochastic component of f's entry profit which I assume to be strictly decreasing in y_g.[15] ϵ_f is an independently distributed, privately observed shock, i.e., $\epsilon_1 \perp \epsilon_2$.

[14] The adjectives "best" and "worst" are appropriate in contexts where π_f is increasing in the actions of its complements and decreasing in the actions of its substitutes. More generally, they are only meant to describe the strategy profiles by f's competitors that maximize/minimize f's strategy.

[15] As usual, dependence of R_f on (x, θ, ξ) is omitted for brevity.

It is easy to show that this is an SMSG. To see this, note that $\mathcal{Y}_f = \{0, 1\}$ is a complete lattice, π_f order upper semi-continuous and supermodular in y_f (trivially so, since y_f is discrete and univariate), and π_f has DD in (y_f, y_g). This is:

$$\pi_f(1, y_{-f}, \epsilon_f) - \pi_f(0, y_{-f}, \epsilon_f) = R_f(y_g) + \epsilon_f$$

is decreasing in y_g.

Given an entry probability for firm g, the optimal strategy for firm f takes the form of a threshold strategy which maps one-to-one into an entry probability. So, without loss of generality, we can think of strategies as being entry probabilities.

Figure 2 depicts the implications of Theorem 1 in the entry-probability space (P_1, P_2). The left panel depicts how the first ISD iteration is conducted. Consider the worst case for firm 1, i.e., $P_2 = 1$. In that case, 1's optimal entry probability is $P_1^{1,L}$, which is the lowest possible entry probability firm 1 can rationally choose. Theorem 1 guarantees that all $P_1 < P_1^{1,L}$ are dominated. Similarly, consider the best case for firm 1, i.e., $P_2 = 0$. Here firm 1's optimal entry probability is $P_1^{1,H}$, which is the highest possible entry probability firm 1 can rationally choose. Again, Theorem 1 guarantees that any $P_1 > P_1^{1,H}$ is dominated. An analogue argument allow us to pin down $P_2^{1,L}$ and $P_2^{1,H}$.

The panel in the middle collects the outcome of this first iteration into $P^{1,L} = (P_1^{1,L}, P_2^{1,L})$ and $P^{1,H} = (P_1^{1,H}, P_2^{1,H})$, and highlights in green the set Σ^1_{ISD} which contains all strategies that survive one ISD round. Letting Σ^1_{ISD} become the strategy set of a new game, with new best and worst cases for each firm, we can conduct a another ISD round following the same process outlined above.

The panel on the right shows the result of repeating this best response iteration until convergence. The limiting strategies P^L and P^H pin down the set Σ_{ISD} which contains all strategies that survive ISD.

2.3.2 Complete Information Entry Game

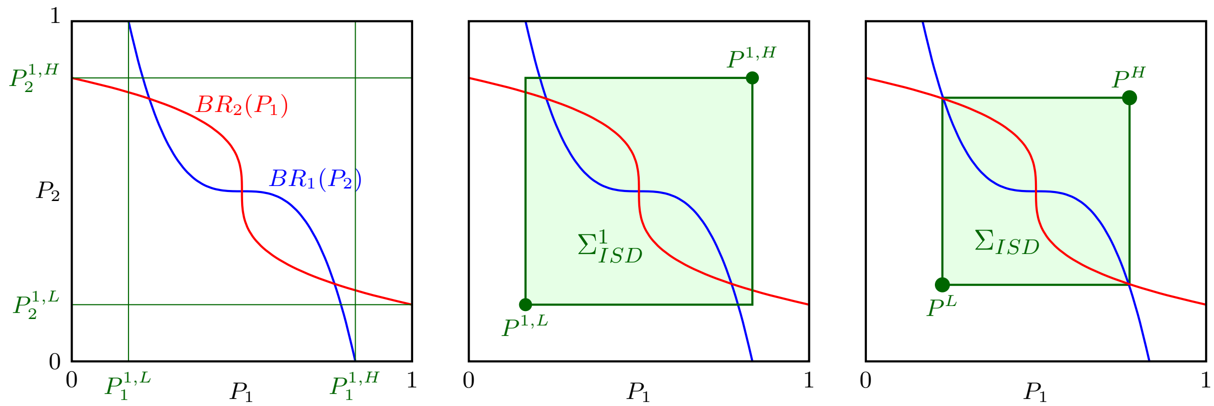
Consider the same example as above only now ξ_f, which is publicly observed, plays the role of ϵ_f. This is:

$$\pi_f(y_f, y_g; \xi_f) = y_f \left(R_f(y_g) + \xi_f\right)$$

There are three possible best response functions for firm f, depending on the realization of ξ_f. One where *entry* is dominant which occurs when $R_f(1) + \xi_f > 0$. One where *no entry* is dominant which occurs if $R_f(0) + \xi_f < 0$. And one where *entry* is only profitable as a monopolist, which occurs if $R_f(1) + \xi_f < 0 < R_f(0) + \xi_f$.

Each realization of (ξ_1, ξ_2) triggers one of nine possible games, one for each combination of best responses for each firm. Figure 3 depicts each of these combinations. For example, if $R_f(1) + \xi_f > 0$ for $f = 1, 2$, i.e., region ⑨, then *entry* is a dominant strategy for both firms. If $R_1(0) + \xi_1 < 0$ and $R_2(1) + \xi_2 < 0 < R_2(0) + \xi_2$, i.e., region ②, *no entry* is dominant for firm 1, and firm 2 chooses

Figure 2: Two Player Game Best Responses in Probability Space



Note: P_f represents the entry probability of firm f and $BR_f(P_{-f})$ represents the best response, in probability, of firm f to P_{-f}. The figure shows the process of ISD iterations for a game with multiple equilibria. The left panel shows the lowest/highest strategies associated to the first ISD round. The middle panel shows the set of strategies that survive the first round. The right panel shows the result of repeating this iterative process until convergence.

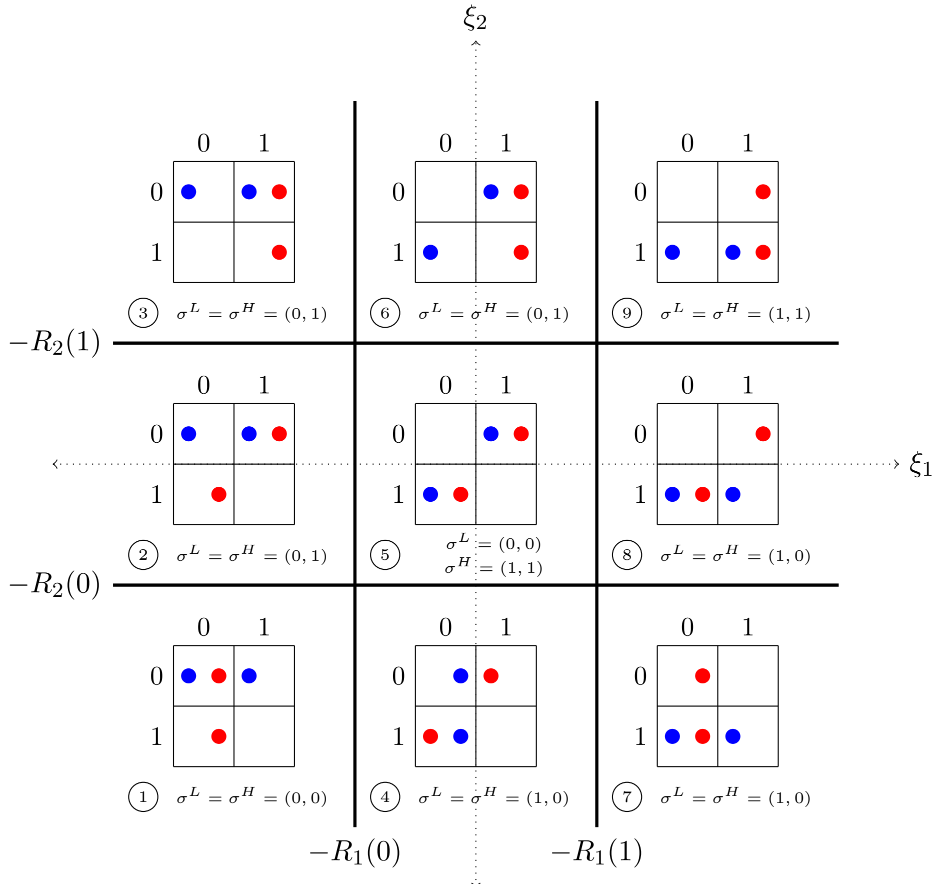
entry only as a monopolist.

To see the implications of Theorem 1 consider $\xi \in$ ② and set $\sigma^{0,L} = (0,0)$ and $\sigma^{0,H} = (1,1)$. Firm 1's best case occurs when $\sigma_2 = \sigma_2^{0,L} = 0$ and its best response is $\sigma_1^{1,H} = 0$. Similarly, firm 1's worst case occurs when $\sigma_2 = \sigma_2^{0,H} = 1$ and its optimal reaction against this is $\sigma_1^{1,L} = 0$. Theorem 1 guarantees that all $\sigma_1 < \sigma_1^{1,L}$ and all $\sigma_1 > \sigma^{1,H}$ are dominated. The same exercise for firm 2 reveals that its best response to $\sigma_1 = \sigma_1^{0,L} = 0$ is $\sigma_2^{1,H} = 1$, while the best response to $\sigma_1 = \sigma_1^{0,H} = 1$ is $\sigma_2^{1,L} = 0$. Putting these together we get $\sigma^{1,L} = (0,0)$ and $\sigma^{1,H} = (0,1)$.

Now we repeat the ISD iteration using $\sigma^{1,L}$ and $\sigma^{1,H}$ as starting points. We have reached convergence for firm 1 who can only respond with *no entry*, so we set $\sigma_1^{2,L} = \sigma_1^{2,H} = 0$. For firm 2, the best and worst cases coincide at $\sigma_1 = 0$, hence we set $\sigma_1^{2,H} = \sigma_1^{2,L} = 1$ which is 1's best response to *no entry*. By Theorem 1 all $\sigma_2 < \sigma_2^{2,L}$ and all $\sigma_2 > \sigma_2^{2,H}$ are dominated. Collecting the updated extreme strategies for each firm we get $\sigma^{2,L} = \sigma^{2,H} = (0,1)$, so we have reach convergence and $\sigma^L = \sigma^H = (0,1)$. The game is dominance solvable, and the ISD iterations successfully identifies the solution.

It is easy to verify that ISD iterations pin down a unique prediction for all games depicted in Figure 3, except for $\xi \in$ ⑤. In this case ISD has no bite and ISD iterations are reflect this fact. Like before, set $\sigma^{0,L} = (0,0)$ and $\sigma^{0,H} = (1,1)$. The best case for firm 1 is $\sigma_2 = \sigma_2^{0,L} = 0$ and its best response is $\sigma_1^{1,H} = 1$. Similarly, the worst case for firm 1 is $\sigma_2 = \sigma_2^{0,H} = 1$ and its best response is $\sigma_1^{1,L} = 0$. Clearly $\sigma_1^{0,H} = \sigma_1^{1,H}$ and $\sigma_1^{0,L} = \sigma_1^{1,L}$, so ISD did not rule out any strategy. Furthermore, the same holds for firm 2, hence $\sigma^{1,L} = \sigma^{0,L}$ and $\sigma^{1,H} = \sigma^{1,L}$ and we have reached convergence after one iteration without ruling out anything.

Figure 3: Game Matrices with Best Responses for Values of ξ



Note: In each region 1,...,9, (ξ_1, ξ_2) generates a different class of games, in the sense that within each region all values of ξ generate the same best responses for both players. The red dots represent the best response of firm 1, and the blue dots represent the best response of firm 2. At the bottom of each region I write the corresponding values of σ^L and σ^H.

3 ISD Bounds and Identified Set

Here I show that ISD implies bounds on the distribution of outcomes generated by the model, and use these bounds to build an identified set for the parameters of interest.

3.1 ISD Bounds

The main assumption behind ISD Bounds in SMSGs, Assumption 1 below, says that for (almost) all possible values of (x, θ, ξ) the (x, θ, ξ)-game is an SMSG. This assumption implies that the results of Theorem 1 hold for all (x, θ, ξ). Importantly, this assumption *does not say* that the set of complements and substitutes of each firm has to be the same for all (x, θ, ξ). This is an important source of flexibility if the researcher does not want to impose the nature of strategic interactions between players, and rather wants this to be revealed by the data (as in Ciliberto and Jäkel (2021)).[16]

Assumption 1 (SMSG Assumption)**.** *The (x, θ, ξ)-game is an SMSG for almost every $(x, \theta, \xi) \in \mathcal{X} \times \Theta \times \Xi$.*

To put bounds on the distribution of outcomes, we need to specify how outcomes are generated in the first place. Here, since we have assumed that firms behave optimally relative to their expectations, this implies the need for a constrain on expectations. Assumption 2, below, constrains expectations to be consistent with ISD.

Assumption 2 (ISD-Consistent Beliefs)**.** *For every (x, θ, ξ)-game and for each pair of firms f and g, f believes that g plays a strategy $\rho_g^f \in \Sigma_{g,ISD}$.*

Two comments about Assumption 2. First, ISD-Consistent beliefs are allowed to depend on (x, θ, ξ). This must be so, since Σ_{ISD} also depends on (x, θ, ξ). Second, rational expectations are ISD-Consistent, so Assumption 2 allows for equilibrium play. More generally, ISD-Consistent beliefs need not satisfy rational expectations, so Assumption 2 allows for non-equilibrium play. The ISD Bounds I derive below are robust to such suboptimal play, as long as the beliefs that generate this behavior are ISD-Consistent.

With this, we are in a position to move to Theorem 2, which uses Theorem 1 to derive bounds on the distribution of outcomes implied by ISD. In particular, the bounds exploit the fact that in every SMSG a strategy σ_f that survives ISD lies between σ_f^L and σ_f^H. To move in this direction define the following probabilities (making dependence on (x, θ, ξ) explicit):

$$\begin{aligned} P^L(y|x,\theta) = & \int_{\mathcal{E}} \int_{\Xi} 1\{\sigma^L(\epsilon; x, \theta, \xi) \geq y\} dG(\epsilon; x, \theta, \xi) dH(\xi; x, \theta) \\ P^\rho(y|x,\theta) = & \int_{\mathcal{E}} \int_{\Xi} 1\{\sigma^\rho(\epsilon; x, \theta, \xi) \geq y\} dG(\epsilon; x, \theta, \xi) dH(\xi; x, \theta) \\ P^H(y|x,\theta) = & \int_{\mathcal{E}} \int_{\Xi} 1\{\sigma^H(\epsilon; x, \theta, \xi) \geq y\} dG(\epsilon; x, \theta, \xi) dH(\xi; x, \theta) \end{aligned} \tag{4}$$

[16]This contrasts with bounds like the ones in Aradillas-Lopez (2011) which requires that the research have *a priori* knowledge of the nature of strategic interaction between players.

where H is the distribution of the common-knowledge unobservables, ξ. With this, we are in a position to show Theorem 2.

Theorem 2 (ISD Bounds in SMSGs)**.** *Consider a class of games indexed by (x, θ, ξ) satisfying Assumption 1 and say for every distinct pair of firms f and g, ρ_g^f is ISD-Consistent as per Assumption 2. The following holds:*

$$P^L(y|x,\theta) \leq P^\rho(y|x,\theta) \leq P^H(y|x,\theta) \tag{5}$$

for all firms and almost all y, x, and θ.

Proof. Fix (x, θ, ξ) and fix ISD-Consistent beliefs for each player. Since σ_f^ρ is a best response to a strategy/belief that survives ISD it must itself survive ISD, hence by Theorem 1 $\sigma^L \leq \sigma^\rho \leq \sigma^H$. Fix an action profile $y \in \mathcal{Y}$ and a shock ϵ, it is easy to see that (making explicit the dependence on (x, θ, ξ)):

$$\mathbb{1}\left\{\sigma^L(\epsilon; x, \theta, \xi) \geq y\right\} \leq \mathbb{1}\left\{\sigma^\rho(\epsilon; x, \theta, \xi) \geq y\right\} \leq \mathbb{1}\left\{\sigma^H(\epsilon; x, \theta, \xi) \geq y\right\}$$

Integrating over ϵ and ξ yields (5).[17] □

Let us make two noteworthy points regarding ISD Bounds. First, as in Aradillas-Lopez and Tamer (2008), Aradillas-Lopez (2010) and Molinari and Rosen (2008), analogous ISD Bounds may be built using k ISD rounds, for $k < \infty$. In fact, it is easy to see that the more ISD rounds one uses, the tighter the bounds. Second, and relatedly, bounds based on k ISD rounds hold for k'-ISD-consistent beliefs if $k' \geq k$. In other words, for k-ISD-rounds bounds to hold we only need that beliefs are k'-ISD-Consistent with $k' \geq k$. This observation is important, as researchers may wish to explore the identifying power of different rounds of ISD. I formalize this below.

Remark 1. *Let $P_f^{k,L}$ and $P_f^{k,H}$ be the ISD Bounds that result from k ISD rounds (defined analogously to* (4)*). For any ISD-Consistent beliefs, ρ:*

$$P^{k,L}(y|x,\theta) \leq P^{k+1,L}(y|x,\theta) \leq P^\rho(y|x,\theta) \leq P^{k+1,H}(y|x,\theta) \leq P^{k,H}(y|x,\theta)$$

for all $k = 0, 1, 2, \ldots$.

Furthermore, say ρ is k'-ISD-Consistent (defined analogously to Assumption 2) with $k' \geq k$. The following holds:

$$P^{k,L}(y|x,\theta) \leq P^\rho(y|x,\theta) \leq P^{k,H}(y|x,\theta)$$

for all $k = 0, 1, 2, \ldots$.

All results below hold for k'-ISD-Consistent beliefs and ISD Bounds that result from k ISD rounds, with $k' \geq k$.

[17] Note that beliefs ρ depend on ξ, so for the term in the middle we are effectively integrating over beliefs in different games.

It is also worth noting that *firm-level bounds*, i.e., bounds on the distribution of y_f, can be derived using the same argument we used for the outcome-level bounds in Theorem 2. Furthermore, when y_f is binary (say, $y_f = 0$ or $y_f = 1$) for all f, we can reduce the inequalities in (5) to a single inequality per firm. This is an important case, as bounds at the firm level are easier to compute and require less data to estimate precisely. Naturally, firm-level bounds are wider than outcome level bounds, as they bind only the marginal distribution, rather than the joint, so estimation based on firm level bounds will result in larger identified sets. I formalize firm-level bounds in the remark below.

Remark 2. *If $\mathcal{Y}_f = \{0,1\}$ for all f, the following **firm-level ISD Bounds** hold for almost all (x,θ):*

$$P_f^L(x,\theta) \leq P_f^\rho(x,\theta) \leq P_f^H(x,\theta)$$

where $P_f^\rho(x,\theta) = P_f^\rho(1|x,\theta)$ is the probability of $y_f = 1$ given beliefs ρ. P_f^L and P_f^H are defined analogously.

Finally, building on results by Garrido (2021), a similar argument to the ones used in Theorems 1 and 2 can be used to build bounds on equilibrium prices for Bertrand pricing games between multi-product firms with nested demand. As opposed to ISD Bounds, however, the these bounds do not have an ISD interpretation to them. I show how these bounds are built in Appendix E.

3.2 Two Entry-Game Examples Continued

Before building the identified set, let us explore how Theorem 2 produces ISD Bounds for the entry games presented in sections 2.3.1 and 2.3.2.

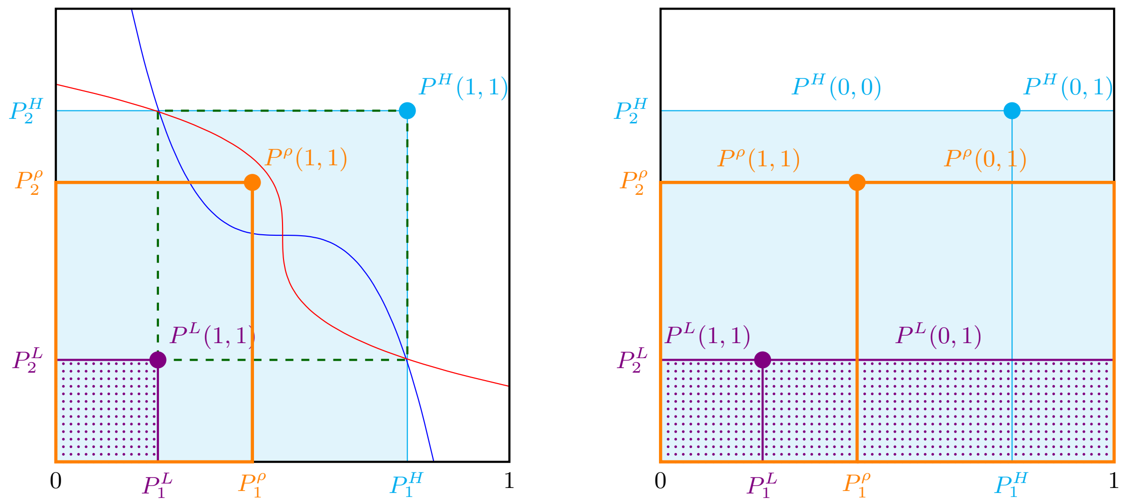
3.2.1 Independent Private Information Entry Game Revisited

Consider the independent private information game introduced in 2.3.1, and assume that there are no common knowledge unobservables, $\xi = \varnothing$. Furthermore, say we want to see the implications of Theorem 2 for outcome $y = (1,1)$. The right panel of Figure 4 depicts said implications.

To see this, note that beliefs ρ are ISD-Consistent, as they lie in the dashed green square that represents the set of strategies that survive ISD. Given the independent ϵ_f assumption, the probability that $\sigma^\rho \geq (1,1)$ is given by the orange rectangle. This rectangle is bounded from below by the probability that $\sigma^L \geq (1,1)$, which corresponds to the purple dotted area, and from above by the probability that $\sigma^H \geq (1,1)$ which corresponds to the light-blue area. Although harder to visualize, similar bounds can be build for the rest of the outcomes.

The right panel of figure Figure 4 shows the result of the same exercise for outcome $(0,1)$. The purple dotted area represents the probability that $\sigma^L(\epsilon) \geq (0,1)$, the light blue area represents the probability that $\sigma^H(\epsilon) \geq (0,1)$ and the areas inside the orange square represent the probability that

Figure 4: ISD Bounds in a 2×2 Incomplete Information Entry Game



Note: P_f represents the entry probability of firm f. The blue (red) line represents the optimal entry probability of firm 1 (2) given the entry probability of firm 2 (1). These lines are omitted from the right panel to minimize clutter. In the left (resp. right) panel, the purple dotted are represent the lower ISD Bound for outcome $(1,1)$ (resp. $(0,1)$), and the lightblue area represents the respective upper ISD Bound. The areas inside orange squares represent $Pr(\sigma^\rho(\epsilon) \geq y)$ for some arbitrary ISD-Consitent beliefs ρ.

$\sigma^\rho(\epsilon) \geq (0,1)$. As before, for any ISD consistent ρ, the probability according to ρ lies between the upper and lower ISD Bounds.

Note that in this example firms are wrong about their competitors' strategies, i.e., $\rho_g^f \neq P_g^\rho$. As a result, the distribution over outcomes under ρ is not compatible with a Bayes Nash Equilibrium. Still, since beliefs are ISD-Consistent, ISD Bounds hold.

3.2.2 Complete Information Example

Consider the complete information example introduced in 2.3.2. As depicted in Figure 3, every realization of (ξ_1, ξ_2) triggers a different SMSG, and for each (ξ_1, ξ_2) the set of strategies that survive ISD is pinned down by different extreme strategies σ^H and σ^L.

For $\xi \notin ⑤$ the game is dominance solvable. Hence, for any such ξ the only ISD-Consistent beliefs are $\rho(\xi) = \sigma^L(\xi) = \sigma^H(\xi)$. Furthermore, the corresponding vector of best responses is $\sigma^\rho(\xi) = \sigma^L(\xi) = \sigma^H(\xi)$. For $\xi \in ⑤$, in contrast, ISD has no bite so any belief is ISD-Consistent, and the most that we can say about firms' behavior is: $\sigma^L(\xi) = (0,0) \leq \sigma^\rho(\xi) \leq \sigma^H(\xi) = (1,1)$.

As before, say we want to build bounds for outcome $(1,1)$. To compute the lower bound we need to find the values of ξ for which $\sigma^L(\xi) \geq (1,1)$, which occurs only in region ⑨. Similarly, to compute the upper bound we need to find the values of ξ for which $\sigma^H(\xi) \geq (1,1)$. This occurs in regions ⑤ and ⑨. Hence, ISD Bounds for outcome $(1,1)$ are:

$$\begin{aligned} P^L(0,1) &= Pr\big(\xi \in ⑨\big) \\ P^H(0,1) &= Pr\big(\xi \in ⑤ \cup ⑨\big) \end{aligned}$$

3.3 Identified Set

In this subsection I derive the ISD identified set. To this end, consider the following assumption on the data generating process.

Assumption 3 (Data Generating Process (DGP))**.** *There is a real parameter vector θ^0, and real ISD-Consistent beliefs, ρ^0 for every x and ξ. Furthermore, observed outcomes are drawn from the distribution $P^0(y|x) \equiv P^{\rho^0}(y|x, \theta^0)$, where, as in equation* (4)*, P^ρ is a complementary cumulative distribution function.*

Assumption 3 says that the model is correctly specified, and that the realization of each game comes from beliefs that are ISD-Consistent. Then, by Theorem 2 ISD Bounds (expression (5)) hold for $\theta = \theta^0$ and $\rho = \rho^0$. Hence, if θ *does not satisfy*:

$$P^H(y|x, \theta) \leq P^0(y|x) \leq P^L(y|x, \theta) \tag{6}$$

for some y and some x, then $\theta \neq \theta^0$. This intuition leads us to define the ISD identified set as follows.

Definition 6 (ISD Identified Set)**.** *The ISD Identified Set, Θ_{ISD}, is the collection of all $\theta \in \Theta$ such that θ satisfies* (6)*, i.e.:*

$$\Theta_{ISD} = \{\theta \in \Theta : \theta \text{ satisfies } (6), \forall x \in \mathcal{X}, \forall y \in \mathcal{Y}\}$$

It is easy to see that an analogous identified set can be derived using the firm-level bounds from Remark 2. In the next section I give some insights on the question of how informative are ISD Bounds.

4 Identifying Power of ISD Bounds in Binary Games

A common concern when dealing with set identification is how informative is the identified set. In this section I give some insights into this question by providing two results. First, I show that for binary games with independent private information, i.e., $\epsilon_f \perp \epsilon_g$ for all $f \neq g$, Θ_{ISD} is sharp, meaning for all θ in the identified set we can find ISD-Consistent beliefs ρ such that the model under (θ, ρ) matches the DGP. Second, I provide necessary and sufficient conditions to rule out the existence of parameters value for which ISD never has any bite, and therefore trivially lie in the identified set.[18]

4.1 Sharp Identified Sets

For the purpose of this model, the sharp identified set is the collection of θ's for which there is some set of beliefs ρ such that the resulting distribution over outcomes matches the DGP. Since the sharp

[18] A third (negative) result, relegated to Appendix C.3, shows a form of non-identification in entry games with linear strategic effects.

identified set is defined relative to a DGP, whether the ISD identified set is sharp or not will depend on the beliefs the researcher assumes generate the data. For example, if the researcher assumes that the beliefs in the DGP satisfy rational expectations, i.e, data comes from (Bayes) Nash equilibrium play, then Θ_{ISD}, which is built under the weaker notion of ISD, will not be sharp.

The main result of this section is that in binary games, when the beliefs in the DGP are assumed to be ISD-Consistent and there is either independent private information or perfect information, then the ISD Identified Set is sharp. Below, I provide the definition of the sharp identified set, and formally state this result.

Definition 7 (Sharp Identified Set)**.** *Let P^0 be a DGP satisfying Assumption 3. The sharp identified set, Θ_{sharp}, is the collection of $\theta \in \Theta$ for which there is an ISD-Consistent ρ for each (x, ξ) such that the implied distribution over outcome matches P^0. Formally:*

$$\Theta_{\text{sharp}} = \{\theta \in \Theta : \exists \rho \text{ satisfying Assumption 2 s.t. } P^{\rho}(y|x, \theta) = P^0(y|x) \text{ for all } x \text{ and } y\} \tag{7}$$

where P^{ρ} and P^0 are the distributions over outcomes under (θ, ρ) and under (θ^0, ρ^0) respectively.

Theorem 3. *For any class of binary games (i.e., $\mathcal{Y}_f = \{0, 1\}$ for all $f \in \mathcal{F}$) indexed by (x, θ, ξ) and satisfying Assumptions 1 and 2, and any corresponding DGP satisfying Assumption 3, if either:*

- *$\epsilon_f \perp \epsilon_g$ for all $f \neq g$, and ϵ_f has a continuous distribution with full support in $\mathbb{R}$. Or,*
- *$\epsilon_f = \varnothing$ for all $f \in \mathcal{F}$, i.e., firms have complete information.*

Then, the ISD identified set is sharp, i.e., $\Theta_{ISD} = \Theta_{\text{sharp}}$.

Proof. See appendix C.1. □

The proof of Theorem 3 relies on showing that Σ_{ISD} contains *only* strategies that survive ISD.[19] To get an idea of how this works, consider the case of independent private information without unobserved heterogeneity, i.e., $\xi = \varnothing$, and note that this implies that for all (x, θ), all strategies between σ_f^L and σ_f^H are the best response to some ISD-Consistent belief, ρ^f (see Figure 2). Hence, all probabilities in the interval $[P_f^L, P_f^H]$ can be generated by an appropriate belief. By definition, any $\theta \in \Theta_{ISD}$ satisfies $P_f^0 \in [P_f^L, P_f^H]$, so by the argument above we can find beliefs such that $P_f^{\rho} = P_f^0$. Finally, since the ϵ_f's are independent this implies that $P^{\rho} = P^0$, which in turn implies $\theta \in \Theta_{\text{sharp}}$.

The independent private information assumption in Theorem 3 is less constraining than what it may seem at a first glance. This assumption is mainly ruling out situations where parties are differentially informed, e.g., one party has full information while the others only have access to independent private shocks. Put differently, Theorem 3 applies to any model in which the informational value of the ϵ_f's can be decomposed into publicly and privately observed components,

[19]Theorem 1 shows that Σ_{ISD} contains all strategies that survive ISD, but it does not rule out that some of its elements do are dominated by another strategy in Σ_{ISD}.

like, for example, if $\epsilon_f = \tau_f + \tilde{\epsilon}_f$ for all f, where τ_f is public and $\tilde{\epsilon}_f$ is private. In such case, we simply need to group τ_f's with ξ as unobserved public information, and redefine the private information as $(\tilde{\epsilon}_f)_f$. Furthermore, Theorem 3 applies in to games with perfect information, in which case $\epsilon_f = \varnothing$ for all f.

4.2 (Non-)Trivial ISD Bounds

Tamer (2003) and Aradillas-Lopez and Tamer (2008) use an "identification-at-infinity" argument to show that point identification can be recovered if there is enough independent variation in the firms' profit shifters. For example, consider a two-firm entry game and say a profit shifter of firm 2 goes to infinity (resp. negative infinity) so that the firm will always (resp. never) choose to enter regardless of the value of unobservables.[20] In such case, since firm 2's decision is a foregone conclusion, firm 1's problem reduces to a single agent problem. As a result, provided certain regularity conditions hold, firm 1's parameters are point identified. This result stresses the importance having wide variation in profit shifters to tighten the identified set.

Here I provide a complementary result, Theorem 4, which gives necessary and sufficient conditions for the existence of some θ that generates trivial ISD Bounds in binary games, i.e., for some θ, $P^L(y|x,\theta) = 0$ and $P^H(y|x,\theta) = 1$ for all x and all $y > (0)_{\forall f}$.[21] Any such θ lies in Θ_{ISD} not because it fits the data well, but because it does not rule anything out. The conditions for Theorem 4 emphasize the importance of cross-sectional variation in variables that uniformly shift profits across markets, such as *market size*, for tightening the ISD Bounds. In this sense, the conditions for Theorem 4 are complementary to the conditions for point identification emphasized by Tamer (2003) and Aradillas-Lopez and Tamer (2008), in that the two emphasize different sources of variation that help with narrowing down the identified set.

To move in this direction, let us first specify the details of the game for which this result holds. This specification is quite flexible and covers many models considered in the empirical literature. Consider a class of binary SMSGs indexed by (x, θ, ξ) where payoffs are additively separable:

$$\pi_f = R_f(y_f, y_{-f}; x, \theta_{f,R}) + y_f\left(\theta_{cons,f} + \theta_{sc,f}\left(\sqrt{\theta_{w,f}}\xi_f + \sqrt{1-\theta_{w,f}}\epsilon_f\right)\right)$$

where $\theta = (\theta_{R,f}, \theta_{cons,f}, \theta_{sc,f}, \theta_{w,f})_{f\in\mathcal{F}}$ is the vector of parameters of interest, and where $\theta_{R,f}$ is a vector of parameters that controls R_f, $\theta_{cons,f}$ is a constant profit term from choosing $y_f = 1$, $\theta_{sc,f}$ is a scale parameter, and $\theta_{w,f}$ is a weight parameter that controls the share of total variance of composite shock that comes from ξ_f.[22] Finally, let us define f's marginal profit:

$$\Delta R_f(y_{-f}; x, \theta_R) = R_f(1, y_{-f}; x, \theta_R) - R_f(0, y_{-f}; x, \theta_R)$$

[20] There is an implicit assumption here that the profit shifter that goes to infinity (negative infinity) has a positive effect on profits.

[21] Note that by construction $P^L((0)_{\forall f}|\theta) = P^H((0)_{\forall f}|\theta) = 1$, so the ISD Bound for $y = (0)_{\forall f}$ is always trivially satisfied.

[22] In most empirical applications $\theta_{w,f}$ is assumed to be known. Theorem 4 still holds under this assumption.

which is useful in stating and proving Theorem 4, below.

Theorem 4. *There exists a $\theta \in \Theta$ such that $P^L(y|x,\theta) = 0$ and $P^H(y|x,\theta) = 1$ for all $y > (0)_{\forall f}$ and all x, if and only if R_f satisfies:*

$$\sup_{x \in \mathcal{X}} \left\{ \min_{y_{-f} \in \mathcal{Y}_f} \Delta R_f(y_{-f}; x, \theta_{f,R}) \right\} < \inf_{x \in \mathcal{X}} \left\{ \max_{y_{-f} \in \mathcal{Y}_f} \Delta R_f(y_{-f}; x, \theta_{f,R}) \right\} \tag{8}$$

for all $\theta_{f,R}$ and all $f \in \mathcal{F}$.

Proof. See appendix C.2 □

To get an intuition for this result, consider a symmetric two-firm incomplete information entry game with payoffs:

$$\pi_f = \begin{cases} R^{mon}(x) - \theta_{ec} + \theta_{sc}\epsilon_f & \text{if} \quad y_f = 1,\ y_{-f} = 0 \\ R^{duo}(x) - \theta_{ec} + \theta_{sc}\epsilon_f & \text{if} \quad y_f = 1,\ y_{-f} = 1 \\ 0 & \text{if} \quad y_f = 0 \end{cases}$$

where θ_{ec} and θ_{sc} represents the entry cost and the scale parameter, respectively. And where, $R^{mon} > R^{duo}$ are known monopolic and duopolistic profit functions. If $R^{duo}(x) < R^{mon}(x')$ for all x and x', then we can set $\theta_{ec} = (\inf_x R^{mon}(x) + \sup_x R^{duo}(x))/2$ and for any $\theta_{sc} > 0$ we have:

$$\frac{R^{duo}(x) - \theta_{ec}}{\theta_{sc}} < 0 < \frac{R^{mon}(x) - \theta_{ec}}{\theta_{sc}}, \forall x$$

letting $\theta_{sc} \to 0$ makes being a monopolist (duopolist) arbitrarily profitable (unprofitable) simultaneously for all x, which implies that we can make ISD Bounds arbitrarily uninformative. If, in contrast, there are x and x' such that $R^{mon}(x) < R^{duo}(x')$, then any $(\theta_{ec}, \theta_{sc})$ that satisfies the inequality above for x will violate it for x', hence is not possible to make being a monopolist (duopolist) arbitrarily profitable (unprofitable) for all x simultaneously.

As mentioned above, Theorem 4 stresses the importance of cross-sectional variation in narrowing down the identified set. In particular, it highlights the role of variables that uniformly shift profits across markets, such as *market size*. Intuition suggests that the more variation in market size we have, i.e., the larger $R^{duo}(x') - R^{mon}(x) > 0$, the more values of θ are ruled out by ISD Bounds.

5 Monte Carlo Exercises and Bound Comparison

Here I provide some Monte Carlo exercises to study the performance of ISD Bounds. First, in Section 5.1, I provide Monte Carlo experiments for the case of incomplete information with unobserved heterogeneity. Second, in Section 5.2, I focus on a entry game of complete information and compare ISD Bounds to previous bounds proposed in the literature

5.1 Incomplete Information Entry Game with Unobserved Heterogeneity

Here I study the performance of ISD Bounds in games of incomplete information. To this end, consider the independent private information entry game described in section 2.3.1 and 3.2.1, except that the number of firms is arbitrary, and there is unobserved heterogeneity in the form of firm specific unobserved profit shifters $\xi_f \sim N(0,1)$ for all f. Let the profit function be:[23]

$$\pi_f = y_f \left(\frac{x_f}{(2 + \sum_{g \neq f} y_g)^2} - \theta_{ec} + \theta_{sc} \left(\sqrt{\theta_w} \xi_f + \sqrt{1-\theta_w} \epsilon_f \right) \right)$$

where $\epsilon_f \sim N(0,1)$ for all f, and where θ_w controls the amount of randomness (from the perspective of the econometrician) that comes from private information vs. unobserved heterogeneity.

I set the parameters to $\theta_{ec}^0 = 1$, $\theta_{sc}^0 = 1$ and conduct separate experiments for the case without unobserved heterogeneity, $\theta_w^0 = 0$, and for the case with unobserved heterogeneity, $\theta_w^0 = 0.5$. Furthermore, I let $\mathcal{X} = \{0.5, 2.5, 4.5\}^{|\mathcal{F}|}$. Figure 5 depicts the the best responses generated in the two player case, for each possible $x \in \mathcal{X}$, and for $\xi_1 = \xi_2 = 0$.[24] The case without (with) unobserved heterogeneity is depicted on the left (right) panel. The green dots represent the Bayes Nash equilibria of each game, while the green areas represent the strategies that survive ISD.

Figure 5 shows that when $\theta_w = 0$. although there is substantial variation in x, no game generates multiple equilibria, and the all games are dominance solvable. For every x ISD Bounds pin down a unique distribution over outcomes, i,e, $P^L = P^H$, and give us point identification. In contrast, when $\theta_w = 0.5$ the DGP is able to generate multiple equilibria for large values of x and point identification is not guaranteed. Note that the best responses in the two cases differ only because the variance of the private information component is lower in when $\theta_w = 0.5$, so firms' best responses are (slightly) more sensitive to changes in the competitors' entry probabilities.

For each case, I simulate 100 samples of $M = 4000$ markets each. For each market, I draw an $x \in \mathcal{X}$ using a uniform distribution and I draw a ξ_f for each firm from a standard normal. Then, for each (x, ξ) I find a Bayes Nash Equilibrium under θ^0.[25] This process yields a sample $(Y_m, X_m)_{m=1}^M$. For inference I follow Fan and Yang (2022) in using the test proposed by Andrews and Soares (2010) (AS) (see Appendix D for details).

Let $\hat{\theta}_{ec}$ and $\hat{\theta}_{sc}$ be guesses for θ_{ec}^0 and θ_{sc}^0, respectively. Keeping $\hat{\theta}_{sc} = \theta_{sc}^0$ I perform the AS test for each sample and for each $\hat{\theta}_{ec}^0$ in an equally spaced grid going from 0 to $2\theta_{ec}^0$, and compute the share of samples for which the null $(\hat{\theta}_{ec}, \theta_{sc}, \theta_w) \in \Theta_{ISD}$ gets rejected. I conduct an analogue exercise for θ_{sc}.

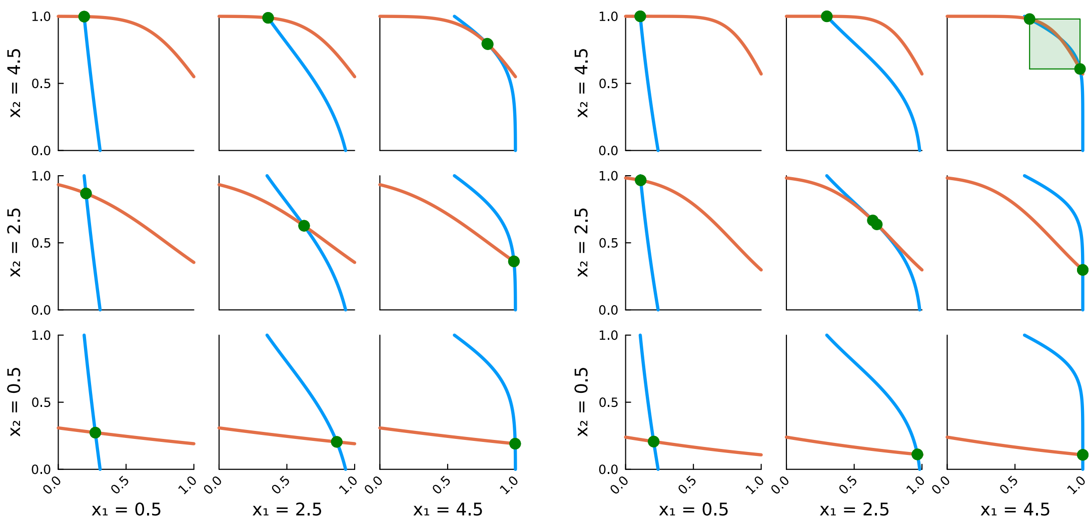
The results of this exercise can be found in Figures 6 and 7 for the case without and with unobserved heterogeneity respectively. In each case the horizontal axis of each plot represents the ratio of the guess to the true parameter, e.g., $\hat{\theta}_{ec}$. The identified sets for θ_{ec} and θ_{sc} are very

[23]The function $x_f/(2 + \sum_{f' \neq f} y_{f'})$ is a reduced for profit from a Cournot competition second stage.

[24]The condition $\xi_1 = \xi_2 = 0$ is only relevant for the case with unobserved heterogeneity. When there is no unobserved heterogeneity $\theta_w = 0$, the ξ_f's get multiplied by zero.

[25]I start the solution algorithm from a randomly chosen vector of entry probabilities, so there is not much that I can say about the actual Equilibrium Selection Mechanism behind the simulated data.

Figure 5: Incomplete Information Best Responses for values of x (Left: $\theta_w = 0$. Right: $\theta_w = 0.5$)



Note: Two-way best responses for each $x \in \mathcal{X}$. Each column (row) depicts the games for a fixed value of x_1 (x_2). The horizontal (vertical) axis of each plot represents the entry probability of firm 1 (resp. 2) and the blue (resp. orange) line represents its best response. Green dots represent the Bayes Nash equilibria, and the light green areas represents the strategies that survive ISD.

informative about the underlying parameters. The parameters seem to be (close to) point identified. This should not be surprising in light of the "identification at infinity" arguments advanced by Tamer (2003) and Ciliberto and Tamer (2009), and more recently Aradillas-López and Rosen (2022).

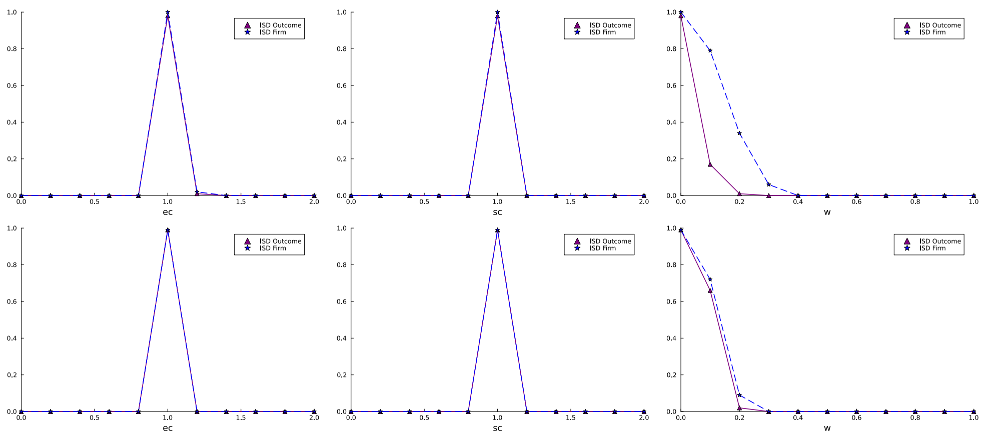
In both figures I include the confidence set on θ_w for completeness. The figures show that ISD Bounds do a reasonably good job of distinguishing the extreme cases. When there $\theta_w^0 = 0$ the bounds quickly rule out values too different from zero, but when $\theta_w^0 = 0.5$ the bounds seem to only rule out the extreme cases, $\theta_w = 1$ and $\theta_w = 0$, and have a hard time ruling out intermediate values of θ_w. The implication is that the set of distributions over outcomes allowable by ISD is not very sensitive to changes in θ_w.

5.2 Complete Information Probability ISD Bounds Comparison

Tamer (2003) and Ciliberto and Tamer (2009) (CT) pioneered the probability bounds approach to set identification for discrete games of complete information. This approach has also been studied by Aradillas-Lopez and Tamer (2008) and Fan and Yang (2022) (FY). In this subsection I study how ISD Bounds compare to CT and FY Bounds. To this end, I consider a complete information entry game, in the spirit of CT, and provide comparisons between CT, FY and ISD Bounds.

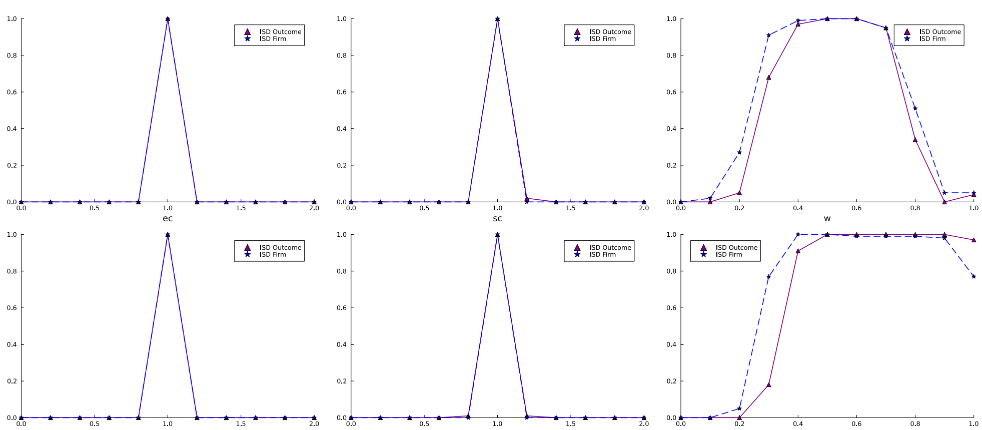
In our notation, CT Bounds assume that beliefs, ρ, satisfy rational expectations so that the data is generated by Nash equilibria. Under this assumption, if y is the unique (pure strategy)

Figure 6: Share of not Reject for $|\mathcal{F}| = 2$ (top row) and $|\mathcal{F}| = 3$ (bottom row) Under Incomplete Information and $\theta_w = 0$



Note: The Left, middle and right columns show the results for θ_{ec}, θ_{sc}, and θ_w, respectively. The top (bottom) row represents the 2 (3) player case. In each case, the x-axis depicts values of $\hat{\theta}_{ec}$ (left) and $\hat{\theta}_{sc}$ (middle) and $\hat{\theta}_w$. The y-axes show the share of samples for which the null that the corresponding parameter is equal to the parameter in the DGP is not rejected.

Figure 7: Share of not Reject for $|\mathcal{F}| = 2$ (top row) and $|\mathcal{F}| = 3$ (bottom row) Under Incomplete Information and and $\theta_w = 0.5$



Note: The Left, middle and right columns show the results for θ_{ec}, θ_{sc}, and θ_w, respectively. The top (bottom) row represents the 2 (3) player case. In each case, the x-axis depicts values of $\hat{\theta}_{ec}$ (left) and $\hat{\theta}_{sc}$ (middle) and $\hat{\theta}_w$. The y-axes show the share of samples for which the null that the corresponding parameter is equal to the parameter in the DGP is not rejected.

Table 1: CT vs. ISD vs. FY Bounds for $(0,1)$

Bound Type	Lower Bound	Upper Bound
CT $(0,1)$	$Pr\left(\xi \in ② \cup ③ \cup ⑥\right)$	$Pr\left(\xi \in ② \cup ③ \cup ⑤ \cup ⑥\right)$
ISD $y=(0,1)$	$Pr\left(\xi \in ② \cup ③ \cup ⑥ \cup ⑨\right)$	$Pr\left(\xi \in ② \cup ③ \cup ⑤ \cup ⑥ \cup ⑨\right)$
ISD $y_1=0$	$Pr\left(\xi \in ① \cup ② \cup ③ \cup ⑥\right)$	$Pr\left(\xi \in ① \cup ② \cup ③ \cup ⑤ \cup ⑥\right)$
ISD $y_2=1$	$Pr\left(\xi \in ② \cup ③ \cup ⑥ \cup ⑨\right)$	$Pr\left(\xi \in ② \cup ③ \cup ⑤ \cup ⑥ \cup ⑨\right)$
FY $y_1=0$	$Pr\left(\xi \in ① \cup ② \cup ③\right)$	$Pr\left(\xi \in ① \cup ② \cup ③ \cup ④ \cup ⑤ \cup ⑥\right)$
FY $y_2=1$	$Pr\left(\xi \in ③ \cup ⑥ \cup ⑨\right)$	$Pr\left(\xi \in ② \cup ③ \cup ⑤ \cup ⑥ \cup ⑧ \cup ⑨\right)$

equilibrium then $\sigma^\rho = y$, and if $\sigma^\rho = y$ then y must be in *an* equilibrium. Hence:

$$\begin{array}{ccccc} \mathbb{1}\{\{y\} = NE(x,\theta,\xi)\} & \leq & \mathbb{1}\{y = \sigma^\rho(x,\theta,\xi)\} & \leq & \mathbb{1}\{y \in NE(x,\theta,\xi)\} \\ P^L_{CT}(y|x,\theta) & \leq & P^\rho(y|x,\theta) & \leq & P^H_{CT}(y|x,\theta)) \end{array}$$

where $NE(x,\theta,\xi)$ is the set of pure strategy equilibria of the (x,θ,ξ)-game, and where the second line comes from integrating over ξ, where $P^L_{CT}(y|x,\theta)$ and $P^H_{CT}(y|x,\theta))$ are the integrals of the LHS and the RHS respectively, and they represent the lower and the upper CT probability bounds. Similarly, FY bounds are based on the idea that players will never choose a dominated action and will always choose a dominant action. Hence:

$$\begin{array}{ccccc} \mathbb{1}\{\{y_f\} = \Sigma^1_{f,ISD}(x,\theta,\xi)\} & \leq & \mathbb{1}\{y = \sigma^\rho(x,\theta,\xi)\} & \leq & \mathbb{1}\{y \in \Sigma^1_{f,ISD}(x,\theta,\xi)\} \\ P^L_{FY}(y_f|x,\theta) & \leq & P^\rho_f(y_f|x,\theta) & \leq & P^H_{FY}(y_f|x,\theta)) \end{array}$$

where $\Sigma^1_{f,ISD}$ is the set of player-f strategies that survive one ISD round, as per Theorem (1).

Table 1 compares CT Bounds, ISD Bounds, firm-level ISD Bounds (as defined in Remark 2) and FY Bounds for the outcome $(0,1)$. For CT Bounds, ISD Bounds, and firm-level ISD Bounds, the "difference" between the upper and the lower bound is equal to region ⑤, hence for outcome $(0,1)$ both bounds are equally informative. For FY Bounds, in contrast, the "difference" between the lower and the upper bounds is larger as it include areas where a single ISD round is not able to rule out anything for one player, but iterative application of would ISD result more precise prediction.

The difference between CT Bounds and ISD Bounds becomes apparent in Table 2, where I consider outcome $(0,0)$. Ignoring mixed strategies, there is no value of ξ for which $(0,0)$ is one of many equilibrium outcomes, so there is no difference between CT upper and lower bounds. In contrast, $y = (0,0)$ survives ISD in regions ① and ⑤, so the upper ISD Bound is larger than the upper CT Bound. Here, again, FY bounds are wider than either CT, ISD, and firm-level ISD

Table 2: CT vs. ISD Bounds for $(0,0)$

Bound Type	Lower Bound	Upper Bound
CT $(0,0)$	$Pr\left(\xi \in ①\right)$	$Pr\left(\xi \in ①\right)$
ISD $(0,0)$	$Pr\left(\xi \in ①\right)$	$Pr\left(\xi \in ① \cup ⑤\right)$
ISD $y_1 = 0$	$Pr\left(\xi \in ① \cup ② \cup ③ \cup ⑥\right)$	$Pr\left(\xi \in ① \cup ② \cup ③ \cup ⑤ \cup ⑥\right)$
ISD $y_2 = 0$	$Pr\left(\xi \in ① \cup ④ \cup ⑦ \cup ⑧\right)$	$Pr\left(\xi \in ① \cup ④ \cup ⑤ \cup ⑦ \cup ⑧\right)$
FY $y_1 = 0$	$Pr\left(\xi \in ① \cup ② \cup ③\right)$	$Pr\left(\xi \in ① \cup ② \cup ③ \cup ④ \cup ⑤ \cup ⑥\right)$
FY $y_2 = 0$	$Pr\left(\xi \in ① \cup ④ \cup ⑦\right)$	$Pr\left(\xi \in ① \cup ② \cup ④ \cup ⑤ \cup ⑦ \cup ⑧\right)$

Bounds.

Let us move to the Monte Carlo experiments. Consider the entry game introduced in sections 2.3 and 3.2, with $|\mathcal{F}| = 2, 3$, and assume that firm's get profits:

$$\pi_f(y_f, y_{f'}; x_f, \theta, \xi_f) = y_f \left(\frac{x_f}{(2 + \sum_{f' \neq f} y_{-f})^2} - \theta_{ec} + \theta_{sc}\xi_f \right)$$

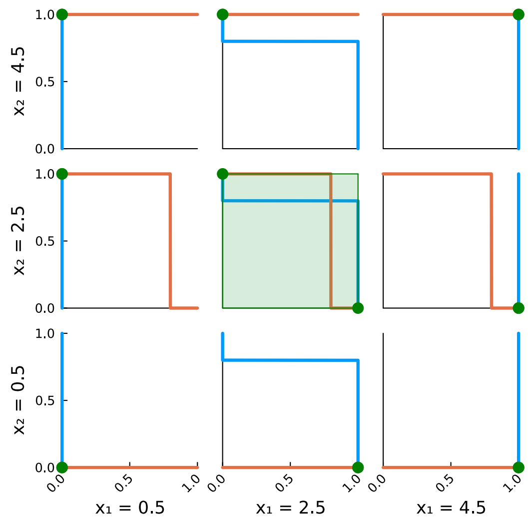
with $\theta = (\theta_{ec}, \theta_{sc})$, and where $\xi \sim N(0,1)$.

For simulations I set $(\theta^0_{ec}, \theta^0_{sc}) = (1,1)$ and generate 100 Monte Carlo Samples consisting of 4000 markets each. For each market I draw $X_m \in \{0.5, 2.5, 4.5\}^{|\mathcal{F}|}$ from a uniform distribution, and draw for each firm an ξ_{fm} from a standard normal. For each (X_m, ξ_m) I simulate the complete information game assuming that the data come from a randomly chosen pure strategy Nash equilibrium.

Figure 8 depicts the the best responses generated in the two player case for each possible $x \in \mathcal{X}$, when $\xi_1 = \xi_2 = 0$. If $x_f = 0.5$, *no entry* is dominant for f, whereas if $x_f = 4.5$ then *entry* is dominant. For all the plots in the borders, at least one firm has a dominant strategy and the game is dominance solvable. When $x_f = 2.5$ for both firms, everything survives ISD. The figure shows that the range of x_f is wide enough to produce every possible game, and that the ISD has no bite for many of the games that this DGP generates.

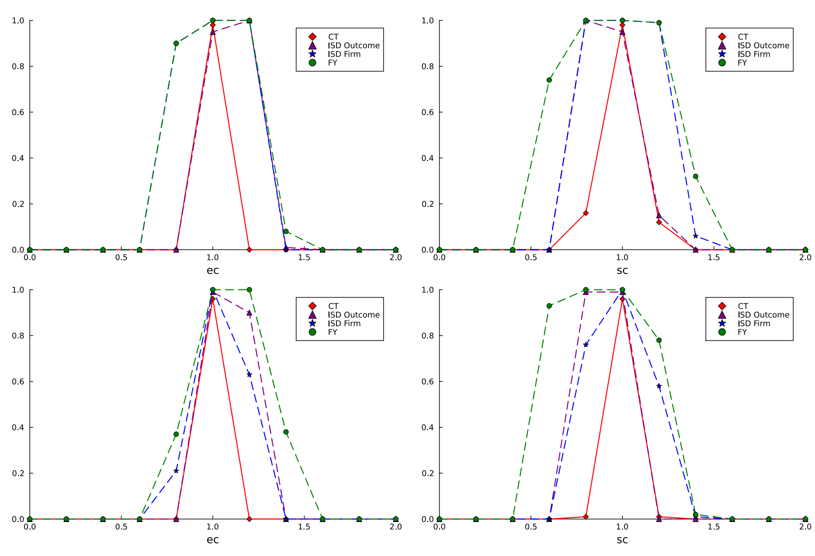
As before, for each Monte Carlo sample I compute the AS (see Appendix D) and compute the share of samples for which the hypothesis $\theta = \theta_0$ is not rejected. As expected, CT Bounds provide the smallest confidence set, followed by outcome-level ISD Bounds, then firm-level ISD Bounds, and finally FY bounds. Regardless of the type of bounds, the confidence sets for both parameters seem to be large larger with 3 firms rather than 2. This is to be expected, since intuition suggests that the set strategies that survives ISD increases with the number of players.

Figure 8: Complete Information Two-way Best Responses for Values of $x \in \mathcal{X}$



Note: Two-way best responses for each $x \in \mathcal{X}$. Each column (row) depicts the games for a fixed value of x_1 (x_2). The horizontal (vertical) axis of each plot represents the entry probability of firm 1 (resp. 2) and the blue (resp. orange) line represents its best response. Green dots represent the (pure strategy) Nash equilibria, and the light green areas represents the strategies that survive ISD.

Figure 9: $Pr(\hat{\theta} \in \Theta_{ISD})$ for $|\mathcal{F}| = 2$ (top) and $|\mathcal{F}| = 3$ (bottom) under Complete Information and Unobserved Heterogeneity



Note: The left column shows the results for θ_{ec} and the right column for θ_{sc}. The top row shows the results for $|\mathcal{F}| = 2$, and the bottom row for $|\mathcal{F}| = 3$. For each column, the horizontal axis represents $\hat{\theta}_{ec}$ and $\hat{\theta}_{sc}$. Finally, in all subplots the vertical axis represent the percentage of samples where the null hypothesis that $(\theta_{ec}, \theta_{sc}) = (1, 1)$ is *not rejected*.

6 Airline Industry Background and Data

6.1 Industry Background and the JetBlue/Spirit Merger

Over the past fifteen years the U.S. Airline Industry has seen a marked trend towards higher concentration. Just between 2010 and 2013 three mega-mergers took place in the U.S.: Delta/Northwest (2010) United/Continental (2012), and American/US Airways (2013). At their respective times, each of these mergers produced the largest carrier in the world. Furthermore, American Airlines remains the largest carrier to this day.[26]

More recently, in 2020, American and JetBlue created *Northeastern Alliance* with the alleged goal of realizing economies of scope through code sharing and network externalities. The alliance was contested by the Department of Justice on the grounds that it reduced competition, and on May 2023 a federal judge ruled that the alliance violated antitrust law and forced the parties to end it.

Against this backdrop, on March of 2022 JetBlue offered to buy Spirit for \$3.8 billion. The Department of Justice sued to block the acquisition, alleging that it would increase prices and harm consumers, while JetBlue and Spirit defended the merger on the grounds that it would allow them to realize economies of scope, expand their network and lower their prices. On January 2024, the merger was blocked by a federal judge.

To evaluate this merger, and to put ISD Bounds to test, in the next section I propose a model that allows me to capture the main pro and anti-competitive effects through which the merger is likely to affect consumer welfare. On the anti-competitive side, the merger is likely to result in higher prices through decreased competition in markets where JetBlue and Spirit overlap. On the pro-competitive side, if a larger network allows firms to realize economies of scope that reduce marginal costs or entry costs, then the price increase in overlapping markets will be ameliorated, and entry of the merged entity into new markets will result in increased competition.

Figure 10 and Table 3, below, use data from the DB1B dataset for the first quarter of 2019 (see Section 6.2) to compute some basic details about the merging parties. Figure 10 shows the network of each airline, while Table 3 shows the number of markets (non-directional airport pairs) where the each firm was active, as well as the number of markets where each firm was a potential entrant, i.e., had activity at both endpoints.[27]

The table reveals that JetBlue's and Spirit's networks overlap in 30 markets and that they jointly serve 257, hence they overlap in roughly 12% of the markets. Additionally, the table reveals that the carriers are simultaneously potential entrants in 317 different markets, these are markets where the number of potential entrants would decrease due to the merger. At the same time, the joint firm would be a potential entrant in 1281 markets, which is 209 additional markets than if we count

[26]Although national concentration after these mergers increased, this did not necessarily translate into increased concentration at the route level. For example, the Delta/Northwest merger did not raise opposition from the Department of Justice because the carriers were direct competitors only in a handful of markets.

[27]The number of *Active* markets in Table 3 does not match the number of *Direct Routes* in Figure 10, because *Active* markets include non-direct routes.

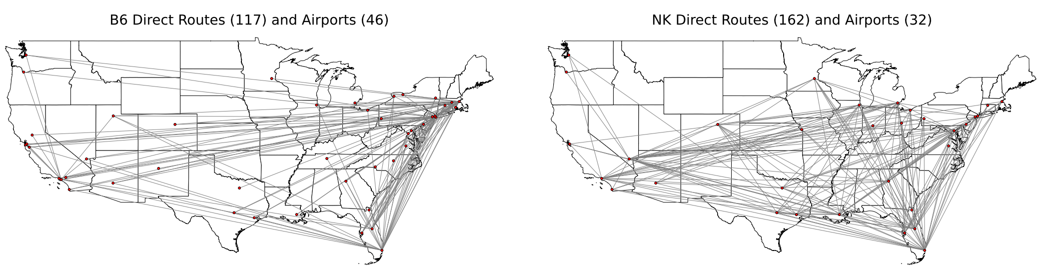
the status quo of level $1072(= 860 + 529 - 317)$.[28]

Table 3: Markets where JetBlue (B6) and Spirit (NK) are Active/Players

Carrier	*Active*	*Player*
B6	118	1016
NK	169	534
Overlap	30	381
Joint	257	1281

Note: A firm is *active* in a market if it serves more than 100 passengers and its market share is more than 1%. A firm is a *player* in a market if it is *active* in at least one other market at each endpoint.

Figure 10: Direct Routes for JetBlue (B6) and Spirit (NK).



6.2 Data Description

My main data source is the Origin and Destination Survey (DB1B) collected by the Bureau of Transportation Statistics (BTS). The data consists of a sample of 10% of all trips taken within the U.S. in a given quarter/year. For each trip, it contains the price of the ticket as well as the origin and destination airports, and all layover airports. The DB1B is a widely used data source in the airline literature (e.g., Berry (1992), Ciliberto and Tamer (2009), Aguirregabiria and Ho (2012)).

I use the DB1B data set for the first quarter of 2019, and supplement it with information on airport locations (city) from the BTS, as well as population data from the Census Bureau. I keep the airports located at the 70 top MSAs in terms of population, which yields a total of 80 airports over 66 MSAs.[29] Table 4 presents the list of the top 15 MSAs ranked by the population, while Table 5 shows some airport summary statistics.

A market corresponds to a non-directional airport pair regardless of the number of stops. With 80 airports, this would imply $3160(= 80 \cdot 79/2)$ markets, however, I drop airport pairs that lie in the same MSA (e.g. JFK and La Guardia), and hand-drop airport pairs that are too close to have flights between them (e.g., JFK-PHL, PHL-BWI, BWI-DCA) leaving a total of 3078 markets.

[28] Consider a market that connects airports A and B. If pre-merger JetBlue has operations only on A, and Spirit has operations only in B, then neither firm is a potential entrant to this market. After the merger, however, the merged entity is a potential entrant because it has operations at both end points.

[29] Some MSAs are lost because there is no airport attached to them, or because the DB1B does not show flights involving airports in these MSAs and airports in other top-80 MSAs.

Table 4: Top 15 Cities by MSA Population

Airports	*City*	*State*	*Population*
$EWR-JFK-LGA$	$NewYork$	NY	19.2
$LAX-LGB$	$LosAngeles$	CA	13.2
$MDW-ORD$	$Chicago$	IL	9.5
$DAL-DFW$	$Dallas$	TX	7.6
$HOU-IAH$	$Houston$	TX	7.1
$DCA-IAD$	$Washington$	DC	6.3
$FLL-MIA$	$Miami$	FL	6.2
PHL	$Philadelphia$	PA	6.1
ATL	$Atlanta$	GA	6.0
$PHX-AZA$	$Phoenix$	AZ	4.9
BOS	$Boston$	MA	4.9
$OAK-SFO$	$SanFrancisco$	CA	4.7
ONT	$Riverside$	CA	4.7
DTW	$Detroit$	MI	4.3
SEA	$Seattle$	WA	4.0

Note: Top 15 cities in terms of MSA population (millions) and their airports.

In terms of carriers, I keep American (AA), JetBlue (B6), Delta (DL), Spirit (NK), United (UA), and Southwest (WN). I group the remaining airlines under a fictional Low Cost Carrier (LCC). Table 6 presents summary statistics, across airports, of the number of direct destinations served by each carrier, and Table 7 presents summary statistics for each carrier at the market level.

Table 5: Airport Summary Statistics

Stat	*Population*	*Carriers*	*Destinations*
$Mean$	3.8	5.3	34.2
SD	4.1	1.7	18.8
Min	0.8	1.0	1.0
$q25$	1.3	5.0	18.8
$Median$	2.2	6.0	35.0
$q75$	4.8	7.0	49.2
Max	19.2	7.0	69.0

Note: Airport summary statistics. *Population* is measured in millions of people, *Carriers* represents the number of carriers that have a flight arriving to/departing from an airport, and *Direct Destinations* represents the number of unique airports that can be reached from an airport via a direct flight.

Finally, Table 8 shows the distribution of number of players across markets. Each row represents said distribution conditional on the corresponding carrier being a player. It is noteworthy that in aggregate the legacy airlines are the unique player in 103 markets, and non legacy carriers are the unique player in 249 markets. In these markets carriers face a single agent problem, so for all parameter values the lower and higher ISD bounds concide, i.e., $P^L = P^H$. AS such, these markets will be particularly helpful in pinning down the parameter values. In fact, in principle one could

Table 6: Direct Destinations Across Airports for Each Carrier

Carrier	*min*	*p25*	*mean*	*p50*	*p75*	*max*	*SD*
AA	0.0	4.0	10.4	7.0	9.0	66.0	14.4
B6	0.0	0.0	2.9	1.0	3.0	36.0	6.1
DL	0.0	3.0	10.6	6.0	10.0	68.0	13.9
LLC	0.0	3.0	9.4	7.0	11.0	51.0	10.1
NK	0.0	0.0	4.1	0.0	5.2	27.0	6.9
UA	0.0	3.0	9.2	5.5	9.0	62.0	13.4
WN	0.0	7.0	25.2	25.5	41.5	59.0	19.8

Note: Summary statistics (across airports) of the number of direct flights offered by each carrier.

recover point identification from these markets alone.

7 Parametrized Airline Model

The model consists of a two stage game. In the first stage firms simultaneously and independently choose whether to enter a market or not. In the second stage firms compete in prices. In what follows, I present the building blocks of the model in reverse order.

7.1 Second Stage: Price Competition

The stage starts with all carriers and consumers in market m observing the outcome of the entry stage, $y_m = (y_{fm})_{f\in\mathcal{F}_m} \in \{0,1\}^{|\mathcal{F}_m|}$, where $\mathcal{F}_m$ is the number of players/potential entrants. From among the carriers that entered the market, i.e., $y_{fm} = 1$, consumer i chooses which one to fly with, $f \in \mathcal{F}$, or the outside option $f = 0$. By choosing carrier f consumer i gets:

$$\begin{aligned} u_{ifm} &= \beta' x^d_{fm} - \lambda p_{fm} + \vartheta_{fm} + \varphi_{ifm} && \text{if } f \in \mathcal{F} \\ u_{i0m} &= \varphi_{i0m} && \text{if } f = 0 \end{aligned}$$

Where x^d_{fm} is a vector of exogenous demand/utility shifters, the impact of which is controlled by the vector of parameters β. p_{fm} is the price charged by airline f in market m and λ is the price sensitivity. Finally, ϑ_{fm} is an unobserved (by the econometrician) taste shifter, φ_{ifm} is an idiosyncratic taste shock.

The distribution of $(\varphi_{ifm})_{f\in\mathcal{F}\cup\{0\}}$ is such that the induced demand is nested logit with nesting parameter $\tau \in (0,1]$, where one nest contains the outside option, and the other nest contains all products.[30] Under this assumption, the market share of firm f is:

$$s_{fm} = \frac{e^{(\beta' x^d_{fm} - \lambda p_{fm} + \vartheta_{fm})/\tau}}{D_m} \frac{D^\tau_m}{1 + D^\tau_m}$$

[30]Cardell (1997) showed that there exist a random variable, ι_{ifm}, such that $\varphi_{ifm} = \iota_{ifm} + \tau\tilde{\varphi}_{ifm}$ generates a nested logit distribution with nesting parameter τ, where φ_{ifm} is an type-I extreme value shock.

Table 7: Carrier Level Summary Statistics

Carrier	*Statistic*	*Player*	*Active*	*Act|Pl*	*Passengers*	*Price*
AA	*Mean*	0.75	0.31	0.41	1596.0	2.4
	SD	0.43	0.46	0.49	2544.6	0.6
B6	*Mean*	0.33	0.04	0.12	3719.0	1.9
	SD	0.47	0.19	0.32	3321.8	0.6
DL	*Mean*	0.75	0.27	0.36	1650.0	2.4
	SD	0.43	0.44	0.48	2398.5	0.6
LCC	*Mean*	0.73	0.13	0.18	1733.0	1.5
	SD	0.44	0.34	0.38	2363.8	0.5
NK	*Mean*	0.17	0.05	0.32	1183.0	1.1
	SD	0.38	0.23	0.47	1078.0	0.1
UA	*Mean*	0.7	0.18	0.25	1926.0	2.5
	SD	0.46	0.38	0.43	2922.2	0.6
WN	*Mean*	0.61	0.34	0.56	1908.0	1.8
	SD	0.49	0.47	0.5	2449.1	0.4

Note: Carrier level means and standard deviations. Player (Active) represents the share of markets where the carrier is a potential entrant (is active, i.e., $Y_{fm} = 1$). Act|Pl represents the share of markets where the carrier is active conditional on it being a player. Passengers and Price correspond to the number of passengers and the price in active markets.

where D_m is the logit denominator:

$$D_m = \sum_{f:y_{fm}=1} e^{(\beta' x^d_{fm} - \lambda p_{fm} + \vartheta_{fm})/\tau}$$

The parameter τ controls the substitution between the flying with any carrier and the outside option. When $\tau = 1$, the model reduces to a standard multinomial logit, whereas when τ approaches zero the substitution between the flying and the outside option vanishes.

Carriers that did not enter the market during the entry stage do nothing. Carriers that did enter the market choose prices simultaneously and independently. Carrier f in market m chooses p_{fm} to solve:

$$\max_{p_{fm}} \quad size_m (p_{fm} - c_{fm}) s_{fm}$$

where $size_m$ is the size of market m, and c_{fm} is the constant marginal cost of production which I parametrize as:

$$\log(c_{fm}) = \eta' x^c_{fm} + \upsilon_{fm}$$

where x^c_{fm} are cost shifters of firm f in market m, η is a vector of parameters that controls the effect of the cost shifters, and υ_{fm} is a common knowledge random shock which is unobserved by the econometrician.

Table 8: Distribution of Number of Players Across Markets

Players	*AA*	*B6*	*DL*	*LCC*	*NK*	*UA*	*WN*	*Total*
0.0	0	0	0	0	0	0	0	150
1.0	36	25	67	206	0	0	18	352
2.0	140	65	82	40	7	85	91	255
3.0	252	45	263	183	12	254	122	377
4.0	263	62	277	244	23	216	119	301
5.0	838	101	838	815	75	818	775	852
6.0	538	473	544	514	172	535	500	546
7.0	245	245	245	245	245	245	245	245
Total	2312	1016	2316	2247	534	2153	1870	3078

Note: For a given carrier, say AA, the corresponding column shows the frecquency of "number of players" across markets in which AA is active. Similarly, the last row shows the number of markets in which AA is a player. The last column shows how many markets have any given number of players. Note that, becuase of double (and triple, quadruple, etc.) counting, the last row is not the sum of the rest of the numbers in the row, but rather the sum divided by the number of players.

7.2 Entry Game

Potential entrants in market m. $\mathcal{F}_m$, simultaneously choose whether to enter market or not. Each carrier has ISD-Consistent beliefs, $\rho_m^f \in \Sigma_{-fm,ISD}$, regarding its competitors strategies and chooses $y_{fm} \in \{0,1\}$ to maximize its interim profit:

$$\Pi_{fm} = y_{fm}\left(R_f(\rho_m^f; x_m) - z'_{fm}\theta_{ec,f} + \theta_{sc,f}\left(\sqrt{\theta_w}\cdot\xi_{fm} + \sqrt{1-\theta_w}\cdot\epsilon_{fm}\right)\right) \tag{9}$$

where $R_f(\rho_m^f; x_m)$ is f's expected variable profit of firm f, with the expectation taken over ρ_m^f, and the demand and marginal cost shocks $(\vartheta_g, \upsilon_g)_{\forall g}$. The vector z_{fm} contains entry cost shifters controlled by the parameter vector $\theta_{ec,f}$, while $\theta_{sc,f}$ is the scale parameter of the joint profit shock, $\sqrt{\theta_w}\xi_{fm} + \sqrt{1-\theta_w}\epsilon_{fm}$. As before, ξ_{fm} is public information and ϵ_{fm} is private information, both of which follow a standard normal distribution, and $\theta_w \in [0,1]$ controls the amount of variation that comes from each source.

Garrido (2021) and Nocke and Schutz (2018) show that entry always decreases equilibrium profits in a (nested) logit Bertrand pricing game. As a result, Π_f exhibits decreasing in (y_{fm}, y_{-fm}), which guarantees that the game is in fact an SMSG for all $\theta = ((\theta_{ec,f}, \theta_{sc,f})_f, \theta_w)$, all (x_m, z_m), and all ξ_m. Since many beliefs may be ISD-Consistent, our solution concept does not pin down a unique distribution over outcomes for all observables, θ and ξ. Hence, ISD Bounds are necessary for estimation.

8 Estimation and Results

I estimate the game in two stages. First I estimate the demand and cost parameters that determine the pricing stage outcomes using standard IO techniques. Then, I use these estimates to compute the expected variable profits, R_{fm}, under every possible market structure, $y_m \in \mathcal{Y}_m$ and use these expected variable profit estimates to estimate the entry stage parameters.

8.1 Pricing Game Estimation

For demand estimation, I use the nested logit inversion from Berry (1994) to get the following estimable expression which I estimate using 2SLS.

$$\log\left(\frac{s_{fm}}{s_{0m}}\right) = \beta' x_{fm}^d - \lambda p_{fm} + (1-\rho)\log\left(\frac{s_{fm}}{1-s_{0m}}\right) + \vartheta_{fm}$$

Demand shifters in x_{fm}^d include *distance* between airports (measured in thousands of kilometers), *distance squared*, carrier dummies, and as in Aguirregabiria and Ho (2012) I control for *hub size*, i.e., the geometric mean of the total population served by direct flights at both end-points by carrier f, as well as origin/destination airport dummies. Table 9 shows some summary statistics of the demand (and marginal cost) shifters.

To deal with the endogeneity of p_{fm} and $\log\left(\frac{s_{fm}}{1-s_{0m}}\right)$, I use the number of active carriers, the number of direct flights, whether an endpoint is a *hub*, and a dummy for monopolists markets. Intuitively, given that firms do not learn about demand and marginal cost shocks until after entry decisions are made, all of these variables are orthogonal to the shocks in the pricing game.[31]

With the demand parameter estimates at hand I back out the marginal costs that rationalize the observed prices, c_{fm}, and use these to estimate the marginal cost equation via OLS. Marginal cost shifters in x_{fm}^c include *distance*, *distance squared*, the geometric mean across end points of the number of direct destinations for a given carrier, i.e., *carrier_ airport_ ndest*, and airport and carrier dummies. See Table 9 for summary statistics.

The demand and marginal cost estimates can be found in Table 10. Parameter estimates have all the expected signs and are, in magnitude, in line with the previous research. It is noteworthy the effect of carrier dummies in demand and marginal are what one would expect. While legacy airlines, i.e., American, Delta, and United, are perceived as higher quality than the low cost counterparts, i.e., Jet Blue, Spirit, Southwest, and LCC, they also have higher marginal costs.

8.2 Entry Game Estimation

[TBA]

8.3 Counterfactuals: Spirit/JetBlue Merger

[TBA]

[31]Bontemps et al. (2023) and Aguirregabiria and Ho (2012) use similar instruments.

Table 9: Demand and Marginal Cost Controls Summary Statistics

$variable$	$Statistic$	AA	$B6$	DL	LCC	NK	UA	WN
p	$Mean$	2.5	1.8	2.6	1.4	1.1	2.6	2.0
	$Median$	2.4	1.7	2.6	1.1	1.1	2.5	2.0
	SD	0.5	0.5	0.6	0.6	0.1	0.5	0.4
$distance$	$Mean$	2.0	2.0	1.9	2.0	1.9	2.2	1.9
	$Median$	1.8	1.7	1.6	1.7	1.7	2.1	1.7
	SD	1.1	1.2	1.1	1.0	0.9	1.1	1.0
$distance_{sq}$	$Mean$	5.0	5.6	4.8	5.0	4.3	6.0	4.7
	$Median$	3.1	3.0	2.6	3.0	2.9	4.3	3.0
	SD	4.8	5.8	4.8	4.8	3.8	5.2	4.4
hub_{size_AH}	$Mean$	71.7	53.6	66.2	55.3	68.8	76.4	83.1
	$Median$	68.6	55.2	61.2	52.1	63.0	72.4	77.7
	SD	27.3	19.8	27.2	26.3	23.4	27.3	29.9
$ndest$	$Mean$	13.6	9.8	14.2	15.8	13.5	13.8	20.2
	$Median$	11.8	8.8	12.7	14.5	12.7	13.5	18.4
	SD	8.8	4.7	8.7	7.0	4.4	8.5	8.7

Note: Carrier specific mean/median/s.d. for demand and marginal cost variables. *Price* is measures in of dollars, *distance* in thousands of kilometers, *hub size* is the geometric mean, across end-points, of the aggregate population (in millions) connected by directs flights, and *carrier airport destinations* is the geometric mean of the number of direct destinations at each end point.

9 Closing Remarks

I provided probability bounds on (the distribution of) outcomes of games, and showed that they pin down an identified set for the parameters of interests. The bounds are based on an ISD argument (ISD Bounds), so they are robust to multiple equilibria both in pure and mixed strategies, as well as to non-equilibrium play as long as beliefs are ISD-Consistent. As opposed to previous bounds proposed in the literature, ISD Bounds can accommodate games of discrete or continuous strategies of any dimensionality, and allow for any informational structure regarding the players' private shocks (e.g., complete information, independent private information, privileged information), and they are informative about the underlying informational structure. i.e., different informational structures will produce different bounds.

To maximize the bite of ISD Bounds I introduce the Strategically Monotonic Supermodular Games, i.e., games where payoffs are supermodular on own actions, and exhibit either increasing differences or decreasing differences between own and competitors' actions. I argue that for these games ISD is informative, in that it rules out large swaths of the strategy set, and useful, in that the bounds are easy to compute.

In Monte Carlo simulations, I show that ISD Bounds are informative about the parameters of interest. Furthermore, I show that the bounds are able to inform about the relative degree of private information vs. unobserved heterogeneity in the underlying DGP.

Finally, in an application to the airline industry I show that ISD Bounds are practical and

Table 10: Second Stage Nested Logit

	(1) log_csos	(2) log_mc
p	-2.038***	
	(0.300)	
log(s_f|In)	0.184**	
	(0.068)	
distance	0.422***	0.196***
	(0.092)	(0.013)
distance_sq	-0.028**	0.007**
	(0.013)	(0.003)
hub_size_AH	0.028***	
	(0.001)	
AA	0.000	0.000
	(.)	(.)
B6	0.516**	-0.358***
	(0.192)	(0.017)
DL	0.676***	0.089***
	(0.070)	(0.006)
LCC	-1.282***	-0.777***
	(0.280)	(0.020)
NK	-2.491***	-1.044***
	(0.348)	(0.015)
UA	0.124**	0.023**
	(0.041)	(0.007)
WN	-0.953***	-0.292***
	(0.154)	(0.007)
ndest		0.003***
		(0.000)
Constant	0.152	0.247
	(0.738)	(0.179)
Observations	8120	8120
AirportFE	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

informative about the parameters of interest. I take advantage of these estimates to evaluate the proposed merger between JetBlue and Spirit and find, in counterfactual simulations, that [TBA]

References

Ackerberg, D. A. and G. Gowrisankaran (2006). Quantifying equilibrium network externalities in the ach banking industry. *The RAND Journal of Economics 37*(3), 738–761.

Aguirregabiria, V. and C.-Y. Ho (2012, May). A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics 168*(1), 156–173.

Aguirregabiria, V. and P. Mira (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica 70*(4), 1519–1543.

Aguirregabiria, V. and P. Mira (2019). Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity. *Quantitative Economics 10*(4), 1659–1701.

Andrews, D. W. K. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

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

Aradillas-Lopez, A. (2011). Nonparametric probability bounds for nash equilibrium actions in a simultaneous discrete game. *Quantitative Economics 2*(2), 135–171.

Aradillas-López, A. and A. M. Rosen (2022). Inference in ordered response games with complete information. *Journal of Econometrics 226*(2), 451–476.

Aradillas-Lopez, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business & Economic Statistics 26*(3), 261–283.

Atal, J. P., J. I. Cuesta, and M. Sæthre (2022). Quality regulation and competition: Evidence from pharmaceutical markets.

Bajari, P., H. Hong, and S. P. Ryan (2010). Identification and estimation of a discrete games of complete information. *Econometrica 78*(5), 1529–1568.

Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica 79*(6), 1785–1821.

Berry, S., A. Eizenberg, and J. Waldfogel (2016). Optimal product variety in radio markets. *The RAND Journal of Economics 47*(3), 463–497.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica 60*(4), 889–917.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Bontemps, C., C. Gualdani, and K. Remmy (2023). Price competition and endogenous product choice in networks: Evidence from the us airline industry. *crctr224_2023_400, University of Bonn and University of Mannheim, Germany*.

Bresnahan, T. F. and P. C. Reiss (1990). Entry in monopoly markets. *The Review of Economic Studies 57*(4), 531–553.

Bresnahan, T. F. and P. C. Reiss (1991). Entry and competition in concentrated markets. *Journal of Political Economy 99*(5), 977–1009.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory 13*(2), 185–213.

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

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

de Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Draganska, M., M. Mazzeo, and K. Seim (2009, Jun). Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *QME 7*(2), 105–146.

Eizenberg, A. (2014, 03). Upstream Innovation and Product Variety in the U.S. Home PC Market. *The Review of Economic Studies 81*(3), 1003–1045.

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

Fan, Y. and C. Yang (2020, May). Competition, product proliferation, and welfare: A study of the us smartphone market. *American Economic Journal: Microeconomics 12*(2), 99–134.

Fan, Y. and C. Yang (2022, June). Estimating discrete games with many firms and many decisions: An application to merger and product variety. Working Paper 30146, National Bureau of Economic Research.

Garrido, F. (2021). An aggregative approach to price equilibrium among multi-product firms with nested demand. *SSRN*.

Grieco, P. L. E. (2014). Discrete games with flexible information structures: an application to local grocery markets. *The RAND Journal of Economics 45*(2), 303–340.

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

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

Li, S. Y., J. Mazur, Y. Park, J. W. Roberts, A. Sweeting, and J. Zhang (2018, January). Endogenous and selective service choices after airline mergers. Working Paper 24214, National Bureau of Economic Research.

Magnolfi, L. and C. Roncoroni (2020). Estimation of discrete games with weak assumptions on information. The warwick economics research paper series (twerps), University of Warwick, Department of Economics.

Mazzeo, M. J. (2002). Product choice and oligopoly market structure. *The RAND Journal of Economics 33*(2), 221–242.

Milgrom, P. and J. Roberts (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica 58*(6), 1255–1277.

Molinari, F. and A. M. Rosen (2008). [the identification power of equilibrium in simple games]: Comment. *Journal of Business & Economic Statistics 26*(3), 297–302.

Nocke, V. and N. Schutz (2018). Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica 86*(2), 523–557.

Otsu, T. and M. Pesendorfer (2022, 04). Equilibrium Multiplicity in Dynamic Games: Testing and Estimation. *The Econometrics Journal*. utac006.

Otsu, T., M. Pesendorfer, and Y. Takahashi (2016). Pooling data across markets in dynamic markov games. *Quantitative Economics 7*(2), 523–559.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015). Moment inequalities and their application. *Econometrica 83*(1), 315–334.

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

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

Uetake, K. and Y. Watanabe (2013, 12). Estimating supermodular games using rationalizable strategies. *Advances in Econometrics 31*, 233–247.

Uetake, K. and Y. Watanabe (2020). Entry by merger: Estimates from a two-sided matching model with externalities. *SSRN*.

Van Zandt, T. and X. Vives (2007). Monotone equilibria in bayesian games of strategic complementarities. *Journal of Economic Theory 134*(1), 339–360.

Wollmann, T. G. (2018, June). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review 108*(6), 1364–1406.

A Proof of Theorem 1 and Best Response Iteration

A.1 Proof of Theorem 1

Here I prove Theorem 1. To this end, it is useful to show lemma 1, below, which states that conditions the define an SMSG (Definition 5), hold for the game written with interim profits.

Lemma 1 (Interim SMSG)**.** *Let the (x, θ, ξ)-game be an SMSG. Then (omitting dependence on (x, θ, ξ) for brevity):*

1.a. ***Complete Lattice Strategy Set:*** *The strategy set Σ_f, together with the partial order "$\geq$" is a complete and compact lattice for all $f \in \mathcal{F}$, where $\sigma_f \geq \sigma'_f \Leftrightarrow \sigma_f(\epsilon_f) \geq \sigma'_f(\epsilon_f)$ for all $\epsilon_f \in \mathcal{E}_f$.*

1.b. ***Order Upper-Semi Continuity:*** *The interim profit function, Π_f, is order upper semi-continuous. This is, for any totally ordered set $C \subset \mathcal{Y}_f$:*

$$\limsup_{y_f \in C, y_f \downarrow \inf(C)} \Pi_f(y_f, \sigma_{-f}, \epsilon_f) \leq \Pi_f(\inf(C), \sigma_{-f}, \epsilon_f)$$
$$\limsup_{y_f \in C, y_f \uparrow \sup(C)} \Pi_f(y_f, \sigma_{-f}, \epsilon_f) \leq \Pi_f(\sup(C), \sigma_{-f}, \epsilon_f)$$

for all $\sigma_{-f} \in \Sigma_{-f}$, all $f \in \mathcal{F}$.

1.c. ***Supermodularity:*** *The interim profit function Π_f is supermodular in y_f for all σ_{-f}.*

1.d. ***Strategic Monotonicity:*** *For all $\epsilon_f \in \mathcal{E}_f$, and all $f, f' \in \mathcal{F}$, if $f' \in C(f)$ then Π_f has ID in $(y_f, \sigma_{f'})$, and if or $f' \in S(f)$, then Π_f has DD in $(y_f, \sigma_{f'})$.*

Proof. Fix an arbitrary SMSG. I begin by showing that Σ_f, together with the partial order $\leq$, where $\sigma_f \leq \sigma'_f \Leftrightarrow \sigma_f(\epsilon_f) \leq \sigma'_f(\epsilon_f)$, for all ϵ_f,[32] conform a complete lattice.

Take two strategies σ_f and σ'_f. By definition, for all ϵ_f, $\sigma_f(\epsilon_f), \sigma'_f(\epsilon_f) \in \mathcal{Y}_f$, hence $\sup\{\sigma_f(\epsilon_f), \sigma'_f(\epsilon_f)\} \in \mathcal{Y}_f$ and $\inf\{\sigma_f(\epsilon_f), \sigma'_f(\epsilon_f)\} \in \mathcal{Y}$, for all ϵ, which implies $\sup\{\sigma_f, \sigma'_f\}, \inf\{\sigma_f, \sigma'_f\} \in \Sigma_f$.

This shows that Σ_f is a lattice. The argument for completeness is analogous. Consider a collection of strategies $\tilde{\Sigma}_f \subseteq \Sigma_f$, and let $\tilde{\mathcal{Y}}_f(\epsilon_f) = \{y_f \in \mathcal{Y}_f : \sigma_f(\epsilon_f) = y_f \text{ for some } \sigma_f \in \tilde{\Sigma}_f\}$. Since $\tilde{\mathcal{Y}}_f(\epsilon_f) \subseteq \mathcal{Y}_f$, and $\mathcal{Y}_f$ is a complete lattice, then $\sup\{\tilde{\mathcal{Y}}_f(\epsilon_f)\}, \inf\{\tilde{\mathcal{Y}}_f(\epsilon_f)\} \in \mathcal{Y}_f$ for all ϵ_f, which implies $\sup\{\tilde{\Sigma}_f\}, \inf\{\tilde{\Sigma}_f\} \in \Sigma_f$.

To see that Π_f is order upper semi-continuous simply fix a strategy for f's competitors σ_{-f}. By order upper semi-continuity of π_f, for any ϵ_f and any totally ordered set $C \subset \mathcal{Y}_f$:

$$\limsup_{y_f \in C, y_f \downarrow \inf(C)} \pi_f(y_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) \leq \pi_f(\inf(C), \sigma_{-f}(\epsilon_{-f}), \epsilon_f)$$
$$\limsup_{y_f \in C, y_f \uparrow \sup(C)} \pi_f(y_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) \leq \pi_f(\sup(C), \sigma_{-f}(\epsilon_{-f}), \epsilon_f)$$

[32] I slightly abuse notation by using "$\leq$" to denote the standard vector inequality and the partial order in Σ.

Integrating over ϵ_{-f}.

$$\limsup_{y_f \in C, y_f \downarrow \inf(C)} \Pi_f(y_f, \sigma_{-f}, \epsilon_f) \leq \Pi_f(\inf(C), \sigma_{-f}, \epsilon_f)$$
$$\limsup_{y_f \in C, y_f \uparrow \sup(C)} \Pi_f(y_f, \sigma_{-f}, \epsilon_f) \leq \Pi_f(\sup(C), \sigma_{-f}, \epsilon_f)$$

as desired.

To see that Π_f is supermodular consider any two actions y_f and y'_f, and fix an arbitrary strategy for f's competitors σ_{-f}. By supermodularity of π_f, for any ϵ_f:

$$\begin{aligned} & \pi_f(\sup\{y_f, y'_f\}, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) + \pi_f(\inf\{y_f, y'_f\}, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) \\ \geq \; & \pi_f(y_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) + \pi_f(y'_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) \end{aligned}$$

which, integrating over ϵ_{-f}, yields:

$$\Pi_f(\sup\{y_f, y'_f\}, \sigma_{-f}, \epsilon_f) + \Pi_f(\inf\{y_f, y'_f\}, \sigma_{-f}, \epsilon_f) \geq \Pi_f(y_f, \sigma_{-f}, \epsilon_f) + \Pi_f(y'_f, \sigma_{-f}, \epsilon_f)$$

Finally, I show that if π_f has ID in (y_f, y_{-f}), then Π_f has ID in (y_f, σ_{-f}) (the proof for the DD case is analogous). Fix actions $y'_f \geq y_f$ and a pair of strategies for f's competitors, $\sigma'_{-f} \geq \sigma_{-f}$. By ID of π_f, for any ϵ_f:

$$\begin{aligned} & \pi_f(y'_f, \sigma'_{-f}(\epsilon_{-f}), \epsilon_f) - \pi_f(y'_f, \sigma'_{-f}(\epsilon_{-f}), \epsilon_f) \geq \\ & \quad \pi_f(y'_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) - \pi_f(y'_f, \sigma_{-f}(\epsilon_{-f}), \epsilon_f) \end{aligned}$$

Integrating over ϵ_{-f},

$$\Pi_f(y'_f, \sigma'_{-f}, \epsilon_f) - \Pi_f(y_f, \sigma'_{-f}, \epsilon_f) \geq \Pi_f(y'_f, \sigma_{-f}, \epsilon_f) - \Pi_f(y_f, \sigma_{-f}, \epsilon_f)$$

as desired. □

Having shown Lemma 1, we are in a position to show Theorem 1, which I restate below.

Theorem 1. [ISD in SMSGs] *Let the (x, θ, ξ)-game be an SMSG and say $\sigma \leq \sigma'$ if and only if*

$\sigma(\epsilon) \leq \sigma'(\epsilon)$ for all ϵ. Consider the following sequence:

$$
\begin{array}{rcl}
\textbf{Set up:} & & \\
\Sigma^0_{ISD} & = & \Sigma \\
\sigma^{H,0}_f & = & \{\sup\{\mathcal{Y}_f\} : \epsilon_f \in \mathcal{E}_f\} \\
\sigma^{L,0}_f & = & \{\inf\{\mathcal{Y}_f\} : \epsilon_f \in \mathcal{E}_f\} \\
\Sigma^k_{ISD} & = & \{\sigma \in \Sigma : \sigma^{L,k} \leq \sigma \leq \sigma^{H,k}\} \\
\mathcal{Y}^k_f(\epsilon_f) & = & \{y_f \in \mathcal{Y}_f : \sigma^{L,k}_f(\epsilon_f) \leq y_f \leq \sigma^{H,k}_f(\epsilon_f)\} \\
\textbf{Best/Worst Case:} & & \\
\sigma^{B,k}_{-f} & = & \left(\sigma^{H,k}_{C(f)}, \sigma^{L,k}_{S(f)}\right) \\
\sigma^{W,k}_{-f} & = & \left(\sigma^{L,k}_{C(f)}, \sigma^{L,k}_{S(f)}\right) \\
\textbf{Update:} & & \\
\sigma^{H,k}_f & = & \sup\left\{ \underset{y_f \in \mathcal{Y}^{k-1}_f(\epsilon_f)}{argmax} \quad \Pi_f(y_f, \sigma^{B,k-1}_{-f}, \epsilon_f) : \epsilon_f \in \mathcal{E}_f \right\} \\
\sigma^{L,k}_f & = & \inf\left\{ \underset{y_f \in \mathcal{Y}^{k-1}_f(\epsilon_f)}{argmax} \quad \Pi_f(y_f, \sigma^{W,k-1}_{-f}, \epsilon_f) : \epsilon_f \in \mathcal{E}_f \right\}
\end{array} \tag{3}
$$

The following holds:

1.a. *For each $k = 1, 2, \ldots$, all $\sigma_f \ngeq \sigma^{k,L}_f$ and all $\sigma_f \nleq \sigma^{k,H}_f$ are dominated relative to Σ^{k-1}_{ISD}.*

1.b. *The set Σ^k_{ISD} contains all strategies that survives k ISD rounds.*

1.c. *For $k \to \infty$, $(\sigma^{k,L}, \sigma^{k,H}) \to (\sigma^L, \sigma^H)$ with $\sigma^L \leq \sigma^H$. Furthermore, the set:*

$$\Sigma_{ISD} = \{\sigma \in \Sigma : \sigma^L \leq \sigma \leq \sigma^H\}$$

contains all strategies that survive ISD.

1.d. *If $\sigma^L = \sigma^H$ then the game is dominance solvable, and this strategy profile is the unique (Bayes) Nash Equilibrium.*

Proof. I start by generalizing Lemma 1 of Milgrom and Roberts (1990) to the case of Strategic Monotonicity. Consider an SMSG and let $\tilde{\Sigma}(s^L, s^H) = \{\sigma \in \Sigma : s^L \leq \sigma \leq s^H\}$ for some pair of strategy profiles $s^L \leq s^H$ in Σ. Let $\lambda^L_f(\sigma_{-f})$ and $\lambda^H_f(\sigma_{-f})$ be f's lowest and highest best responses to σ_{-f} in $\tilde{\Sigma}_f(s^L_f, s^H_f)$.[33] Furthermore, let $\lambda^L_f(\epsilon_f; \sigma_{-f})$ and $\lambda^H_f(\epsilon_f; \sigma_{-f})$ be these strategies evaluated at ϵ_f. Finally let $\sigma^B_{-f} = (s^H_{C(f)}, s^L_{S(f)})$ be the "best case" for firm f, i.e., the case where f's complements are playing their highest possible strategy and f's substitutes are playing their lowest possible strategy. I argue that any $\sigma_f \in \tilde{\Sigma}_f(s^L_f, s^H_f)$, such that $\sigma_f \nleq \lambda^H_f(\sigma^B_{-f})$ is strictly dominated (relative to $\tilde{\Sigma}_{-f}(s^L_{-f}, s^H_{-f})$) by $\inf\{\sigma_f, \lambda^H_f(\sigma^B_{-f})\}$.

[33] By Assumption 1 and Lemma 1, these are guaranteed to exist.

If there is no $\sigma_f \in \tilde{\Sigma}_f(s_f^L, s_f^H)$ such that $\sigma_f \not\leq \lambda_f^H(\sigma_{-f}^B)$, the statement is trivially true and we are done.

Say a $\sigma_f \not\leq \lambda_f^H(\sigma_{-f}^B)$ exists. By definition there is at least one ϵ_f such that $\sigma_f(\epsilon_f) \not\leq \lambda_f^H(\epsilon_f; \sigma_{-f}^B)$, so for such ϵ_f, $\sigma_f(\epsilon_f) \geq \inf\{\sigma_f(\epsilon_f), \lambda_f^H(\epsilon_f; \sigma_{-f}^B)\}$. Then, for any $\sigma_{-f} = (\sigma_{C(f)}, \sigma_{S(f)}) \in \tilde{\Sigma}_{-f}(s^L, s^H)$:

$$
\begin{aligned}
& \Pi_f\Big(\sigma_f(\epsilon_f), (\sigma_{C(f)}, \sigma_{S(f)}), \epsilon_f\Big) - \Pi_f\Big(\inf\Big\{\sigma_f(\epsilon_f), \lambda_f^H(\epsilon_f; \sigma_{-f}^B)\Big\}, (\sigma_{C(f)}, \sigma_{S(f)}), \epsilon_f\Big) \\
< \; & \Pi_f\Big(\sigma_f(\epsilon_f), (s_{C(f)}^H, \sigma_{S(f)}), \epsilon_f\Big) - \Pi_f\Big(\inf\Big\{\sigma_f(\epsilon_f), \lambda_f^H(\epsilon_f; \sigma_{-f}^B)\Big\}, (s_{C(f)}^H, \sigma_{S(f)}), \epsilon_f\Big) \\
< \; & \Pi_f\Big(\sigma_f(\epsilon_f), (s_{C(f)}^H, s_{S(f)}^L), \epsilon_f\Big) - \Pi_f\Big(\inf\Big\{\sigma_f(\epsilon_f), \lambda_f^H(\epsilon_f; \sigma_{-f}^B)\Big\}, (s_{C(f)}^H, s_{S(f)}^L), \epsilon_f\Big) \\
\leq \; & \Pi_f\Big(\sup\big\{\sigma_f(\epsilon_f), \lambda_f^H(\epsilon_f; \sigma_{-f}^B)\big\}, \sigma_{-f}^B, \epsilon_f\Big) - \Pi_f\Big(\lambda_f^H(\epsilon_f; \sigma_{-f}^B), \sigma_{-f}^B, \epsilon_f\Big) \\
\leq \; & 0
\end{aligned}
$$

where the first inequality uses the fact Π_f has ID in $(y_f, \sigma_{C(f)})$, and the second inequality comes from the fact that Π_f has DD in $(y_f, \sigma_{S(f)})$. The third comes from supermodularity of Π_f and from substituting $\sigma_{-f}^B = (s_{C(f)}^H, s_{S(f)}^L)$, while the fourth inequality follows from fact that $\lambda_f^L(\epsilon_f; \sigma_{-f}^B)$ maximizes Π_f given σ_{-f}^B and ϵ_f. It follows that:

$$
\Pi_f(\sigma_f(\epsilon_f), \sigma_{-f}, \epsilon_f) < \Pi_f(\inf\{\sigma_f(\epsilon_f), \lambda_f^H(\epsilon_f; \sigma_{-f}^B)\}, \sigma_{-f}, \epsilon_f)
$$

for all $\sigma_{-f} \in \tilde{\Sigma}_{-f}(s_{-f}^L, s_{-f}^H)$.

Letting $\sigma_{-f}^W = (s_{C(f)}^L, s_{S(f)}^H)$ be the "worst case," for f, an analogue argument shows that $\sigma_f \not\geq \lambda_f^L(\sigma_{-f}^W)$ is strictly dominated by $\sup\{\sigma_f, \lambda_f^L(\sigma_{-f}^W)\}$. From these two results, it follows that every strategy in $\tilde{\Sigma}_f(s^L, s^H) \setminus \tilde{\Sigma}_f(\lambda^L(\sigma^W), \lambda^H(\sigma^B))$ is strictly dominated. This concludes the generalization of Lemma 1 from Milgrom and Roberts (1990) to the case of strategic monotonicity.

The statements of the Theorem follow immediately. For part 1.a., set $\Sigma_{ISD}^k = \tilde{\Sigma}(\sigma^{k,L}, \sigma^{k,H})$ and $\Sigma_{ISD}^{k+1} = \tilde{\Sigma}(\lambda^L(\sigma^{k,W}), \lambda^H(\sigma^{k,B}))$. The result above implies that for every f, every strategy in $\Sigma_{f,ISD}^k \setminus \Sigma_{f,ISD}^{k+1}$ is strictly dominated relative to Σ_{ISD}^k, as desired.

Part 1.b. follows by definition of Σ_{ISD}^k.

Part 1.c. follows from the fact that $\sigma^{k,L} \leq \sigma^{k,H}$ for all k, and the fact that $\sigma^{k,L}$ is increasing, and $\sigma^{k,H}$ decreasing, in k.

Part 1.d. is straightforward. Say $\sigma^L = \sigma^H$, then for every firm $\sigma_f^L = \sigma_f^H$ and $\sigma_{-f}^W = \sigma_{-f}^L$. By definition σ_f^L is a best response to σ_{-f}^W, so it is a best response to σ_{-f}^L, so σ^L is a (Bayesian) Nash equilibrium. □

A.2 Applying ISD

Here I show how to apply ISD to an SMSG for two special cases of interest: the pure ID case, where all players are each others' complements, i.e., $C(f) = \mathcal{F} \setminus \{f\}$ for all f; and the pure DD case, where all players are each others' substitutes, i.e., $S(f) = \mathcal{F} \setminus \{f\}$ for all f. The pure ID case encompasses coordination games, while the pure DD case encompasses games with strategic substitution (like the entry game in the example).

A.2.1 Pure ID Case

In this case, the best response iteration that converges to (σ^L, σ^H) follows directly from Milgrom and Roberts (1990) and Van Zandt and Vives (2007). The details of the sequence are outlined in (10).

To get an intuition, consider the case where y_f is univariate, and start from f's "best case," i.e., $\sigma_{-f}^{0,H}(\epsilon_{-f}) = \sup\{\mathcal{Y}_{-f}\}$ for all ϵ_{-f}. By ID, f's best response to $\sigma_{-f}^{0,H}$, i.e., $\sigma_f^{1,H}$, is the largest strategy that f can optimally choose, and it strictly dominates all $\sigma_f > \sigma_f^{1,H}$. Since this holds for all f, all strategy profiles $\sigma > \sigma^{1,H}$ are eliminated by $\sigma^{1,H}$. Iterating over this procedure yields the largest strategy profile not eliminated by ISD, σ^H. An analogous sequence, starting from $\sigma^{0,L}$, yields σ^L.

ISD sequence for the ID case.

$$
\begin{array}{rcl}
\multicolumn{3}{l}{\textbf{Set-up}} \\
\hline
\sigma^{0,L} & = & \{\inf\{\mathcal{Y}\} : \epsilon \in \mathcal{E}\} \\
\sigma^{0,H} & = & \{\sup\{\mathcal{Y}\} : \epsilon \in \mathcal{E}\} \\
\Sigma_{ISD}^{k} & = & \left\{\sigma \in \Sigma : \sigma^{k,L} \leq \sigma \leq \sigma^{k,H}\right\} \\
\mathcal{Y}_f^k(\epsilon_f) & = & \left\{y_f \in \mathcal{Y}_f : \sigma_f^{k,L}(\epsilon_f) \leq y_f \leq \sigma_f^{k,H}(\epsilon_f)\right\} \\
\\
\multicolumn{3}{l}{\textbf{ISD Step}} \\
\hline
\sigma_f^{k,L} & = & \left\{\inf\left\{\underset{y_f \in \mathcal{Y}_f^{k-1}(\epsilon_f)}{\operatorname{argmax}} \quad \Pi_f(y_f, \sigma_{-f}^{k-1,L}, \epsilon_f)\right\} : \epsilon_f \in \mathcal{E}_f\right\} \\
\sigma_f^{k,H} & = & \left\{\sup\left\{\underset{y_f \in \mathcal{Y}_f^{k-1}(\epsilon_f)}{\operatorname{argmax}} \quad \Pi_f(y_f, \sigma_{-f}^{k-1,H}, \epsilon_f)\right\} : \epsilon_f \in \mathcal{E}_f\right\}
\end{array}
\tag{10}
$$

A.2.2 Pure DD Games

The intuition for the pure DD case is similar. Consider the case of univariate y_f for all f, and start from f's "best case," i.e., $\sigma_{-f}^{0,L}(\epsilon_{-f}) = \inf\{\mathcal{Y}_{-f}\}$ for all ϵ_{-f}, and its "worst case," i.e., $\sigma_{-f}^{0,H}(\epsilon_{-f}) = \sup\{\mathcal{Y}_{-f}\}$ for all ϵ_{-f}. In the best case, DD implies that $\sigma_f^{1,H}$ is the largest strategy that player f could plausibly choose, hence any $\sigma_f > \sigma_f^{1,H}$ is dominated. Since this is true for all f, we can discard all $\sigma > \sigma^{H,1}$. Similarly, in the worst case, $\sigma_f^{1,L}$ is the smallest best response that f could plausibly choose, hence any $\sigma_f < \sigma_f^{1,L}$ is strictly dominated by $\sigma_f^{1,L}$. Since this is true for all f we can safely discard all $\sigma < \sigma^{1,L}$. Putting these two arguments together, we build a new game with strategy set $\Sigma^1 = \{\sigma \in \Sigma : \sigma^{1,L} \leq \sigma \leq \sigma^{1,H}\}$. Finally, applying this argument iteratively, yields the extreme strategy profiles σ^L and σ^H.

ISD sequence for the ID case.

$$
\begin{array}{rcl}
\multicolumn{3}{l}{\textbf{Set-up}} \\
\hline
\sigma^{0,L} & = & \{\inf\{\mathcal{Y}\} : \epsilon \in \mathcal{E}\} \\
\sigma^{0,H} & = & \{\sup\{\mathcal{Y}\} : \epsilon \in \mathcal{E}\} \\
\Sigma_{ISD}^k & = & \left\{\sigma \in \Sigma : \sigma^{k,L} \leq \sigma \leq \sigma^{k,H}\right\} \\
\mathcal{Y}_f^k(\epsilon_f) & = & \left\{y_f \in \mathcal{Y}_f : \sigma_f^{k,L}(\epsilon_f) \leq y_f \leq \sigma_f^{k,H}(\epsilon_f)\right\} \\
\\
\multicolumn{3}{l}{\textbf{ISD Step}} \\
\hline
\sigma_f^{k,L} & = & \left\{\inf\left\{\underset{y_f \in \mathcal{Y}_f^{k-1}(\epsilon_f)}{\operatorname{argmax}} \quad \Pi_f(y_f, \sigma_{-f}^{k-1,H}, \epsilon_f)\right\} : \epsilon_f \in \mathcal{E}_f\right\} \\
\sigma_f^{k,H} & = & \left\{\sup\left\{\underset{y_f \in \mathcal{Y}_f^{k-1}(\epsilon_f)}{\operatorname{argmax}} \quad \Pi_f(y_f, \sigma_{-f}^{k-1,L}, \epsilon_f)\right\} : \epsilon_f \in \mathcal{E}_f\right\}
\end{array}
\tag{11}
$$

B Outcome Probability Bounds

Theorem 2 derives bounds on the distribution y_f. Analogue bounds can be built for the distribution of y rather than y_f. For binary games with independent private information, Theorem 3 implies that there is no informational gain from moving from bounds over y_f rather than y. Furthermore, empirical moment inequalities based on y are like less precise than those based on y_f. As a result, it is hard to justify the move.

Nevertheless, if we drop the independent private information assumption, it may be that bounds on y are tighter than bounds on y_f. For this reason, here I show how one would build bounds over y. To this end,consider the following probabilities:

$$
\begin{aligned}
P^L(y|x,\theta) &\equiv \int_\Xi \int_\mathcal{E} \mathbb{1}\{\sigma^L(\epsilon; x, \theta, \xi) \geq y\} dG(\epsilon|x,\theta,\xi) dH(\xi|x,\theta) \\
P^\rho(y|x,\theta) &\equiv \int_\Xi \int_\mathcal{E} \mathbb{1}\{\sigma^\rho(\epsilon; x, \theta, \xi) \geq y\} dG(\epsilon|x,\theta,\xi) dH(\xi|x,\theta) \\
P^H(y|x,\theta) &\equiv \int_\Xi \int_\mathcal{E} \mathbb{1}\{\sigma^L(\epsilon; x, \theta, \xi) \geq y\} dG(\epsilon|x,\theta,\xi) dH(\xi|x,\theta)
\end{aligned}
\tag{12}
$$

With this, I present *Outcome Probability ISD Bounds* as Corollary to Theorem 2.

Corollary 1. [Outcome Probability ISD Bounds in SMSGs] *Consider a class of games indexed by* (x, θ, ξ) *satisfying Assumption 1. Furthermore, let* $H(\cdot|x, \theta)$ *be the distribution of* ξ. *Say for every* f *and* g, ρ_g^f *is ISD-Consistent as per Assumption 2. The following holds:*

$$P^L(y|x,\theta) \leq P^\rho(y|x,\theta) \leq P^H(y|x,\theta) \tag{13}$$

for all y, x, *and* θ.

Proof. Fix (x, θ, ξ) and a set of ISD-Consistent beliefs, ρ. Because ρ is ISD-Consistent, then each firm's best response also survives ISD, hence $\sigma^L \leq \sigma^\rho \leq \sigma^H$. It is easy to see that the following inequality holds (making explicit the dependence on (x, θ, ξ)):

$$\mathbb{1}\left\{\sigma^L(\epsilon; x, \theta, \xi) \geq y\right\} \leq \mathbb{1}\left\{\sigma^\rho(\epsilon; x, \theta, \xi) \geq y\right\} \leq \mathbb{1}\left\{\sigma^H(\epsilon; x, \theta, \xi) \geq y\right\}$$

Integrating over ϵ and ξ yields the desired expression. □

With this, we can define the identified set based on Outcome Probability ISD Bounds as follows:

Definition 8 (Outcome Probability ISD Identified Set)**.** *The Outcome Probability ISD Identified Set,* $\Theta_{ISD}^{Outcome\ Pr.}$, *is the collection of all* $\theta \in \Theta$ *such that:*

$$\Theta_{ISD}^{Outcome\ Pr.} = \{\theta \in \Theta : P^L(y|x,\theta) \leq P^\rho(y|x,\theta) \leq P^H(y|x,\theta), \forall x \in \mathcal{X}, \forall y \in \mathcal{Y}\}$$

We can also use Theorem 1 to build bounds over outcomes rather than their distribution. This bounds are likely much less informative than the bounds proposed in Theorem 2, as they involve integrating over the implied distribution of y. I add them here, nevertheless, for the sake of completeness. To move in this direction, consider the following expressions:

$$\begin{aligned} y^L(x,\theta) &\equiv \int_\Xi \int_\mathcal{E} \sigma^L(\epsilon; x, \theta, \xi) dG(\epsilon|x,\theta,\xi) dH(\xi|x,\theta) \\ y^\rho(x,\theta) &\equiv \int_\Xi \int_\mathcal{E} \sigma^\rho(\epsilon; x, \theta, \xi) dG(\epsilon|x,\theta,\xi) dH(\xi|x,\theta) \\ y^H(x,\theta) &\equiv \int_\Xi \int_\mathcal{E} \sigma^H(\epsilon; x, \theta, \xi) dG(\epsilon|x,\theta,\xi) dH(\xi|x,\theta) \end{aligned} \tag{14}$$

An immediate corollary of Theorem 1 is that $y^L(x,\theta) \leq y^\rho(x,\theta) \leq y^H(x,\theta)$. Letting $y^0(x) = y^{\rho^0}(x, \theta^0)$ be the expected outcome under the DGP, conditional on x, and applying the same logic we used in Definitions 6 and 8 we can define the following identified set:

Definition 9 (Level ISD Identified Set)**.** *The Level ISD identified set is the collection of all* $\theta \in \Theta$ *such that:*

$$\Theta_{ISD} = \{\theta \in \Theta : y^L(x,\theta) \leq y^0(x) \leq y^H(x,\theta), \forall x \in \mathcal{X}\}$$

C Identification Results

C.1 Proof of Theorem 3

Theorem 3. *For any class of binary games (i.e., $\mathcal{Y}_f = \{0,1\}$ for all $f \in \mathcal{F}$) indexed by (x, θ, ξ) and satisfying Assumptions 1 and 2, and any corresponding DGP satisfying Assumption 3, if either:*

- $\epsilon_f \perp \epsilon_g$ *for all* $f \neq g$, *and* ϵ_f *has a continuous distribution with full support in* $\mathbb{R}$. *Or,*
- $\epsilon_f = \varnothing$ *for all* $f \in \mathcal{F}$, *i.e., firms have complete information.*

Then, the ISD identified set is sharp, i.e., $\Theta_{ISD} = \Theta_{\text{sharp}}$.

Proof. For ease of exposition, throughout this proof I omit dependence on x.

That $\Theta_{\text{sharp}} \subseteq \Theta_{ISD}$ follows by definition. To see this say $\theta \in \Theta_{\text{sharp}}$, then for every ξ there is an ISD-Consistent ρ such that $P^\rho(y|\theta,\xi) = P^0(y|\xi)$. Because ρ is ISD-Consistent $P^L(y|\xi,\theta) \leq P^\rho(y|\xi,\theta) \leq P^H(y|\xi,\theta)$. Since this holds for all ξ, $\theta \in \Theta_{ISD}$.

Now I show the inclusion in the other direction. To this end, note that both under independence of the ϵ_f's and under perfect information, firms actions are independent conditional on ρ, i.e., $\sigma_f^\rho(\epsilon_f;\theta,\xi) \perp \sigma_g^\rho(\epsilon_g;\theta,\xi)$.[34] Hence, we can write $P^0(y|\xi)$ as the product of the marginals (recall that P^0 is a complementary cumulative distribution), i.e.,

$$P^0(y|\xi) = \prod_{f \in \mathcal{F}} \left(P_f^0(\xi)\right)^{y_f} \left(1 - P_f^0(\xi)\right)^{1-y_f}$$

where P_f^0 is the marginal probability that $y_f = 1$ in the DGP. With this, we can write any P^0 as the sum of P_f^0's as follows:

$$P^0(y|\xi) = \sum_{\tilde{y} \geq y} \prod_{f \in \mathcal{F}} \left(P_f^0(\xi)\right)^{\tilde{y}_f} \left(1 - P_f^0(\xi)\right)^{1-\tilde{y}_f}$$

Hence, to show that Θ_{ISD} is sharp it suffices to show that for any $\theta \in \Theta_{ISD}$, any ξ, and any f, there are ISD-Consistent beliefs such that $P_f^\rho(\theta,\xi) = P_f^0(\xi)$.

For the imperfect information case, I do this by showing that $P_f^\rho(\theta,\xi)$ is a continuous function of ρ^f whose minimum is $P_f^L(\theta,\xi)$ and its maximum is $P_f^H(\theta,\xi)$. Then, the result follows by the intermediate value theorem.

That the minimum and maximum values of $P_f^\rho(\theta,\xi)$ are $P_f^L(\theta,\xi)$ and $P_f^H(\theta,\xi)$, respectively, follows from the definition of sequence (3) in Theorem (1). In particular, P_f^ρ is minimal when $\rho^f = P_{-f}^W(\theta,\xi)$, and it is maximal when $\rho^f = P_f^B(\theta,\xi)$, where P_f^W and P_f^B are f's best and worst cases as defined in sequence (3).

[34] If ϵ_f is correlated to ϵ_g then f infers information about g's action through ϵ_f. In such case, even conditional on ρ, σ_f and σ_g are correlated through the correlation between the shocks.

To see that P_f^ρ is continuous, note that f's interim payoff given ρ^f, i.e.:

$$\Pi_f(y_f, \rho^f, \epsilon_f; \theta, \xi) = \sum_{y_{-f} \in \mathcal{Y}_{-f}} \left[\prod_{g \neq f} (\rho_g^f)^{y_g} (1 - \rho_g^f)^{1-y_g} \right] \pi_f(y_f, y_{-f}, \epsilon_f; \theta, \xi)$$

is continuous in ρ^f, hence f's best response is:

$$P_f^\rho(\theta, \xi) = \int_{\epsilon_f \in \mathcal{E}_f} \mathbb{1} \left\{ \Pi_f(1, \rho^f, \epsilon_f; \theta, \xi) \geq \Pi_f(0, \rho^f, \epsilon_f; \theta, \xi) \right\} dG_f$$

which is also continuous in ρ^f, as desired.

Now consider the complete information case, i.e., $\epsilon = \varnothing$. Given ρ^f, f's interim payoff is:

$$\Pi_f(y_f, \rho^f; \theta, \xi) = \sum_{y_{-f} \in \mathcal{Y}_{-f}} \left[\prod_{g \neq f} (\rho_g^f)^{y_g} (1 - \rho_g^f)^{1-y_g} \right] \pi_f(y_f, y_{-f}; \theta, \xi)$$

If $\Pi_f(0, \rho^f; \theta, \xi) > \Pi_f(1, \rho^f; \theta, \xi)$ for all ISD-Consistent ρ^f, i.e., $y_f = 0$ is dominant, then:

$$P_f^L(\theta, \xi) = P_f^\rho(\theta, \xi) = P_f^0(\xi) = P_f^H(\theta, \xi) = 0$$

Similarly, if $\Pi_f(0, \rho^f; x, \theta, \xi) < \Pi_f(1, \rho^f; x, \theta, \xi)$ for all ISD-Consistent ρ^f, i.e., $y_f = 1$ is dominant, then:

$$P_f^L(\theta, \xi) = P_f^\rho(\theta, \xi) = P_f^0(\xi) = P_f^H(\theta, \xi) = 1$$

Finally, if $\Pi_f(0, \rho^f; \theta, \xi) = \Pi_f(1, \rho^f; \theta, \xi)$ for some ISD-Consistent ρ^f, i.e., no strategy is dominant nor dominated, then $P_f^L(\theta, \xi) = 0$ and $P_f^H(\theta, \xi) = 1$. For such ρ^f any P_f^ρ is a best response. In particular $P_f^\rho(\xi, \theta) = P_f^0(\xi)$ is optimal. Putting these three cases together we can always find ρ such that $P_f^\rho(\xi, \theta) = P_f^0(\xi)$, hence $\theta \in \Theta_{\text{sharp}}$, as desired. □

C.2 Proof Theorem 4

Theorem 4. *There exists a $\theta \in \Theta$ such that $P^L(y|x, \theta) = 0$ and $P^H(y|x, \theta) = 1$ for all $y > (0)_{\forall f}$ and all x, if and only if R_f satisfies:*

$$\sup_{x \in \mathcal{X}} \left\{ \min_{y_{-f} \in \mathcal{Y}_f} \Delta R_f(y_{-f}; x, \theta_{f,R}) \right\} < \inf_{x \in \mathcal{X}} \left\{ \max_{y_{-f} \in \mathcal{Y}_f} \Delta R_f(y_{-f}; x, \theta_{f,R}) \right\} \tag{8}$$

for all $\theta_{f,R}$ and all $f \in \mathcal{F}$.

Proof. Consider the sequence of best response iterations defined in (3) and let:

$$\begin{aligned}\Delta_f^B(x,\theta_{R,f}) &\equiv \Delta R_f(\sigma_{-f}^{0,B};x,\theta_{R,f})\\ \Delta_f^W(x,\theta_{R,f}) &\equiv \Delta R_f(\sigma_{-f}^{0,W};x,\theta_{R,f})\end{aligned}$$

Note that since $\sigma^{0,B}$ and $\sigma^{0,W}$ are deterministic strategies (see the definition of sequence (3)), we can do without the expectation operator. Furthermore, note that by increasing differences/decreasing differences, Δ_f^B is the maximal ΔR_f and Δ_f^W is the minimal ΔR_f over σ_{-f}.

To show necessity, say $\theta_{R,f}$ satisfies:

$$\sup_{x\in\mathcal{X}}\left\{\Delta_f^W(x,\theta_{R,f})\right\} < \inf_{x\in\mathcal{X}}\left\{\Delta_f^B(x,\theta_{R,f})\right\}$$

and set:

$$\theta_{cons,f} = -\frac{1}{2}\left(\sup_{x\in\mathcal{X}}\left\{\Delta_f^W(x,\theta_{R,f})\right\} + \inf_{x\in\mathcal{X}}\left\{\Delta_f^B(x,\theta_{R,f})\right\}\right)$$

By construction, we can write:

$$\begin{array}{rcccl}\Delta_f^W(x,\theta_{R,f}) & < & -\theta_{cons,f} & < & \Delta_f^B(x,\theta_{R,f})\\ \Delta_f^W(x,\theta_{R,f})+\theta_{cons,f} & < & 0 & < & \Delta_f^B(x,\theta_{R,f})+\theta_{cons,f}\end{array}$$

for all x, where the second line simply re-arranges some terms. The left-hand side (right-hand side) of this inequality represents the marginal profit of going from $y_f = 0$ to $y_f = 1$ under $\sigma_{-f}^{W,0}$ (resp. $\sigma_{-f}^{B,0}$). Hence:

$$\begin{aligned}\sigma_f^{1,H}(\epsilon_f) &= \mathbb{1}\left\{\Delta_f^B(x,\theta_{R,f})+\theta_{cons,f}+\theta_{sc,f}\left(\sqrt{\theta_{w,f}}\xi_f+\sqrt{1-\theta_{w,f}}\epsilon_f\right)>0\right\}\\ \sigma_f^{1,L}(\epsilon_f) &= \mathbb{1}\left\{\Delta_f^W(x,\theta_{R,f})+\theta_{cons,f}+\theta_{sc,f}\left(\sqrt{\theta_{w,f}}\xi_f+\sqrt{1-\theta_{w,f}}\epsilon_f\right)>0\right\}\end{aligned}$$

which induces entry probabilities equal to:

$$\begin{aligned}P_f^{1,H}(x,\theta,\xi) &= 1-G_f\left(-\frac{1}{\sqrt{1-\theta_w}}\left(\frac{\Delta_f^B(x,\theta_{R,f})+\theta_{cons,f}}{\theta_{sc,f}}+\sqrt{\theta_{w,f}}\xi_f\right)\right)\\ P_f^{1,L}(x,\theta,\xi) &= 1-G_f\left(-\frac{1}{\sqrt{1-\theta_w}}\left(\frac{\Delta_f^W(x,\theta_{R,f})+\theta_{cons,f}}{\theta_{sc,f}}+\sqrt{\theta_{w,f}}\xi_f\right)\right)\end{aligned}$$

where G_f is the marginal distribution of ϵ_f. By construction $(\Delta_f^B+\theta_{cons,f})/\theta_{sc,f} > 0$ for all x, so taking the limit as $\theta_{sc,f}\to 0$ implies $P_f^{1,H} = P^{0,H} = 1$. A similar argument shows that $\theta_{sc,f}\to\infty$ implies $P_f^{1,L} = P_f^{0,L} = 0$. We have reached convergence after one iterations, so $P_f^L(x,\theta,\xi) = 0$ and $P_f^H(x,\theta,\xi) = 1$ for all x and ξ.

For sufficiency I show the contra-positive. Fix R_f and assume that for all $\theta_{R,f}$ there exist x and

x' such that:

$$\Delta_f^B(x, \theta_{R,f}) < \Delta_f^W(x', \theta_{R,f})$$

For any θ we can write:

$$\frac{\Delta_f^W(x, \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}} < \frac{\Delta_f^B(x, \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}} < \frac{\Delta_f^W(x', \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}} < \frac{\Delta_f^B(x', \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}}$$

where the left-hand side represents f's best case (normalized by $\theta_{sc,f}$) marginal profit under x, the term in the center represents f's worst case marginal profit under x' and the right-hand side represents f's best case marginal profit under x'.

If all the expressions in the inequality above are positive (negative), then ISD Bounds are not trivial since ϵ_f has positive density on all $\mathbb{R}$. If the last two terms are positive and the first two terms are negative, then $P_f^H(x, \theta, \xi) < 1$ for some ξ and $P_f^L(x', \theta, \xi) > 0$ for some ξ, so bounds are not trivial.

Finally, say that:

$$\frac{\Delta_f^W(x', \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}} < 0 < \frac{\Delta_f^B(x', \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}}$$

then $P^H(x, \theta, \xi) < 1$. Similarly if:

$$\frac{\Delta_f^W(x, \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}} < 0 < \frac{\Delta_f^B(x, \theta_{R,f}) + \theta_{cons,f}}{\theta_{sc,f}}$$

then $P^L(x, \theta) > 0$

This exhaust all possibilities, hence there is no way to simultaneously make $P^L(x, \theta, \xi) = P^L(x', \theta, \xi) = 0$ and $P^H(x, \theta, \xi) = P^H(x', \theta, \xi) = 1$, as desired.

$\square$

C.3 Non-Identification in Models with Linear Competitive Effect

Here I present a non-identification result that arises in entry games when competitive effects are linear. For example, as in $\pi_f(y_f, y_{-f}, \epsilon_f; x, \theta) = y_f(x'\theta_x - \theta_{st} \sum_{g \neq f} y_g + \epsilon_f)$, where θ_{st} is the strategic effect.[35]

The problem arises because given taking $\theta_{st,f} \to \infty$ makes the profit of *entry* go to negative infinity whenever $\rho_g^f > 0$ for some g, but it does not affect the profit of entry when $\rho_g^f = 0$.

Consider a two firm independent private information entry game like the one described in 2.3.1.

[35] Note that this specification covers many examples of games that have been studied in prior the prior empirical literature such as Bresnahan and Reiss (1990), Berry (1992), Tamer (2003), Ciliberto and Tamer (2009), among others.

Letting P_{-f} represent the entry probability of f's competitor we can write f's profit as:

$$\Pi_f = y_f \left(x'\theta_x - \theta_{st} y_{-f} P_{-f} + \epsilon_f \right)$$

where $\theta = (\theta_x, \theta_{st})$ are the parameters of interest, and $\theta^0 = (\theta_x^0, \theta_{st}^0)$ are the "real" parameter values. As $\theta_{st} \to \infty$ the profit of being a duopolist goes to negative infinity, hence if $P_{-f} = 1$ we get $\lim_{\theta_{st}\to\infty} P_f^L \to 0$. In contrast, if $P_{-f} = 0$ the monopolistic profit remains unchanged even if $\theta_{st} \to \infty$, hence: $\lim_{\theta_{st}\to\infty} P_f^H = 1 - G(-x'\theta_x)$.

To see why this is a problem, let ρ^0 be the real ISD-Consistent beliefs. It is easy to see that for $P_f^0 = 1 - G(-\theta_x^{0\prime} x + \theta_{st} y_{-f})$ the following holds:

$$P_f^L \leq P_f^0 \leq \lim_{\theta_{st}\to\infty} P_f^H$$

hence clearly $P_f^L(x,\theta) \leq P_f^0(x) \leq P_f^H(x,\theta)$.

As mentioned above, the key driver of this result is that monopolistic profits do not depend on γ, so by increasing γ we can make duopolistic profits infinitely unattractive while keeping monopolistic profits unchanged. Figure 11, below, shows mutual best responses for $R^{mon} = 1$, $\delta_0 = 1$, $\beta_0 = 0$ and different values of γ. The red dot represents the entry probabilities when there is no strategic interaction, i.e., $\gamma = 0$, the purple dot represents the data generating process y_0 which results from $\gamma_0 = 1$, and the green dot/area represents the strategies that survive ISD. For $\gamma \in \{0.5, 1, 2\}$ the game has a unique equilibrium so only one strategy profile survives ISD, i.e., $\underline{y} = \overline{y}$. Furthermore, for $\gamma = 0.5$, we have $\underline{y} > y_0$ so $\gamma = 0.5$ is not in the identified set. Similarly, for $\gamma = 2$, $y_0 > \overline{y}_f$, so $\gamma = 2$ is not in the identified set either.

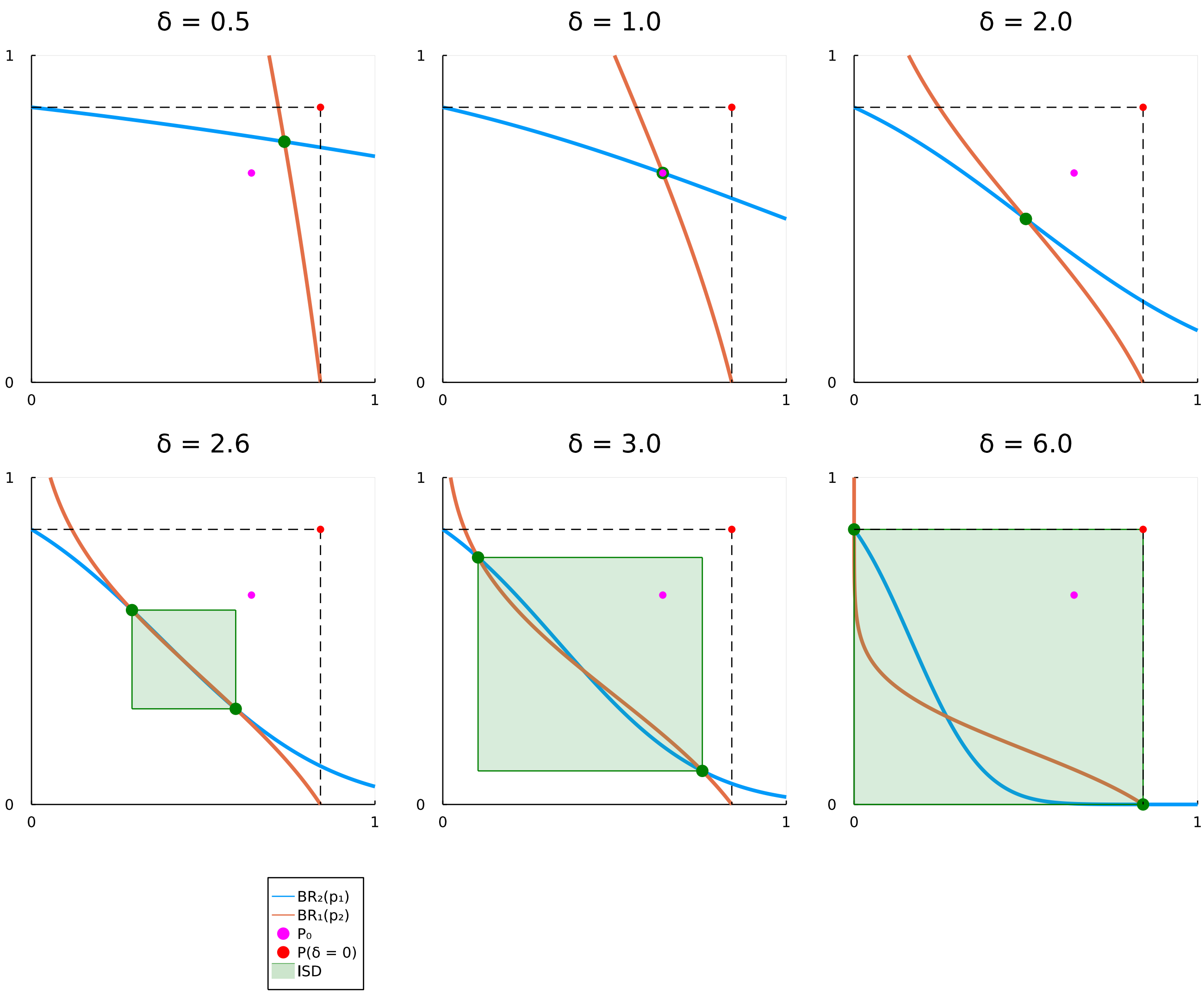
As γ grows, the game starts exhibiting multiple equilibria. For $\gamma = 2.6$ the ISD set is no longer singleton, i.e., $\sigma^L < \sigma^H$, however $\sigma^H < \sigma_0$ so $\gamma = 2.6$ is not in the identified set. When $\gamma = 3$ the ISD set has grown larger and now $\sigma^L < \sigma_0 < \sigma^H$, so $\gamma = 3$ belongs to the identified set. Finally, for $\gamma = 6$, the competitive effect is so large that for a large enough entry probability of firm 1, firm 2 enters with close zero probability (and vice versa). Clearly, $\gamma = 6$ is in the identified set.

D Entry Game Estimation

Assuming that firms have ISD Consistent beliefs, ISD Bounds imply the following conditional moment inequalities for all f and all x:

$$E\left[\sigma_f^0(\epsilon_f; x, \theta, \xi) - P_f^H(x,\theta)|x\right] \leq 0$$
$$E\left[P_f^L(x,\theta) - \sigma_f(\epsilon_f^0; x, \theta, \xi)|x\right] \leq 0$$

Figure 11: Two-way Best Response and ISD Set for Values of δ



Letting $h_l(x)$ be a collection of $l = 1, \ldots, L$ non-negative random variables we can write the following unconditional moment inequatlities:

$$E\left[(\sigma_f^0(\epsilon_f; x, \theta, \xi) - P_f^H(x, \theta))h(x)\right] \leq 0$$
$$E\left[(P_f^L(x, \theta) - \sigma_f(\epsilon_f^0; x, \theta, \xi))h(x)\right] \leq 0$$

Taking a simulated sample $(Y_m, X_m)_{m=1}^M$, the empirical analogue of the unconditional moment function is:

$$\psi_{fml}^H(\theta) = (Y_{fm} - P_f^H(X_m, \theta))h_l(X_m)$$
$$\psi_{fml}^L(\theta) = (P_f^L(X_m, \theta) - Y_{fm})h_l(X_m)$$

and we can compute the sample mean for each f, l:

$$\overline{\psi}_{fl}^H(\theta) = \frac{1}{M}\sum_{m=1}^{M} \psi_{fml}^H(\theta)$$
$$\overline{\psi}_{fl}^L(\theta) = \frac{1}{M}\sum_{m=1}^{M} \psi_{fml}^L(\theta)$$

Let $\psi_m(\theta) = ((\psi_{fml}^H)_{\forall f,l}, (\psi_{fml}^L)_{\forall f,l})$ be a column vector of size $\mathcal{L} = 2 \cdot |\mathcal{F}| \cdot L$ that stacks ψ_{fml}^H and ψ_{fml}^L for all f and l, for.a given market m. With this, we can write the estimator of the variance-covariance matrix as:

$$\hat{V}(\theta) = \frac{1}{M}\sum_{m=1}^{M}(\psi_m(\theta) - \overline{\psi}(\theta))(\psi_m(\theta) - \overline{\psi}(\theta))'$$

where $\overline{\psi}(\theta) = ((\overline{\psi}_{fl}^H)_{\forall f,l}, (\overline{\psi}_{fl}^L)_{\forall f,l})$ is the vector that stacks the cross-market means.

Andrews and Soares (2010) propose a T-statistic:

$$T(\theta) = S(\sqrt{M}\overline{\psi}_(\theta), \hat{V}(\theta))$$

where for a vector W of length N and an $N \times N$ matrix Z, S is defined as:

$$S(W, Z) = \sum \frac{[W_n]_+^2}{Z_{n,n}}$$

where $[]_+$ is a function that takes the maximum between zero and its argument, W_n is the n'th term of W and $Z_{n,n}$ is n'th term of the diagonal of W.

Andrews and Soares (2010) compare this T-statistic with a critical value that is built as follows. Let $\hat{D}(\theta) = Diag(\hat{V}(\theta))$, let $\hat{\Omega}(\theta) = \hat{D}^{-1/2}\hat{V}(\theta)\hat{D}(\theta)^{1/2}$, and note that $\hat{\Omega}(\theta)$ is a correlation matrix.

The critical value of AS is the $1-\alpha$'th percentile of the random variable:

$$S(\Omega_0^{1/2} W + \eta, \Omega_0), \text{ where } W \sim N(0_{\mathcal{L}}, I_{\mathcal{L}})$$

where $\eta = \left(\frac{M}{\log(M)}\right)^{1/2} \hat{D}^{-1/2}[\overline{\psi}(\theta)]_-$, where $[\overline{\psi}(\theta)]_-$ corresponds to the result of applying $[\cdot]_-$ to $\overline{\psi}(\theta)$ element by element. As AS and FY, I compute this critical value by drawing 500 realization of W and backing out the 95'th percentile of the simulated vector.

D.1 Monte Carlo Experiments

Here I define the h functions I use in the Monte Carlo experiments. To this end, let $q_{fm} = \frac{x_f}{\prod_{g \neq f} x_g^{1/(|\mathcal{F}|-1)}}$ and let $q = (q_{fm})_{\forall f,m}$. This variable is meant to capture the relative "strength" of firm f relative to its competitors. Moreover, let $l(i,f)$ be a one-to-one function that maps a firm and an index i, to an h-function index, l. The functions I use are:

$$\begin{aligned}
h_{l(i,f)}(X_m) &= \mathbb{1}\{q_{fm} \text{ above the 80'th percentile of } q\} && \text{if } i=1 \\
h_{l(i,f)}(X_m) &= \mathbb{1}\{q_{fm} \text{ between 60'th and 80'th percentile of } q\} && \text{if } i=2 \\
h_{l(i,f)}(X_m) &= \mathbb{1}\{q_{fm} \text{ between 40'th and 60'th percentile of } q\} && \text{if } i=3 \\
h_{l(i,f)}(X_m) &= \mathbb{1}\{q_{fm} \text{ between 20'th and 50'th percentile of } q\} && \text{if } i=4 \\
h_{l(i,f)}(X_m) &= \mathbb{1}\{q_{fm} \text{ below the 20'th percentile of } q\} && \text{if } i=5 \\
h_{l(i,f)}(X_m) &= q_{fm} && \text{if } i=6 \\
h_{l(i,f)}(X_m) &= \frac{1}{q_{fm}} && \text{if } i=7
\end{aligned}$$

So, we have a total of $\mathcal{L} = 2 \cdot 7 \cdot |\mathcal{F}|$ moment inequalities for estimation.

D.2 Empirical Exercise

To estimate the entry game parameters, θ, I use a collection of functions $h_k(X_m)$ to transform moments conditional on $X_m = (X_m, w_m)$ into unconditional moments. In particular, I use functions of variables that shift entry costs, and function of the firm type which, as mentioned in the main text, serve as sufficient statistic of the expected profits R_f. Note, that I do not use demand and marginal cost shifters as instruments in the entry game stage, as these variables have less information about expected profits than firm types.

I use the following functions:

$$
\begin{aligned}
h_{0f}(X_m) &= 1 \\
h_{1f}(X_m) &= \mathbb{1}\{w_{fm} \leq \text{20th percentile over } m \text{ of } w_{fm}\} \\
h_{2f}(X_m) &= \mathbb{1}\{w_{fm} \geq \text{80th percentile over } m \text{ of } w_{fm}\} \\
h_{3f}(X_m) &= \mathbb{1}\{w_{fm} \text{ is between the 20th and 80th percentile over } m \text{ of } w_{fm}\} \\
h_{4f}(X_m) &= hub_{fm} \\
h_{5f}(X_m) &= direct_{fm} \\
h_{6f}(X_m) &= \mathbb{1}\{distance_m \geq \text{median distance}\} \\
h_{7f}(X_m) &= \mathbb{1}\{distance_m < \text{median distance}\}
\end{aligned}
$$

This gives a total of $K = 7 \times 8 = 56$ moment inequalities.

E Bounds in Oligopolistic Pricing Games with Nested Logit Demand

In this section I show how one can build bounds, similar to ISD Bounds, for oligopolistic pricing games among multi-product firms facing nested logit demand.

Consider a multi-product firm oligopoly facing a two-level nested logit demand as described in figure 12. The nesting structure is such that there is a single high-level nest (the inside option) which represents the choice between buying and not buying, and multiple low-level nests/segments, indexed by $b = 1, \ldots, B$. Each nest represents different market segments. Let N represent the set of all products in the market and N_b the set of products in market segment b. As is standard, every product belongs to exactly one nest/segment, so that the collection of nests $(N_b)_b$ is a partition of N.

Every consumer k purchases at most one one product. Consumer k buying product j at a price p_j gets a utility of:

$$
\begin{aligned}
u_{ji} = \alpha_j - \lambda p_j + \epsilon_{ji}, \forall j \in N \\
u_{0i} = \epsilon_{0i}, \text{ for } j = 0
\end{aligned}
$$

where α_j is the product quality, λ is the price sensitivity parameter, and ϵ_j is an idiosyncratic taste shock whose distribution generates a nested logit demand. Furthermore, product $j = 0$ represents the outside option, whose utility has been normalized to zero. Given these assumption, demand for product j is:

$$
s_j(p_j, \mathbf{I}) = \frac{e^{(\alpha_j - \lambda p_j)/\tau_b}}{e_b^I} \frac{e^{\tau_b I_b/\gamma}}{e^W} \frac{e^{\gamma W}}{1 + e^{\gamma W}}
$$

where $\mathbf{I} = (I_b)_{b=1,\cdots,B}$, and

$$\begin{aligned} I_b &= \log\left(\sum_{j\in b} e^{(\alpha_j-\lambda p_j)/\tau_b}\right) \\ W &= \log\left(\sum_{j\in b} e^{\tau_b I_b/\gamma}\right) \end{aligned}$$

represent de inclusive value of nest b, and the inclusive value of the inside option, respectively. Additionally, $\tau_b \in (0,\gamma]$ is the nesting parameter of nest b and it controls the substitution between products in b and products in other nests. In particular, $\tau_b \to 0$ implies little substitution between b and other nests, where as $\tau_b = \gamma$ implies that consumers do not perceive products in b as closer substitutes to each other than to products not in b. Similarly, $\gamma \in (0,1]$ is the nesting parameter of the "inside nest," and controls substitution between buying and not buying.

Producers face a product specific marginal cost c_j, and own a set of products N_b^f in each nest b. Each good is produced by exactly on firm, so the collection $(N_b^f)_f$ is a partition of N_b. Firms simultaneously choose prices to maximize profits:

$$\pi^f(\mathbf{p}^f,\mathbf{I}) = \sum_{b=1}^{B}\sum_{j\in N_b^f}(p_j - c_j)s_j(p_j,\mathbf{I})$$

In Theorem 3.1 of Garrido (2021) I show that this is an aggregative game, meaning that firm optimal behavior can be represented as responding to an aggregator level $\mathbf{I}$. Let $\boldsymbol{\rho}^f(\mathbf{I}) = (\rho_j(\mathbf{I}))_{j\in N_b^f,\forall b}$ represent f's optimal prices given an aggregator $\mathbf{I}$, and let $\boldsymbol{\rho}(\mathbf{I}) = (\boldsymbol{\rho}^f(\mathbf{I}))_{\forall f}$. An equilibrium can be represented as an aggregator, $\mathbf{I}$, such that the aggregate behavior of firms given $\mathbf{I}$ generates that same aggregator. Formally, an equilibrium corresponds to a value of $\mathbf{I}$ such that:

$$\mathbf{I} = \boldsymbol{\Gamma}(\mathbf{I}) \equiv \left\{\log\left(\sum_{j\in N_b} e^{\alpha_j-\lambda\rho_j(\mathbf{I})}\right) : b\in B\right\}$$

In Theorems 3.2 and 3.3 of Garrido (2021) I show that there exist $\underline{\mathbf{I}}$ and $\overline{\mathbf{I}}$, with $\underline{\mathbf{I}} \leq \overline{\mathbf{I}}$, such that both $\underline{\mathbf{I}}$ and $\overline{\mathbf{I}}$ are equilibrium aggregators, and any other equilibrium aggregator, $\mathbf{I}$, satisfies $\underline{\mathbf{I}} \leq \mathbf{I} \leq \overline{\mathbf{I}}$. Furthermore, I show that $\boldsymbol{\rho}(\mathbf{I})$ is monotonic on $\mathbf{I}$.

Theorem 3.4 of the paper uses this fact to provide the following equilibrium finding algorithm. Let $\underline{\mathbf{H}}_0$ be the aggregator that results from all firms charging monopoly prices, and define the sequence $\underline{\mathbf{H}}_{n+1} = \boldsymbol{\Gamma}(\underline{\mathbf{H}}_n)$. Theorem 3.4 of the paper shows this sequence in increasing and that $\lim_{n\to\infty}\underline{\mathbf{H}}_n = \underline{\mathbf{I}}$. Similarly, let $\overline{\mathbf{H}}_0$ be the aggregator that results from all firms charging $p_j = c_j$ for all $j \in N$, and define the sequence $\overline{\mathbf{H}}_{n+1} = \boldsymbol{\Gamma}(\overline{\mathbf{H}}_n)$. Theorem 3.4 of the paper shows that this sequence is increasing and $\lim_{n\to\infty}\overline{\mathbf{H}}_n = \overline{\mathbf{I}}$.

An immediate corollary of this result is that the sequence $\overline{\mathbf{p}}_n = \boldsymbol{\rho}(\overline{\mathbf{H}}_n)$ is decreasing in n and converges to $\overline{\mathbf{p}} = \boldsymbol{\rho}(\overline{\mathbf{I}})$. Similarly, the sequence $\underline{\mathbf{p}}_n = \boldsymbol{\rho}(\underline{\mathbf{H}}_n)$ is increasing in n and converges to $\underline{\mathbf{p}} = \boldsymbol{\rho}(\underline{\mathbf{I}})$. Finally, any equilibrium vector of prices $\mathbf{p}^\star$ satisfies $\underline{\mathbf{p}} \leq \mathbf{p}^\star \leq \overline{\mathbf{p}}$. In what follows, I use this result to derive bounds on the distribution of prices that can be used for identification.

To understand how this works let $\alpha_j = x_j^{\alpha\prime}\beta^\alpha + \xi_j^\alpha$, where x_j is a vector of observable product characteristics, β^α is a vector of parameters, and ξ_j^α is a quality shock observed by all firms but not by the econometrician. Similarly, let $c_j = x_j^{c\prime}\beta^c + \xi_j^c$ where x_j^c is a vector of observable cost shifters, β^c is a vector of parameters, and ξ_j^c is an unobservable cost shifter. Furthermore, assume that both ξ_j^α and ξ_j^c follow distributions H^α and H^c both of which are known up to a vector of parameters θ^α and θ^c. Finally, let us collect the parameters of the game in $\theta = (\beta^\alpha, \beta^c, (\tau_b)_b, \gamma, \theta^\alpha, \theta^c)$.

Fix an (x, θ, ξ)-game, where $x = (x_j^\alpha, x_j^c)_{\forall j}$ and $\xi = (\xi_j^\alpha, \xi_j^c)_{\forall j}$. From above, for any equilibrium $\mathbf{p}^\star(x, \theta, \xi)$, the following hold:

$$\underline{\mathbf{p}}(x, \theta, \xi) \leq \mathbf{p}^\star(x, \theta, \xi) \leq \overline{\mathbf{p}}(x, \theta, \xi)$$

As before, the inequality above holds for all ξ, so integrating over ξ we get:

$$\underline{\mathbf{p}}(x, \theta) \leq \mathbf{p}^\star(x, \theta) \leq \overline{\mathbf{p}}(x, \theta)$$

This inequality holds for all θ. In particular, it holds for $\theta = \theta^0$. Letting $\mathbf{p}_0(x) = \mathbf{p}^\star(x, \theta^0)$ it is easy to see that any θ that violates:

$$\underline{\mathbf{p}}(x, \theta) \leq \mathbf{p}_0(x) \leq \overline{\mathbf{p}}(x, \theta) \tag{15}$$

cannot be equal to θ^0. This reasoning yields the following identified set.

Definition 10 (Nested Logit Identified Set)**.** *Consider an oligopolistic pricing model among multi-product firms facing nested logit demand as described above. The identified set for θ corresponds to all values of θ satisfy equation* (15)*. This is:*

$$\Theta_I = \left\{\theta \in \Theta : \underline{\mathbf{p}}(x, \theta) \leq \mathbf{p}_0(x) \leq \overline{\mathbf{p}}(x, \theta), \forall x \in \mathcal{X}\right\}$$

Figure 12: Nesting Structure