

Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets

Paul L. E. Grieco*

November 9, 2009

Abstract

Empirical investigation of discrete-choice games requires assumptions about payoff functions and player information sets. In practice, applied researchers have focused on the estimation of payoff functions using strict informational assumptions. In this paper, I propose a flexible information structure that nests the two most common informational assumptions: pure complete and incomplete information. As in other models of discrete-choice games, the parameters of player payoff functions are point identified if the model covariates have sufficiently rich support. In addition, the model provides testable restrictions on the information structure of the data generating process. I apply the model to study the impact of supercenters on entry and exit patterns of grocery stores and I show that the model can produce useful bounds on counterfactual outcomes. I find that models that account for only incomplete information are excluded from the confidence set of the general model. Moreover, a flexible information structure enhances the credibility of empirical studies of games by allowing for levels of uncertainty between players that are consistent with the data but are assumed away under stronger assumptions.

*Northwestern University, p-grieco@northwestern.edu. I gratefully acknowledge the help and support of my dissertation committee: Aviv Nevo, Rob Porter, Mark Satterthwaite and Elie Tamer. The data for this project was generously provided by Paul Ellickson. I also benefited from many conversations and discussions of earlier drafts with Guy Arie, Ivan Canay, Ulrich Doraszelski, Jon Gemus, Gautam Gowrisankaran, Kei Kawai, Michael Mazzeo, Mauricio Valera, and Chris Vickers. I would also like to thank the Robert Eisner Memorial Graduate Fellowship for financial support, and the Zienna Corporation for providing me with the KNITRO optimization package for use in this project. All remaining errors are my own.

1 Introduction

Game-theoretic models are frequently employed to study strategic interaction between agents. Empirical research has focused on estimating payoff functions while maintaining strong assumptions regarding the information structure of the game. Relaxing assumptions on the information structure can enhance the credibility of game-theoretic empirical analysis in discrete games. In this paper, I investigate the use of a flexible information structure that nests the most common assumptions in the literature. Despite this added flexibility, the parameters of the players' payoff functions remain point identified under standard rich support assumptions. Given the flexibility of the model, multiple equilibria are likely to arise in applications. I incorporate an approach to equilibrium selection that avoids parametric assumptions and lets selection depend on public information that is not observed by the econometrician. I apply the method to data on the entry and exit patterns of grocery stores. The model provides useful bounds to equilibrium outcomes. In addition, the empirical analysis indicates that the informational assumptions of traditional models may drive their conclusions.

Assumptions about the information structure determine how the structural errors of a game are accounted for in players' information sets. In an oligopoly setting, firms typically have some private information about their own costs. In addition, some determinants of firm demand are commonly observed by the game's players but not by the econometrician. However, most applied entry papers fall into one of two frameworks. Under the complete information framework, each agent's payoff function is perfectly known to his opponents but not the econometrician. In the incomplete information framework, each agent receives a private shock that is unknown to both opponents and the econometrician. Abstracting away from either public or private unobservables will result in a misspecified model. Moreover, assumptions about the information structure have implications for counterfactual analysis. Consider two firms contemplating entry into a market where only one can operate profitably. If the firms are uncertain about each other's actions, the market may be unserved for several periods even though it is profitable for a single entrant. In contrast, a pure strategy complete information model will predict that one firm immediately enters the market whenever a monopolist can make positive profits. A structural model that makes

incorrect assumptions about the information structure is unable to produce robust predictions about the environment under study.

My method uses the available data to make inferences about agents' information sets. Assume the econometrician observes some variable, such as an independent cost shifter, that directly affects the profits of only one firm in the market. Differences in rival firms' response to variation in this variable can be used to separate complete and incomplete information. In a complete information game with pure strategies, firms know their opponents' actions when making their own entry decisions. Therefore, variation in the cost shifter should not directly affect rival firms' actions because their information set includes the firms true action. By contrast, when firms have private information, they must base their decisions on a noisy signal of opponents' actions rather than actual actions. In this case, variation in the cost shifter will impact rivals' actions because it shifts their expectations about the entry of the initial firm. The relative degree of correlation between the rival's action, the firm's action, and the firm's cost shifter contains information the relationship between public and private information.¹ These relationships can be used to test whether commonly assumed information structures are consistent with the data.

Point identification of firm payoff functions as well as tests against pure complete and pure incomplete information assumptions rely on the econometrician observing markets where entry by some firm is nearly certain. In practice, researchers may be unable to observe such markets and must instead work with the data at hand. I characterize the identified set without assuming a rich support for covariates and show the data still contains valuable information in this case. While the model may be only partially identified, the data can still provide informative bounds on the data generating process. This analysis leads to the inference technique employed in the empirical portion of this paper.

I apply this method to data on the entry and exit decisions of grocery stores to derive informative bounds on several counterfactuals of interest. I compute a 95 percent confidence set for the parameters of the full model. For purposes of comparison, I also estimate the model under

¹If error terms of firm profit functions are uncorrelated across firms, this test is a simple regression. The identification section discusses how to point identify correlation between firms' unobserved public errors under the complete information assumptions.

complete and incomplete information assumptions assuming point identification. The full model nests the two restricted models. In this empirical setting, the incomplete information model is excluded from the confidence interval implied by the full model, while a pure complete information model is inside the 95 percent confidence region generated by the full model. This finding is consistent with the explanation that unobserved heterogeneity that is observed by both players is an important determinant of outcomes. However, a parameterization where more than half of the unobserved variation is generated by a private error component also fits the observed data, so it would be incorrect to argue that the data indicate that uncertainty between players is unimportant. I further show that using the complete information model alone in counterfactual analysis can lead to misleading results. In summary, the results of both the complete and incomplete information analysis are not robust to the allowing for more flexible information structures.

By allowing for both a public and private structural error, I unify two strands of the literature on discrete-choice games. An early model of strategic interaction was proposed by Bjorn and Vuong (1984) in the context of spousal labor-force decisions. Bresnahan and Reiss (1990, 1991a,b) considered a complete information framework. They analyzed identification and estimation, including issues associated with the existence of multiple equilibria. Seim (2006) introduced the incomplete information framework to empirical work, showed that the existence of equilibria could be guaranteed, and employed numerical simulation to show that multiple equilibria do not arise in a number of interesting applications. Because the incomplete information model is easily adapted to dynamics through the use of Makrov perfect equilibrium, it is used in estimating infinite-horizon dynamic games based on period profits (Aguirregabiria and Mira, 2007; Bajari, Benkard, and Levin, 2007). While the techniques developed here are applied to more computationally tractable static games, an extension to infinite-horizon dynamic games is conceptually straightforward.

My empirical application contributes to the literature on the effect of supercenters on traditional grocery stores. Several studies have investigated how traditional grocery stores are affected by supercenter entry. Rather than study the effect of competition from supercenters on firm prices or quality, I examine the decision to enter or exit a local-grocery-store market, and how that decision is affected by the presence of a supercenter in the market's vicinity. Because most customers

face a relatively small choice set when choosing where to do their grocery shopping, the effect of firm openings and closings may have a much larger impact on consumer welfare than competitive responses in product price or quality. Supercenters are commonly believed to hold a significant cost advantage over traditional grocery stores due to their scale and integrated distribution networks. However, the locational convenience of a local grocery store, and perhaps higher-quality service and more customized product lines, may provide some insulation from competition with supercenters. How much insulation is provided is the key empirical question of this paper.

When analyzing the entry and exit decisions of grocery stores, it is important to account for both public and private unobservable information. Rural grocery firms and their markets differ in ways that are observable to all players but unobservable to the econometrician. There are differences across markets in local terrain, zoning regulations, local tastes, location availability, as well as differences among firms (e.g., the quality of their produce and level of customer service). However, firms and potential entrants also keep some cost information private from their competitors—such as that regarding management expertise, outside opportunities, and ease of integrating into a distribution network. By allowing both a public and private error term for each player, the model avoids a potential source of misspecification.

Given the confidence set for the model parameters, I simulate a confidence interval for firm values and compare the value of a rural supermarkets in different market configurations. I find that entry by supercenter outside, but within 20 miles, of a local monopolist's market has a smaller impact than would entry by a local rival. While supercenters appear to be associated with a decrease in the store's expected profits, and appear to lower the number of grocery stores in surrounding markets, the effects are small. Indeed, I cannot reject the possibility that supercenters *increase* long term profits of local grocery stores by discouraging local entry. I interpret this as evidence that location and format-based differentiation partially insulates rural stores from competition with supercenters. While a full demand model is outside the scope of this paper, my results indicate how consumer choice sets are likely to evolve as a result of supercenter entry. Because I find a relatively small reaction in the number of expected local stores, the crowding-out effect of supercenter entry is unlikely to offset the static benefits of supercenters highlighted by Hausman and Leibtag (2005).

Of course, further research on the welfare effect of endogenous store entry and exit is warranted.

The following section provides a brief review of the literature on discrete games. Section 3 provides background information on the supercenter format and the retail grocery industry. Section 4 introduces the model used in this paper, while Section 5 provides a discussion and uses numerical examples to illustrate how the model incorporates public and private information. I study the identification of the model in Section 6, and show how to conduct inference on the model in Section 7. I then turn to the application. Section 8 introduces the data and performs some descriptive analysis. The results of the application of the full structural model and several counterfactual experiments are presented in Section 9. The final section concludes by reiterating that allowing for flexible information structures improves the credibility of empirical investigations of discrete games. All proofs are presented in the Appendix.

2 Literature Review: Discrete Games

The earliest example of the estimation of a game is Bjorn and Vuong (1984) who extended the selection model of Heckman (1978) to a game setting where a married couple make joint labor force participation decisions. The topic was introduced to the industrial organization literature by the papers of Bresnahan and Reiss (1990, 1991a,b), which employed the complete information framework. The incomplete information framework was introduced by Seim (2001). A third approach, recently suggested by Pakes, Porter, Ho, and Ishii (2006), avoids modeling the information sets entirely and instead attempts to estimate the parameters of the agents' objective functions using the assumption that agents are payoff maximizing, a necessary but not sufficient condition for equilibrium.

2.1 Complete Information Models

Most early papers on the estimation of games followed the complete information framework. Because of the difficulties implied by multiple equilibria, authors adopted the strategy of finding a statistic that was uniquely predicted by the model and using this statistic for estimation. Bresnahan and Reiss (1991b) use the assumption that unobserved determinants of profits vary at the market

level, rather than at the firm level to estimate entry thresholds based on observing the equilibrium number of firms in the market. This model conveniently reduces to an ordered probit. Berry (1992) makes use of a simulation based estimator that relaxes the firm homogeneity assumptions of earlier applications, the specification of the model implies that the number of firms in equilibrium will be unique, and he uses this outcome to estimate the model.

Several authors have provided extensions to the basic complete information model where a selection mechanism is imposed by the researcher in order to complete the model. Mazzeo (2002) adopts a model of endogenous type choice which further extends the scope of allowable heterogeneity by letting the competitive effect of entry vary by firm type. Jia (2008) uses the theory of super-modular games to speed the computation of equilibrium for a large action space, and selects the maximal equilibrium for a game as the model's unique prediction. Like Mazzeo, she argues for the robustness of her results by using different selection mechanisms and comparing estimation results.

Tamer (2003) explored what could be learned from a model admitting multiple equilibria by examining the full outcome distribution. He shows that even in the presence of multiplicity one can derive bounds on equilibrium outcome probabilities that can be used in estimation. Ciliberto and Tamer (2007) apply the bounds estimation to the airline industry and compare the technique to that used by Berry (1992).

Recent work has extended the literature to allow for mixed strategies. Bajari, Hong, and Ryan (2007) assume a parametric form for the equilibrium selection probabilities to estimate a model allowing for mixed strategy profiles. Beresteanu, Molchanov, and Molinari (2008) have proposed using the theory of random sets to generate approximate bounds on outcome probabilities allowing for arbitrary equilibrium selection when mixed strategies are allowed.

2.2 Incomplete Information Models.

The first papers to make use of the incomplete information framework in the estimation of games highlighted computational reasons for doing so rather than modeling realism. Seim (2001, 2006) introduced a location choice model where agents observe the same market characteristics and receive their own privately observed cost shock for each location. Seim assumes uniqueness of equilibrium

in her model and tests this assumption through simulations where she calculates equilibria from several initial guesses to see the algorithm always converges to the same point.² Seim’s model includes a special market-level unobservable that drops out of the fixed point calculation, and allows for a random number of potential entrants. Other models of incomplete information have avoided the use of market level heterogeneity (Augereau, Greenstein, and Rysman, 2006; Vitorino, 2008).

Other studies have taken the incomplete information assumption as a jumping off point to other identification results. Aradillas-Lopez (2007) estimates a fully semi-parametric model, eliminating the distributional assumption on the error term. Sweeting (2007) shows that with enough observations of each player to observe strategies, multiple equilibria in an incomplete information game can be used to aid identification in a coordination game with strategic complementarities.

2.3 Agent-Level Optimality

In contrast to the frameworks presented above, Pakes, Porter, Ho, and Ishii (2006, PPHI) take a decidedly different approach. PPHI avoid modeling the player’s information set by allowing for a structural error that is the difference between the model’s predicted payoff and the payoff expectation of the player given the agent takes a particular action. This error absorbs all direct effects on the player’s decision that is unknown to the econometrician. Its distribution and how it enters into opponents information sets is left unspecified. Rather than analyze the game as a whole, their analysis focuses on the decision problem of each player individually. They assume optimal decision making on the part of each player—a necessary condition for equilibrium—and derive a set of moment conditions to use for estimation. In order to control for endogeneity, PPHI assume that the structural error can either be differenced out across choices or that a valid instrument is available. Inference is then performed using moment inequalities.

This approach may be attractive when the environment under consideration is complex and solving for equilibrium would be prohibitively burdensome. Ishii (2005) used this framework to model the capacity decisions of banks determining how to construct their ATM networks. Ho

²While computationally compelling, this exercise does not constitute a proof. Furthermore, the likelihood function used by Seim in estimation is only valid for parameterizations that admit a unique equilibrium.

(2008) applies this setup to a network bargaining game between hospitals and insurance providers.

To contrast, the model proposed in this paper assumes parametric distributions for structural errors and explicitly model's agents' beliefs based on a common prior assumption. This approach entails substantial computational cost, but brings the benefit of clearly interpretable estimates and exploits the information contained in the joint distribution of agent decisions.³

3 Background: Supercenters and the Retail Grocery Industry

The supercenter format has dramatically altered the retail grocery industry and poses a substantial challenge to existing grocery retailers. A supercenter combines a discount store, a grocery store, and possibly several other retail services (pharmacy, tires, gas, etc.) into a single store of roughly 175,000 square feet.⁴ The format was initiated in France by Carrefour and first used in the United States by Meijer, a Michigan-based firm operating in the Midwest. By far, the largest supercenter chain is Wal-Mart, which accounted for roughly two-thirds of all US supercenters in 2000. Kmart and Target also operate supercenter chains throughout the United States. Key characteristics of a supercenter are a vertically integrated distribution network, aggressive pricing strategies, and an emphasis on low costs. Because of their size, supercenters tend to be located away from population centers and draw their customers from a wider area than traditional grocery stores. Nationwide, supercenters have grown from a 2 percent share of all grocery sales in 1994 to a 13.5 percent share in 2005; their share is expected to continue to expand to over a fifth of all grocery sales by 2010 (Martinez, 2007). In 2003, Wal-Mart surpassed Kroger as the nations largest grocery retailer.

Several studies have examined the competitive effect of supercenters on traditional grocery stores. Hausman and Leibtag (2005) use a national panel of households to study consumer welfare effects of supercenters. They find that supercenters both offer consumers lower prices on products and induce other grocery retailers to lower their prices, providing both a direct and indirect positive effect on consumer welfare. Basker and Noel (2008) analyze store-level price data from 175 US cities and find that Wal-Mart's prices on average are 10 percent lower than its competitors and that Wal-

³For example, the Nash equilibrium assumption implies that agent's are best responding to each other's strategies, so we can make use of the fact that all agents in the same market are playing the same equilibrium profile.

⁴Singh, Hansen, and Blattberg (2006) provide a full discussion of the supercenter format.

Mart entry causes prices competitors to lower their prices by 1-1.2 percent. Singh, Hansen, and Blattberg (2006) investigate supercenter's effect on grocery store revenue. Using a frequent-shopper database from a single grocery store before and after entry by a Wal-Mart supercenter 2 miles away, they find that supercenter entry caused a 17 percent decline in sales revenue. Their analysis further indicates that there is substantial heterogeneity in consumers' reactions to supercenter entry. Matsa (2009) shows that traditional grocery stores improve their quality of service by reducing the probability that items are out of stock in response to supercenter entry. All these studies focus on competitive effects of supercenters through adjustments in price or quality of service, rather than entry or exit.

In the popular press, the possibility of supercenters crowding out small grocery stores and reducing consumer utility has received much greater attention than the effects of supercenters on pricing, which are unambiguously welfare improving.⁵ It is also tempting to draw parallels between the decline of single-store discounters and the future of traditional supermarkets.⁶ A significant fall-off in the number of rural grocery stores could generate a significant decrease in consumer welfare if some consumers have high travel costs.⁷ A mass closure of traditional grocery stores could also have significant health implications; public health physicians have reported a link between residents' distance traveled to a grocery store and obesity, which is itself connected to a large number of ailments (Inagami, Cohen, Finch, and Asch, 2006). These negative effects would be most acutely felt in rural areas, where the density of grocery retailers is already low. At a Rural Grocery Store Summit sponsored by Kansas State University and the United States Department of Agriculture, competition with "big box" retailers was listed as "Challenge #1" in the summit minutes.⁸ Using the methods developed in this paper, I am able to examine how

⁵For example, in 2004 the National Trust for Historic Preservation placed the entire state of Vermont in its register of endangered sites because of Wal-Mart's plans to begin supercenter expansion into the state, Vermont was already home to standard Wal-Mart stores ("Preservationists call Vermont Endangered, by Wal-Mart," *New York Times*, May 25, 2004). For other examples see <http://walmartwatch.com>.

⁶Jia (2008) has found that the growth of chain discount stores (Wal-Mart and Kmart) explains 40 to 50 percent of the decline in the number of small discount stores. The data for her study ends prior to the opening of Wal-Mart's first supercenter.

⁷Basker and Noel (2008) report that in 2002, 77 percent of Americans in their dataset live within 5 miles of a grocery store, only 14 percent are within 5 miles of a supercenter. Although supercenters are expanding, they remain much more dispersed than traditional grocery stores, whose market radius is commonly thought to be 3-5 miles.

⁸Rural Grocery Store Initiative, <http://www.ruralgrocery.org/events/>, accessed August 12, 2009.

serious a threat supercenters are both to traditional grocery stores (through their effect on expected profits) and on consumers (through their effect on whether traditional grocery stores will remain open within a market). While previous studies have looked at the effect of supercenters on prices and revenue in traditional grocery stores, I will consider the effect of supercenters on entry and exit of grocery stores in small markets. In contrast to more urban areas, access to nearby source of food is an active concern in rural markets. There is also a concern that the presence of vertically integrated chains syphon profits out of small communities that would otherwise be captured by local entrepreneurs. To answer both of these questions, we must study whether or not local grocery stores will be crowded out by supercenters rather than how supercenters will affect pricing or product line strategies. Grocery stores make entry and exit decisions in view of long run profit opportunities. I employ a model that estimates the present discounted value of a grocery store given current market characteristics based on observed entry and exit decisions.

4 A Model of Entry and Exit with Public and Private Errors

I model firms' joint entry and exit decisions over a five year period. A firm will operate only if it believes long run profits will be positive. The firms payoff functions include two structural errors, one that is publicly observed by all players, and one which is known only to one player. This private error accounts for uncertainty about rival players, while the public error accounts for firm level unobserved heterogeneity.

Since the seminal paper of Bresnahan and Reiss (1990), static models have routinely been used to analyze decisions that take place within a dynamic environment. I interpret the payoff function as the sum of profits accrued during the period and the present discounted value of the firm at the end of the period.

4.1 Model Setup

I assume firms decide whether or not to operate in a market in the next 5 years. This time period is reasonable because it may take several years to open a grocery store, and stores will operate with several year leases. There is a high-degree of year to year persistence in the data, which is also an

indicator that a long period length is appropriate. At the start of the period there are two potential firms competing in the grocery retail sector in each market.⁹ The researcher and all players jointly observe a state variable $x = (x_0, x_1, x_2)$, where x_0 are variables that may affect all players and x_i are variables that only affects the payoffs of player i . The common variables include, for example, the town characteristics such as population. An example of a firm-specific variable is whether or not the firm is operating at the start of the period or not, which determines whether or not it must pay an entry cost to operate.

In addition to the observable vector x , firms view shocks which are not observed by the econometrician, $\epsilon = (\epsilon_1, \epsilon_2)$. While ϵ_i contributes only to firm i 's payoff function, the ϵ_1 and ϵ_2 can be correlated with each other. Finally, each firm also observes its own private shock, ν_i , drawn from a distribution that is commonly known to all firms and is assumed to be independent of all other variables in the model. Independence of ν_i is a strong assumption, but it is necessary to ensure that player i 's beliefs about equilibrium play are not dependent on ν_i and to guarantee the existence of an equilibrium in cutoff strategies. The key difference between the public and private shocks is that a firm's strategy will be conditional on its opponents public shock, but not its opponent's private shock.

After observing the shocks, firms decide whether or not to operate. I assume they base their decision on equilibrium strategies. In the event that there are multiple equilibria, equilibrium is selected via a public randomizing device, which is also unobserved by the econometrician. Firms receive their payoffs based on the outcome profile. The econometrician directly observes firm actions (y_1, y_2) . Firms choose to either operate, $y_i = 1$ or not operate, $y_i = 0$ based on their expected profits. If the firms choose to operate, they will play again in the next period. If they decide not to operate, they will be replaced in the following period by a new potential entrant. Figure 1 provides an overview of the model timing.

⁹While adding additional players is conceptually straightforward, it is computationally intensive, and I concentrate on markets that are small enough such that more than two active market participants are unlikely.

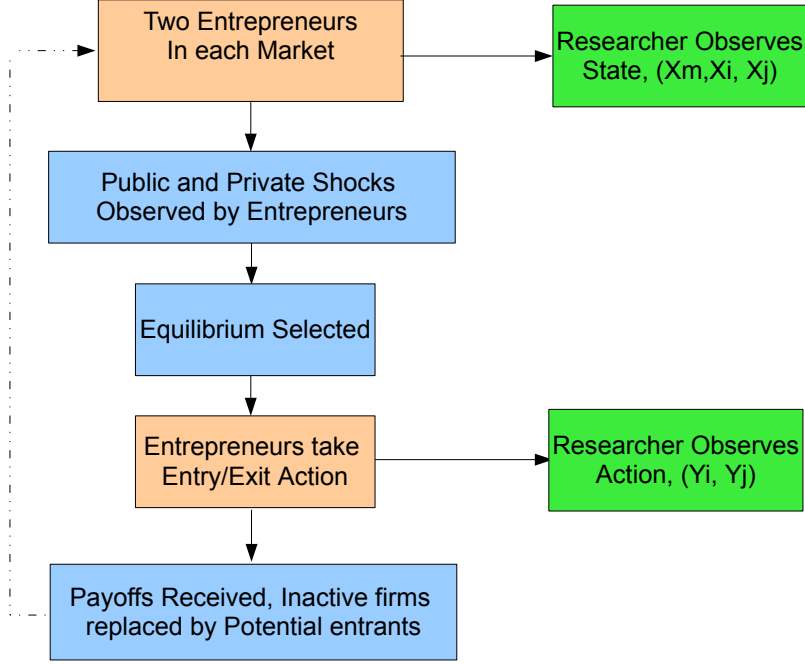


Figure 1: Timing of the game. Orange boxes represent actions taken by entrepreneurs, blue boxes are all other events. Green boxes represent points of data collection.

4.2 Firm Payoffs and Error Distributions

Payoffs for the period game are a function of observable characteristics, the actions of the players, and the unobservable shocks. The payoff received at the end of the period can be thought of as the ex-post value of the firm.

$$V_i(a_i, a_{-i}; x) = \begin{cases} \mu(x_0, x_i; \theta_\mu) + \epsilon_i + \nu_i & \text{if } y_i = 1, y_{-i} = 0 \\ \mu(x_0, x_i; \theta_\mu) - \delta(x_0; \theta_\delta) + \epsilon_i + \nu_i & \text{if } y_i = 1, y_{-i} = 1 \\ 0 & \text{if } a_i = 0 \end{cases} \quad (1)$$

Where μ and δ are known functions of the observable characteristics up to the finite dimensional parameters θ_μ and θ_δ , μ represents the baseline profits within this market for a monopolist, while

δ parameterizes the “competition effect” or the reduction in profits due to an opponent operating in the market. I normalize the payoff of a firm that opts not to operate to zero.

The public shocks, ϵ_1 and ϵ_2 , are drawn from a bivariate normal distribution with a common variance component σ_ϵ^2 and a correlation coefficient ρ . The assumption that σ_ϵ^2 is the same for both players may be relaxed, but is reasonable for the application below. Each firm’s private shock is drawn from a normal distribution with variance σ_ν^2 . Because of the private shock, firm i is uncertain as to firm j ’s exact type, and whether or not it will operate. Firm i must use observable characteristics plus the public shock to form beliefs about firm j ’s entry probability.

4.3 Equilibrium

From the perspective of the players, who observe both elements of ϵ , the model is a game of incomplete information. I confine the analysis to pure strategy equilibria—a function from the firm “type” ν_i to an action $y_i \in \{0, 1\}$.¹⁰

I now derive necessary and sufficient conditions for an equilibrium strategy. For expositional clarity I suppress the dependence on covariates x in this section. A firm’s optimal strategy is a cutoff in ν_i . Increasing ν_i unambiguously increases the expected profits of entry (action 1), so a strategy in which the firm operates at $\nu_i' < \nu_i''$ but not at ν_i'' is clearly sub-optimal. Therefore, an optimal strategy must be of the form,

$$s_i(\nu_i; \theta, \epsilon) = \begin{cases} 1 & \text{if } \nu_i \geq \chi_i(\epsilon; \theta) \\ 0 & \text{otherwise} \end{cases}$$

Where $\chi_i(\epsilon; \theta)$ is the entry cutoff for agent i . This is convenient because we can associate optimal strategies with their cutoffs, which are real numbers rather than functions. Furthermore, there is a simple expression for player j ’s beliefs about player i ’s probability of entry under the common

¹⁰Milgrom and Weber (1985) have shown that in incomplete information games such as the one considered here where player types are conditionally independent, equilibrium must exist and every mixed strategy equilibrium has a nearby “purification” pure strategy under which the agents distribution of observed behavior and expected payoffs are identical.

prior assumption,

$$\rho_i(\chi_i, \epsilon; \theta) = \int s_i(\nu_i, \epsilon; \theta) d\Phi(\nu_i) = 1 - \Phi\left(\frac{\chi_i(\theta, \epsilon)}{\sigma_\nu}\right) \quad (2)$$

Players select their cutoffs to optimize their expected payoff given their beliefs about the actions of other players. Let $\chi_i^b(\chi_j, \epsilon; \theta)$ denote player i 's best response to player j playing the cutoff χ_j . Player i 's best response is to adopt the cutoff where he is exactly indifferent between his two actions. I.e, firm i 's best response is to operate when when $\nu_i \geq \chi_i^b(\chi_j, \epsilon; \theta)$. The following equation defines agent i 's best response cutoff, $\chi_i^b(\cdot)$ as a function of opponent strategies and the publicly observed ϵ .

$$\chi_i^b(\chi_j, \epsilon; \theta) = -(\mu_i(\theta) + \epsilon_i) + \rho_j(\chi_j, \epsilon; \theta)\delta_i(\theta) \quad (3)$$

Equation (3) gives the condition that player i is optimally responding to his opponent strategies given his beliefs, while (2) ensures that beliefs are rational given the agents' common prior. Therefore any joint set of cutoffs $\chi = (\chi_1, \chi_2)$ that satisfies these equations for all players represents an equilibrium. Because there is a simple one-to-one mapping between χ_i and ρ_i , we can alternatively describe the equilibrium in terms of either cutoffs or entry probabilities.

4.4 Multiple Equilibria and Equilibrium Selection

Although it is well known that games of incomplete information can admit multiple equilibria, the empirical issues associated with the possibility of multiple equilibria have received more attention in the context of complete information games. Using the necessary and sufficient conditions derived in the previous section, I can numerically solve for the equilibrium set for the incomplete information game given θ and the realization of public shocks ϵ as the solution set to a system of nonlinear equations. Let this set be denoted,

$$\mathcal{E}(\epsilon, x, \theta) = \{\chi : \forall i, \chi_i = \chi_i^b(\chi_{-i}, \epsilon, x; \theta)\}$$

Each equilibrium implies a multinomial distribution across agent actions. If the equilibrium set was a singleton everywhere, a unique observable outcome distribution for the model could be obtained by integrating over the observable shock ϵ . However, when there are multiple equilibria,

the model does not provide a unique distribution over actions, so the model is incomplete. Instead, we can derive a set of restrictions on the outcome distribution which can be used to test the model parameters. Clearly, in order to build restrictions based on the actions resulting from some element in the equilibrium set, it will be necessary to compute all elements of the set. In general games, it may be difficult to verify that all the elements of $\mathcal{E}(\epsilon, x, \theta)$ have been computed.¹¹ In the simple 2-player model presented here, however, finding all equilibria can be done by checking for intersections of the firm's best response functions. The precise method I employ to find the equilibrium set is detailed in Appendix C.

If there is more than one equilibrium, I assume that an equilibrium is chosen via some public coordination device. The resulting outcome distribution is a mixture across the outcome distributions implied by each element of equilibrium set.

Assumption 1. *An equilibrium is selected on the basis of a public coordination device that may depend on (ϵ, x, θ) but is independent of ν .*

The public coordinating device must be independent of ν (like all other variables in the model) to prevent information leakage, which would cause agents' beliefs about other players to depend on private information.

The public coordinating device picks out the equilibrium to be played from $\mathcal{E}(\epsilon, x, \theta)$. Once agents observe the device it is clearly not optimal for the a player to unilaterally deviate to play any strategy other than the one the device has selected. However, economic theory gives us no model for the equilibrium selection device, I assume only that it selects strategy profiles from the equilibrium set.

A selection mechanism is a function from that maps the space of strategy profiles to the probability that that particular strategy profile is played, and a selection mechanism to be valid when only equilibrium strategy profiles are played with positive probability.¹²

Definition 1. *Let λ be a selection mechanism, where $\lambda^e(\epsilon, x, \theta)$ is the probability that that strategy*

¹¹Algorithms which attempt to compute all equilibria of a game in various settings have been proposed by Wilson (1971), Garcia and Zangwill (1979), Kalbala and Tesfatsion (1991), Bajari, Hong, Krainer, and Nekipelov (2008), Grieco (2008), and many others.

¹²Conditional the publicly observed information, strategies can be described as a vector of cutoffs, so in a two player game the selection mechanism is a function from \mathbb{R}^2 to $[0, 1]$.

profile e is played when the game is defined by (ϵ, x, θ) . The following conditions must hold for a selection mechanism to be valid,

1. If $e \notin \mathcal{E}(\epsilon, x, \theta)$, $\lambda^e(\epsilon, x, \theta) = 0$, and
2. $\sum_{e \in \mathcal{E}(\epsilon, x, \theta)} \lambda^e(\epsilon, x, \theta) = 1$.
3. $\lambda^e(\epsilon, x, \theta) \geq 0$

The equilibrium selection mechanism is an infinite dimensional parameter that completes the model. In other words, given a valid selection mechanism and a full set of model parameters, the model predicts a unique probability distribution over actions, and it is possible to write the model likelihood. If only $\theta = (\theta_\mu, \theta_\delta, \rho, \sigma_\epsilon^2)$ is specified, then a set of probability distributions are available and the model is only partially determined. Nonetheless, I can still use the data to test whether a given θ could have produced the observed data. This is because we can still enforce the restriction that the selection mechanism is valid. I use $\Lambda(\theta)$ to denote the set of valid selection mechanisms for a fixed θ . The inference procedure proposed in Section 7 will maximize over this set to determine whether the model parameters θ are consistent with the data.

5 Impact of Information Structure

The model I proposed in Section 4 allows for flexible assumptions about the information structure. Before turning to identification, it is important to understand how differences in the information structure will impact observable outcomes. This section presents numerical examples to highlight how the nature of the information structure affects the competition between firms and how the model can be used to explain the observed data. In Section 5.1, I plot equilibrium entry probabilities according to changes in variance of the private error term while holding other parameters fixed in order to indicate how the level of uncertainty affects equilibrium strategies. Section 5.2, illustrates how the information structure affects the measure of markets that have multiple equilibria.

The two structural shocks, ϵ and ν , affect strategies and observed action distributions in different ways. While both firm's strategies condition on ϵ_i , only firm i 's action is determined by

ν_i . Abstracting away from either shock simplifies how the model interprets observed actions. Under the assumptions of the incomplete information framework, econometricians often assume they can directly measure equilibrium strategies in the data (e.g., Bajari, Hong, Krainer, and Nekipelov, 2008).¹³ In contrast, when a complete information shock is present, the observed distribution is understood as outcomes from a mixture of games determined according to the distribution of ϵ . However, the equilibrium concept implies that given ϵ the distribution of outcomes is degenerate. The general model allows for the observed data distribution to be a mixture of games that imply non-degenerate equilibrium outcome distributions. While changes in the distribution of ν affect the shape of the equilibrium correspondence, changes in the distribution of ϵ affect the distribution of games over the correspondence.

5.1 Firm reactions to the level of uncertainty

The firms within the model are able to observe ϵ before taking their decision, so the equilibrium entry probability correspondence $\mathcal{E}(x, \epsilon; \theta)$ is a function of the public error. Taking the prior distribution of ν and opponent strategies as given, firms choose entry strategies that are cutoffs in their own private shock. When the variance of the private error component is zero, firms are playing a game of complete information.

To illustrate the effect of player uncertainty on equilibrium entry probabilities, I compute numerical examples within a very simple framework. I consider examples with two observably identical players and no covariates. I consider the case where for both firms $\mu_i = \delta_i = 1$. So in a complete information game, entering as a monopolist is profitable if $\epsilon_i > -1$, and as a duopolist is profitable if $\epsilon > 0$.

First, we examine particular points on the equilibrium correspondence and show how players' action probabilities change with σ_ν . Figure 2 displays how equilibrium entry probabilities vary with σ_ν for four different public error outcomes. The observed entry probabilities for the game could be any mixture of these equilibrium outcome distributions according to the density of ϵ . The parameterizations are chosen to be near the boundary of between one and three equilibria in the

¹³This requires the additional assumption that all markets with the same covariate choose to play the same equilibrium.

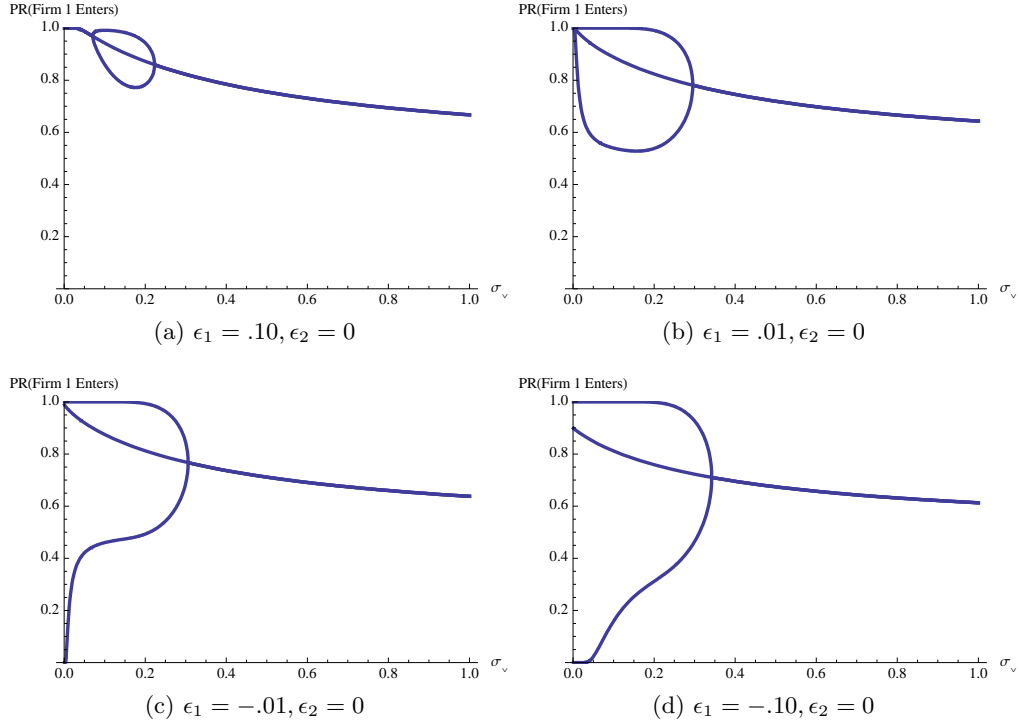


Figure 2: Equilibrium entry probability of player 1 in a symmetric 2-player entry game as the level of uncertainty changes. For both players, $\mu_i = \delta_i = 1$.

complete information game.

Figure 2 illustrates the limit result of Harsanyi (1973): The limit of the set of entry probabilities as $\sigma_\nu \rightarrow 0$ is the equilibrium set of the complete information game. However, even for moderate levels of σ_ν , the equilibrium probabilities are substantially different from those of the complete information game. Figures 2a and 2b provide counterexamples to the commonly held intuition that adding uncertainty to a game tends to reduce the cardinality of the equilibrium set. As the level of uncertainty about player types increases from zero, the cardinality of this set changes from 1 to 3 and back to 1 again.

Assumptions about the information structure affect the estimation of the model. First, examine the assumption of the pure incomplete information framework: estimating the game within in this framework eliminates ϵ from the model and attributes all variation in the data to the private error, ν . Rather than integrating over the equilibrium correspondence the incomplete information estimation searches for the single point on the correspondence which best explains the data. If there

is even a small amount of public information unobserved by the econometrician, this approach biases the degree of player uncertainty in the game upwards. Since σ_ν is fixed according to a scale normalization, overestimating the amount of private information in the game will cause the researcher to underestimate the parameters of the objective function, which will in turn bias the change in entry probabilities in reaction to a counterfactual change in the market observables.

The independence of the private error term places a heavy restriction on the observed distribution of pure incomplete information models: Player actions must be independent of conditional on observables. Including the public information unobservable relaxes this restriction and delivers a significantly more flexible model. From this perspective the role of the public unobservable is similar to the use of random coefficients to relax the independence of irrelevant alternatives assumption inherent in a multinomial logit discrete choice model (Berry, 1994; Berry, Levinsohn, and Pakes, 1995).

The complete information framework assumes that $\sigma_\nu = 0$, so the equilibrium entry probabilities given ϵ are those on the right hand side of Figure 2, and observed entry probabilities given x are derived from integrating over ϵ . When the game is restricted to pure complete information strategies, the equilibrium entry probabilities are either 0 or 1, so integration over $\mathcal{E}(\epsilon, x, \theta)$ amounts to dividing the correspondence into regions within which it is constant. If there are multiple equilibria, for a wide range of ϵ , the market will offer only weak restrictions on the observed outcome distribution entering with probability 1 and entering with probability 0 are both equilibrium outcomes. In the next subsection, I will look how the information structure alters the amount of multiplicity in the model.

5.2 Relationship between Information Structure and Multiple Equilibria

The issue of multiple equilibria is closely related to the information structure. In the general model, for a given set of covariates, different draws of ϵ result in a different game being played. The extent to which multiplicity causes an identification problem is related to the *proportion* of markets with multiple equilibria. In contrast, the game played within the incomplete information framework is the same for all observably identical markets, so a given observed market either has multiple

equilibria or not. In addition, if there are multiple equilibria in an incomplete information game, the bounds on entry probabilities may still be tight (as in Figure 2b), whereas the multiple equilibria of a complete information game usually include extreme outcomes, such as entering with probability 0 or 1.

Morris and Shin (2000) have argued that adding incomplete information can reduce the degree of multiplicity in a model and that such assumptions are appropriate because they begin to relax the common knowledge assumptions of the game. In this section I present numerical evidence that the complete information assumption may exaggerate the multiple equilibria problem by restricting the model to those parameterizations where multiple equilibria are most common. To make this point, Figure 3, presents a graph of the indicating the region of multiplicity in ϵ -space for the model where $\mu = .5$ and $\delta = 1$. This figure can be thought of as a generalization of Figure 1 in Bresnahan and Reiss (1991a) that presents a figure of the region of multiplicity within the complete information framework. The variance of the incomplete information shock is increasing across the four panels of the figure. The distribution of ϵ will determine the density of markets in ϵ -space and the proportion of markets where multiple equilibria exist—but it plays no direct role in Figure 3 itself. The limit result of Harsanyi (1973) is apparent as σ_ν becomes small, the region of multiplicity closely resembles the "box" of multiplicity in the complete information game studied by Bresnahan and Reiss (1991a) and Tamer (2003).

Two striking observations can be drawn from Figure 3. First, the size of the region of multiplicity shrinks as σ_ν increases. Second, multiplicity in the presence of uncertainty is much more likely when the two firms are symmetric—that is, markets along the line $\epsilon_1 = \epsilon_2$ continue to exhibit multiplicity even when σ_ν is relatively high. While these figures do not constitute a proof, these results are stable across several different parameterizations of the model.

Intuition for the decrease in the size of the region of multiplicity can be found by considering players' best response functions. The shape of the players' best response functions corresponds to the cumulative density function (CDF) of ν . Within the set of rationalizable strategies, this CDF becomes closer and closer to linear as the variance of ν is increased. Since approximately linear best response functions are likely to intersect only once, multiplicity becomes more rare as σ_ν increases.

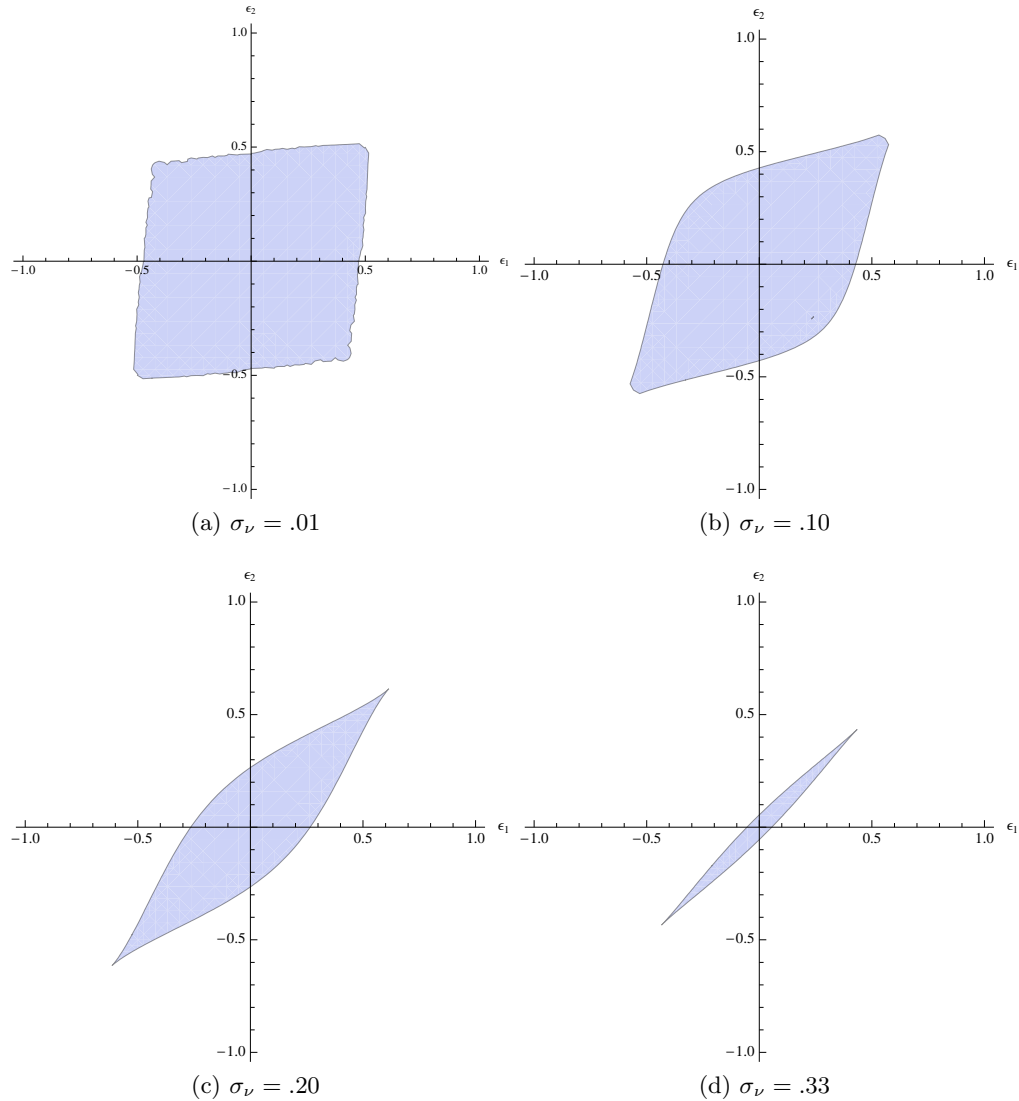


Figure 3: Multiplicity in a 2-player game correspondence $\mathcal{E}(\cdot, \theta)$ varying the degree of incomplete information by panel. There are multiple equilibria in the shaded region, 1 equilibrium in the unshaded region. Axes correspond to (ϵ_1, ϵ_2) . For both players $\mu = 0.5$ and $\delta = 1$. The region of multiplicity for the limiting complete information game is the box $[-.5, .5] \times [-.5, .5]$ (cf. Bresnahan and Reiss, 1991a, Figure 1).

The symmetry result is also intuitive. If one firm is seen to be more profitable ex ante, this firm will be expected to enter with higher probability, which in turn lowers the expected profits of its rival. In this way, strong firms increase their advantage over weaker firms on account of expectations. Expectations are shifted towards the stronger firm entering, while equilibrium beliefs under which the weak firm enters with high probability are more difficult to rationalize.

The result that even a small amount of uncertainty can drastically reduce the amount of multiplicity within the model means that, even without specifying the mixing distribution, the bounds on entry probabilities from a model with a moderate amount of multiplicity may be much tighter than the bounds of a similar complete information model. For counterfactuals, this is welcome news, the generalization from complete information to the general model is unlikely to exacerbate the multiplicity issue when we use the model to perform counterfactuals. For estimation, the result is more muted. Because the model is at the height of its flexibility under the complete information assumptions, it may be difficult to reject the complete information paradigm. As I show in the following section, identification of the model, and testability of the complete information assumption relies heavily on variation in the covariates.

6 Identification

The full model presented here nests the complete and incomplete information frameworks as end-points of a continuum of possible information structures. I study identification under rich support assumptions in Section 6.1. Tamer (2003) showed that the payoff function of a complete information game was point identified if the covariates have a rich support, and I show that this result extends to the general model. However, I am unable to achieve point identification of the information structure itself. Instead, I show that both the complete information and incomplete information assumptions commonly used in the literature are testable against the general framework. That is, these assumptions provide a set of restrictions which can be checked in the observed data.

Without assuming covariates have a rich support, the model is only partially identified. In Section 6.2 I derive the identified set for the model parameters without assuming a rich support.¹⁴

¹⁴Bajari, Hahn, Hong, and Ridder (2008) study a similar model and argue that it may not be point identified

Even though the model is set identified, this may have little practical effect on the results of the estimation if the identified set is small.¹⁵ Section 7 will provide techniques to perform inference on the model and conduct counterfactual analysis using confidence regions for the parameters of interest without assuming point identification. The alternative to inference under partial identification is to add assumptions that are capable of point identifying the model. The mis-specified complete information or incomplete framework, while point identified, may produce results that are driven by these assumptions alone.

As is typical of discrete outcome models, we must provide a location and scale normalization. The location normalization is due to the fact we only observe differences when latent variables cross some threshold. It is implicitly recognized by assuming that the payoff for non-action is always zero. The need for a scale normalization is also due to the threshold crossing nature of the problem, since all parameters can be scaled by a constant to reproduce identical outcome probabilities. The scale normalization can be accomplished by fixing one term one variance term in the model. In the application, I use the normalization that $\sigma_\epsilon^2 + \sigma_\nu^2 = 1$.¹⁶ We must also assume a set of exclusion restrictions in order to separately identify $\mu_i(\cdot)$ from $\delta_i(\cdot)$. These exclusion restrictions were formally included in the model by the partitioning of x in Section 4. Bajari, Hong, and Ryan (2007) discuss the need for exclusion restrictions to identify complete information games and Bajari, Hong, Krainer, and Nekipelov (2008) provide a similar discussion in the context of incomplete information games. These arguments carry over directly into the model of this paper.

6.1 Identification with Rich Support Assumptions

In this section, we analyze the 2 player model under the assumption that the covariates have a rich support. With this assumption, I first show that the parameters of firm payoff functions in the general model are point identified. Intuitively, this is because in the limit, we can observe markets where firm's probability of entry given its covariates is 1 or 0 regardless of whether we condition

without parametric restrictions on the selection mechanism. In this section I explore what can be learned from the model without imposing point identifying assumptions.

¹⁵See Honoré and Tamer (2006) for an illustration of this point in the context of a single agent dynamic model.

¹⁶Because players identities are randomly assigned, I assume that the variance of the error terms is the same for both players.

on the publicly observed shock. In these cases, agents always take opponent entry as given, and the problem reduces to a standard threshold crossing model. I then show that the implications of both the complete information model and the incomplete information model are testable. That is, data generated by a model that includes a public and private information shock will be inconsistent with the implications of the complete information and the incomplete information frameworks.

Rich support assumptions are a common tool in the identification of games. The assumption used here is similar o that of Tamer (2003), which analyzes identification in the context of a pure strategies complete information game.

Assumption 2. *We observe a random sample of M markets $\{(y_m, x_m)\}$, $m = (1, \dots, M)$ where $y_m = (y_{1m}, y_{2m})$ are binary indicators of whether firm i operates in the market and $x_m = (x_{1m}^*, x_{1m}, x_{2m}^*, x_{2m})$ are covariates.¹⁷*

Assumption 3. *The components of the payoff functions are linear, and for each firm i there is at least one covariate x_i^* which is excluded from the other firms payoff function and the determinants of the competition effect. Given this assumptions we can rewrite the payoff functions as follows, for $i = \{1, 2\}$,*

$$\pi_i(y_i, y_{-i}; x, \theta) = \begin{cases} x_i^* \theta_{i\mu}^* + x_i \theta_{i\mu} + \nu_i + \epsilon_i & \text{if } y_i = 1, y_{-i} = 0 \\ x_i^* \theta_{i\mu}^* + x_i \theta_{i\mu} + x_i \theta_{i\delta} + \epsilon_i + \nu_i & \text{if } y_i = 1, y_{-i} = 1 \\ 0 & \text{if } a_i = 0 \end{cases} \quad (4)$$

The variables x_1 and x_2 may have common elements. The matrices (x_1^, x_1) and (x_2^*, x_2) have full column rank.*

Assumption 4. *For firms 1 and 2, the a regressor the density of x_i^* conditional on all other covarites is everywhere positive for $i = \{1, 2\}$.*

Assumption 5. *The unobserved error terms (ϵ, ν) are assumed to be independent of x , and ϵ is independent of ν . The distributions of the error terms are parameterized as follows: $(\epsilon_1, \epsilon_2) \sim BVN(\sigma_\epsilon^2, \sigma_\epsilon^2, \rho)$ and $\nu_i \sim N(0, \sigma_\nu^2)$. For a scale normalization, we assume that $\sigma_\epsilon^2 + \sigma_\nu^2 = 1$.*

¹⁷I suppress the market subscript when it is unnecessary from context.

With these restrictions the parameters of the model are $\theta = (\theta_{i\mu}^*, \theta_{i\mu}, \theta_{i\delta}, \sigma_\epsilon^2, \rho)$ for $i \in \{1, 2\}$ and the equilibrium selection mechanism $\lambda(\cdot)$.

Theorem 1. *If Assumptions 1 through 5 hold, then the parameters of the payoff function $(\theta_{i\mu}^*, \theta_{i\mu}, \theta_{i\delta})$ are point identified.*

Theorem 1 generalizes the result of Tamer (2003) for complete information games to the general model with an unspecified information structure. The argument of the proof concentrates on those markets where strategic interaction is not a factor, and so the information structure plays no role in the identification argument. We next show that assumptions of the commonly used pure strategy complete information model can be tested against the general model.

Complete information models make the assumption that all unobserved variation is observable to all players, eliminating ν from the model while retaining ϵ . Within the context of the general model, this is equivalent to maintaining the assumption that $\sigma_\epsilon^2 = 0$. If it is assumed that players play pure strategies, then we will show below that for every x , that the probability that either both firms enter or neither firm enters is independent of the selection mechanism. By examining these statistics for a single market, we can identify the only remaining parameter, ρ under the pure strategies complete information assumption. The only unidentified parameter is the selection mechanism. However, there are also an infinite number of restrictions implied by the complete information model which are independent of selection, and these can be used to test the complete information assumption on the information structure.

Theorem 2. *If Assumptions 1 through 5 hold, the pure strategies complete information framework is testable.*

Theorem 2 provides a set of equality restrictions with which to test the pure strategies complete information assumption. The key feature of these equalities is that they do not depend on the selection mechanism, λ , which is not identified. If the degree of private information is very small, than the complete information assumption is mild due to Harsanyi's limit result. However, Figure 2 indicates that entry probabilities for a particular point on the equilibrium correspondence change substantially when even a small amount of incomplete information is introduced in the model. The

following theorem shows that a similar test can be devised for the incomplete information model.

Theorem 3. *If Assumptions 1 through 5 hold, the pure incomplete information framework is testable.*

If the only unobserved variation is generated by privately observed structural errors, which are by independent across firms by assumption, then it must be that players actions will be independent conditional on the other observables. This requirement is very strict, and so the it would appear that the incomplete information model will be rejected in a many cases. The full model of this paper relaxes this conditional independence implication by adding public firm-level heterogeneity while still allowing firm uncertainty to play a role in determining equilibrium.

6.2 Identified Set

For expositional simplicity I first consider in Section 6.2.1 identification in the context of an incomplete information game where no there is no public information unobservable. In Section 6.2.2 I will derive the identified set for the general model with both complete and incomplete information structural errors.

6.2.1 Identified Set with Only Incomplete Information

Consider the following model,

$$\pi_i(y_i, y_{-i}; x, \theta) = \begin{cases} \mu(x_0, x_i; \theta_\mu) + \nu_i & \text{if } y_i = 1, y_{-i} = 0 \\ \mu(x_0, x_i; \theta_\mu) - \delta(x_0; \theta_\delta) + \nu_i & \text{if } y_i = 1, y_j = 1 \\ 0 & \text{if } a_i = 0 \end{cases} \quad (5)$$

With the distributional assumption that $\nu_i \sim N(0, 1)$ is independent of x and i.i.d. across players. Note that the location, scale, and exclusion restriction assumptions are already built into the model. The payoff of the action 0 is always 0, the variance of the error is scaled to 1, and x_i affects the payoff of only player i .

Section 4 derived the necessary and sufficient condition equilibrium strategies and entry probabilities for this model (the only difference is that ϵ is eliminated from the equilibrium constraints).

An equilibrium can be expressed either as a vector of cutoffs $(\chi_1(x, \theta), \chi_2(x, \theta))$ or entry probabilities $(\rho_1(x, \theta), \rho_2(x, \theta))$ which are related through the one-to-one mapping (2). It is more convenient to describe the equilibrium in terms of the entry probability profile $\rho(x, \theta)$. As we have seen above, this model may have multiple equilibria, hence the solution to the model is a correspondence $\mathcal{E}(x, \theta)$, which maps a set of covariates and the parameters to a finite set of equilibrium strategies. Let $\#\mathcal{E}(x, \theta)$ be the cardinality of this set and let $\rho^e(x, \theta)$ index an element of $\mathcal{E}(x, \theta)$ where $1 \leq e \leq \#\mathcal{E}(x, \theta)$. In our model, where 2 players are making a binary decision and ν is assumed to be normally distributed, the maximum number of equilibria possible is 3. In a game with more players or a larger choice set, the number of equilibria may be substantially larger.

For the purposes of identification, we can treat the vector of conditional outcome probabilities $P(y|x)$ as observed. Every equilibrium profile implies a multinomial distributions over outcomes. For a profile $e \in \mathcal{E}(x, \theta)$ let $P^e(\theta, x)$ be the resulting multinomial distribution over outcomes. In our two player model there are four possible outcomes, in a general binary choice model with N players there are 2^N possible outcomes where N is the number of players. Suppose we knew with certainty that a particular equilibrium e of the set $\mathcal{E}(x, \theta)$ is played, then the probability of observing outcome y can be written,

$$P_y^e(x, \theta) = \prod_{i=1}^2 \rho_i^e(x, \theta)^{\mathbf{1}[y_i=1]} (1 - \rho_i^e(x, \theta))^{\mathbf{1}[y_i=0]} \quad (6)$$

Let $P^e(x, \theta)$ be the vector of outcome probabilities indexed by y for equilibrium e of the set $\mathcal{E}(x, \theta)$. Note that the independence of the agents decision conditional on x is implied by the fact that the only structural error within the incomplete information model is independent across players and privately observed.

If equilibrium were unique, then we could connect $P^e(x, \theta)$ directly to the observed data, since this is the only possible outcome consistent with equilibrium. When equilibrium is not unique, our assumptions imply that the observed outcome distribution is some mixture of equilibrium strategies according to the equilibrium selection mechanism $\lambda(x, \theta)$. For a given selection mechanism, we have can denote a unique prediction for the model,

$$P_y(x, \theta, \lambda(x, \theta)) = \sum_{e \in \mathcal{E}(x, \theta)} \lambda^e(x, \theta) P_y^e(x, \theta) \quad (7)$$

If we had a parametric model for λ , we could estimate the selection mechanism using the likelihood function implied by (7). Such a strategy is pursued in the complete information context by Bajari, Hong, and Ryan (2007). Another common assumption in the context of incomplete information games is that the same equilibrium is always played in observationally equivalent markets. However, economic theory tells us nothing about the selection of an equilibrium to play, so I do not wish to impose strong restrictions on the selection mechanism. In Appendix A, I use a numerical example to show how incorrect assumptions about the selection mechanism result in a mis-specified model and biased estimates of the model parameters.

Instead, I allow λ to be any a valid mixture across equilibria and derive the sharp identified set implied by the model. That is we use the model of the equilibrium correspondence given θ and the fact that we know that $\lambda^e(x, \theta) \geq 0$ and $\sum_{e \in \mathcal{E}} \lambda^e(x, \theta) = 1$ to derive the identified set.

Theorem 4. *If assumptions 1-3, and 5 hold. The sharp identified set of θ for the incomplete information model is,*

$$\Theta_I = \left\{ \theta \in \Theta : \forall x \in X, \exists \tilde{\lambda} \in [0, 1]^{\bar{E}} \text{ s.t. } \begin{aligned} P(\cdot|x) &= \sum_{e \in \mathcal{E}(x; \theta)} \tilde{\lambda}^e P^e(x, \theta), \\ \sum_{e \in \mathcal{E}(x; \theta)} \tilde{\lambda}^e &= 1 \end{aligned} \right\} \quad (8)$$

Where \bar{E} is a constant which represents the largest possible number of equilibria the model admits almost everywhere over $X \times \Theta$.

To gain some intuition into how (8) restricts the identified set, consider testing whether θ is the true parameter. For a given x , a 4x1 vector $P(y|x)$ is observed. For a given θ , we know that there are generically either 1 or 3 equilibria based on the assumed normality of the private error term. Suppose $\mathcal{E}(x, \theta)$ is a singleton, then the equilibrium mixing distribution is degenerate and x provides 3 restrictions with which to test the parameter θ . In contrast, suppose there are three equilibria in $\mathcal{E}(x, \theta)$. Then the observed outcome distribution given θ is the true parameter may be any valid probability mixture across the three equilibria, so $\lambda(x, \theta)$ is overidentified given following

set of 4 linear constraints and 3 unknowns.

$$\begin{bmatrix} P((0,0)|x) \\ P((0,1)|x) \\ P((1,0)|x) \\ 1 \end{bmatrix} = \begin{bmatrix} P_{(0,0)}^1(x, \theta) & P_{(0,0)}^2(x, \theta) & P_{(0,0)}^3(x, \theta) \\ P_{(0,1)}^1(x, \theta) & P_{(0,1)}^2(x, \theta) & P_{(0,1)}^3(x, \theta) \\ P_{(1,0)}^1(x, \theta) & P_{(1,0)}^2(x, \theta) & P_{(1,0)}^3(x, \theta) \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda^1(x, \theta) \\ \lambda^2(x, \theta) \\ \lambda^3(x, \theta) \end{bmatrix} \quad (9)$$

Since we have an extra restriction to identify the mixing distribution, market types with multiple equilibria provide a single restriction which can be used to verify whether θ is the true parameter. We can rule out any θ for which the restriction that $\lambda(x, \theta) \geq 0$ is not satisfied.

When we observe variation in x , we test θ by aggregating the restrictions across all market types. Suppose M different market types are observed, $X = \{x^1, \dots, x^M\}$, so $P(y|x)$ is a vector of length $4M$. At worst, there are multiple equilibria in all of the markets, in this case, there are M restrictions available to identify the components of θ .

6.2.2 Identified Set of the Full Model

I now return to the full model presented in Section 4 by adding firm specific complete information shocks that are observed to the players but unobserved by the econometrician. This model can be understood as a generalization of the incomplete information model with firm level unobserved heterogeneity. As a result, player strategies and equilibrium selection must both be modeled as functions of the publicly observed error ϵ . If selection were independent of complete information, then the researcher could deal with this additional unknown by simply integrating it out of the final likelihood function. However, the potential relationship between the complete information shock and the selection mechanism is not restricted by our assumptions.¹⁸ Therefore, selection mechanism is now potentially a function of ϵ , and as a consequence is an infinite dimensional parameter.

Since players know ϵ prior to forming strategies, the introduction of unobserved public shock has no effect on how equilibrium is constructed, it merely requires that the econometrician con-

¹⁸One alternative would be to adopt one of several equilibrium refinement concepts to reduce the number of valid solutions. Kajii and Morris (1997) have explored the possibility of using robustness to incomplete information as an equilibrium refinement. Even if a refinement is applied, an equilibrium selection mechanism will be necessary unless the refinement can be shown to produce solution set that is always single valued.

sider a distribution of possible games rather than a single one. The solution to the model is now a correspondence mapping ϵ to a set of equilibria $\mathcal{E}(\cdot, x, \theta)$. The mapping from equilibria inherits the additional argument ϵ but is otherwise unchanged. Identification therefore will be based on the mapping from $(\epsilon, x, \theta, \lambda)$ to outcome distributions conditional on x , and comparing these distributions with the observed data.

By adding heterogeneity to the model, it must be true that this identification set is larger than the one presented in the earlier section where we assumed the researcher had perfect knowledge up to private shocks. The two models are nested when since the former is a special case of the latter when the distribution of ϵ is degenerate. The observed distribution of outcomes should now integrate out unobserved heterogeneity.

$$P_y(x, \theta, \lambda(\cdot)) = \int \sum_{\epsilon \in \mathcal{E}(\epsilon, x, \theta)} \lambda^\epsilon(\epsilon, x, \theta) P_y^\epsilon(\epsilon, x, \theta) dF(\epsilon; \theta) \quad (10)$$

We can go no further in simplifying this expression since economic theory offers little direction as to how the selection mechanism is related to ϵ . Nonetheless, we have nontrivial restrictions on the identified set.

Theorem 5. *If assumptions 1-3 and 5 hold, the identified set of the full model is,*

$$\Theta_I = \left\{ \begin{array}{l} \theta \in \Theta : \forall x \in X, \exists \lambda_x(\epsilon) \in [0, 1]^{\bar{E}} \text{ s.t. } P(y|x) = \int \sum_{\epsilon \in \mathcal{E}(\epsilon, x, \theta)} \lambda_x^\epsilon(\epsilon) P_y^\epsilon(\epsilon, x, \theta) dF(\epsilon), \\ \forall \epsilon, x : \sum_{\epsilon \in \mathcal{E}(\epsilon, x, \theta)} \lambda_x^\epsilon(\epsilon) = 1 \end{array} \right\} \quad (11)$$

The function $\lambda(\cdot)$ is restricted to the set $\Lambda(\theta)$: it must be a valid mixing distribution between equilibria. So meaningful inference can still be performed. Indeed, the distribution implied by λ must be degenerate whenever equilibrium is unique, which may occur with high probability for some parameterizations.

Suppose for some θ we observe a market which implies a unique for all values of ϵ , this market provides 3 equality restrictions on the data with which to test θ , the same number supplied in the incomplete information case when equilibrium for a market type is unique. The more such markets are observed the more restrictions can be collected to test the model parameters. Even if

all market types exhibit multiple equilibria with high probability, nontrivial restrictions on θ can still be found.

The identified set can be equivalently defined as the maximizers of the model likelihood. In the next section, we will use this fact to make inference on the model.

7 Inference

The previous section showed that with complete knowledge of the outcome distribution we can reduce the set of parameter values that may have generated the data to a non-trivial set, if not to a single point. In this section, we use a random sample markets to make inference on the set of models that plausibly could have generated the data. Chernozhukov, Hong, and Tamer (2007) have shown how to perform inference on a set identified model by using the empirical analog of a function that attains its optimum only on the identified set. I use the log-likelihood of the model. This section proceeds in three steps. First, I show how the likelihood function of the model can be used as an objective function conduct set inference on θ . Second, I show that a computable empirical analogue of this function converges to the true model likelihood. Finally, I show that the weighted bootstrap can be used to test the null hypothesis that a particular model is in the identified set.

7.1 The Model Likelihood

I assume that I observe data from n independent markets $\{x_i, y_i\}_{i=1}^n$ each of which has comes from a data generating process defined by the true parameters (θ^0, λ^0) .¹⁹ Since the true model is complete in the sense that it includes both a well defined model and a valid equilibrium selection mechanism, it maps on to a unique point in the space of outcome distributions, $P^0 \in \mathcal{P}$. The partial identification problem arises because there may be multiple models (θ, λ) which generate

¹⁹In this section, i is used to index observations rather than players. I.e., y_i is a vector recording each player's action.

P^0 besides true model. The likelihood function for the model can be written,

$$L(\theta, \lambda) = E [\log(P_y(x, \theta, \lambda))] \quad (12)$$

Following the usual likelihood arguments, the likelihood function will be maximized at (θ^0, λ^0) , since $P(\cdot, \theta^0, \lambda^0) = P^0$. The maximizer is *not* assumed to be unique, i.e., there may exist (θ', λ') such that $L(\theta^0, \lambda^0) = L(\theta', \lambda')$.²⁰ I treat the selection mechanism λ as an infinite dimensional nuisance parameter and focus on the model parameters θ as the object of interest. The profiled likelihood function associates each θ with the most favorable selection valid selection mechanism in terms of likelihood.²¹

$$L(\theta) = \max_{\lambda \in \Lambda(\theta)} E [\log(P_y(x, \theta, \lambda))] \quad (13)$$

The following Lemma states that $L(\theta)$ is a continuous function that, if it were known, could be used to locate the identified set.

Lemma 1. (i) $L(\theta) = \operatorname{argmax}_{\theta' \in \Theta} L(\theta')$ for all $\theta \in \Theta_I$ as defined by (11), and $L(\theta) < \operatorname{argmax}_{\theta' \in \Theta} L(\theta')$ for all $\theta \in \Theta \setminus \Theta_I$, and (ii) $L(\theta)$ is continuous in θ .

7.2 Empirical Analogue for the Log-Likelihood

Using $L(\theta)$ for inference is infeasible because we do not perfectly observe the outcome distribution and so are unable to evaluate the expectation. Furthermore computation of $L(\theta)$ involves an optimization over the infinite dimensional selection mechanism. I now define a feasible empirical analogue that will allow me to conduct inference on the identified set. In order to make the empirical likelihood feasible to compute, I use a numerical approximation to convert the infinite dimensional optimization problem into a finite dimensional one.

To numerically approximate the integral over ϵ in the definition of $P_y(x, \theta, \lambda)$ I assume that a numerical approximation technique is employed by which the integrand is evaluated at a finite

²⁰In the event that the maximum is unique, the model is point identified. In this case the inference procedure presented in this section is still valid.

²¹An alternative argument for the inference method is that I am minimizing the Kullback-Leibler divergence between P^0 and the set of probability distributions which can be generated by θ for any selection mechanism. Using this explanation, the maximization over λ is part of the divergence metric between a point and a set.

number of R sample points from the distribution of ϵ .

$$P_y^R(x, \theta, \lambda) = \frac{1}{R} \sum_{r=1}^R \sum_{e \in \mathcal{E}(\epsilon, x, \theta)} \lambda^e(\epsilon_r, x) P_{y_i}^e(\epsilon_r, x, \theta) f(\epsilon_r; \theta) \quad (14)$$

Where $f(\epsilon_r; \theta)$ is the density of ϵ_r given the parameterization θ . By the law of large numbers, as the number of sample points $R \rightarrow \infty$, $P^R(x, \theta, \lambda) \rightarrow P(x, \theta, \lambda)$ in probability. The method for selecting sample points may be either through simulation using random or pseudo-random sequences or through deterministic sequences such as Halton sequences. In practice, I use Halton sequences, which have been shown to have desirable fast-convergence properties vis-a-vis pseudo-random sequences in numerical simulations.²² Since the number of sample points used is under the control of the researcher, we assume that $R \rightarrow \infty$ as $n \rightarrow \infty$, and that hence simulation error is asymptotically negligible. Using this approximation for the integral,

$$L_n(\theta) = \max_{\lambda \in \Lambda(\theta)} \frac{1}{n} \sum_{i=1}^n [\log(P_{y_i}^R(x_i, \theta, \lambda))] \quad (15)$$

The empirical analogue $L_n(\cdot)$ converges uniformly to $L(\cdot)$. That is, $\sup_{\theta \in \Theta} |L_n(\theta) - L(\theta)| = o_p(1)$. To see this, recall that we assume that both covariates and outcomes have discrete support and define the maximum likelihood estimator for the joint outcome distribution as,

$$\hat{P}^n(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}[x_i = x, y_i = y]$$

Clearly \hat{P}^n is a consistent estimator for P^0 . By re-arranging terms, we can write $L_n(\cdot)$ as a function of \hat{P}^n ,

$$L_n(\theta) = \max_{\lambda \in \Lambda(\theta)} \sum_{x \in X} \sum_{y \in Y} \hat{P}^n(x, y) \log(P_y^R(x, \theta, \lambda))$$

we can also write,

$$L(\theta) = \max_{\lambda \in \Lambda(\theta)} \sum_{x \in X} \sum_{y \in Y} P^0(x, y) \log(P_y(x, \theta, \lambda))$$

Since $\hat{P}^n \rightarrow P^0$ in probability, and since by assumption $R \rightarrow \infty$ as $n \rightarrow \infty$, for all θ , $L_n(\theta) \rightarrow L(\theta)$

²²For details about Halton sequences, see Bhat (2001).

by the continuous mapping theorem.²³ This is a sufficient condition for uniform convergence of $L_n(\cdot)$ to $L(\cdot)$.

Computing L_n involves a maximization over the set $\Lambda(\theta)$. However, to evaluate the maximand of L_n it is only necessary to evaluate the likelihood on the set of R sample points, not every point in ϵ space. This reduces the the maximization problem over $\Lambda(\theta)$ to a finite, but high dimensional constrained optimization problem. Fortunately, the form of this problem is tractable for modern nonlinear optimization packages even with a very high dimensional parameter space. Appendix C appendix provides more computational detail on the optimization step.

Informally, the uniform convergence result indicates that points that are nearly maximizers of $L_n(\cdot)$ will be nearly maximizers of $L(\cdot)$ with “nearly” growing stricter as n increases. Since $\operatorname{argmax}_{\theta \in \Theta} L(\theta) = \Theta_I$, this implies that parameter values that nearly maximize $L_n(\cdot)$ will be more likely to be in the identified set than those that do not. Put more formally,

Theorem 6. *For all $\theta \in \Theta_I$, $\max_{\theta' \in \Theta} L_n(\theta') - L_n(\theta) = o_p(1)$.*

The next section uses this result to propose a hypothesis test for the null hypotheses that a given θ is included in the identified set.

7.3 Likelihood Ratio Test and the Weighted Bootstrap

Inference will be done via a likelihood ratio test on the statistic $n(\max_{\theta' \in \Theta} L_n(\theta') - L_n(\theta))$. To see why, suppose momentarily that θ^0 was point identified, we could write the likelihood test in the usual way, testing each θ against the (unique) maximizer of L_n and approximating the asymptotic distribution using an estimator for the Fisher information matrix.²⁴ This would be valid because the maximizer and its Hessian are consistent estimators for the (assumed) unique maximizer of \mathcal{L} and its Hessian respectively. If θ^0 is only set identified, the regularity conditions that support the classical asymptotic distribution do not hold. In particular the information matrix is degenerate on

²³The continuity of $L_n(\cdot)$ in θ for each n follows the argument for the continuity of $L(\theta)$ in Lemma 1.

²⁴The classical expansion is of the likelihood ratio statistic when the model is point identified and $\hat{\theta} = \operatorname{argmax} L_n(\theta)$ is $2n(\max_{\theta' \in \Theta} L_n(\theta') - L_n(\theta)) = n(\hat{\theta} - \theta)\Psi(\hat{\theta} - \theta) + o_p(1)$, where Ψ is the Fisher information matrix, the right hand side can be shown to have a chi-squared distribution. The statistic is multiplied by 2 because the right hand side is the quadratic term in a Taylor expansion. I drop the 2 because I approximate this distribution using the weighted bootstrap rather than derive an analytic approximation.

the interior of the identified set. To relax the assumption that the model is point identified, Liu and Shao (2003) have derived the asymptotic distribution in the set identified case under generalized regularity conditions that use the Hellinger distance between probability distributions rather than the Euclidean distance in Θ as the basis for an asymptotic expansion. However, their analytic approximation is quite cumbersome to apply in this case, particularly because L_n itself involves a high dimensional optimization problem due to the selection mechanism. Instead, I approximate this distribution and use it for inference using the weighted bootstrap.

The weighted bootstrap was introduced by Ma and Kosorok (2005) for use in attaining confidence intervals for the finite dimensional parameters in a semi-parametric estimation of a point identified problem. The technique was shown to be valid in partially identified semi-parametric problems by Chen and Tamer (2009). The weighted bootstrap is attractive in our problem because of the straightforwardness of its implementation and the ease with which validity can be shown.²⁵

We wish to test the null hypothesis $H_0 : \theta = \theta^0$, under the null hypothesis,

$$\lim_{n \rightarrow \infty} P \left(n \left(\sup_{\theta' \in \Theta} L_n(\theta') - L_n(\theta) \right) \geq c_{1-\alpha} \right) \geq 1 - \alpha \quad (16)$$

Where $c_{1-\alpha}$ is the $1 - \alpha$ quantile of the distribution of $n(\sup_{\theta' \in \Theta} L_n(\theta') - L_n(\theta))$. If this quantile were known, we could use (16) as a test of H_0 that is consistent at level α . To gain a feasible test, we replace $c_{1-\alpha}$ with a consistent estimator by using the weighted bootstrap to simulate draws from the asymptotic distribution. The weighted bootstrap uses B different weighted likelihood functions, which share the same asymptotic distribution as the standard likelihood function, to approximate the distribution of the likelihood ratio test statistic. The *weighted* likelihood function is defined as a function of weights $w = (w_1, \dots, w_n)$ which are independent of the data and distributed such that $E(w_i) = 1$, $V(w_i) = 1$ and $w_i > 0$.²⁶ We can then compute the weighted likelihood function

²⁵See Romano and Shaikh (2006, 2008) for more on subsampling in this context. Several authors have studied other inference techniques in the context of moment inequality models (Andrews and Soares, 2007; Bugni, 2007; Canay, 2008). These techniques are not immediately applicable in our context since the objective function does not take the form of a finite weighting of moment inequalities.

²⁶I use the standard exponential distribution to produce these weights, some experimentation using the log-normal distribution produced very similar results.

as,

$$L_n(\theta, w) = \max_{\lambda \in \Lambda(\theta)} \frac{1}{\sum w_i} \sum_{i=1}^n w_i [\log(P_{y_i}^R(x_i, \theta, \lambda))]$$

I approximate the quantiles of $\sup_{\theta' \in \Theta} L_n(\theta') - L_n(\theta)$ by choosing B sets of weights and computing the quantiles of $\{\sup_{\theta' \in \Theta} L_n(\theta', w^b) - L_n(\theta, w^b)\}_{b=1}^B$. Note that this requires an optimization step for weighted bootstrap draw. The key to validity of the weighted bootstrap is the independence of the bootstrap weights from the data. This allows one to show that the unweighted and all of the weighted estimators share an identical asymptotic distribution as $n \rightarrow \infty$.²⁷

Using the weighted bootstrap to approximate the distribution of the likelihood ratio statistic, we follow the procedures of Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2008) to derive confidence sets for the identified set and the identifiable parameter. The computational details of these procedures are discussed in Appendix C.

8 Data and Preliminary Analysis

8.1 Data

My primary data source is annual extracts of the Trade Dimension TDLinx database of all grocery store locations in the United States.²⁸ To define markets, assume that a local grocery-store is in competition with other grocery stores located roughly within the same ZIP code. Since ZIP codes themselves are route assignments and not geographic areas, the US Census bureau has developed ZCTA Code Tabulation Areas (ZCTAs) which are generalized geographic representations of ZCTA codes. Roughly speaking, ZCTAs map each census block to the ZCTA code of the majority of residents.²⁹ I use the address information on each store to link that store to a year 2000 ZCTA.

ZCTAs provide a reasonable approximation of a grocery store catchment area in the markets that I study. In their study of the loyalty card program of a grocery store located in a “small East

²⁷This is in contrast to the standard bootstrap, where validity would require that the distribution of $\mathcal{L}_n(\theta, \cdot)$ be smooth in P .

²⁸This data was graciously provided by Paul Ellickson. Various extracts of the Trade Dimensions database have been used in several empirical studies investigating retail industries (Ellickson, 2007; Beresteanu and Ellickson, 2006; Holmes, 2008; Orhun, 2005).

²⁹For a full description of the creation of ZCTAs, see US Census, <http://www.census.gov/geo/zip/zcta.html>, accessed August 1, 2009.

Coast town”, Singh, Hansen, and Blattberg (2006) found that the average customer lived 3.5 miles from the store while 78 percent of customers lived within 5 miles of the store. A five mile radius translates roughly to a catchment area of 80 square miles. The average land area of a market as defined in the data is 144 square miles, with larger markets in the West than on the East Coast.

To be included in the dataset as a rural grocery market a ZCTA must (i) have fewer than 15,000 people, (ii) have a population density of less than 750 people per square mile, (iii) have had at least one grocery store in operation between 1994 and 2006, and (iv) have had no more than two grocery stores in simultaneous operation between 1995 and 2006.³⁰ The restrictions on the number of stores active in the ZCTA is necessary to maintain computational feasibility and to eliminate those ZCTAs which are not viable locations for grocery stores. Including markets where three stores are open simultaneously would expand the data set by 9 percent. Including these markets does not appear to affect the descriptive results.

Examining the data geographically, there are fewer markets in the West than the East due to differences in population density and larger ZCTA size. There are also relatively fewer markets near a supercenter in the west. Expansion of supercenters into the West, led by Wal-Mart, has taken place since 2000, however the density of supercenters remains much sparser than in the East. In the East, markets near a supercenter are fairly interspersed with those that are not. There is some degree of clustering because a single supercenter affects all markets within a 20 mile radius. It does not appear that distance to a supercenter varies across geographical regions.

Table 1 displays summary statistics on the markets in the data set. Distance to a supercenter is calculated on the basis of ZCTA geographic centroids provided by the US Census. The expansion of supercenters is apparent from the evolution of the distance from the markets to a supercenter over time. Between 1998 and 2002, the median distance to a supercenter decreased by 10 miles. A market is considered to be in the vicinity of a supercenter if the minimum distance to a supercenter is less than 20 miles in 2000. My results are unchanged by using 1998 or 2002 as the base year, or

³⁰For comparison, the US Census designates a census block as “urban” if it has a population density of more than 1000 people per square mile and surrounding blocks have a density of at least 500 per square mile. All census blocks that are not urban are classified as “rural”, US Census, http://www.census.gov/geo/www/ua/ua_2k.html, accessed August 1, 2009.

	Quantiles			Mean	Std Dev
	0.25	0.5	0.75		
Population	2,720	4,582	7,261	5,299.06	3,313.78
Distance to Supercenter					
1998	16.16	29.80	72.36	71.57	108.64
2000	13.57	22.97	46.14	49.71	76.55
2002	12.06	19.54	34.68	31.31	37.48
South				0.34	0.47
West				0.14	0.34

Table 1: Summary statistics of market characteristics. Mean distance to a supercenter excludes outliers more than 500 miles from a supercenter. These are accounted for in the quantile calculation. Total number of markets is 5,893

by extending the supercenter radius to 25 miles.³¹ The average distance to a supercenter in 2000 for those markets who were within the 20 mile radius is 12.3 miles, with a standard deviation of 4.5 miles. Only 14 percent of the data comes from the Mountain West and West Coast, where zip codes are larger, population is sparser, and there are fewer supercenters. The Southeast, which is considered the stronghold of Wal-Mart, the nations largest supercenter chain, contains 34 percent of the markets.

8.2 Preliminary Analysis

Table 2 presents the distribution of active firms in each market in 2002. The first two columns condition the distribution on whether or not there is a supercenter in the vicinity of the market. The table indicates that the presence of a supercenter is associated with fewer duopolies in the market and more monopolies while the number of unserved markets increases only very slightly. This is consistent with the presence of a supercenter negatively affecting traditional grocery stores in small markets.

Table 3 presents the transition matrix of the number of firms in a market between 1998 and 2002 conditional on whether or not there is a supercenter in the vicinity. The transition matrix appears to shift towards less entry and more exit when a supercenter is nearby, but the effect is

³¹Because I use distance between zip code centroids, a smaller distance cutoff may be subject to a significant amount of measurement error. I have also experimented with using two distance bands of 15 and 30 miles and found qualitatively similar results.

Active Firms	Supercenter		Overall
	Yes	No	
0	23.03	23.26	23.16
1	64.45	61.38	62.70
2	12.52	15.36	14.14
N	2,540	3,353	5,893

Table 2: Distribution of number of active firms in 2002 by whether the market was within 20 miles of a supercenter in 2000 (percent).

	Supercenter			No Supercenter		
	0	1	2	0	1	2
0	70.03 (1.19)	28.97 (1.18)	1.01 (0.26)	71.79 (1.21)	27.34 (1.20)	0.87 (0.25)
1	10.02 (0.53)	84.34 (0.65)	5.64 (0.41)	11.94 (0.65)	83.06 (0.75)	5.00 (0.43)
2	0.57 (0.26)	25.98 (1.49)	73.45 (1.50)	0.56 (0.32)	30.06 (1.97)	69.39 (1.99)

Table 3: Transition matrix on the number of firms in the market. Each entry is the percent chance that a market with the row number of firms in 1998 has the column number of firms in 2002. Standard errors in parenthesis.

mild, and not likely to result in a dramatic decline in grocery store availability. These descriptive results indicate the magnitude of the supermarket effect, while the structural model explores the underlying forces that generate this effect.

Table 3 also aggregates firms across all market sizes. As expected, market size is a strong determinant of the number of firms, seen in Table 4. To get a sense of how entry and exit patterns are affected by competition once observable market characteristics are controlled for, Tables 5 and 6 present probit regressions on the propensity of firms to enter and exit, respectively. These regressions do not control for the simultaneous entry and exit decisions of rival firms. Thus the results should not be interpreted causally. The presence of competitors is associated with less entry and more exit, as one would expect. Also, having a supercenter nearby is associated with less entry and more exit, although the association is stronger for exit than entry.

To examine whether geographic variation effects the parameters of interest, I consider the use of region dummies for the South and West. A West dummy might control for the larger zip codes or other region wide characteristics of the West. The results in Tables 5 and 6 show that the

	Percent Near Supercenter	Number Active Stores		
		0	1	2
Pop 0-3k	28.43			
Supercenter		45.67	51.71	2.62
No supercenter		43.16	54.05	2.78
Pop 3-6k	54.14			
Supercenter		24.30	69.44	6.26
No supercenter		15.35	71.16	13.49
Pop 6k+	43.84			
Supercenter		11.91	65.98	22.11
No supercenter		7.51	58.56	33.93

Table 4: Distribution of number of active firms in 2002 by market size and whether or not a supercenter was within 20 miles in 2000 (percent).

effect of region dummies is statistically insignificant and their inclusion has almost no effect on the parameters of interest. Therefore, region dummies are not included in the structural model.

Second, In the structural model I discretize the population data. To validate the appropriateness of the discretization, Tables 5 and 6 include log population as a control in columns I and II and then the discretization used in the full model in columns III and IV. I include dummies for whether population is higher than 3,000 or 6,000.³² The discretization has little effect on the estimated coefficient of supercenter presence or the existence of local competition. The effect of population is non-monotonic, but these results should not be given a causal interpretation. The cutoffs for the dummies were chosen to approximate the one-third and two-thirds quantiles of the data. Probit regressions such as those in Tables 5 and 6 are robust to slight changes in these cutoffs.

Column IV adds an interaction term between the presence of a competitor and population. It appears that the degree of competition between firms varies with market size. I therefore include an interaction between the competition effect and market size in the structural model.³³

The structural model of the following section assumes that entry and exit rates in markets near a supercenter are independent of the supercenter's age. One might be concerned that entry and exit rates immediately following entry by a supercenter are different from the rates that are

³²Both coefficients should be added to compute the total effect of population when the market is larger than 6,000 persons.

³³I have also experimented with interactions between supercenter presence and market size in the entry and exit probits, but these have little effect.

maintained once the supercenter is established in the region. To see whether this is true in the data, I add a dummy for recent supercenter entry in Column V of Tables 5 and 6. This coefficient on recent supercenter entry is not statistically significant for either entry or exit. The positive relationship between exit and supercenters does not appear to depend on whether the supercenter was a recent entrant. Although the addition of the supercenter entry dummy does appear to alter affect the relationship between entry and supercenters, this difference between the coefficients is not statistically significant.

9 Results from the Structural Model

I now apply the model described in Section 4 to data on entry and exit patterns of grocery stores. I assume that entrepreneurs choose whether or not to be active (either remain open or enter the market) in order to maximize expected firm value. I focus on entry and exit within each market for the five year period between 1998 and 2002. A two-player game is assumed to be played in each market. Unfilled slots are assumed to be occupied by potential entrants. State variables at the market level are whether or not a supercenter is present within 20 miles in 2000, and dummies for population greater than 3,000 or 5,000 people. Firm level state variables are whether or not the firm was operating in 1998. The outcome variables are whether or not each firm was operating in 2002. I assume that opening a store affects only a firm's costs and not demand, so it can be used as an exclusion restriction.

The baseline profit function is,

$$\mu_i(x) = \mu_0 + \mu_1 \mathbf{1}[Pop > 3k] + \mu_2 \mathbf{1}[Pop > 6k] + \mu_3 \mathbf{1}[Supercenter < 20mi] - \mu_4 \mathbf{1}[Inactive \text{ in } 1998]$$

We will refer to μ_4 as the entry cost, since it measures the extra costs of opening a new grocery store. Note that we model this as a cost, so the expected sign of this parameter is positive. The effect of competition within the local market is captured by $\delta(\cdot)$ which is also assumed to be linear,

$$\delta_i(x) = \delta_0 + \delta_1 \mathbf{1}[Pop > 3k] + \delta_2 \mathbf{1}[Pop > 6k] + \delta_3 \mathbf{1}[Supercenter < 20mi]$$

	I	II	III	IV	V
Supercenter within 20 mi, 2000	-0.0201 (0.0426)	-0.0133 (0.0447)	-0.0342 (0.0451)	-0.0317 (0.0452)	-0.0065 (0.0489)
Supercenter Entry 1998-2000					-0.0965 (0.0729)
Log Population	0.2660 (0.0278)	0.2620 (0.0279)			
Pop over 3k			-0.4013 (0.0515)	-0.3369 (0.0797)	-0.3348 (0.0798)
Pop over 6k			0.5902 (0.0553)	0.4623 (0.0721)	0.4581 (0.0722)
Competitor Active in 1998	-0.4836 (0.0440)	-0.4775 (0.0441)	-0.4753 (0.0446)	-0.7064 (0.0932)	-0.7090 (0.0932)
(Pop > 3k)x(Comp. 98)				-0.0984 (0.1039)	-0.0989 (0.1039)
(Pop > 6k)x(Comp. 98)				0.3585 (0.1185)	0.3631 (0.1186)
South Region		0.0638 (0.0465)	0.0589 (0.0466)	0.0609 (0.0467)	0.0574 (0.0468)
West Region		0.0953 (0.0614)	0.0593 (0.0619)	0.0640 (0.0621)	0.0674 (0.0622)
Log-Likelihood	-2265.25	-2263.57	-2249.18	-2244.32	-2243.44
N	6496	6496	6496	6496	6496

Table 5: Probit regressions on firm entry between 1998 and 2002. These results do not control for endogeneity of decisions between small grocery stores.

	I	II	III	IV	V
Supercenter within 20 mi, 2000	0.1361 (0.0452)	0.1238 (0.0469)	0.1284 (0.0472)	0.1340 (0.0473)	0.1538 (0.0510)
Supercenter Entry 1998-2000					-0.0828 (0.0804)
Log Population	-0.2516 (0.0332)	-0.2517 (0.0332)			
Pop over 3k			0.1589 (0.0522)	0.1364 (0.0667)	0.1377 (0.0667)
Pop over 6k			-0.4541 (0.0633)	-0.4840 (0.0721)	-0.4843 (0.0721)
In duopoly in 1998	0.1900 (0.0491)	0.1901 (0.0492)	0.1901 (0.0499)	-0.0864 (0.1561)	-0.0843 (0.1561)
(Pop > 3k)x(Duopoly 98)				0.0818 (0.1057)	0.0854 (0.1058)
(Pop > 6k)x(Duopoly 98)				0.2756 (0.1707)	0.2710 (0.1708)
South Region		-0.0114 (0.0485)	-0.0080 (0.0486)	-0.0079 (0.0486)	-0.0096 (0.0487)
West Region		-0.0786 (0.0711)	-0.0721 (0.0711)	-0.0670 (0.0712)	-0.0646 (0.0712)
Log-Likelihood	-2021.46	-2020.84	-2023.52	-2021.39	-2020.86
N	5290	5290	5290	5290	5290

Table 6: Probit regressions on firm exit between 1998 and 2002. These results do not control for endogeneity of decisions between small grocery stores.

Note that the only exclusion restriction for this model is the absence of the activity dummy from $\delta(\cdot)$. The effect on profits from opening a store are assumed to be due to construction and start-up costs, rather than due to low demand for a new store when a rival is active. I do *not* require that the costs of opening the store be paid for by the entrepreneur in the opening period. (I.e, he may take out loans to cover the cost.)

The information structure is parameterized by the variables σ_ϵ^2 and ρ .³⁴ The error structure of most entry game models is nested by the framework presented in this paper. The earliest entry games assumed that markets were subject to a single profit shock that is homogeneous across firms and publicly known to the firms. Bresnahan and Reiss (1991b) showed that with appropriate distributional assumptions this model reduces to an ordered probit. The flexible model of this paper nests these assumptions when, $\sigma_\epsilon^2 = 1$ and $\rho = 1$. The more general complete information model that allows for firm-level heterogeneity in the error term is equivalent to assuming $\epsilon^2 = 1$ within the full model, but allowing ρ to take any value. Most applications of the complete information model have restricted themselves to pure strategy equilibria (e.g., Berry, 1992; Mazzeo, 2002; Ciliberto and Tamer, 2007), so I take the pure strategies assumption as part of the complete information framework. The incomplete information framework is given when by the full model when $\sigma_\epsilon^2 = 0$. A market-level public shock could be added to an incomplete information game, which would be accomplished within my framework by assuming that $\rho = 1$ and allowing σ_ϵ to vary.³⁵

The main parameters of interest for this study are the effect of supercenters on local grocery store profits and the competitive effect of other local grocery stores on store profits. I assume the presence of a supercenter is exogenous to the game being played by local grocery stores. The assumption that national level firms decisions are exogenous to those of local stores greatly simplifies the model, and has been employed in other studies of strategic interaction between local firms (Akerberg and Gowrisankaran, 2006). Supercenters' entry decisions are based on their entire catchment area, rather than the isolated markets under study here, so they are not strongly influenced by the

³⁴Recall that I use the scale normalization $\sigma_\epsilon^2 + \sigma_\nu^2 = 1$.

³⁵In empirical work, Seim (2006) uses a two-stage game where potential entrants first decide whether or not to enter based on a publicly observable market level shock, and then decide on their location given the number of firms who entered. The location decision is an incomplete information game. The two-stage structure is distinct from simply assuming $\rho = 1$ in the model presented here.

presence or absence of local stores. Furthermore, it often takes several years for supercenters to gain zoning approval and build their stores, so their presence is known to the local firms at the time they take operation decisions.

9.1 Confidence Set for the Structural Parameters

For comparison purposes, I estimate the model under the incomplete information and complete information assumptions that are commonly employed in the entry literature. For these models, I use traditional inference methods that rely on the assumption that the model is point identified. For the full model, I do not assume point identification and instead employ the inference techniques described in Section 7. These techniques do not provide point estimates. Instead, I report the 95 percent confidence sets for the identified set and a 95 percent confidence region for the true parameter. The first covers the entire identified set at a confidence level of 95 percent, while the second covers the true parameter values with a confidence level of 95 percent. Table 7 presents the results. Because the inference procedure for the full model yields a joint confidence region, I report projections of this region onto parameter axes. For this reason, Table 7 exaggerates the size of the confidence sets. Many parameter values within the cartesian product of these intervals are outside the confidence set. The counterfactuals in the next section operate using the true confidence region, a subset of the “box” reported in Table 7.

The full model nests both the complete and the incomplete information frameworks. Therefore, we can use the full model to test the other two. The incomplete information model is rejected at the .05 level, while the complete information model cannot be rejected. I compare the intervals from the complete information framework and the full model to see the impact of the complete information assumption within this application. Substantial differences between the two confidence sets indicate that the results of the complete information model are misleading if the complete information assumption does not hold.

As in most discrete choice models, it is difficult to interpret the parameter confidence intervals. The counterfactual calculations presented in the next section clarify the implications of the model. Nonetheless, the results indicate that the presence of a supercenter has a mild negative effect on

the value of a grocery store. In contrast, the entry cost is large and positive. The baseline effect of population on monopoly profits is monotonically increasing, while its effect on competition is ambiguous. The non-monotonic effect of population in the entry and exit probit regressions of Tables 5 and 6 may be due to the failure to control for the endogeneity of rival activity.

9.2 Counterfactual Experiments

Using the confidence region for the identifiable parameter, I construct bounds for counterfactual statistics. Such as the change in firm value from an exogenous change in market structure, or the difference in the expected number of grocery stores between markets of different sizes. These counterfactuals are functions of both the parameters of the model, θ , and the selection mechanism, λ . A very conservative method for deriving bounds would be to allow the selection mechanism to take any form in the counterfactual experiments. Instead, I restrict the bounds to the set of selection methods which fit the data using a criterion based on the likelihood function. The details of this procedure are presented in Appendix D.

9.2.1 Effects on Firm Valuations

I now consider the effect of changing market structure on firm valuations. For comparison purposes, I also present bounds based on the complete information model.³⁶ The bounds of the incomplete information model and the full model are not directly comparable because the assumptions of the incomplete information model are rejected. Because the incomplete information model is rejected by the full model, bounds based on the incomplete information model are not directly comparable to the full model. Firm valuations are the expected payoff from operating the firm before shocks are revealed. The formula for the expected value of the firm at the start of the period is,

³⁶Bounds for the complete information model are constructed by drawing from the asymptotic distribution of $\hat{\theta}$, this assumes that there is sufficient variation in covariates to point identify θ under the complete information restriction. This is the standard assumption used in the literature.

Parameter	Incomplete Info.		Complete Info.		Full Model	
	Identified Set	Identifiable Parameter	Identified Set	Identifiable Parameter	Identified Set	Identifiable Parameter
Monopoly Profits, $\mu(\cdot)$						
Constant	-1.177 (0.053)	-1.217 (0.053)	-1.615 -0.800]	[-1.493 -0.947]		
$\mathbf{1}[Pop > 3k]$	0.250 (0.056)	0.181 (0.057)	[-0.081 0.897]	[0.020 0.871]		
$\mathbf{1}[Pop > 6k]$	0.444 (0.064)	0.513 (0.057)	[0.140 1.467]	[0.329 1.228]		
$\mathbf{1}[Supercenter < 20mi]$	-0.052 (0.049)	-0.062 (0.051)	[-0.337 0.252]	[-0.272 0.186]		
Entry Cost, $\mathbf{1}[Inactive \text{ in } 1998]$	2.140 (0.034)	2.213 (0.036)	[1.543 2.512]	[1.657 2.391]		
Competition Effect, $\delta(\cdot)$						
Constant	-0.788 (0.120)	-0.557 (0.091)	[-2.065 -0.106]	[-1.531 -0.242]		
$\mathbf{1}[Pop > 3k]$	0.132 (0.122)	0.107 (0.098)	[-0.780 0.956]	[-0.339 0.823]		
$\mathbf{1}[Pop > 6k]$	0.392 (0.125)	0.120 (0.091)	[-0.586 0.969]	[-0.368 0.725]		
$\mathbf{1}[Supercenter < 20mi]$	-0.109 (0.081)	-0.056 (0.068)	[-0.638 0.336]	[-0.498 0.264]		
Public v. Private Info, σ_e^2	0.000	1.000	[0.293 1.000]	[0.398 1.000]		
Correlation of Public Info, ρ	n.a.	-0.132 (0.031)	[-0.990 0.833]	[-0.952 0.795]		
Log-Likelihood	-4283.4	-4168.6	-4218.8 to -4161.6			
In CR of Identified Set, $H_0 : \theta \in \Theta_I$	No	Yes				
In CR of Identifiable Parameter, $H_0 : \theta = \theta^*$	No	Yes				

Table 7: Preliminary results from the structural model. The incomplete information model has a unique equilibrium at the objective optimum, so results are reported assuming uniqueness. For the complete information model, I assume that a pure equilibrium is chosen with probability .5 to resolve multiplicity with this assumption the model is point identified. For the full, this table reports projections of the 95 percent joint confidence region onto the parameter axes. The true confidence region is a subset of the box reported in the table.

$$E[V_i(x; a)|x; \theta, \lambda] = \int \sum_{\epsilon \in \mathcal{E}(x, \epsilon)} \lambda^\epsilon(x, \epsilon) (\rho_i^\epsilon(x, \epsilon; \theta) (\mu(x) - \rho_{-i}^\epsilon(x, \epsilon)\delta(x) + E[\nu_i|\nu_i \geq \chi_i^\epsilon(x, \epsilon)] + \epsilon_i)) dF(\epsilon) \quad (17)$$

I numerically approximate (17) by simulating the distribution of ϵ . The object of interest is the relative change in firm value of moving a firm from state x to state x' , i.e., $\frac{E[V_i(x'; y)] - E[V_i(x; a)]}{E[V_i(x; y)]}$.

First, consider the effect of adding a supercenter in the vicinity of a market where none had existed. Bounds for the effect on firm value are presented in Table 8. The bounds indicate that supercenters may decrease expected long-run firm profits by up to 25 percent but may also generate up to a 10 percent increase in profits. The upper bound under the complete information model is substantially lower. The difference is a result of relaxing assumptions about the information structure. The upper bound is generated by parameters for which the net effect on $\mu(\cdot) + \delta(\cdot)$ is negative as a result of supercenter entry.³⁷ One might expect lower long-run profits as a result. However, the entry of a supercenter lowers the equilibrium probability of rival activity, an effect that is favorable to firms conditional on being active. In the complete information framework, firms can condition on their rival's public shock and avoid negative outcomes, so the benefits from less rival activity only appear on the margin. On the other hand, if the firms are uncertain about rival entry, a reduction in entry probabilities reduces the chance of simultaneous entry resulting in negative payoffs. Lowering the probability of negative profit outcomes can substantially benefit firms by alleviating their coordination problem. In contrast, the complete information model assumes the coordination problem away. Since negative profit outcomes never occur under the complete information assumptions, that model is unable to capture the full benefit of reducing rival entry when firms are uncertain.

Furthermore, in duopoly markets, the lower bound of the supercenter effect for the full model is substantially lower than the bound calculated using complete information assumptions. In the complete information case, firms are able to avoid negative profit outcomes, so the decrease in profits comes from there being fewer profitable opportunities and profits given those opportunities

³⁷Although the confidence set includes parameters where the net effect is slightly positive, these do not generate either the upper or lower bounds of the confidence intervals.

Effect of Supercenter Entry on Firm Value (Percent)		
	Complete Info.	Full Model
Pop 0-3k		
Monopoly	[-13.9 3.7]	[-22.7 11.4]
Duopoly	[-13.7 0.2]	[-23.9 5.6]
Pop 3k-6k		
Monopoly	[-12.6 3.0]	[-14.9 9.9]
Duopoly	[-13.5 -1.1]	[-18.9 7.2]
Pop 6k+		
Monopoly	[-10.9 2.3]	[-12.7 7.9]
Duopoly	[-12.4 -1.5]	[-20.3 4.7]

Table 8: Effect on firm value of grocery store if Supercenter enters within 20 miles (percent), 95 percent confidence intervals for the complete information model and the full model.

being less positive. In contrast, negative profit outcomes will occur if firms are uncertain. The full model includes the possibility that the harsher environment resulting from the supercenter entry makes profits even more negative in the event of ex-post regret. This result is particularly stark if firms begin the period as a duopoly because both firms are likely to choose to operate.

In sum, the pure strategy complete information model abstracts away from uncertainty about opponents. This ignores the coordination problem that firms face when entry by both will result in negative payoffs.³⁸ Allowing for uncertainty between firms systematically widens the bounds on the effect of supercenter entry. Since neither the data, nor economic intuition rules out the possibility of a substantial private component in player objective functions, the bounds using the complete information model are driven by prior assumptions.

Table 9 bounds the effect on firm value for a monopolist who experiences entry by a local rival. In contrast to the effect of supercenter entry, this effect is unambiguously negative. Moreover, it can be devastating. The lower bound is a decrease in expected profit in the range of 60 percent. When restricted to complete information, the lower bound is only a 25-30 percent loss. The difference is again that with complete information firms are able to avoid situations which result in negative profits. Under the complete information model, if two firms in the market is unprofitable, exactly

³⁸The coordination problem can arise due to the play of mixed strategies, however one might reason that mixed strategies would be avoided in this case because they result in lower expected profits for both players.

Effect of Local Entry on Firm Value (Percent)		
	Complete Info.	Full Model
Pop 0-3k		
Supercenter	[-29.8 -16.5]	[-75.3 -15.1]
No Supercenter	[-26.9 -13.1]	[-63.1 -12.4]
Pop 3k-6k		
Supercenter	[-27.0 -9.4]	[-60.2 -10.6]
No Supercenter	[-29.3 -14.7]	[-69.4 -9.6]
Pop 6k+		
Supercenter	[-25.2 -10.6]	[-66.0 -6.1]
No Supercenter	[-24.5 -8.1]	[-57.4 -5.2]

Table 9: Effect on firm value of monopolist grocery store of another grocery store entering its market (percent), 95 percent confidence intervals.

one will quickly exit, leaving the other to enjoy monopoly status. The effect of local competition on long run firm value is strongly negative under both the full model and the complete information model.

Tables 8 and 9 give the impression that entry by a local grocery store is more harmful than non-local supercenter entry. However, the bounds overlap and do not rule out the possibility that monopolists might prefer facing local competition to competition from a supercenter. I use the model to examine this question directly by computing bounds for the following statistic,

$$\frac{E[V_i|Monopoly, Supercenter] - E[V_i|Duopoly, NoSupercenter]}{E[V_i|Monopoly, NoSupercenter]}.$$

The sign of this statistic indicates whether a firm prefers to face a supercenter or local competition; its magnitude measures the strength of preference as a percentage of the firm value. Bounds for this statistic are presented in Table 10. In both small and large market categories, firms unambiguously prefer supercenter competition to local competition. The sign is technically ambiguous in medium sized markets, however results are again heavily tilted towards preferring supercenter competition. Differentiation on the basis of location and store type is effective at blunting the cost advantages of supercenters over local grocery stores.

Difference in Store Value between Supercenter versus Local Entry		
	Complete Info.	Full Model
Pop 0-3k	[5.0 28.9]	[4.2 63.8]
Pop 3k-6k	[2.6 25.7]	[-1.2 61.5]
Pop 6k+	[-1.1 24.8]	[1.5 55.3]

Table 10: The difference between the value of a store following supercenter entry versus local grocery store entry as a percentage of monopolist store value. A positive number implies that a monopolist would prefer supercenter entry to local grocery entry, a negative number implies the opposite, 95 percent confidence intervals.

9.2.2 Availability of Grocery Stores

The previous subsection analyzed the impact of supercenter entry on the prospects of local firms. I am also concerned with the effect of supercenters on consumers. Focusing on Wal-Mart, Hausman and Leibtag (2005) proposed a model whereby supercenter entry increases consumer welfare directly by offering consumers an additional choice and indirectly due to lower prices at all firms due to increased competition. However, if supercenter entry alters the choice set by inducing closure of local stores, some consumers will be worse off. Total consumer welfare could then fall due to supercenter entry. To determine the overall effect on consumer welfare, It is necessary to examine the extent to which supercenter entry endogenously reduces the set of choices available to consumers. This initial step is undertaken this section.

I assume that exogenous market characteristics, including the presence of a supercenter, are constant over time.³⁹ With this assumption, the structural model produces a Markov chain that governs transitions over the number of stores in the market. Table 11 presents the confidence bounds on the stationary distribution of this Markov chain by market type using the full model.⁴⁰ Across market sizes, it appears that the presence of a supercenter shifts the bounds on the proportion of unserved markets up, and the bounds on the proportion of markets served by two local firms down. While the bounds are wide, the effect of supercenters on the long run distribution of local grocery

³⁹This assumption is not needed for estimation. I estimate firms' expectations of their long run profits based on their entry and exit actions. Firms' expectations of how the exogenous variables will change in the future is accounted for in their expectations of long run profits.

⁴⁰The bounds from the complete information model are not presented in this section as they do not provide any additional insights beyond the results of the full model.

	Number of Grocery Stores		
	0	1	2
Pop 0-3k			
Supercenter	[27.9 66.8]	[33.0 70.7]	[0.1 6.1]
No Supercenter	[24.8 50.5]	[49.0 70.9]	[0.4 10.7]
Pop 3k-6k			
Supercenter	[14.1 39.0]	[57.5 80.4]	[1.6 13.0]
No Supercenter	[8.0 32.5]	[60.3 83.6]	[1.9 18.7]
Pop 6k+			
Supercenter	[3.0 13.5]	[57.4 81.9]	[15.0 34.2]
No Supercenter	[2.0 10.0]	[48.0 77.2]	[18.9 48.7]

Table 11: Stationary distribution of the number of firms in a market by market type (percent), 95 percent confidence intervals.

stores is seems mild.

Comparing these results with the observed distribution of stores across markets (Table 4), I find that the 2000 distribution of stores is well within the bounds of the steady state distribution. Note that this need not be the case, the results of Table 11 are derived from entry and exit patterns while those of Table 4 come from the static distribution of stores. This is consistent with the view that a major shift in the availability of local grocery stores is not underway.

An area of specific concern is that supercenters could cause a large increase in the number of markets that are unserved by local grocery stores, spawning “food deserts.” Table 11 provides upper bounds on the proportion of unserved markets. Market size is a much more important determinant of unserved markets than the presence of a supercenter, and the majority of markets will be served by at least one grocery store in large and medium markets. In small markets, where the proportion of unserved markets is already high (Table 4), the proportion of unserved markets can only be bound below two-thirds. It appears that the smallest markets are barely able to meet the minimum scale for even a single grocery store. The presence of a supercenter exacerbates the problem, but low population root issue. Restrictions on supercenters would not be likely to remove the threat of small markets becoming “food deserts.”

While Table 11 compares stationary distributions of local stores with and without supercenters nearby. Table 12 examines the effect of introducing a supercenter in the vicinity of the market on the stationary distribution. This table largely confirms that supercenter entry is likely to lead to a

	Number of Grocery Stores		
	0	1	2
Pop 0-3k	[-10.6 18.9]	[-18.6 10.5]	[-5.3 1.3]
Pop 3k-6k	[-6.7 14.5]	[-10.3 7.0]	[-12.6 2.0]
Pop 6k+	[-1.9 7.4]	[-4.9 23.7]	[-28.1 5.4]

Table 12: Effect of supercenters on stationary distribution of the number of grocery stores (percent), 95 percent confidence intervals.

	Expected Number of Grocery Stores	Supercenter Effect (Percent)
Pop 0-3k		
Supercenter	[0.33 0.76]	[-36.6 17.0]
No Supercenter	[0.50 0.82]	
Pop 3k-6k		
Supercenter	[0.65 0.94]	[-25.7 7.6]
No Supercenter	[0.74 1.03]	
Pop 6k+		
Supercenter	[1.04 1.27]	[-22.9 5.0]
No Supercenter	[1.13 1.45]	

Table 13: The expected number of stores by market size and presence of supercenter and the effect of adding a supercenter to the market on the expected number of stores, 95 percent confidence intervals.

decrease in the number of duopolies, although we cannot rule out the possibility that the number of duopolies increases slightly. This possibility arises because the data does not reject the possibility that supercenter entry can have a strong negative effect on baseline firm profits, $\mu(\cdot)$, but softens the competition effect between firms, $\delta(\cdot)$. If this is the case, the softening of competition leads to an increase in the proportion of duopoly markets.⁴¹ The effect of supercenters on the number of unserved markets is ambiguous, but the majority of the confidence set is positive. In areas of the identified set where supercenter entry corresponds to decreases in the number of unserved markets, the effect of supercenter entry on baseline profits is slightly positive, but the effect of entry on the competition effect is strongly negative.

While a full analysis of consumer welfare is outside the scope of this model, consumer welfare can be roughly related to the number of options the consumer has when shopping for groceries.

⁴¹For this part of the identified set, supercenter entry substantially reduces firm values, because of the overall negative effect on baseline profits and because firms are more likely to be in the duopoly state.

The expected number of stores in a market gives some indication of how much choice a typical consumer will have, and how this level of choice is affected by supercenter entry. Bounds on the expected number of stores in each market type are presented in Table 13 as well as bounds on the supercenter effect on the expected number of stores in a market in percentage terms. These results indicate a downward shift in the bounds on the number of local stores available to consumers as a result of a supercenter being nearby, although the bounds overlap. The interval for the supercenter effect is mostly negative. This echoes the descriptive results from Section 8 and other indicators in this section that the effect of supercenter entry mildly reduces the number of local stores. However, seems unlikely that this decrease in the expected number of local grocery options offsets the benefits from adding the supercenter option to consumers' choice sets. Of course, a full analysis of consumer welfare is outside the scope of this paper.

Allowing for a flexible information structure is important to the robustness of the counterfactual analysis presented in this section. The amount of uncertainty firms have about rivals is directly related to the amount of excess entry—when firms make negative profits, but consumers have more choices—and the number of underserved markets—where entry would be profitable but is avoided for fear of encountering a rival. The complete and incomplete information frameworks assign all variation in the data to either firm heterogeneity or firm uncertainty as a starting point of the analysis. In contrast, the full model generates counterfactual bounds based on allowing all mixtures of these sources of variation which are consistent with the observed data.

10 Conclusion

Earlier studies of discrete games have assumed that unobserved factors of the game are either publicly observed by all players (complete information) or known only to a single player (incomplete information). I provide a more general model, which nests these assumptions and parameterizes the extent to which unobservable elements of a firm's profit opportunities are publicly known. Using this model, information in the data can be used to make inference on the extent to which variation in firm actions is due to public or private information. The usual assumptions made by both the pure complete and pure incomplete information frameworks are testable using my model.

By using inference techniques that avoid point identification assumptions, I construct bounds on model statistics without imposing ad hoc assumptions on the information structure. Comparing my method to existing methods shows that assumptions about the information structure are not innocuous and they systematically constrict the confidence bounds for several statistics of interest.

The method presented in this paper can be used to conduct robust but non-trivial inference on questions of substantial public policy importance. The growth of the supercenter format has led industry observers to ask how supercenters alter the grocery market, and whether they are likely to displace local grocery stores. I analyze the impact of supercenters on small localized markets, using a flexible information structure. For grocery-store owners, I find that entry by a supercenter is far less detrimental than entry by a local competitor, and that if supercenter entry is effective at suppressing the probability of entry by a local challenger, it may actually increase long-run profits for incumbents. This outcome is the product of a reasonable economic model that fits the data well, but is artificially ruled out by the complete information framework. From the consumer's perspective, I find that supercenters likely cause a slight reduction in the expected number of grocery stores within a market, but that this effect is small when compared to other factors, such as the size of the market.

The empirical results show that placing strong assumptions on the information structure of a game has real consequences. The incomplete information framework is rejected when tested against the general model. While the complete information framework is not rejected, bounds produced using this framework are driven by ad hoc assumptions on the information structure. This leads the researcher to make overly strong conclusions. Moreover, the empirical exercise shows that inference techniques that are robust to partial identification can be used to derive meaningful confidence bounds for many statistics of interest.

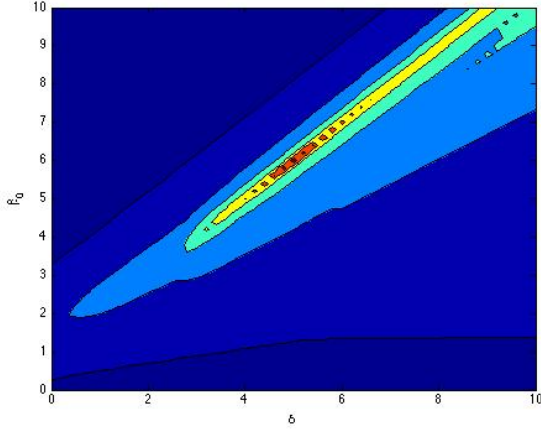
A Equilibrium Selection Assumptions and Misspecification

In Section 6.2.1, I have derived the identified set for an pure incomplete information model assuming a non-parametric equilibrium selection mechanism. In this appendix, I take use an incomplete information model to examine the effect of a stronger selection assumption frequently applied in the literature. Many papers in the incomplete information framework have assumed that a unique equilibrium is selected in all markets conditional on the observed variables Seim (2006); Vitorino (2008); Bajari, Hong, Krainer, and Nekipelov (2008); Bajari, Benkard, and Levin (2007) state this assumption or a stronger unique equilibrium assumption explicitly while it is sometimes left implicit in other work. With this assumption it is no longer necessary to treat the observed outcome distribution within a mixture model, and several authors have noted that this assumption can be employed to estimate an incomplete information model without solving for equilibrium (Aguirregabiria and Mira, 2007; Bajari, Benkard, and Levin, 2007; Bajari, Hong, Krainer, and Nekipelov, 2008). First, consider the case when equilibrium is unique everywhere, so the unique selection assumption is correct *a fortiori*. In this case $\#\mathcal{E}(x, \theta) = 1$, so the mixture model presented in this paper collapses and the model is locally point identified with or without the unique selection assumption. Second, consider the case of multiple equilibria, but unique selection is correct. In this case, the unique selection assumption will achieve local point identification, while the nonparametric selection model presented here may be partially identified. Finally, we consider the case when the assumption is incorrect. In this case the unique selection assumption is mis-specified.

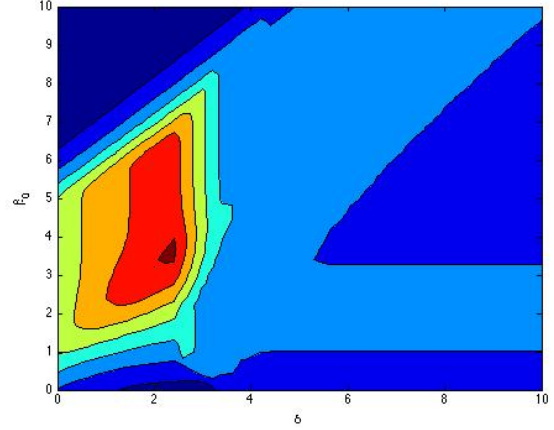
To test the extent of mis-specification, we construct a simple 2-firm model where firms are either high cost or low cost. Payoffs are specified as

$$\pi_i(a_i, a_{-i}; x_i, x_{-i}) = \mathbf{1}[a_i = 1](\beta_0 + \beta_1 \mathbf{1}[x_i = H] - \delta \mathbf{1}[a_{-i} = 1] + \nu_i)$$

Where we assume $\nu_i \sim N(0, 1)$. Within this model there are three market types (L, L) , (H, L) and (H, H) . We assume that actions are directly observed, so there are four possible outcomes $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$. Let the true parameter values be denoted $\theta^* = (\beta_0^*, \beta_1^*, \delta^*) = (6, -2, 5)$. With this parameterization there is a single equilibrium for the (L, L) , and (L, H) market types



(a) Nonparametric Selection, $Q_I(\cdot)$



(b) Unique Selection Conditional on Observables, $Q_U(\cdot)$

Figure 4: Contour map of the identification function for a 2-player model under nested selection assumptions. The identified set is the set of points for which this function equals zero. Under identification assuming a nonparametric selection mechanism (a), the identified set is the point corresponding to the true model. The identified set under unique selection (b) is empty and the optimum is inconsistent with the true model. See text for model specification. Contour regions are constructed using a log scaling.

and three equilibria for the (H, H) market type. We let the mixing distribution be $\lambda^* = (.4, .2, .4)$ where the equilibrium with weight .2 is the unstable but symmetric equilibrium, and the two with probability .4 are the stable “label switching” equilibria.

Consider the following function,

$$Q_I(\theta) = \sum_{x \in X} \alpha_x \min_{\lambda_x} \|P(y(\theta^*, \lambda^*)|x) - \lambda'_x P(x, \theta)\| \quad (18)$$

Where λ is appropriately restricted according to (8) and $P(y(\theta^*, \lambda^*)|x)$ is the observed outcome distribution given the true parameters and selection mechanism, α_x is some positive weighting and $\|\cdot\|$ is an appropriate vector norm. It is straightforward to see that the identified set (8) is equivalent to the set of values that set Q to zero. The we can rewrite identified set of our model as $\Theta_I = \{\theta \in \Theta : Q_I(\theta) = 0\}$.

We can now compute the identified set by computing Q across a grid of points in the parameter space. We find that the only point which sets Q_I equal to zero is θ^* , so the model is point identified.

A contour map of the minimum of Q_I in the (β_0, δ) plane is shown in Figure 4a. We use an even weighting across X and the square norm. Contour regions are drawn according to a log scaling. While Figure 4a depicts sharp curvature of the identified function at the optimum, we should note that this is dependent on the model, the chosen weighting, and the chosen norm.

Next, consider the identified set if we add the (incorrect) assumption of unique selection conditional on unobservables. We can construct the identified set under this assumption using the function,

$$Q_U(\theta) = \sum_{x \in X} \alpha_x \min_{e \in \mathcal{E}(x, \theta)} ||P(y(\theta^*, \lambda^*)|x) - P^e(x, \theta)|| \quad (19)$$

The identified set under the unique selection assumption is the set $\{\theta \in \Theta : Q_U(\theta) = 0\}$. Computing this function across the same grid as before, we find that this set is empty, indicating that the model is mis-specified. Since it is common for a structural model to be mis-specified, the usual practice is to take the minimizer of the identification function as the object of interest. I plot the contour map of Q_U in Figure 4b. This figure shows that not only is the minimizer of Q_S substantially different θ^* but θ^* appears unlikely to lie in the confidence region of the optimum of this function for substantial data sets.⁴² This finding should give pause to researchers considering the unique selection assumption without a strong indication that it in fact holds.⁴³

B Proofs

Theorem 1 If Assumptions 3 and 4 hold, then the parameters of the payoff function $(\theta_{i\mu}^*, \theta_{i\mu}, \theta_{i\delta})$ are point identified.

Proof. The proof is similar to Theorem 1 of Tamer (2003). Without loss of generality consider player 1's action, the argument for player 2 is symmetric. Player 1's best response function is,

$$\chi_i(\chi_{-i}, \epsilon, x) = -(x_i^* \theta_{i\mu}^* + x_i \theta_{i\mu} + \epsilon_i) + \rho_2(\chi_{-i}, \epsilon, x) (x_i \theta_{i\delta}) \quad (20)$$

⁴²This of course will depend on the amount of data available and the chosen weighting and norm. The key message from the figure is that there are a large set of models that appear to fit the data better than the true model when the unique selection assumption is enforced but incorrect.

⁴³A case where unique selection is intuitively more likely would be when data comes from the same agents repeating the game many times (although this may bring dynamic complications into the game). The assumption appears less credible to hold when we pool data across many markets.

Where ρ_2 is player i 's rational belief about the rival firm's probability of entry based on its given strategy χ_{-i} . This probability is derived according to (2).

Assume without loss of generality that $\theta_{2\mu} > 0$.⁴⁴ Then $\lim_{x_2^* \rightarrow -\infty} P(y_2 = 1|x, \epsilon) = \lim_{x_2^* \rightarrow -\infty} P(y_i = 1|x) = 0$, because from (20), $\chi_2 \rightarrow \infty$ as $x_2^* \rightarrow \infty$ when holding all other parameters fixed. Since by the Bayesian Nash equilibrium assumption, $\rho_2(\infty, \epsilon; \theta) = 0$, the probability of firm 1 entering is

$$\lim_{x_2^* \rightarrow -\infty} P(y_1 = 1|x) = E[P(\nu \geq \chi_1(\chi_2, \epsilon, x))] = P(\epsilon_i + \nu_i \geq -(x_1^* \theta_{1\mu}^* + x_1 \theta_{1\mu}))$$

By our scale assumption on the distribution of $\epsilon_i + \nu_i$, this is a simple linear probit model, so $(\theta_{i\mu}^*, \theta_{i\mu})$ is identified.⁴⁵ To identify the parameters of θ_δ , note that, $\lim_{x_2^* \rightarrow \infty} P(y_2 = 1|x, \epsilon) = \lim_{x_i^* \rightarrow \infty} P(y_2 = 1|x) = 1$, so we have

$$\lim_{x_2^* \rightarrow \infty} P(y_1 = 1) = P(\epsilon_i + \nu_i \geq -(x_1^* \theta_{1\mu}^* + x_1 \theta_{1\mu} + x_1 \theta_{1\delta}))$$

Since $(\theta_{1\mu}^*, \theta_{1\mu})$ are already identified by the preceding argument, we treat these parameters as known, the result is a linear probit model with a constant adjustment, so $\theta_{i\delta}$ is identified as well. \square

Theorem 2 If Assumptions 1 through 5 hold, the assumptions of the pure strategies complete information framework are testable.

Proof. The complete information assumption fixes σ_ϵ^2 at 1, so the only remaining parameters to identify are ρ and $\lambda(\cdot)$. Given the pure strategies assumption, player 1 knows y_2 with certainty when making his own entry decision (and vice versa). Given this, if we observe either both firms entering or neither firms entering, we can infer that the strategy was generated by a model with a unique equilibrium.⁴⁶

$$P(y_1 = 1, y_2 = 1, x; \theta) = \int_{\epsilon_1 = -(x_1^* \theta_{1\mu}^* + x_1 \theta_{1\mu} + x_1 \theta_{1\delta})}^{\infty} \int_{\epsilon_2 = -(x_2^* \theta_{2\mu}^* + x_2 \theta_{2\mu} + x_2 \theta_{1\delta})}^{\infty} dF(\epsilon_1, \epsilon_2; \rho)$$

⁴⁴The same argument with signs reversed can be carried out if $\theta_{2\mu} < 0$.

⁴⁵The argument is symmetric for player 2.

⁴⁶This argument assumes that $x_i \theta_{i\delta} < 0$ is zero for both firms, a similar argument holds with the sign reversed.

Since the parameters of the objective function are identified by Theorem 1, ρ is the only free parameter on the right hand side. Because this expression is monotonically increasing in ρ , ρ is identified by observing the left hand side for a single market. After identifying ρ from a single market, $P(y_1 = 1, y_2 = 1|x)$ is known for all x , since it is independent of the selection mechanism. Therefore can test the assumptions of the pure strategy complete information model by checking to see whether $P(y_1 = 1, y_2 = 1|x)$ implied by the pure strategies complete information model is consistent with the observed distribution across all markets. \square

Theorem 3 If Assumptions 1 through 5 hold, the assumptions of the pure incomplete information framework are testable.

Proof. Under the assumptions for the incomplete information model, $\sigma_\epsilon = 0$ by assumption, so ρ drops out of the model. This implies that player's observed actions are independent conditional on the observed covariates,

$$y_1 \perp y_2 | x_1^*, x_1, x_2^*, x_2$$

We can use the data to test whether or not this restriction holds. \square

Theorem 4 If assumptions 1-3, and 5 hold. The sharp identified set of θ for the incomplete information model is,

$$\Theta_I = \left\{ \theta \in \Theta : \forall x \in X, \exists \tilde{\lambda} \in [0, 1]^{\bar{E}} \text{ s.t. } \begin{aligned} P(\cdot|x) &= \sum_{e \in \mathcal{E}(x;\theta)} \tilde{\lambda}^e P^e(x, \theta), \\ \sum_{e \in \mathcal{E}(x;\theta)} \tilde{\lambda}^e &= 1 \end{aligned} \right\} \quad (21)$$

Where \bar{E} is a constant which represents the largest possible number of equilibria the model admits almost everywhere over $X \times \Theta$.

Proof. We are only interested in λ at the true value of θ , and we have assumed that $P(y|x)$ is observed for identification purposes. Therefore, we treat λ as a restricted nuisance parameter. Given θ this parameter essentially indexes the set of possible observed distributions that are consistent with the model. The conditional outcome vector $P(y|x)$ describes all the restrictions on the model

we observe from the data. By definition, $\theta \in \Theta_I$ if and only if it can be paired with some valid mixture $\tilde{\lambda}$ across equilibrium outcomes to generate the outcome distribution. \square

Theorem 5 If assumptions 1-3, and 5 hold. The identified set of the full model is,

$$\Theta_I = \left\{ \begin{array}{l} \theta \in \Theta : \forall x \in X, \exists \lambda_x(\epsilon) \in [0, 1]^{\bar{E}} \text{ s.t. } P(y|x) = \int \sum_{e \in \mathcal{E}(\epsilon, x, \theta)} \lambda_x^e(\epsilon) P_y^e(\epsilon, x, \theta) dF(\epsilon), \\ \forall \epsilon, x : \sum_{e \in \mathcal{E}(\epsilon, x, \theta)} \lambda_x^e(\epsilon) = 1 \end{array} \right\} \quad (22)$$

Proof. The proof of this theorem is a trivial extension of the argument from Theorem 4. \square

Lemma 1 (i) $L(\theta) = \operatorname{argmax}_{\theta' \in \Theta} L(\theta')$ for all $\theta \in \Theta_I$ as defined by (11), and $L(\theta) > \operatorname{argmax}_{\theta' \in \Theta} L(\theta')$ for all $\theta \in \Theta \setminus \Theta_I$, and (ii) $L(\theta)$ is continuous in θ .

Proof. The proof of (i) trivially follows from the definition of Θ_I . To see (ii), define the set of probability distributions that are consistent with a given set of model parameters θ as,

$$\mathcal{P}(\theta) = \left\{ \begin{array}{l} P(\cdot|\cdot) \in \mathcal{P} : \forall x \in X, \exists \lambda_x(\epsilon) \in [0, 1]^{\bar{E}} \text{ s.t. } P(y|x) = \int \sum_{e \in \mathcal{E}(\epsilon, x, \theta)} \lambda_x^e(\epsilon) P_y^e(\epsilon, x, \theta) dF(\epsilon), \\ \forall \epsilon, x : \sum_{e \in \mathcal{E}(\epsilon, x, \theta)} \lambda_x^e(\epsilon) = 1 \end{array} \right\} \quad (23)$$

It is clear that a discontinuity in $L(\theta)$ implies a discontinuity in $\mathcal{P}(\theta)$, so we show that $\mathcal{P}(\theta)$ is continuous in θ . For each point in $P \in \mathcal{P}(\theta)$ there exists a selection mechanism λ which produces the outcome distribution P . Consider a differential change in θ to θ' . At every regular point (x, ϵ) on the equilibrium correspondence, the cardinality of the equilibrium sets $\mathcal{E}(x, \epsilon, \theta)$ and $\mathcal{E}(x, \epsilon, \theta')$ is the same, so the same selection mechanism can be used. Irregular points on the correspondence are a set of measure zero and can be ignored. Now consider each individual regular equilibrium, by the implicit function theorem there is an equilibrium in the set $\mathcal{E}(x, \epsilon, \theta')$ in the neighborhood of every regular equilibrium on the set $\mathcal{E}(x, \epsilon, \theta)$. This follows from continuity of the equilibrium conditions in θ and continuity of the bivariate normal distribution in its parameters. By using the selection mechanism $\lambda(x, \epsilon)$ and matching the equilibria of $\mathcal{E}(x, \epsilon, \theta')$ to their “nearby” points on $\mathcal{E}(x, \epsilon, \theta)$. The resulting outcome distribution P' must be in the neighborhood of P . Since this can be accomplished for any $P \in \mathcal{P}$ and for any $\theta \in \Theta$, continuity of $\mathcal{P}(\cdot)$ and $L(\cdot)$ follows. \square

Theorem 6 For all $\theta \in \Theta_I$, $\max_{\theta' \in \Theta} L_n(\theta') - L_n(\theta) = o_p(1)$.

Proof. We have,

$$\max_{\theta' \in \Theta} L_n(\theta') - L_n(\theta) \rightarrow_p \max_{\theta' \in \Theta} L(\theta') - L(\theta) = 0$$

Where the probability limit follows from uniform convergence of $L_n(\cdot)$ to $L(\cdot)$ and the equality follows from Lemma 1 and $\theta \in \Theta_I$. □

C Computational Appendix

C.1 Computing the Equilibrium Set

In this section I present a method to approximate all equilibrium in the set $\mathcal{E}(\epsilon, x, \theta)$. For simplicity, we suppress covariates in this section and treat μ and δ as constants. Moreover, because we deal with each ϵ draw independently, the notation of this appendix suppresses the dependence of the strategies (χ_1, χ_2) and beliefs (ρ_1, ρ_2) on x and ϵ .

The set of equilibria is equivalent to the set of all solutions to the system of equations,

$$\begin{aligned}\chi_1 &= -(\mu_1(\theta) + \epsilon_1) + \rho_2(\chi_2; \theta)\delta_1(\theta) \\ \chi_2 &= -(\mu_2(\theta) + \epsilon_2) + \rho_1(\chi_1; \theta)\delta_2(\theta)\end{aligned}\tag{24}$$

Where χ_i is a cutoff strategy for entry and ρ_i is agent i 's probability of entry based given he is using the strategy χ_i .

$$\rho_i(\chi; \theta) = \int \mathbf{1}[\nu_i \geq \chi] dF(\nu_i; \theta)$$

We need only search for equilibrium strategies within the set of rationalizable strategies. Because agents beliefs about the probability of entry are bounded between 0 and 1, the set of rationalizable strategies for player i is,

$$\Psi_i = [\min(-(\mu_i(\theta) + \epsilon_i), -(\mu_i(\theta) + \epsilon_i) + \delta_i(\theta)), \max(-(\mu_i(\theta) + \epsilon_i), -(\mu_i(\theta) + \epsilon_i) + \delta_i(\theta))]$$

Since equilibrium strategies must be rationalizable, we can confine our search for equilibrium cutoffs

for player i to Ψ_i . We then search for equilibria using the following algorithm:

1. For a grid of points $\chi_1^p \in \Psi_1$:

(a) Compute player 2's best response given player 1 uses the strategy χ_1^p ,

$$\chi_2^{bp} = -(\mu_2(\theta) + \epsilon_2) + \rho_1(\chi_1^p; \theta)\delta_2(\theta)$$

(b) Compute player 1's best response given player 2 uses the strategy χ_2^{bp}

$$\chi_1^{bp} = -(\mu_1(\theta) + \epsilon_1) + \rho_2(\chi_2^{bp}; \theta)\delta_1(\theta)$$

(c) Compute $z^p = \chi_1^p - \chi_1^{bp}$

2. Wherever z^p and z^{p+1} are opposite signs, use Newton's method starting (χ_1^b, χ_2^{bp}) to solve (24).

3. If $|z^p| < |z^{p-1}|$ and $|z^p| < |z^{p+1}|$, use Newton's method starting (χ_1^b, χ_2^{bp}) to solve (24).

If (χ_1^b, χ_2^{bp}) is an equilibrium, then $z^p = 0$. The vector of points $\{z^p\}$ is a discretization of a continuous function. The algorithm locates equilibria by searching near the zeros of this function, which is much more efficient than simple multi-starting. Finding all equilibria depends on using a fine enough discretization of the rationalizable set. Clearly, there is a tradeoff between accuracy and computation time. The results of this paper are robust to changing the coarseness of the discretization of the rationalizable set.

C.2 Numerical Approximation of the Profiled Likelihood Function

This appendix provides details on the optimization problem which is used to calculate $L_n(\theta)$. Given a value for λ , the maximand $L_n(\theta, \lambda)$ is calculated using standard numerical simulation techniques.

We can write the maximand as,

$$L_n(\theta, \lambda) = \frac{1}{n} \sum_{i=1}^n [\log(P_{y_i}^R(x_i, \theta, \lambda))]$$

where,

$$P_y^R(x, \theta, \lambda) = \frac{1}{R} \sum_{r=1}^R \sum_{e \in \mathcal{E}(\epsilon_r, x, \theta)} \lambda^e(\epsilon_r, x) P_{y_i}^e(\epsilon_r, x, \theta) f(\epsilon_r; \theta)$$

Where we have selected R sample points over which to approximate the integral over ϵ . These sample points could be chosen in several different ways, including monte carlo simulation. I have chosen to use Halton sequences to approximate this integral.

Our task is to profile λ out of this function. To accomplish this, we need to find a "most favorable" selection mechanism given θ . Let λ_θ be any element of the set of maximizers of the likelihood for a fixed θ .

$$\lambda_\theta \in \operatorname{argmax}_{\lambda(\cdot)} L_n(\theta, \lambda(\cdot))$$

It is clear that all λ which are equal on the sample points chosen for the numerical approximation of P^R evaluate to the same likelihood, so λ_θ is not be uniquely defined. However, maximizing λ over the set of sample points will yield an appropriate value for the purpose of approximating $L(\theta)$.

For each sample point I calculate the equilibrium set $\mathcal{E}(\epsilon_r, x, \theta)$. For each equilibrium, I assign each an index e , and a mixing probability $\lambda_{r,x,e}$. We then optimize the following constrained maximization problem over the vector of mixing probabilities,

$$\begin{aligned} \lambda_\theta^R = \operatorname{argmax}_{\lambda \in [0,1]^{R \times X \times E}} \sum_{i=1}^N \sum_{y \in Y} \mathbf{1}[y_i = y] \log \left(R^{-1} \sum_{r=1}^R \sum_{e \in \mathcal{E}(\epsilon_r, x_i, \theta)} \lambda_{r,x_i,e} P_y^e(\epsilon_r, x_i, \theta) f(\epsilon_r; \theta) \right) \quad (25) \\ \text{s.t. } \forall r, x : \sum_{e \in \mathcal{E}(\epsilon_r, x, \theta)} \lambda_{r,x,e} = 1 \end{aligned}$$

While high-dimensional and somewhat daunting in appearance, the optimization problem in (25) is a concave objective with linear constraints, and can be handled by modern nonlinear solvers for R in the hundreds. Furthermore, whenever $\mathcal{E}(\epsilon_r, x, \theta)$ is unique, the selection mechanism is degenerate and there is no need to optimize over the selection mechanism. This leads to a dramatic reduction in the number of unknowns in this problem in many cases. Efficient computation of (25) is important because this problem must be solved for each θ we wish to test during simulated annealing.

We can now evaluate the profiled likelihood statistic for each θ by plugging λ_θ^R back into the full likelihood function.

$$\mathcal{L}_N(\theta) = L_N(\theta, \lambda_\theta^R)$$

C.3 Weighted Bootstrap Algorithm for Confidence Sets

This appendix describes the implementation of the weighted bootstrap to derive the confidence region for the identified set and the confidence region for the identifiable parameter. Let the objective function be defined as

$$L_n(\theta) = \max_{\lambda} n^{-1} \sum_{i=1}^n \log p(z_i, \theta, \lambda)$$

The *weighted* likelihood function is a function of weights $w = (w_1, \dots, w_n)$.

$$L_n(\theta, w) = \max_{\lambda} n^{-1} \sum_{i=1}^n w_i \log p(z_i, \theta, \lambda)$$

The key observation is that, under appropriate distributional assumptions on w , the weighted likelihood will have the same asymptotic distribution as the “standard” likelihood. Bootstrapping the weighted likelihood amounts to evaluating the function for different sets of weights. The quantiles found by bootstrapping the weighted likelihood will approximate the quantiles of the standard likelihood.

First, normalize the likelihood function at its maximum, to get a set objective function in the form of CHT,

$$Q_n(\theta) = \max_{\theta'} L_n(\theta') - L_n(\theta)$$

$$Q_n(\theta, w) = \max_{\theta'} L_n(\theta', w) - L_n(\theta, w)$$

For the weighted bootstrap procedure, draw $B = 200$ sets of weights where $E[w_i^b] = 1$ and $V[w_i^b] = 1$ which are independent of the data. Since the weighted likelihood and the unweighted likelihood have the same asymptotic distribution we can use the quantiles of $\{Q_n(\theta, w^b)\}_{b=1}^B$ to estimate the cutoff for the confidence region. Ideally, we would use all points in Θ for this procedure,

however this is clearly computationally infeasible. Instead we will use simulated annealing to select a large number of points which adequately cover the parameter space near its minimum. Some tuning of the jump distance and the temperature of the simulated annealing algorithm may be needed to ensure adequate coverage.

1. From multiple (25) start points, run the simulated annealing algorithm on $Q_n(\cdot)$ for many (10000) iterations each. Save all points.
2. Define the starting cutoff c_0 and the starting set of points $\mathcal{S}_0 = \{\theta : nQ_n(\theta) \leq c_0\}$. In practice we will use the set of points S_0 which are in \mathcal{S}_0 and have been visited by the simulated annealing. The starting cutoff must be decreasing in n at a slow enough rate a la CHT. An extreme alternative is to let $c_0 = \infty$, which implies $\mathcal{S}_0 = \Theta$ and S is simply all points visited by simulated annealing.
3. For each point in S_0 , compute $\{Q_n(\cdot, w^b)\}$ note that we only need to solve the model 1 time for each θ and then can compute the likelihood for each weight sample.
4. Iterate the following until $|c_{\ell-1} - c_\ell| < \varepsilon$

(a) Compute:

$$c_{\ell+1} = \inf\left\{x : \frac{1}{B} \sum_{b=1}^B \mathbf{1}[\max_{\theta \in S_\ell} nQ_n(\theta, w^b) \leq x] \geq 1 - \alpha\right\}$$

(b) Define:

$$S_{\ell+1} = \{\theta \in S_\ell : nQ_n(\theta) \leq c_{\ell+1}\}$$

5. Let $\hat{c} = c_\ell$, $\hat{\Theta}_I = \{\theta : nQ_n(\theta) \leq \hat{c}\}$. Use S_ℓ to construct confidence intervals for statistics of interest. Such confidence intervals are valid assuming the number of simulated annealing iterations goes to ∞ with n .

To find the confidence set for the identifiable parameter we individually test the hypothesis that $\theta \in \Theta_I$ for each point. The collection of all points that are not rejected is denoted $\bar{\Theta}$, the confidence set for the identifiable parameter. It is easy to see that any point in $\bar{\Theta}$ must also be in $\hat{\Theta}_I$ —if we cannot reject that θ is the true parameter, we clearly cannot exclude it from our coverage

region for the identified set. Therefore, we can restrict the hypothesis test to all points in $\hat{\Theta}_I$. The test is conducted for a given θ by computing,

$$c(\theta) = \inf\left\{x : \frac{1}{B} \sum_{b=1}^B \mathbf{1}[nQ_n(\theta, w^b) \leq x] \geq 1 - \alpha\right\}$$

The hypothesis is rejected if $nQ_n(\theta) > c(\theta)$. So the confidence set for the identifiable parameter is,

$$\bar{\Theta} = \{\theta \in \hat{\Theta}_I : nQ_n(\theta) \leq c(\theta)\}$$

Again, we use the set of points visited by simulated annealing which pass the above condition to construct confidence intervals for the statistics of interest.

D Selection and Counterfactual Bounds

This appendix outlines the procedure used to produce the counterfactual bounds for the full model presented in Section 9. I am interested in counterfactual statistics such as entrepreneurs' willingness to pay for a firm given observable characteristics at the start of the period, or the effect of placing a supercenter near a market where one does not exist. Since the model is only identified up to scale, I concentrate on comparisons between two observable markets characterized by x and x' . Counterfactuals are described by a smooth function $f(x, x', \theta_0, \lambda_0)$, where (θ_0, λ_0) are the true parameters.⁴⁷

To produce confidence intervals for a statistic $f(x, x', \theta_0, \lambda_0)$, I can draw model parameters from the identified set $\hat{\Theta}$. However, determining the selection mechanisms to use in counterfactuals is less straightforward. The most conservative method would be to derive bounds assuming only that some valid selection mechanism is used. Under this procedure the bounds on the effect of moving

⁴⁷In this paper, I do not consider policy experiments, which would be the reaction of firms to a change in the true parameters from θ_0 to some other θ . A procedure very similar to the one presented here could be employed to bound the marginal effect of a change in a policy parameter.

from state x to x' are,

$$\left[\inf_{\theta \in \hat{\Theta}} \inf_{\lambda \in \Lambda(\theta)} f(x, x', \theta, \lambda), \sup_{\theta \in \hat{\Theta}} \sup_{\lambda \in \Lambda(\theta)} f(x, x', \theta, \lambda) \right]$$

These bounds are too conservative in practice for two reasons. First, they conflate the uncertainty about the model parameters and a change in the selection mechanism from its most to its least favorable outcome from the perspective of $f(\cdot)$. Within the confidence set, some market types result in multiple equilibria with a probability of more than .45.⁴⁸ With such a high degree of multiplicity, the effect of dramatic changes in the selection mechanism will swamp information in the model about $f(\cdot)$. Second, fixing $\theta \in \Theta_I$, it may be that only a small subset of the available selection mechanisms in $\Lambda(\theta)$ actually fit the observed data. Intuitively, the researcher could draw from $\Lambda(\theta)$, check how well the proposed model fits the data by computing its likelihood, and then construct bounds using only those selection mechanisms which are plausible given the data according to the likelihood function. In other words for each $\theta \in \Theta_I$, the researcher constructs the set⁴⁹

$$\tilde{\Lambda}(\theta) = \{\lambda \in \Lambda(\theta) : \sup_{\theta' \in \Theta} L_N(\theta') - L_N(\theta, \lambda) < \kappa_N\}$$

and then report as the bounds,

$$\left[\inf_{\theta \in \hat{\Theta}} \inf_{\lambda \in \tilde{\Lambda}(\theta)} f(x, x', \theta, \lambda), \sup_{\theta \in \hat{\Theta}} \sup_{\lambda \in \tilde{\Lambda}(\theta)} f(x, x', \theta, \lambda) \right]$$

The theoretical difficulty with this method lies in the choice of κ_N . If $\kappa_N \rightarrow 0$ as $N \rightarrow \infty$, then $\tilde{\Lambda}(\theta)$ is a consistent estimator for the set of selection functions that maximize the likelihood when the model is specified by θ . However, theory does not provide a guide on how to choose a κ_N that will ensure the desired coverage probability. Instead, I assume that the rate at which κ_N goes to zero is slow enough that the probability of a false rejection of the selection mechanism is asymptotically negligible. In practice, I choose κ_N to be higher than the \hat{c} cutoff which defines the joint confidence

⁴⁸Formally, $\sup_{\theta \in \hat{\Theta}_I} \max_x \int \mathbf{1}[\#\mathcal{E}(x, \epsilon; \theta) > 1] dF(\epsilon) \geq .45$

⁴⁹I use two methods to draw selection mechanisms from this set, I both randomly perturb the selection mechanism that maximizes $L_n(\cdot)$ and also independently draw multinomial distributions at each evaluation point of the selection mechanism.

set for Θ_I . The resulting bounds, which I present below are insensitive to large perturbations in the choice of κ_N .

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 P. MIRA (2007): “Sequential Estimation of Dynamic Discrete Games,” *Econometrica*, 75(1), 1–53.
- ANDREWS, D. W., AND G. SOARES (2007): “Inference for Parameters defined by Moment Inequalities using Generalized Moment Selection,” Cowles Foundation Discussion Paper No. 1631.
- ARADILLAS-LOPEZ, A. (2007): “Semiparametric Estimation of a Simultaneous Game with Incomplete Information,” Princeton University.
- AUGEREAU, A., S. GREENSTEIN, AND M. RYSMAN (2006): “Coordination versus Differentiation in a Standards War: 56K Modems,” *The RAND Journal of Economics*, 37(4), 887–909.
- BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): “Estimating dynamic models of imperfect competition,” *Econometrica*, 75(5), 1331–1370.
- BAJARI, P., J. HAHN, H. HONG, AND G. RIDDER (2008): “A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models,” U of Minn., UCLA, Stanford, and USC.
- BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2008): “Estimating Static Models of Strategic Interactions,” NBER Working Paper 12013.
- BAJARI, P., H. HONG, AND S. RYAN (2007): “Identification and Estimation of a Discrete Game of Complete Information,” U of Minnesota, Stanford, and MIT.
- BASKER, E., AND M. D. NOEL (2008): “The Evolving Food Chain: Competitive Effects of Wal-Mart’s Entry into the Supermarket Industry,” University of Missouri and University of California San Diego.
- BERESTEANU, A., AND P. B. ELLICKSON (2006): “The Dynamics of Retail Oligopoly,” Duke University.
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2008): “Sharp Identification Regions in Games,” CEMMAP Working Paper CWP 15/08.
- BERRY, S. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(4), 889–917.
- (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25(2), 242–262.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841–890.
- BHAT, C. R. (2001): “Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model,” *Transportation Research Part B: Methodological*, 35(7), 677 – 693.

- BJORN, P., AND Q. VUONG (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," California Institute of Technology.
- BRESNAHAN, T. F., AND P. REISS (1991a): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48(1), 57–81.
- (1991b): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99, 977–1009.
- BRESNAHAN, T. F., AND P. C. REISS (1990): "Entry in Monopoly Markets," *The Review of Economic Studies*, 57(4), 531–553.
- BUGNI, F. A. (2007): "Bootstrap Inference in Partially Identified Models," Northwestern University.
- CANAY, I. A. (2008): "EL Inference for Partially Identified Models: Large Deviations and Bootstrap Validity," Northwestern University.
- CHEN, X., AND E. TAMER (2009): "Inference in Partially Identified Semiparametric Models," Yale University and Northwestern University (in progress).
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(6), 1243–1285.
- CILIBERTO, F., AND E. TAMER (2007): "Market Structure and Multiple Equilibria in Airline Markets," University of Virginia and Northwestern University.
- ELLICKSON, P. B. (2007): "Does Sutton Apply to Supermarkets?," *RAND Journal of Economics*, 38(1), 43–59.
- GARCIA, C. B., AND W. I. ZANGWILL (1979): "An Approach to Homotopy and Degree Theory," *Mathematics of Operations Research*, 4(4), 390–405.
- GRIECO, P. L. E. (2008): "Numerical Approximation of an Equilibrium Correspondence: Method and Applications," Northwestern University.
- HARSANYI, J. C. (1967): "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model," *Management Science*, 14(3), 159–182.
- (1973): "Games with randomly distributed payoffs: A new rationale for mixed strategy equilibria," *International Journal of Game Theory*, 2(1), 1–23.
- HAUSMAN, J., AND E. LEIBTAG (2005): "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart," NBER Working Paper 11809.
- HECKMAN, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46(4), 931–959.
- HO, K. (2008): "Insurer-Provided Networks in the Medical Care Market," Columbia University.

- HOLMES, T. J. (2008): “The Diffusion of Wal-Mart and the Economies of Density,” NBER Working Paper 13783.
- HONORÉ, B. E., AND E. TAMER (2006): “Bounds on Parameters in Panel Dynamic Discrete Choice Models,” *Econometrica*, 74(3), 611–629.
- INAGAMI, S., D. A. COHEN, B. K. FINCH, AND S. M. ASCH (2006): “You are Where You Shop: Grocery Store Location, Weight, and Neighborhoods,” *American Journal of Preventative Medicine*, 31(1), 10–17.
- ISHII, J. (2005): “Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry,” Stanford University.
- JIA, P. (2008): “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, 76(6), 1263–1316, MIT, forthcoming *Econometrica*.
- KAJII, A., AND S. MORRIS (1997): “The Robustness of Equilibria to Incomplete Information,” *Econometrica*, 65(6), 1283–1309.
- KALBALA, R., AND L. TEFATSION (1991): “Solving Nonlinear Equations by Adaptive Homotopy Continuation,” *Applied Mathematics and Computation*, 41, 99–115.
- LIU, X., AND Y. SHAO (2003): “Asymptotics for Likelihood Ratio Tests under Loss of Identifiability,” *The Annals of Statistics*, 31(3), 807–832.
- MA, S., AND M. R. KOSOROK (2005): “Robust semiparametric M-estimation and the weighted bootstrap,” *Journal of Multivariate Analysis*, 96(1), 190 – 217.
- MARTINEZ, S. W. (2007): “The US Food Marketing System: Recent Developments 1997-2006,” USDA Economic Research Service Report 42.
- MATSA, D. A. (2009): “Competition and Product Quality in the Supermarket Industry,” Kellogg School of Management.
- MAZZEO, M. (2002): “Product Choice and Oligopoly Market Structure,” *RAND Journal of Economics*, 33(2), 1–22.
- MILGROM, P. R., AND R. J. WEBER (1985): “Distributional Strategies for Games with Incomplete Information,” *Mathematics of Operations Research*, 10(4), 619–632.
- MORRIS, S., AND H. S. SHIN (2000): “Rethinking Multiple Equilibria in Macroeconomic Modeling,” *NBER Macroeconomics Annual*, 15, 139–161.
- ORHUN, A. Y. (2005): “Spatial Differentiation in the Supermarket Industry,” UC Berkeley.
- PAKES, A., J. PORTER, K. HO, AND J. ISHII (2006): “Moment Inequalities and their Application,” Harvard University and University of Wisconsin.
- ROMANO, J. P., AND A. M. SHAIKH (2006): “Inference for the Identified Set in Partially Identified Econometric Models,” Stanford University and University of Chicago.

- (2008): “Inference for identifiable parameters in partially identified econometric models,” *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.
- SEIM, K. (2001): “Spatial Differentiation and Market Structure in the Video Retail Industry,” Ph.D. thesis, Yale University, New Haven, CT.
- (2006): “An empirical model of firm entry with endogenous product-type choices,” *RAND Journal of Economics*, 37(3), 619–640.
- SINGH, V. P., K. T. HANSEN, AND R. C. BLATTBERG (2006): “Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter,” *Marketing Science*, 25(5), 457–476.
- SWEETING, A. (2007): “Coordination Games, Multiple Equilibria, and the Timing of Radio Commercials,” Duke University.
- TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, pp. 147–165.
- VITORINO, M. (2008): “Empirical Entry Games with Complementarities: An Application to the Shopping Center Industry,” University of Chicago.
- WILSON, R. (1971): “Computing Equilibria of N-Person Games,” *SIAM Journal on Applied Mathematics*, 21(1), 80–87.