Quantitative Economics 7 (2016), 727–780 1759-7331/20160727

Estimation of games with ordered actions: An application to chain-store entry

Andres Aradillas-López
Department of Economics, Pennsylvania State University

Amit Gandhi
Department of Economics, University of Wisconsin–Madison

We study the estimation of static games where players are allowed to have ordered actions, such as the number of stores to enter into a market. Assuming that payoff functions satisfy general shape restrictions, we show that equilibrium of the game implies a covariance restriction between each player's action and a component of the player's payoff function that we call the *strategic index*. The strategic index captures the direction of strategic interaction (i.e., patterns of substitutability or complementarity) as well as the relative effects of opponents' decisions on players' payoffs. The covariance restriction we derive is robust to the presence of multiple equilibria, and provides a basis for identification and estimation of the strategic index. We introduce an econometric method for inference in our model that exploits the information in moment inequalities in a computationally simple way. We analyze its properties through Monte Carlo experiments and then apply our approach to study entry behavior by chain stores where there is both an intensive margin of entry (how many stores to open in a market) as well as the usual extensive margin of entry (whether to enter a market or not). Using data from retail pharmacies we find evidence of asymmetries in strategic effects among firms in the industry that has implications for merger policy. We also find that business stealing effects are less pronounced in larger markets, which helps explain the large positive correlation in entry behavior observed in the data.

Keywords. Static games, multiple equilibria, partial identification, conditional moment inequalities, entry decisions.

JEL classification. C01, C14, C57.

1. Introduction

The econometric analysis and applications of static games has been an increasingly active area of research in the recent past. A partial list of papers would include Bjorn and Vuong (1984), Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Berry (1992), Tamer (2003), Seim (2006), Davis (2006), Berry and Tamer (2006), Pesendorfer and Schmidt-Dengler (2008), Sweeting (2009), Aradillas-Lopez (2010),

Andres Aradillas-López: aaradill@psu.edu
Amit Gandhi: agandhi@ssc.wisc.edu

DOI: 10.3982/QE465

Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2011), Bajari, Hong, Kreiner, and Nekipelov (2009), Bajari, Hong, and Ryan (2005), Ciliberto and Tamer (2009), Kline and Tamer (2010), Gowrisankaran and Krainer (2011), Aradillas-Lopez (2011), De Paula and Tang (2012), Lewbel and Tang (2012), and Grieco (2012). Most of the existing econometric work on static games has been characterized by at least one of two features: (i) a full parameterization of payoff functions with fairly limited forms of strategic effects (e.g., constant strategic effects), and (ii) a limited strategy space, with binary choice games being the most common example. One of the major difficulties with using richer models of strategic interaction in empirical work is that the multiplicity of equilibria can complicate the use of methods that require computing the equilibria in the game. Furthermore, even inferential approaches that rely solely on necessary conditions in equilibrium could also become impractical because characterizing such conditions can be difficult if the game has a rich strategy space.

In this paper we study static games with a rich, possibly unbounded, strategy space that is only required to be ordered in nature (and can be discrete or continuous). Players' payoffs are left nonparametrically specified except for a component that summarizes the strategic interaction effect. This "strategic index" captures the direction of strategic interaction (i.e., patterns of substitutability or complementarity) as well as the relative effects of opponents' strategies on players' payoffs and the potentially continuous variation in these effects with observable covariates (i.e., market size, demographics, etc.) in an empirically flexible way. Instead of fully parameterizing payoff functions we only impose weak shape restrictions on payoffs that are motivated by economic theory. Our main result is showing that these shape restrictions alone are sufficient for doing inference on the strategic index in a way that is fully robust to the presence of multiple equilibria.

Our paper is motivated by the work in De Paula and Tang (2012), who show how to exploit the presence of multiple equilibria[1] to identify the direction of strategic interaction in binary choice games in incomplete information games. Like De Paula and Tang (2012), the object of interest here is an aggregate index that summarizes the strategic interaction effect for each player. We extend the seminal results in De Paula and Tang (2012) in several important ways. First, we go beyond binary choice games and consider games with rich—possibly unbounded—strategy spaces, as long as such spaces satisfy an ordinal property. Second, we impose much fewer restrictions on our strategic interaction functions (indices). For example, we allow for asymmetries in strategic interaction effects across opponents in a more general way. While De Paula and Tang (2012) allow for differential effects, these have to be mediated through a *known* function; in contrast, our results allow differential effects that depend on functions that are not exactly known; in fact our goal is to do inference on such functions. Third, our econometric method explicitly allows for observable payoff shifters with continuous and/or discrete support, whereas De Paula and Tang (2012) require categorical or discrete covariates. Our strategic interaction indices are allowed to depend on observable payoff covariates in a very flexible way.

[1] Multiple equilibria as a source of identification power is also studied in Sweeting (2009) in a fully parametric model.

Our model's testable implications takes the form of a sign restriction on a conditional covariance. By the definition of a covariance, this restriction can be expressed as an inequality involving a nonlinear transformation of conditional moments. Among existing methods for inference with conditional moment inequalities, those that avoid the use of nonparametrically estimated conditional moments and rely instead on spaces of "instrument functions" (Andrews and Shi (2011a, 2011b), Armstrong (2011a, 2011b)) are not directly applicable to our case since they are not designed to handle, in general, nonlinear transformations of a collection of conditional moments. In general a problem like ours requires the use of plug-in nonparametric estimators for the conditional moments involved. Along these lines, the methodology proposed in Chernozhukov, Lee, and Rosen (2011) could potentially be adapted and applied to our problem. Its implementation would require the computation of a supremum of a particular test statistic over a target testing range of the conditioning variables. However, when these include a large number of elements with rich support, approximating this supremum with a reasonable degree of precision would pose a computational challenge. This is the case of our empirical application, where the vector of conditioning covariates includes eight continuously distributed elements. To be able to conduct inference in a setting like ours we propose an inferential approach based on a particular type of one-sided expectation[2] whose construction uses plug-in nonparametric estimators. Unlike existing methods that also rely on one-sided L^p functionals in related problems (Lee, Song, and Whang (2013)), our approach is not based on a *least favorable* configuration and is therefore less conservative when used to construct confidence sets. By design, our method is computationally easy to implement even in the presence of a rich model with multiple conditioning covariates with continuous support. We describe our approach in the main body of the paper and we establish its asymptotic properties in Appendix B.

We apply our approach to study the pattern of entry by the three major national drug store chains (CVS, Walgreens, and Rite Aid) competing in local geographic markets. Our model allows us to study both the extensive-margin decision of whether to enter a market or not, as well as the intensive-margin decision of how many establishments to open in a market. Most papers (see, e.g., Bresnahan and Reiss (1991b), Berry (1992), Seim (2006), and Ciliberto and Tamer (2009)) have modeled entry exclusively as an extensive-margin binary decision and have therefore abstracted away from the intensive margin. Some exceptions to this include Davis (2006) and Gowrisankaran and Krainer (2011) but these papers rely on very strong parametric assumptions and equilibrium selection restrictions.[3] Our application shows that this intensive margin reveals many important features of competition that are obscured by the extensive margin alone. In particular, we find important evidence of asymmetries in the competition among these players that suggests that the least anticompetitive takeover of Rite Aid by one of the competitors (a policy currently under consideration) would be CVS rather than Walgreens. We also find evidence that the strength of strategic interactions diminishes with market size, which

[2] A one-sided expectation refers to an expectation of the form $E[\max\{Z, 0\}]$.

[3] Aradillas-Lopez (2011) also focuses on rich strategy spaces but the goal there is to answer a different question than the one posed here.

plays a central role in explaining the large positive correlation of entry behavior found in the data.

The rest of the paper proceeds as follows. Section 2 describes our general assumptions along with the resulting properties of our model. The observable implications that result from our model are studied in Section 3. Section 4 describes our econometric inferential procedure in semiparametric models and characterizes its asymptotic properties. Section 5 analyzes the properties of our procedure through Monte Carlo experiments. Section 6 applies our approach to entry decisions in the U.S drugstore industry, modeling entry strategies as involving not only a binary choice of entry but also a capacity (number of stores) choice. Section 7 concludes. The proof of our main identification result and a description of the asymptotic properties of our econometric approach are included in the Appendix. An in-depth econometric analysis and several extensions are included in the Supplement, available in a supplementary file on the journal website, http://qeconomics.org/supp/465/supplement.pdf and http://qeconomics.org/supp/465/code_and_data.zip.

2. A static game with a rich strategy space

We now present a nonparametric game with incomplete information and derive its testable implications. Our model nonparametrically generalizes three main features of existing models. First, we allow for a rich action space, which includes binary choice games as a special case. This expands the scope of real world problems that can be studied through our approach. Second, we place no restrictions on the dimension and the "magnitude" of private information or the manner in which private information shifts the payoff function. Third, we isolate a fundamental feature of the game that aggregates the effect that rivals' strategies have on a player's own payoff. However, instead of imposing a full functional form on payoffs, we only place general restrictions regarding the way this index enters a player's payoffs. These restrictions formalize the idea that a larger value of the strategic index, by definition, decreases a player's marginal payoff from increasing its own action. Our main questions would then include how the strategic index changes with the actions of players' rivals (which would determine patterns of strategic substitutability or complementarity as well as the relative impact of rivals' strategies on a given player), as well as how these features depend on observable characteristics of the environment. In the context of entry models the strategic index would capture the competition effect, summarizing how a firm's marginal payoff from increasing its presence in a market is affected by the entry decisions of others. It can also help us learn how these features change from one market to another given the observable market characteristics available to the researcher.

2.1 *Players and actions*

We have $p = 1, \ldots, P$ players ($-p$ denotes the collection of all players except p); each p has a *real-valued* decision variable Y^p, which is *either* binary (i.e., $Y^p \in \{0, 1\}$) or (if it can take on more than two values) *ordinal in nature*, with $Y^p \in \mathcal{A}^p$. The strategy space

$\mathcal{A}^p$ can be unbounded, it can be discrete or continuous (or it can consist of the union of discrete and continuous sets in $\mathbb{R}$), and its ordered elements do not have to be evenly spaced. In fact, our identification results do not require that the econometrician know the exact structure of $\mathcal{A}^p$. The only restriction is that it must possess a natural order. We let $\mathcal{A}^{-p} = \prod^{q \neq p} \mathcal{A}^q$ denote the action space of p's opponents. We use lowercase y^p to denote a potential action (in $\mathcal{A}^p$) for p and use $y^{-p} \equiv (y^q)^{q \neq p}$ to denote a potential action profile (in $\mathcal{A}^{-p}$) for p's opponents. We use uppercase letters (Y^p and $Y^{-p} \equiv (Y^q)^{q \neq p}$) to denote the actions (profiles of actions) actually chosen by players. The game is simultaneous.

2.2 *Payoff functions*

Each player p has a payoff function that indicates the (von Neumann–Morgenstern) utility associated with its choices. The payoff for p if $Y^p = y^p$ and $Y^{-p} = y^{-p}$ is given by

$$\nu^p\left(y^p, y^{-p}; \xi^p\right), \tag{1}$$

where ξ^p denotes p's payoff shifters (other than opponents' choices). For convenience and in accordance with the boundaries of $\mathcal{A}^p$, for any $y^{-p} \in \mathcal{A}^{-p}$ we decree $\nu^p(y^p, \cdot; \cdot) = -\infty$ for any $y^p \notin \mathcal{A}^p$. We will partition p's payoff shifters as

$$\xi^p = \left(X, \varepsilon^p\right),$$

where X is observed by the econometrician and ε^p is not. The dimension of ε^p is left unspecified and we allow ε^p and X to be correlated in an arbitrary way. We will not make assumptions here about the direction in which payoffs shift in response to particular elements of X. Furthermore we will not assume the existence of player-specific observable payoff shifters. Throughout, X will denote the collection of all covariates observable by the researcher.

2.2.1 *Basic restrictions on payoff functions* We assume that payoff functions can be expressed in the following way.

Assumption 1 (Generic expression of payoff functions). *The term ν^p can be expressed as*

$$\nu^p\left(y^p, y^{-p}; \xi^p\right) = \nu^{p,a}\left(y^p; \xi^p\right) - \nu^{p,b}\left(y^p; \xi^p\right) \cdot \eta^p\left(y^{-p}; X\right), \tag{2}$$

where $\nu^{p,b}$ and η^p are real-valued functions or "indices" whose product captures the entire strategic effect of p's opponents on his payoff function.

The key feature about η^p is that it depends on ξ^p solely through X. While strategic interaction effects are allowed to depend on unobservable components of payoff shifters, this dependence must be fully captured by $\nu^{p,b}$.

Expected payoff functions and Assumption 1 We assume Bayesian Nash equilibrium (BNE) behavior here. As a result, we can focus on beliefs for p that can be expressed as probability functions defined over $\mathcal{A}^{-p}$. For any set of beliefs $\sigma^{-p}: \mathcal{A}^{-p} \longrightarrow [0,1]$, the associated expected utility for p of choosing $Y^p = y^p$ is

$$\begin{aligned} \overline{\nu}_\sigma^p(y^p;\xi^p) &= \sum_{y^{-p}\in\mathcal{A}^{-p}} \sigma^{-p}(y^{-p})\cdot \nu^p(y^p,y^{-p};\xi^p) \\ &= \nu^{p,a}(y^p;\xi^p) - \nu^{p,b}(y^p;\xi^p)\cdot \overline{\eta}_\sigma^p(X), \quad \text{where} \\ \overline{\eta}_\sigma^p(X) &= \sum_{y^{-p}\in\mathcal{A}^{-p}} \sigma^{-p}(y^{-p})\cdot \eta^p(y^{-p};X). \end{aligned} \tag{3}$$

A key feature of p's beliefs is that they do not depend on p's own action. This independence is the defining feature of Nash equilibrium as opposed, for example, to correlated equilibrium.

Our model will normalize the "strategic meaning" of the index $\eta^p(y^{-p};X)$ by assuming that $\nu^{p,b}(\cdot;\xi^p)$ is nondecreasing with probability 1 (w.p.1). This in turn will imply that the marginal gain for p of increasing its own strategy is nonincreasing in the expected value of the strategic index η^p.

Assumption 2 (Marginal benefit of Y^p is nonincreasing in η^p). *With probability* 1 *in* ξ^p, *the function* $\nu^{p,b}(\cdot;\xi^p)$ *is nondecreasing over* $\mathcal{A}^p$. *That is, for any* $v > u$ *in* $\mathcal{A}^p$ *we have* $\nu^{p,b}(v;\xi^p) \geq \nu^{p,b}(u;\xi^p)$ *w.p.1.*

Take any pair of actions $v > u$ in $\mathcal{A}^p$. Take any pair of beliefs σ^{-p} and $\sigma^{-p'}$. Then

$$\begin{aligned} &\big[\overline{\nu}_\sigma^p(v;\xi^p) - \overline{\nu}_\sigma^p(u;\xi^p)\big] - \big[\overline{\nu}_{\sigma'}^p(v;\xi^p) - \overline{\nu}_{\sigma'}^p(u;\xi^p)\big] \\ &\quad = \big[\overline{\eta}_{\sigma'}^p(X) - \overline{\eta}_\sigma^p(X)\big]\cdot\big[\nu^{p,b}(v;\xi^p) - \nu^{p,b}(u;\xi^p)\big]. \end{aligned}$$

Therefore by Assumption 2,

$$\begin{aligned} &\overline{\eta}_\sigma^p(X) \geq \overline{\eta}_{\sigma'}^p(X) \\ &\quad\Longrightarrow\quad \overline{\nu}_\sigma^p(v;\xi^p) - \overline{\nu}_\sigma^p(u;\xi^p) \leq \overline{\nu}_{\sigma'}^p(v;\xi^p) - \overline{\nu}_{\sigma'}^p(u;\xi^p) \quad \forall u < v \in \mathcal{A}^p. \end{aligned} \tag{4}$$

The "shape" restriction described in Assumption 2 will be the key to our identification results. It is reminiscent of conditions found in the supermodular game literature (more precisely, it amounts to a supermodularity property for $-\nu^{p,b}$; see Topkis (1998) and Vives (1999)) but our setup extends beyond supermodularity since it allows for very general patterns of pairwise complementarity or substitutability. In this paper we will not make any assumptions[4] regarding how payoffs shift with specific elements in ξ^p.

Observe that given Assumption 2, Y^q is a strategic substitute (complement) for Y^p if $\eta^p(y^{-p};\xi^p)$ is increasing (decreasing) in y^q. Cournot competition (where firms compete in quantities with each other) is a classic case of a game of strategic substitutes. In

[4]If economic theory provides ex ante information about how payoffs should shift with some specific elements in ξ^p, this information could potentially be used to refine the results that follow.

that case $\eta^p(y^{-p}; X)$ would be increasing in each element of y^{-p}. Conversely, if an increase in player q's action Y^q lowers η^p, then by Assumption 2 it increases the marginal gain to player p from increasing its actions and thus Y^q would be a strategic complement for Y^p. Bertrand competition (where firms compete in prices with each other) is a classic case of a game of strategic complements. Note that Assumption 2 allows for any pattern of pairwise complementarity or substitutability between players' strategies. Whether player q's strategy is a complement or a substitute for player p's will be determined by whether the index η^p is decreasing or increasing in y^q.

2.3 *Example: A structural model of imperfect competition*

It is useful to contextualize our setup within a well known structural economic model. Consider a model of Cournot competition between P firms with differentiated products. To avoid confusion with our notation (where we have used p to denote each player and P as the total number of players) let us use script typeface letters to denote prices $\mathcal{P}$ and quantities $\mathcal{Q}$. Suppose the model is described by a linear demand system where

$$\mathcal{Q}^p = \sum_{q=1}^{P} d^{p,q}(\xi^p) \cdot \mathcal{P}^q + f^p(\xi^p) \quad \text{for } p = 1, \dots, P.$$

Suppose $\sum_{q=1}^{P} d^{p,q}(\xi^p) \neq 0$ w.p.1 (an assumption grounded on economic theory). Define $\zeta^p(\xi^p) \equiv f^p(\xi^p) / \sum_{q=1}^{P} d^{p,q}(\xi^p)$. Our assumptions will imply restrictions on the structure of the coefficients $d^{p,q}(\xi^p)$. Specifically, suppose we can express $d^{p,q}(\xi^p) = \phi^p(\varepsilon^p) \cdot a^{p,q}(X)$. The demand system can be expressed as

$$\mathcal{Q}^p = \phi^p(\varepsilon^p) \cdot \sum_{q=1}^{P} a^{p,q}(X) \cdot (\mathcal{P}^q + \zeta^p(\xi^p)) \quad \text{for } p = 1, \dots, P.$$

Let $A(X)$ denote a $P \times P$ matrix where $[A(X)]_{p,q} = a^{p,q}(X)$ and let $D(\phi(\varepsilon))$ denote a $P \times P$ diagonal matrix where $[D(\phi(\varepsilon))]_{p,p} = \phi^p(\varepsilon^p)$. By our above assumption the last matrix is invertible w.p.1. Suppose this is also true for $A(X)$ and denote $[A(X)^{-1}]_{p,q} \equiv b^{p,q}(X)$. Then inverse demands are of the form

$$\mathcal{P}^p = \frac{1}{\phi^p(\varepsilon^p)} \sum_{q=1}^{P} b^{p,q}(X) \cdot \mathcal{Q}^q - \zeta^p(\xi^p) \quad \text{for } p = 1, \dots, P.$$

Denote firm p's cost function as $C^p(\mathcal{Q}^p; \xi^p)$, which can be entirely unrestricted (e.g., it can include a fixed cost and it need not have to display increasing marginal costs). Profit functions are of the form

$$\pi^p(\mathcal{Q}^p, \mathcal{Q}^{-p}; \xi) = \left(\frac{1}{\phi^p(\varepsilon^p)} \sum_{q=1}^{P} b^{p,q}(X) \cdot \mathcal{Q}^q - \zeta^p(\xi^p) \right) \cdot \mathcal{Q}^p - C^p(\mathcal{Q}^p; \xi^p).$$

In a Cournot model firms compete in quantities, so $Y^p = \mathcal{Q}^p$. This model fits our representation of payoffs (profits) in (2). We have $\nu^p(y^p, y^{-p}; \xi) = \nu^{p,a}(y^p; \xi) - \nu^{p,b}(y^p; \xi) \cdot$

$\eta^p(y^{-p}; X)$, where

$$\nu^{p,a}(y^p;\xi) = \left(\frac{1}{\phi^p(\varepsilon^p)} b^{p,p}(X)\cdot y^p - \zeta^p(\xi^p)\right)\cdot y^p - C^p(y^p;\xi^p),$$

$$\nu^{p,b}(y^p;\xi) = \frac{y^p}{\phi^p(\varepsilon^p)}.$$

So as to satisfy Assumption 2 it suffices that the function $\phi^p(\varepsilon^p)$ be of constant sign. Given our structural model, it is natural to assume that $\phi^p(\varepsilon^p)\geq 0$ w.p.1. ($\phi^p(\varepsilon^p)>0$ w.p.1. given our invertibility assumptions). In this case the strategic index would be

$$\eta^p(y^{-p};\xi) = -\sum_{q\neq p} b^{p,q}(X)\cdot y^q;$$

the qth good will be a substitute for the pth good if $b^{p,q}(X)\leq 0$; otherwise it will be a complement. Note that, since $[A(X)^{-1}]_{p,q} = b^{p,q}(X)$, the strategic indices η^p allow us to recover $A(X)$, a key structural component of the model.

Suppose instead that we have a log-linear system of demand,

$$\log(\mathcal{Q}^p) = \sum_{q=1}^{P} d^{p,q}(\xi^p)\cdot\log(\mathcal{P}^q) + f^p(\xi^p) \quad \text{for } p = 1,\dots,P.$$

Now the coefficients $d^{p,q}(\xi^p)$ directly measure elasticities of demand. In this case our assumptions imply a different set of restrictions. We now need $d^{p,q}(\xi^p) = d^{p,q}(X)$ (privately observed shocks ε^p should now be excluded from these elasticities). Suppose $\sum_{q=1}^{P} d^{p,q}(X)\neq 0$ w.p.1 for each p (a reasonable assumption given the homogeneity properties of demand). Define $\lambda^p(\xi^p) = f^p(\xi^p)/\sum_{q=1}^{P} d^{p,q}(X)$. Then the demand system can be rewritten as

$$\log(\mathcal{Q}^p) = \sum_{q=1}^{P} d^{p,q}(X)\cdot\left(\log(\mathcal{P}^q) + \lambda^p(\xi^p)\right) \quad \text{for } p = 1,\dots,P.$$

Let us maintain that the $P\times P$ matrix $D(X)$, where $[D(X)]_{p,q} = d^{p,q}(X)$ is invertible w.p.1, and denote $[D(X)^{-1}]_{p,q}\equiv r^{p,q}(X)$. Inverting the demand system we obtain the inverse demands

$$\mathcal{P}^p = e^{-\lambda^p(\xi^p)}\cdot\prod_{q=1}^{P}(\mathcal{Q}^q)^{r^{p,q}(X)} \quad \text{for } p = 1,\dots,P.$$

Profit functions are now of the form

$$\pi^p(\mathcal{Q}^p,\mathcal{Q}^{-p};\xi^p) = e^{-\lambda^p(\xi^p)}\cdot(\mathcal{Q}^p)^{r^{p,p}(X)+1}\cdot\prod_{q\neq p}(\mathcal{Q}^q)^{r^{p,q}(X)} - C^p(\mathcal{Q}^p;\xi^p).$$

Define $\nu^{p,a}(y^p;\xi^p) = -C^p(y^p;\xi^p)$. For $\nu^{p,b}$ and η^p we can proceed as follows. Satisfying the condition in Assumption 2 depends on the sign of $r^{p,p}(X)+1$. It is easy to see that our payoff representation in (2) and the condition in Assumption 2 will be satisfied if we define

$$\nu^{p,b}\big(y^p;\xi^p\big) = e^{-\lambda^p(\xi^p)} \cdot \big(\mathbb{1}\big\{r^{p,p}(X) \geq -1\big\} - \mathbb{1}\big\{r^{p,p}(X) < -1\big\}\big) \cdot \big(y^p\big)^{r^{p,p}(X)},$$
$$\eta^p\big(y^{-p};X\big) = \big(\mathbb{1}\big\{r^{p,p}(X) \geq -1\big\} - \mathbb{1}\big\{r^{p,p}(X) < -1\big\}\big) \cdot \prod_{q \neq p} \big(y^q\big)^{r^{p,q}(X)}.$$

Suppose $r^{p,p}(X) \geq -1$. Then the qth good is a substitute for the pth good if $r^{p,q}(X) > 0$ and it is a complement otherwise. If $r^{p,p}(X) < -1$, then this holds with the reverse the inequalities. Once again the index $\eta^p(y^{-p};X)$ has a structural interpretation as it contains information about the relative price elasticities in the demand system.

Using the demand systems described above we could also study competition in prices instead of quantities. In that case our assumptions would place restrictions on firms' cost functions while allowing more flexibility in the specification of demand functions compared to the Cournot case (which placed no restrictions on firms' cost functions as we showed above).

2.4 *Strategic interaction features captured by the index η^p*

Given our payoff representation, the overall scale of the strategic effect would be absorbed into the term $\nu^{p,b}$. While the index η^p would not capture the overall scale of strategic interaction it would nevertheless summarize the following key features of strategic interaction in the model.

(i) The directional patterns of strategic interaction between any subset of players: This is captured by the direction in which the strategic indices move in response to rivals' actions.

(ii) The relative magnitude of the effects of strategic interaction between one player and each one of his/her opponents: This is captured by the relative magnitude in which the strategic indices shift in response to each rival's action.

As we illustrated in the previous section, different conjectures involving these strategic features can be incorporated directly into the structure of η^p.

2.5 *Information and behavior*

An incomplete information setting allows us to focus on pure strategy equilibria in which players strictly best respond to each other in equilibrium. Such an equilibrium restriction is natural for empirical work because equilibria can then be interpreted as a steady state outcome, which mixed strategy equilibria do not allow.[5] The empirical appeal of pure strategy equilibria is the key motivation for Harsanyi's well known purification theorem (Harsanyi (1973)). His result showed that when there exists (potentially

[5]Mixed strategies force one to question why it is the case that when a player is indifferent among several strategies, he or she mixes over these strategies in exactly such a way that makes the *other* player indifferent. For a further discussion see Morris (2008).

small) private information about own payoffs in a normal form game, then this ensures that all equilibria generically take this pure strategy form.[6] He modeled private information shocks to be idiosyncratic and hence independent across players. We follow in this approach and assume that the private payoff shocks to firms are independent conditional on publicly observable payoff shifters, but we analyze what happens when this assumption is violated in our Monte Carlo experiment section.

Assumption 3 (Independent private shocks). *The term X is perfectly observed by all players, but ε^p is only privately observed by p. We assume that each ε^p is independent of ε^{-p} conditional on X. The true distribution of $(X, (\varepsilon^p)_{p=1}^P)$ is common knowledge among the players, as are the functional forms of payoff functions $(\nu^p)_{p=1}^P$. Thus, the only source of incomplete information for p is the realization of ε^{-p}.*

The dimension of ε^p is left unspecified and we allow ε^p and X to be correlated in an arbitrary way. A special case of Assumption 3 is one where some player p possesses no private information and $\xi^p = X$ (recall that $\xi^p = (X, \varepsilon^p)$). Thus, a game of complete information would be a special case of our setting as long as $\xi^p = X$ for each p. In this case the only source of unobserved heterogeneity for the econometrician would be the equilibrium selection mechanism. The case $\xi^p = X$ is a very limited, special case of complete information games where the econometrician happens to observe all the payoff shifters in the game. Because this case is not of general interest, our relevant setting should be viewed as that of an incomplete information game.

Independent private shocks (Assumption 3) is a type of condition widely imposed in econometric work on incomplete information games (it is imposed, for example, in De Paula and Tang (2012)). Nevertheless it is an important restriction, leaving out, for example, the presence of market-level unobserved (to the econometrician) payoff shocks. The extent to which Assumption 3 is realistic depends on the richness of the collection of covariates available in X, which would be application-specific. We study the implications of violations to Assumption 3 in two ways. First, using our Monte Carlo experiments we study the properties of our procedure under the presence of correlation between players' private shocks (this is done in the Supplement). Also in the Supplement we briefly model the presence of an unobserved common shock to all players (meant to approximate, e.g., market-level unobserved shocks) and we study conditions under which this may lead to inconsistency of our results. Our arguments there suggest that consistency is more likely to be preserved if actions are strategic complements rather than substitutes. We also briefly outline what it would take to mitigate the influence of an unobserved common shock. With these caveats in mind, Assumption 3 will be maintained for our results.

[6]The second part of his result, the so-called *approachability* party, showed that the set of pure strategy equilibria in the perturbed private information game is arbitrarily close to the set of all mixed strategy equilibria of the corresponding unperturbed complete information game.

2.5.1 *Bayesian Nash equilibrium behavior* We maintain that the outcome observed is the result of a BNE of the underlying game. Given the independent private shock restriction in Assumption 3, any BNE can be characterized as a collection of conditional (on X) probability functions $\{\sigma_*^p(\cdot|X) : \mathcal{A}^p \longrightarrow [0,1]\}_{p=1}^P \equiv \sigma_*(X)$ with corresponding expected utility functions

$$\bar{v}_{\sigma_*}^p(\cdot; \xi^p) = \sum_{y^{-p} \in \mathcal{A}^{-p}} \sigma_*^{-p}(y^{-p}|X) \cdot v^p(\cdot, y^{-p}; \xi^p),$$

where, for each $y^{-p} \equiv (y^q)^{q \neq p} \in \mathcal{A}^{-p}$ and $y^p \in \mathcal{A}^p$,

$$\sigma_*^{-p}(y^{-p}|X) = \prod_{q \neq p} \sigma_*^p(y^q|X) \quad \text{and}$$

$$\sigma_*^p(y^p|X) > 0 \quad \text{only if} \quad y^p \in \operatorname*{argmax}_{y \in \mathcal{A}^p} \bar{v}_{\sigma_*}^p(y; \xi^p).$$

Assumption 4. *The outcome observed is the realization of a BNE, that is,*

$$Y^p \in \operatorname*{argmax}_{y \in \mathcal{A}^p} \bar{v}_{\sigma_*}^p(y; \xi^p) \quad \textit{for some BNE } \sigma_*(X).$$

For a given realization of payoff shifters, multiple BNE may exist and we leave the underlying selection mechanism $\mathcal{S}$ unspecified except for the assumption that it always picks a BNE $\sigma_*(X)$ such that the resulting expected payoff function $\bar{v}_{\sigma_*}^p(\cdot; \xi^p)$ has a unique *optimal choice.*

Assuming pure strategy play in games of incomplete information is not a very restrictive assumption. Recall the above discussion that Harsanyi's purification theorem ensures that the restriction to pure strategy equilibria with unique best responses is generically without loss of generality in (finite) incomplete information games. In more general games of incomplete information where player types are conditionally independent—the type of setting we assume here—Milgrom and Weber (1985) show that every mixed strategy equilibrium has a nearby "purification" pure strategy such that the distributions of players' observed behavior and expected payoffs are identical.[7] The setup described in Assumptions 1–4 encompasses many existing static models of incomplete information as special cases. The two key restrictions are either binary or an otherwise ordinal nature of the strategy space, and conditional independence of players' private information. Examples include, among others, Seim (2006), the binary choice (i.e., two-time period) model in Sweeting (2009), the general setup in De Paula and Tang (2012), binary choice or ordinal choice versions of the incomplete information games studied in Bajari et al. (2009), and of the quantal response equilibrium (QRE) model proposed in McKelvey and Palfrey (1995) and studied further in Haile, Hortacsu, and Kosenok (2008). Our setup can also encompass complete-information games under the restriction $\xi^p =$

[7]Note that Assumption 4 implicitly imposes an additional restriction on payoff functions; namely, the existence of equilibria where each player has a unique best response. Sufficient conditions can be made precise in the context of specific structural models (see our examples in Section 2.3).

X in which case it would encompass, for example, the binary choice games in Bresnahan and Reiss (1990) and Tamer (2003). Beyond the existing literature, our assumptions can handle models under unprecedentedly general assumptions regarding players' payoffs, which include discrete or continuous strategy spaces. While our results may retain some of their validity if there is relatively small correlation in players' private shocks (see our experimental results in Section 5), static games where the strategy space does not have a natural order (e.g., truly multinomial choice games) are entirely outside the scope of our paper.

3. Implications of Assumptions 1–4

3.1 *Properties of players' best responses*

As we stated above, we focus on equilibrium beliefs that yield a unique optimal choice to players. For any such set of beliefs our payoff shape restrictions imply a monotonicity property between optimal actions and the expected value of the strategic index induced by each player's beliefs. We describe this next.

Result 1. *Let σ^{-p} and $\sigma^{-p'}$ denote any pair of beliefs that produce unique expected-payoff-maximizing choices for p given the realization of ξ^p, and let $y^p_\sigma(\xi^p)$ and $y^p_{\sigma'}(\xi^p)$ denote the corresponding optimal choices. If Assumptions 1 and 2 hold, then w.p.1 we have*

$$\text{If } \overline{\eta}^p_\sigma(X) \geq \overline{\eta}^p_{\sigma'}(X), \quad \text{then } \mathbb{1}\{y^p_\sigma(\xi^p) \leq y^p\} \geq \mathbb{1}\{y^p_{\sigma'}(\xi^p) \leq y^p\} \quad \forall y^p \in \mathcal{A}^p.$$

The proof is given in Appendix A.

3.2 *Main result*

Let σ_{*j} and σ_{*k} denote any pair of existing BNE that the selection mechanism $\mathcal{S}$ could choose with positive probability. By Result 1, w.p.1 we must have

$$\text{If } \overline{\eta}^p_{\sigma_{*j}}(X) \geq \overline{\eta}^p_{\sigma_{*k}}(X), \quad \text{then } \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\} \geq \mathbb{1}\{y^p_{\sigma_{*k}}(\xi^p) \leq y^p\} \quad \forall y^p \in \mathcal{A}^p.$$

Our main result will follow from here and the independence condition in Assumption 3.

Theorem 1. *Let y^p be given. If Assumptions 1–4 hold, then w.p.1 in X we have*

$$\begin{aligned} &E[\mathbb{1}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p}; X)|X] \\ &\quad \geq E[\mathbb{1}\{Y^p \leq y^p\}|X] \cdot E[\eta^p(Y^{-p}; X)|X] \quad \forall y^p. \end{aligned} \tag{5}$$

A step-by-step proof of Theorem 1 is included in Appendix A, but we can summarize it as follows. Given X, let J denote the number of BNE $\{\sigma_{*j}(X)\}_j$ that the selection mechanism $\mathcal{S}$ can choose with positive probability, and let $P^{\mathcal{S}}_j(X)$ denote the probability that $\mathcal{S}$ selects the jth BNE ($\sigma_{*j}(X)$), conditional on X. Our assumptions maintain that $\mathcal{S}$ concentrates on BNE that have a unique optimal choice. Denote it as $y^p_{\sigma_{*j}}(\xi^p)$ for the jth BNE. Using iterated expectations, our assumptions (independent private shocks is key) yield

$$\begin{aligned}
&E\big[\mathbb{1}\{Y^p \le y^p\} \cdot \eta^p(Y^{-p}; X)|X\big] - E\big[\mathbb{1}\{Y^p \le y^p\}|X\big] \cdot E\big[\eta^p(Y^{-p}; X)|X\big] \\
&\quad = E_{\xi^p|X}\Bigg[\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \le y^p\} \cdot \overline{\eta}^p_{\sigma_{*j}}(X) \\
&\qquad - \Bigg(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \le y^p\}\Bigg) \times \Bigg(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \overline{\eta}^p_{\sigma_{*j}}(X)\Bigg)\Bigg| X\Bigg].
\end{aligned}$$

The result in Theorem 1 follows because, under our assumptions, the expression inside the expectation operator above is nonnegative w.p.1. To see why, note that this expression can be rewritten as

$$\begin{aligned}
&\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \le y^p\} \cdot \overline{\eta}^p_{\sigma_{*j}}(X) \\
&\qquad - \Bigg(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \le y^p\}\Bigg) \times \Bigg(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \overline{\eta}^p_{\sigma_{*j}}(X)\Bigg) \\
&\quad = \sum_{\ell=1}^{J}\sum_{j=1}^{J} P_\ell^{\mathcal{S}}(X) P_j^{\mathcal{S}}(X) \cdot \underbrace{\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \le y^p\} \cdot \big(1 - \mathbb{1}\{y^p_{\sigma_{*\ell}}(\xi^p) \le y^p\}\big)}_{\text{always} \ge 0} \\
&\qquad \cdot \big(\overline{\eta}^p_{\sigma_{*j}}(X) - \overline{\eta}^p_{\sigma_{*\ell}}(X)\big) \\
&\quad \ge 0.
\end{aligned}$$

The last inequality follows from Result 1. To see why, note that this inequality can be violated if and only if $\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \le y^p\} \cdot (1 - \mathbb{1}\{y^p_{\sigma_{*\ell}}(\xi^p) \le y^p\}) = 1$ and $\overline{\eta}^p_{\sigma_{*j}}(X) - \overline{\eta}^p_{\sigma_{*\ell}}(X) < 0$, which would violate Result 1.

If the underlying game has a unique equilibrium w.p.1—or more generally if it has a degenerate equilibrium selection mechanism–we would have $Y^p \perp Y^{-p}|X$ and therefore any measurable function $g(Y^{-p}; X)$ should satisfy Theorem 1 as an equality. Therefore Theorem 1 provides identification power for η^p only if the underlying game has multiple equilibria for at least a subset of realizations of payoff shifters and if players randomize across such equilibria.

Remark 1. De Paula and Tang (2012) derived the conditions in Theorem 1 for the case of binary choice games. Their result relies essentially on the same conditions as Assumptions 1 and 3, along with Nash equilibrium behavior (Assumption 4 in our case). Our primary contribution is introducing Assumption 2 and showing that it is sufficient to extend the results from binary choice games to richer strategy spaces.

Remark 2. A few relevant results follow immediately from Theorem 1.

(i) *No strategic interaction.* Our assumptions cannot rule out that there is no strategic interaction effect at all, since $\eta^p(Y^{-p}; X) = g(X)$ would satisfy Theorem 1 as an equality (a fact that is also true in De Paula and Tang (2012)). The value of Theorem 1 is its ability to help us test many different conjectures about strategic interaction under the assumption that some interaction effect exists.

(ii) *Rejecting unique equilibria.* As we pointed out above, if the underlying game has a unique equilibrium w.p.1—or more generally if it has a degenerate equilibrium selection mechanism—then any measurable function $g(Y^{-p}; X)$ should satisfy Theorem 1 as an equality. Therefore, if we maintain the assumptions in our model, the existence of some function $g(Y^{-p}; X)$ that violates the result in Theorem 1 would immediately reject the notion that the game has a unique equilibrium w.p.1. In particular, this would reject the assertion that there is no strategic interaction in the model.

(iii) *Rejecting our model when strategic interaction exists.* Suppose we maintain that some form of strategic interaction exists in the model, ruling out the scenario in part (i) of this remark. Then, under the assumptions of our model, there must exist a function $\eta^p(Y^{-p}; X)$ that satisfies the result in Theorem 1. Thus, ruling out the existence of such a function would immediately reject our model. Therefore, under the maintained hypothesis that strategic interaction is present, our model is falsifiable and could be rejected nonparametrically.

The rest of our paper will be devoted to using Theorem 1 to do inference on the strategic index η^p in a context where this index is assumed to belong to a parametric family of functions while leaving every other aspect of the model nonparametric.

3.2.1 *Unconditional covariance* It is important to note that the result in Theorem 1 only restricts the sign of the conditional (on X) covariance between $\mathbb{1}\{Y^p \leq y^p\}$ and the strategic index $\eta^p(Y^{-p}; X)$. The sign of the unconditional covariance is unrestricted by our assumptions. To see this, note that by the so-called law of total covariance we have

$$
\begin{aligned}
&\operatorname{Cov}\big(\mathbb{1}\{Y^p \leq y^p\}, \eta^p(Y^{-p}; X)\big) \\
&\quad = \underbrace{E\big[\operatorname{Cov}\big(\mathbb{1}\{Y^p \leq y^p\}, \eta^p(Y^{-p}; X) | X\big)\big]}_{\geq 0 \text{ by Theorem 1}} \\
&\qquad + \underbrace{\operatorname{Cov}\big(E\big[\mathbb{1}\{Y^p \leq y^p\} | X\big], E\big[\eta^p(Y^{-p}; X) | X\big]\big)}_{\text{sign is unrestricted by our assumptions}}.
\end{aligned}
$$

Thus, depending on how payoffs shift with X (our assumptions are completely silent on this), the unconditional covariance could be positive or negative even if the conditional covariance is nonnegative w.p.1.

4. Inference of strategic interactions in a semiparametric model

The result in Theorem 1 does not rely on a parametric specification for the strategic interaction index η^p. If we assume symmetry of interaction effects so that players care equally about the actions of each opponent, the strategic index could simply be $\eta^p(Y^{-p}; X|\theta^p) = \sum_{q\neq p} \pm Y_q$, where the sign $\pm$ would indicate whether actions are substitutes or complements and this in turn would be identified through the restriction in Theorem 1. The model then would remain fully nonparametric throughout. If we want to allow for asymmetries and more complexity in the interaction effects we need a more flexible characterization of the strategic index. We will focus now on the case where the strategic interaction index η^p is assumed to belong to a parametric family of functions of the form

$$\eta^p\big(Y^{-p}; X|\theta^p\big),$$

with all other elements of the model left nonparametrically specified. In the examples of Section 2.3, this could be done by specifying a parameterization for the matrix $A(X)$ in the case of linear demands, and for the matrix $D(X)$ in the log-linear case. All other components of the structural models would be left unspecified in both cases. Let $\theta = (\theta^p)_{p=1}^{P}$ and let Θ denote the parameter space. The true value of θ will be denoted by θ_0. For given y^p, x, and θ^p define

$$\begin{aligned}
&F_{Y^p}\big(y^p|x\big) = E_{Y^p|X}\big[\mathbb{1}\big\{Y^p \le y^p\big\}|X = x\big],\\
&\lambda^p\big(x; \theta^p\big) = E_{Y^{-p}|X}\big[\eta^p\big(Y^{-p}; x|\theta^p\big)|X = x\big],\\
&\mu^p\big(y^p|x; \theta^p\big) = E_{Y|X}\big[\mathbb{1}\big\{Y^p \le y^p\big\} \cdot \eta^p\big(Y^{-p}; x|\theta^p\big)|X = x\big],\\
&\tau^p\big(y^p|x; \theta^p\big) = F_{Y^p}\big(y^p|x\big) \cdot \lambda^p\big(x; \theta^p\big) - \mu^p\big(y^p|x; \theta^p\big).
\end{aligned}$$

Theorem 1 predicts that for each p,

$$\tau^p\big(y^p|X; \theta_0^p\big) \le 0 \quad \text{w.p.1 in } X\ \forall y^p \in \mathcal{A}^p.$$

The econometrician is not required to know the exact structure of $\mathcal{A}^p$. Since $\mathrm{Supp}(Y^p) \subseteq \mathcal{A}^p$, it is natural to focus on testing the above inequality over $y^p \in \mathrm{Supp}(Y^p)$. For this reason, we choose to test whether the inequality holds over $\mathrm{Supp}(Y^p, X)$. Therefore, our inferential approach is based on the fact that our model predicts

$$\Pr\big(\tau^p\big(Y^p|X; \theta_0^p\big) \le 0\big) = 1. \tag{6}$$

We will propose an inferential method based on the restriction in (6) and refer to the identified set Θ^I as the collection of parameter values that satisfy (6); that is,

$$\Theta^I = \big\{\theta \in \Theta : \Pr\big(\tau^p\big(Y^p|X; \theta^p\big) \le 0\big) = 1\ \forall p = 1, \dots, P\big\}. \tag{7}$$

Note that the restriction in (6) involves inequalities of nonlinear transformations of conditional moments (the conditional covariance involves the product of two conditional expectations). Developing methods for inference with conditional moment inequalities has been an area of active research in the recent past. There are generically speaking two types of methods. The first type avoids having to estimate the conditional expectations involved and relies instead on *instrument functions*. Examples of this approach include

Armstrong (2011a, 2011b) and Andrews and Shi (2011a, 2011b). Suppose $m(W;\theta)$ is a vector of *known* functions such that $E[m(W;\theta_0)|X]\leq 0$ w.p.1. Let $\mathcal{G}$ be a space of measurable, nonnegative functions of X. Then the previous inequality implies that we must have $E[m(W;\theta_0)\cdot g(X)]\leq 0$ for all $g\in\mathcal{G}$. Thus, for a given choice of $\mathcal{G}$ the conditional moment inequality implies unconditional moment inequality restrictions everywhere on $\mathcal{G}$. Cramer–von Mises or Kolmogorov–Smirnov test statistics can be constructed from here. This approach has the great advantage of not having to rely on smoothness assumptions about the conditional moments. However, it is *not applicable* here since our problem involves a nonlinear transformation of conditional moments and therefore it cannot be written as $E[m(W;\theta_0)|X]\leq 0$ for a *known* function $m(\cdot)$.

The second type of approach relies on plug-in estimators of the conditional moments involved. Most of the existing work in this area has been devoted to testing nonparametric restrictions rather than doing inference on a finite dimensional parameter. One notable exception is Chernozhukov, Lee, and Rosen (2011). Based on their approach, we would test whether θ^p satisfies our restrictions for player p over a range $(y^p,x)\in\mathcal{W}$ by using a test statistic of the form $\widehat{\nu}^p_\alpha(\theta^p)=\inf_{(y^p,x)\in\mathcal{W}}[(-\widehat{\tau}^p(y^p|x;\theta^p))+\widehat{k}(\alpha)\cdot\widehat{\sigma}^p(y^p|x;\theta^p)]$, where $\widehat{\sigma}^p$ is an estimator of the standard error of $\widehat{\tau}^p$ and $\widehat{k}(\alpha)$ is a critical value based on the αth quantile of a particular process. We would reject the inequalities for θ^p if $\widehat{\nu}^p_\alpha(\theta^p)<0$ and fail to reject them otherwise.

While this method works in principle, in practice being able to compute the statistic with precision can be a computational challenge when X includes a large number of covariates with rich support. This will be the case in our empirical application where X includes eight such covariates. In this case it is not clear how to do a grid search in eight dimensions so as to compute the test statistic (and approximate the critical value) with a reasonable degree of precision, especially if the parameterization of our strategic index η^p is such that $\tau^p(y|x;\theta^p)$ is nonseparable in θ^p. In such cases the critical value $\widehat{k}(\alpha)$ would also depend on θ^p, further complicating its use for the construction of a confidence set.

Since the instrument-function approach does not apply to our setting and since procedures that rely on computing the supremum over X of a semiparametric test statistic can pose significant computational challenges when X is large (as in our empirical example), we propose a different approach. Our method will be based on an unconditional mean-zero restriction implied by our inequalities. We describe it next.

4.1 *Expressing our inequalities using unconditional mean-zero restrictions*

For a given θ^p consider the one-sided expectation

$$T^p(\theta^p)=E_{Y^p,X}\big[\max\big\{\tau^p\big(Y^p|X;\theta^p\big),0\big\}\big].$$

Note that $T^p(\theta^p)\geq 0$ for any θ^p. For a given $\theta=(\theta^p)_{p=1}^P$ let

$$T(\theta)=\sum_{p=1}^P T^p\big(\theta^p\big).$$

Note that $T(\theta) \geq 0\ \forall\theta$ and $T(\theta) = 0$ if and only if $\theta \in \Theta^I$. Therefore we can reexpress the identified set as

$$\Theta^I = \{\theta \in \Theta : T(\theta) = 0\}.$$

Our method will rely on nonparametric plug-in estimators, focusing on the expectations defined above taken over an inference range where our estimators satisfy uniform asymptotic properties. Let $\mathcal{X} \subset \mathrm{Supp}(X)$ denote a prespecified set such that

$$x^c \in \mathrm{int}\big(\mathrm{Supp}\big(X^c|X^d = x^d\big)\big) \quad \forall\big(x^c, x^d\big) \in \mathcal{X}.$$

We maintain the assumption that $f_X(x) \geq \underline{f} > 0$ for all $x \in \mathcal{X}$. Let $\mathbb{I}_{\mathcal{X}}(x)$ denote a "trimming" function such that $\mathbb{I}_{\mathcal{X}}(x) = 0$ if $x \notin \mathcal{X}$ and $\mathbb{I}_{\mathcal{X}}(x) > 0$ otherwise. Let

$$T^p_{\mathcal{X}}\big(\theta^p\big) = E_{Y^p,X}\big[\max\big\{\tau^p\big(Y^p|X;\theta^p\big), 0\big\} \cdot \mathbb{I}_{\mathcal{X}}(X)\big], \qquad T_{\mathcal{X}}(\theta) = \sum_{p=1}^{P} T^p_{\mathcal{X}}\big(\theta^p\big). \tag{8}$$

The inference range $\mathcal{X}$ will be assumed to be such that the nonparametric estimators involved in our construction have appropriate asymptotic properties uniformly over it. Given our choice of $\mathcal{X}$, we focus attention on the superset of the identified set Θ^I:

$$\Theta^I_{\mathcal{X}} = \big\{\theta \in \Theta : T^p_{\mathcal{X}}\big(\theta^p\big) = 0 \text{ for } p = 1, \dots, P\big\}.$$

Note that $\Theta^I \subseteq \Theta^I_{\mathcal{X}}$. Also note that choosing a very limited inference range $\mathcal{X}$ may result in a loss of identification power if we preclude realizations of X that lead to multiple equilibria (see Remark 2). Under some conditions (e.g., compactness and density uniformly bounded away from zero) we could allow for the inference range $\mathcal{X}$ to correspond to the entire support of X, or we could allow $\mathcal{X}$ to grow with the sample size and cover the entire support of X asymptotically.

4.2 *Summary of econometric methodology*

The details of our econometric methodology are given in Appendix B but we provide a summary here. Our basic setting is one where the researcher observes an iid sample $((Y_i^p)_{p=1}^P, X_i)_{i=1}^n$ produced by a model satisfying our assumptions. We replace the objects in (8) with estimators of the form

$$\widehat{T}^p_{\mathcal{X}}\big(\theta^p\big) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big) \cdot \mathbb{1}\big\{\widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big) \geq -b_n\big\} \cdot \mathbb{I}_{\mathcal{X}}(X_i),$$

$$\widehat{T}_{\mathcal{X}}(\theta) = \sum_{p=1}^{P} \widehat{T}^p_{\mathcal{X}}\big(\theta^p\big),$$

where $b_n \longrightarrow 0$ is a nonnegative sequence going to zero at an appropriate rate. The use of b_n will allow us to deal with the "kink" of the $\max\{0, z\}$ function at $z = 0$ while producing asymptotically pivotal properties. To construct $\widehat{\tau}^p$ we use kernel-based estimators.

In Appendix B we describe conditions under which

$$\widehat{T}_{\mathcal{X}}(\theta) = T_{\mathcal{X}}(\theta) + \frac{1}{n}\sum_{i=1}^{n}\psi(Y_i, X_i; \theta) + \varepsilon_n(\theta), \quad \text{where}$$

$$\sup_{\theta\in\Theta}\bigl|\varepsilon_n(\theta)\bigr| = O_p\bigl(n^{-1/2-\epsilon}\bigr) \quad \text{for some } \epsilon > 0.$$

The "influence function" ψ can be expressed as

$$\psi(Y_i, X_i; \theta) = \sum_{p=1}^{P}\bigl(\max\bigl\{\tau^p\bigl(Y_i^p|X_i; \theta_p\bigr), 0\bigr\}\cdot \mathbb{I}_{\mathcal{X}}(X_i) - T_{\mathcal{X}}\bigl(\theta^p\bigr)\bigr) + \sum_{p=1}^{P}\psi_U^p\bigl(Y_i, X_i; \theta^p\bigr),$$

where ψ_U^p is the leading term in the Hoeffding decomposition of a U-statistic, and it is a function of conditional expectations (projections) and is therefore identified. The function $\psi(Y_i, X_i; \theta)$ is identified and has two key properties:

(i) We have $E[\psi(Y_i, X_i; \theta)] = 0\ \forall\theta\in\Theta$.

(ii) Let

$$\overline{\Theta}_{\mathcal{X}}^{I} = \bigl\{\theta\in\Theta : \tau^p\bigl(Y^p|X; \theta^p\bigr) < 0 \text{ w.p.1. over } \mathcal{X}, \forall p = 1, \dots, P\bigr\}.$$

Then $\psi(Y_i, X_i; \theta) = 0$ w.p.1, $\forall\theta\in\overline{\Theta}_{\mathcal{X}}^{I}$.

The term $\overline{\Theta}_{\mathcal{X}}^{I}$ is the collection of parameter values that satisfy our inequalities as *strict inequalities* w.p.1 over our inference range. Let $\sigma^2(\theta) = \mathrm{Var}(\psi(Y_i, X_i; \theta))$. Based on the properties outlined above, we have

$$\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta) = \sqrt{n}T_{\mathcal{X}}(\theta) + V_n(\theta) + \xi_n(\theta),$$

where $V_n(\theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$ and $\sup_{\theta\in\Theta}|\xi_n(\theta)| = o_p(n^{-\epsilon})$ for some $\epsilon > 0$. Given these features, our statistic will be of the form

$$\widehat{t}_n(\theta) = \frac{\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta)}{\max\bigl\{\kappa_n, \widehat{\sigma}(\theta)\bigr\}},$$

where $\widehat{\sigma}^2(\theta)$ is an estimator of $\sigma^2(\theta)$ and κ_n is a sequence converging to zero at a sufficiently slow rate (it must satisfy $\kappa_n \cdot n^{\epsilon} \longrightarrow \infty$ for any $\epsilon > 0$). Recall from our results described above that $\sup_{\theta\in\overline{\Theta}_{\mathcal{X}}^{I}}|\widehat{T}_{\mathcal{X}}(\theta)| = O_p(n^{-1/2-\epsilon})$ for some $\epsilon > 0$. The use of κ_n allows our statistic to satisfy $\sup_{\theta\in\overline{\Theta}_{\mathcal{X}}^{I}}|\sqrt{n}\cdot\widehat{t}_n(\theta)| = o_p(1)$.

For a desired coverage probability $1-\alpha$, our confidence set (CS) for θ_0 is of the form

$$\mathrm{CS}_n(1-\alpha) = \bigl\{\theta\in\Theta : \widehat{t}_n(\theta) \le c_{1-\alpha}\bigr\},$$

where $c_{1-\alpha}$ is the standard Normal critical value for $1-\alpha$. By the features outlined above our CS will have correct pointwise coverage properties; namely,

$$\inf_{\theta\in\Theta:\theta=\theta_0} \liminf_{n\to\infty} P\big(\theta \in \mathrm{CS}_n(1-\alpha)\big) \geq 1-\alpha.$$

Suppose we generalize our basic setting and assume that $\{((Y_i^p)_{p=1}^P, X_i) : 1 \leq i \leq n\}$ is a triangular array that is rowwise iid with distribution $F_n \in \mathcal{F}$. For our CS to possess correct coverage properties uniformly over $(\mathcal{F}, \Theta)$ we need to equip $\mathcal{F}$ with integrability conditions such that the following statements hold:

(i) A central limit theorem for triangular arrays holds for

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(Y_i, X_i; \theta_n, F_n)}{\sigma(\theta_n, F_n)}$$

for any sequence $F_n \in \mathcal{F}$ and $\theta_n \in \Theta \setminus \overline{\Theta}^I_{\mathcal{X}}(F_n)$.

(ii) The necessary laws of large numbers for triangular arrays hold to ensure that $|\widehat{\sigma}^2(\theta_n) - \sigma^2(\theta_n, F_n)| = o_p(1)$ over any sequence $F_n \in \mathcal{F}$ and $\theta_n \in \Theta$.

We describe such conditions in the Appendix. If they hold, then

$$\liminf_{n\to\infty} \inf_{\substack{\theta\in\Theta:\theta=\theta_0\\ F\in\mathcal{F}}} P_F\big(\theta \in \mathrm{CS}_n(1-\alpha)\big) \geq 1-\alpha.$$

In Appendix B we also study the power properties of our approach. Unlike methods that rely on one-sided L^p functionals (e.g., Lee, Song, and Whang (2011)) our approach is not guided by a *least favorable configuration*. In such settings test statistics are normalized by looking at the largest possible variance that would still be consistent with the inequalities. In our context this would amount to using a test statistic of the form

$$\widetilde{t}_n(\theta) = \frac{\sqrt{n}\widehat{T}_{\mathcal{X}}(\theta)}{\widehat{\Omega}(\theta)},$$

where $\widehat{\Omega}(\theta)$ is the estimator of $\widehat{\sigma}(\theta)$ that would result if the inequalities were *binding* almist surely (a.s.). To construct it we would replace each indicator function $\mathbb{1}\{\tau^p(Y^p|X; \theta^p) \geq 0\}$ with 1. By breaking away from least favorable configurations our procedure is, by construction, less conservative. The cost is having to introduce the tuning parameter κ_n. By design, our methodology is computationally simple to implement even in the presence of a rich parameterization and a large collection of conditioning covariates X. This computational simplicity also enables us to study the sensitivity of our results to various choices of the tuning (bandwidth) parameters involved. Computing the confidence set for different values of these parameters is a computationally costless exercise.

4.3 *On a nonparametric treatment of* η^p

As we stated previously, under assumptions such as symmetry of interaction effects, the strategic index $\eta^p(Y^{-p}; X)$ can be characterized simply as an aggregate function of Y^{-p} without the need for any parameterization. In a more general setting, the computational simplicity of our approach can allow us to handle a very rich and flexible parameterization of the strategic index. However, the use of a parametric approximation may shrink the identified set for η^p in ways that could be difficult to predict (see Ponomareva and Tamer (2011) for a discussion of related issues). To avoid these issues, a fully nonparametric treatment for η^p can be considered, and a sieves approximation (see Chen (2007)) seems appropriate since it allows us to impose shape or sign restrictions in addition to the conditional moment inequalities from Theorem 1. Due to its complexity, a complete characterization of its asymptotic properties is beyond the scope of this paper and is the current focus of ongoing work on a separate, more general paper on the subject of sieves-based nonparametric inference with conditional moment inequalities.

5. Monte Carlo experiments

In this section we apply our inferential approach to a Monte Carlo design described as follows. We consider a model of imperfect competition between three firms.

5.1 *Demand system*

We consider a model of imperfect competition with differentiated products. The system of (inverse) demand functions is of the form

$$\mathcal{P}^1 = \zeta^1 \cdot X_a - (\lambda^1 + \delta^1 \cdot X_b) \cdot Y^1 - (\beta^{12} + \gamma^{12} \cdot X_b) \cdot Y^2 - (\beta^{13} + \gamma^{13} \cdot X_b) \cdot Y^3,$$
$$\mathcal{P}^2 = \zeta^2 \cdot X_a - (\lambda^2 + \delta^2 \cdot X_b) \cdot Y^2 - (\beta^{21} + \gamma^{21} \cdot X_b) \cdot Y^1 - (\beta^{23} + \gamma^{23} \cdot X_b) \cdot Y^3,$$
$$\mathcal{P}^3 = \zeta^3 \cdot X_a - (\lambda^3 + \delta^3 \cdot X_b) \cdot Y^3 - (\beta^{31} + \gamma^{31} \cdot X_b) \cdot Y^1 - (\beta^{32} + \gamma^{32} \cdot X_b) \cdot Y^2.$$

The term Y^p refers to the quantity produced by firm p and X_a and X_b are demand shifters assumed to be observed by the econometrician and the firms. The rest of the terms are parameters that for the purposes of our experiment were chosen to be

$$\zeta^1 = \zeta^2 = \zeta^3 = 2, \qquad \lambda^1 = \lambda^2 = \lambda^3 = 0, \qquad \delta^1 = \delta^2 = \delta^3 = 1,$$
$$\beta^{12} = \beta^{21} = 0, \qquad \beta^{13} = \beta^{31} = \beta^{23} = \beta^{32} = 1,$$
$$\gamma^{12} = \gamma^{21} = \gamma^{23} = \gamma^{32} = 1, \qquad \gamma^{13} = \gamma^{31} = -5.$$

The demand shifters X_a and X_b are independently drawn from the distributions

$$X_a = \exp\{Z_a\}, \quad \text{where } Z_a \sim \mathcal{N}(0, 1), \quad \text{and} \quad X_b \sim U[0, 1].$$

5.1.1 *Strategy space* The *strategy space* is given by

$$\mathcal{Y}^p = \{0, 1, 2, \dots, 10\}.$$

Using our previous notation this means $M^p = 10$ for each p.

5.2 *Cost functions*

Cost functions are of the form

$$C^p(Y^p) = F^p \cdot \mathbb{1}\{Y^p > 0\} + \mu^p \cdot (X_{mc} + \varsigma^p) \cdot Y^p,$$

where X_{mc}, F^p, and ς^p are random cost shifters and μ^p is a parameter. Here X_{mc} is observed by the econometrician, but both F^p and ς^p are only *privately* observed by firm p. The marginal cost parameter μ^p was set to $\mu^p = 1/10$ for each p. The cost shifters F^p and ς^p are independently drawn from the distributions

$$F^p \sim U[1, 2], \qquad \varsigma^p \sim U[0, 1].$$

These private shocks are independent across $p = 1, 2, 3$. The observable cost shifter X_{mc} is drawn from a $U[0, 1]$ distribution, independent of $(F^p, \varsigma^p)^{p=1,2,3}$.

5.3 *Payoff functions*

Firms compete in quantities produced. Given our previous description of demand and costs, firms' profit (payoff) functions are given by

$$\begin{aligned}
&\nu^1(y^1, y^2, y^3; X, \varepsilon^1) \\
&\quad = 2X_a - X_b \cdot (y^1)^2 - F^1 \cdot \mathbb{1}\{y^1 > 0\} \\
&\qquad - \frac{1}{10} \cdot (X_{mc} + \varsigma^1) \cdot y^1 - \underbrace{\big[X_b \cdot y^2 + (1 - 5X_b) \cdot y^3\big]}_{=\eta^1(y^2, y^3; X)} \cdot y^1, \\
&\nu^2(y^2, y^1, y^3; X, \varepsilon^2) \\
&\quad = 2X_a - X_b \cdot (y^2)^2 - F^2 \cdot \mathbb{1}\{y^2 > 0\} \\
&\qquad - \frac{1}{10} \cdot (X_{mc} + \varsigma^2) \cdot y^2 - \underbrace{\big[X_b \cdot y^1 + (1 + X_b) \cdot y^3\big]}_{=\eta^2(y^1, y^3; X)} \cdot y^2, \\
&\nu^3(y^3, y^1, y^2; X, \varepsilon^3) \\
&\quad = 2X_a - X_b \cdot (y^3)^2 - F^3 \cdot \mathbb{1}\{y^3 > 0\} \\
&\qquad - \frac{1}{10} \cdot (X_{mc} + \varsigma^3) \cdot y^3 - \underbrace{\big[(1 - 5X_b) \cdot y^1 + (1 + X_b) \cdot y^2\big]}_{=\eta^3(y^1, y^2; X)} \cdot y^3.
\end{aligned} \tag{9}$$

The game is played simultaneously and, in accordance with the assumptions in previous sections, the outcome is a Bayesian Nash equilibrium induced by degenerate beliefs.

5.3.1 *Strategic indices, substitutability, and complementarity* It is easy to verify that the payoff functions described in (9) satisfy our Assumption 1. The strategic indices are given by

$$\begin{aligned}\eta^1(y^2, y^3; X) &= X_b \cdot y^2 + (1 - 5X_b) \cdot y^3, \\ \eta^2(y^1, y^3; X) &= X_b \cdot y^1 + (1 + X_b) \cdot y^3, \\ \eta^3(y^1, y^2; X) &= (1 - 5X_b) \cdot y^1 + (1 + X_b) \cdot y^2.\end{aligned} \tag{10}$$

The patterns of substitutability/complementarity that emerge from (10) can be summarized as follows:

(i) The (Y^1, Y^3) are *strategic complements* whenever $X_b > \frac{1}{5}$, which occurs with probability 80% since $X_b \sim U[0, 1]$.

(ii) The (Y^1, Y^2) and (Y^2, Y^3) are always *strategic substitutes*.

5.4 *Equilibrium selection rule*

The nature of the strategy space (discrete and bounded at $M^p = 10$) induces the existence of multiple equilibria. Whenever multiple equilibria exist we impose the following equilibrium selection rule.

(i) An equilibrium selection device ξ is randomly drawn from a $[0, 1]$ distribution. This draw is independent from all payoff shifters in the model.

(ii) If $\xi < \frac{1}{3}$, the equilibrium is selected completely at random from the existing equilibria.

(iii) If $\frac{1}{3} \leq \xi < \frac{2}{3}$, the equilibrium selected is the one that yields the largest combined profits for firms 1 and 3.

(iv) If $\xi \geq \frac{2}{3}$, the equilibrium selected is the one that yields the largest profits for firm 2.

As we remarked in the paragraph immediately following Theorem 1, the identification power of our procedure requires a nondegenerate equilibrium selected mechanism. The one described above is a particular instance of a nondegenerate selection mechanism.

In summary, the researcher observes (Y^1, Y^2, Y^3). In addition, the collection of covariates observed by the researcher is

$$X \equiv (X_a, X_b, X_{mc}),$$

and the unobserved payoff shifters are

$$\varepsilon^p \equiv (F^p, \varsigma^p), \quad p = 1, 2, 3.$$

Also in adherence to the assumptions of our model, ε^p is only privately observed by p.

5.5 *Summary of equilibrium features of the experimental data*

The existence of multiple equilibria was prevalent within our designs. Table 1 summarizes the number of equilibria observed in 500,000 simulations.

Table 1. Summary statistics for the cardinality of equilibria in our designs.

% of Games With One BNE	% of Games With Two BNE	% of Games With Three BNE	% of Games With Four or More BNE
24.7%	63.5%	7.7%	3.9%

Note: Results from 500,000 simulations.

Table 2. Average value of $\mathrm{Corr}(Y^p, \eta^p(Y^{-p}; X)|X)$.

Firm 1	Firm 2	Firm 3
−0.716	−0.384	−0.743

Note: Results from 500,000 simulations.

Our design also allows us to corroborate the negative association, conditional on X, between Y^p and the strategic index $\eta^p(Y^{-p} : X)$, which is at the center of our inference and identification results. Table 2 illustrates this important feature.

As a quick gauge of the identification power of our main result we can compare the correlations in Table 2 with those that would result from a function other than the true strategic index η^p. For example, take firm 1 and consider the (incorrect) index $g^1(Y^{-1}; X) = Y^2 + Y^3$. For this function, the average value of $\mathrm{Corr}(Y^1, g^1(Y^{-1}; X)|X)$ is 0.32, which would lead us to reject g^1 as the strategic index η^1. It is also useful here to show that looking at unconditional correlations, or at pairwise correlations between actions, can obscure the true strategic interaction features of the model. Note that in our designs, both Y^1 and Y^3 are always strategic substitutes for Y^2. However, the unconditional correlations are given by

$$\mathrm{Corr}\big(Y^2, Y^1\big) = 0.048, \qquad \mathrm{Corr}\big(Y^2, Y^3\big) = -0.412.$$

The fact that one of these correlations is positive could seem “counterintuitive,” but as the results in this paper demonstrate, this “intuition” is incorrect, as mere pairwise comparisons between strategies and/or unconditional correlations are not the right objects to look at. It is easy to show that slightly different values of the parameters in our Monte Carlo (MC) designs, which would render all actions pairwise substitutes, can also lead to positive pairwise unconditional correlations between *all* actions.[8] Intuitively, the sim-

[8]Take our MC designs and change the signs of γ^{13} and γ^{31} to $\gamma^{13} = \gamma^{31} = 5$. This ensures that all actions are pairwise strategic substitutes. Now reduce (but do not eliminate) the magnitude of the strategic effects by setting $\beta^{13} = \beta^{31} = \beta^{23} = \beta^{32} = 0$, and leave everything else unchanged. Even with the relatively smaller strategic effects, multiple equilibria would still be prevalent in the game, with an average number of equilibria per game being around 2.03. As our results predict, the conditional (on X) covariance between Y^p and $\eta^p(Y^{-p} : X)$ would be largely negative, with the average value of $\mathrm{Corr}(Y^p, \eta^p(Y^{-p}; X)|X)$ being equal to −0.455 for player 1, −0.309 for player 2, and −0.455 for player 3. Due to the smaller strategic interaction effect, these figures are smaller (in absolute value) than those in Table 2, but they are still clearly negative, reflecting the fact that multiple equilibria are still prevalent. The stronger relative effect of nonstrategic shifters in X as determinants of payoffs produces pairwise unconditional correlations that are

plest scenario where this may occur is when the magnitude of the strategic interaction effect is relatively minor compared to the effect of nonstrategic payoff shifters.

5.6 *Inference in Monte Carlo experiments*

Next we describe the econometric analysis of our Monte Carlo data.

5.6.1 *Parametric family for strategic interaction indices* We focus on a parameterization that is compatible with our true strategic indices, described in (10). Specifically we focus on a parametric family of the form

$$\eta^p\big(y^q, y^r; X\big) = \big(\beta^{pq} + \gamma^{pq} \cdot X_b\big) \cdot y^q + \big(\beta^{pr} + \gamma^{pr} \cdot X_b\big) \cdot y^r, \quad (p, q, r) \in \{1, 2, 3\}.$$

Let us group the parameters of the strategic indices as

$$\theta^p \equiv \big(\beta^{pq}, \gamma^{pq}, \beta^{pr}, \gamma^{pr}\big), \quad \text{and} \quad \theta \equiv \big(\theta^1, \theta^2, \theta^3\big).$$

Their *true* values are given by

$$\begin{aligned} &\theta_0^1 = (0, 1, 1, -5), && \theta_0^2 = (0, 1, 1, 1), \\ &\theta_0^3 = (1, -5, 1, 1), && \theta_0 \equiv \big(\theta_0^1, \theta_0^2, \theta_0^3\big). \end{aligned} \tag{11}$$

Parameter space Some of our inference experiments (and choice of tuning parameters) involved searches over a parameter space Θ. The parameter space used throughout was the Cartesian product given by the interval $[0, 4]$ for each β^{pq}, the interval $[-10, 0]$ for γ^{13} and γ^{31}, and the interval $[0, 4]$ for every other γ^{pq}.

5.6.2 *Kernels and bandwidths* To study their finite-sample properties, we employed kernels and bandwidths identical to those that we will use in our empirical application in Section 6. These are described in exact detail in the Supplement. As we describe there,[9] the kernel employed is of order $M = 18$, which is of higher order than needed given that the experimental data include three continuously distributed observable covariates $X \equiv (X_a, X_b, X_{mc})$. We opted to utilize this specific kernel because it is the one employed in our empirical application in Section 6, and we want to study its finite-sample properties in the context of our Monte Carlo experiments. Our target inference range included the entire data.

positive across all actions:

$$\mathrm{Corr}\big(Y^1, Y^2\big) = 0.668, \qquad \mathrm{Corr}\big(Y^1, Y^3\big) = 0.231, \qquad \mathrm{Corr}\big(Y^2, Y^3\big) = 0.668.$$

Furthermore, $\mathrm{Corr}(Y^p, \eta^p(Y^{-p}; X))$, the *unconditional* correlation between Y^p and $\eta^p(Y^{-p}; X)$, is equal to -0.052 for player 1, 0.477 for player 2, and -0.052 for player 3. The fact that this value is positive for player 2 highlights that, as we discussed in Section 3.2.1, while our results imply restrictions on the sign of the conditional covariance between Y^p and $\eta^p(Y^{-p}; X)$, they do not restrict the sign of the unconditional covariance between these variables.

[9]Some of the tuning parameters described in the Supplement involve searches over the parameter space Θ, which was described above for our Monte Carlo experiments.

Table 3. Observed frequency with which θ_0 was INCLUDED in our CS.

Sample Size	Target Coverage: 95% ($c_{1-\alpha} = 1.645$)	Target Coverage: 99% ($c_{1-\alpha} = 2.33$)	95th Percentile Observed Value of Test Statistic	Maximum Observed Value of Test Statistic
$n = 500$	99.9%	100%	1.074	1.344
$n = 1000$	99%	100%	1.291	2.163
$n = 1500$	98%	100%	1.485	2.162
$n = 2000$	94.9%	99.9%	1.647	2.352

5.6.3 *Inference exercises* Next we describe the specific features of our confidence sets (CS) that were studied in our experiments.

(A) Inclusion of the true parameter value θ_0 in our CS Our first exercise is to evaluate the ability of our procedure to include the true parameter value θ_0 in the CS. Given that our Monte Carlo experiments are designed to satisfy the assumptions underlying our econometric procedure, we know that asymptotically θ_0 will be included in our CS with probability at least $1 - \alpha$, where the latter represents our target coverage probability. Our first goal is to determine the ability of our procedure to accomplish this in finite samples. To this end we generated 1000 simulations of the Monte Carlo design described above for sample sizes $n = 500$, $n = 1000$, $n = 1500$, and $n = 2000$ (1000 simulations in each case) and computed the proportion of samples for which θ_0 was accepted into our CS. The results are shown in Table 3.

(B) Exclusion of false parameter values from our CS Our next exercise is aimed at studying the power of our econometric approach to reject false conjectures about strategic interaction. Suppose we maintain the (incorrect) assumption that the goods produced by the three firms are always strategic substitutes. Under this maintained assumption we test the power of our approach to reject the following three conjectures:

(i) *Symmetric and constant strategic interaction effects.* Our first false parameter value is aimed at testing the conjecture that strategic effects are symmetric and constant across markets and across players. Under the maintained assumption of strategic substitutes, this is equivalent to testing whether the parameter value belongs in our CS:

$$\theta_a^p \equiv \left(\beta_a^{pq}, \gamma_a^{pq}, \beta_a^{pr}, \gamma_a^{pr}\right) = (1, 0, 1, 0), \quad \text{with } \theta_a \equiv \left(\theta_a^1, \theta_a^2, \theta_a^3\right).$$

(ii) *Symmetric interaction effects equal to zero if $X_b = 0$.* Our next false tests the conjecture that strategic effects are once again symmetric across players, but they depend on X_b in a way such that if $X_b = 0$, then there is no strategic effect. Under the maintained assumption of strategic substitutes, this is equivalent to testing whether the parameter value belongs in our CS:

$$\theta_b^p \equiv \left(\beta_b^{pq}, \gamma_b^{pq}, \beta_b^{pr}, \gamma_b^{pr}\right) = (0, 1, 0, 1), \quad \text{with } \theta_b \equiv \left(\theta_b^1, \theta_b^2, \theta_b^3\right).$$

Table 4. Observed frequency with which false parameter values were EXCLUDED from our CS.

Sample Size	Target Coverage: 95% ($c_{1-\alpha}=1.645$)	Target Coverage: 99% ($c_{1-\alpha}=2.33$)	95th Percentile Observed Value of Test Statistic	Maximum Observed Value of Test Statistic
	Observed frequency with which θ_a was EXCLUDED from our CS			
$n=500$	54.3%	45.2%	7.324	11.811
$n=1000$	66.3%	57.7%	16.537	15.598
$n=1500$	77.6%	70.0%	13.977	21.931
$n=2000$	81.4%	76.4%	17.410	22.769
	Observed frequency with which θ_b was EXCLUDED from our CS			
$n=500$	57.5%	48.3%	7.746	11.498
$n=1000$	69.2%	61.1%	11.652	15.088
$n=1500$	79.7%	73.3%	15.113	19.838
$n=2000$	84.6%	79.1%	18.199	22.386
	Observed frequency with which θ_c was EXCLUDED from our CS			
$n=500$	54.6%	46.2%	7.488	9.431
$n=1000$	67.4%	59.7%	11.056	15.718
$n=1500$	78.8%	71.7%	14.653	20.546
$n=2000$	83.0%	77.3%	17.774	20.966

(iii) *Symmetric interaction effects with $\beta^{pq}=\gamma^{pq}$*. As a third case we aim to test the conjecture that strategic effects are symmetric across players and they satisfy $\beta^{pq}=\gamma^{pq}$ for each p and q. Under the maintained assumption of strategic substitutes, this amounts to testing whether the parameter value belongs in our CS:

$$\theta_c^p \equiv \left(\beta_c^{pq}, \gamma_c^{pq}, \beta_c^{pr}, \gamma_c^{pr}\right) = (1,1,1,1), \quad \text{with } \theta_c \equiv \left(\theta_c^1, \theta_c^2, \theta_c^3\right).$$

Asymptotically each one of these parameter values will be excluded from our CS with probability 1. Our goal here is to study the ability of our econometric procedure to reject these false conjectures by excluding the corresponding parameter values from our CS in *finite* samples. To this end we generated 1000 simulations of the Monte Carlo design described above for sample sizes $n=500$, $n=1000$, $n=1500$, and $n=2000$ (1000 simulations in each case) and computed the proportion of samples for which θ_a, θ_b, and θ_c were *rejected* from our CS. The results are summarized in Table 4.

(C) Exploring other features and conjectures of our CS Suppose we maintain the following (correct) conjectures about θ_0:

(i) We have $\beta^{pq}=\beta^{qp}$ and $\gamma^{pq}=\gamma^{qp}$ for each p, q. We impose this restriction on the parameter space Θ.

(ii) We have that (Y^1, Y^2) and (Y^2, Y^3) are always strategic substitutes. This can be captured by imposing the following restrictions on Θ:

$$\beta^{12}\geq 0, \gamma^{12}\geq 0, \qquad \beta^{23}\geq 0, \gamma^{23}\geq 0, \quad \text{with } \left|\beta^{12}\right|+\left|\gamma^{12}\right|>0, \left|\beta^{23}\right|+\left|\gamma^{23}\right|>0.$$

(iii) We have that (Y^1, Y^3) are strategic substitutes for "small" values of X_b but they are strategic complements for "large" values of X_b. However, we do not know the corresponding threshold for X_b that induces complementarity. This can be captured by imposing the following restrictions on Θ:

$$\beta^{13} > 0, \qquad \gamma^{13} < 0.$$

Given these maintained assumptions, suppose we want to study the following two strategic interaction features:

- The relative effects of firms 1 and 3 on firm 2.
- How large X_b has to be to induce strategic complementarities between firms 1 and 3.

Both features will be analyzed by estimating a CS after imposing the restrictions on Θ described above. Our construction involved a grid search over two million points in Θ. The results described below represent the results for a randomly drawn sample generated according to our previous description. Our target coverage probability was 95%.

Relative effects of firms 1 and 3 on firm 2 We have maintained that both Y^1 and Y^3 are always strategic substitutes for Y^2. We want to explore which one of the two rivals of firm 2 has a larger strategic effect on firm 2's payoff. In this sense, we want to study which one of firm 2's rivals is a closer competitor to firm 2. Given our parameterization of the strategic indices, this question involves a comparison between $\beta^{21} + \gamma^{21} \cdot X_b$ and $\beta^{23} + \gamma^{23} \cdot X_b$. So as to aggregate X_b, we will compare[10]

$$E_{X_b}[\beta^{21} + \gamma^{21} \cdot X_b] = \beta^{21} + \gamma^{21} \cdot \frac{1}{2}$$

and

$$E_{X_b}[\beta^{23} + \gamma^{23} \cdot X_b] = \beta^{23} + \gamma^{23} \cdot \frac{1}{2}.$$

Note that the true values of these quantities are

$$E_{X_b}[\beta^{21} + \gamma^{21} \cdot X_b] = \frac{1}{2} \quad \text{and} \quad E_{X_b}[\beta^{23} + \gamma^{23} \cdot X_b] = \frac{3}{2}$$

and therefore firm 3 is the true closest competitor to firm 2. Figure 1 shows that our CS overwhelmingly reflects this key property. It also illustrates how our CS contains the true value of these parameters.

Strategic complementarities between firms 1 and 3 We have maintained the conjecture that Y^1 and Y^3 are strategic complements for "large" values of X_b. Let x^* denote the value such that Y^1 and Y^3 are strategic complements whenever $X_b > x^*$ and are substitutes otherwise. This threshold is given by

$$x^* = -\frac{\beta^{13}}{\gamma^{13}}.$$

[10]Recall that $X_b \sim U[0, 1]$.

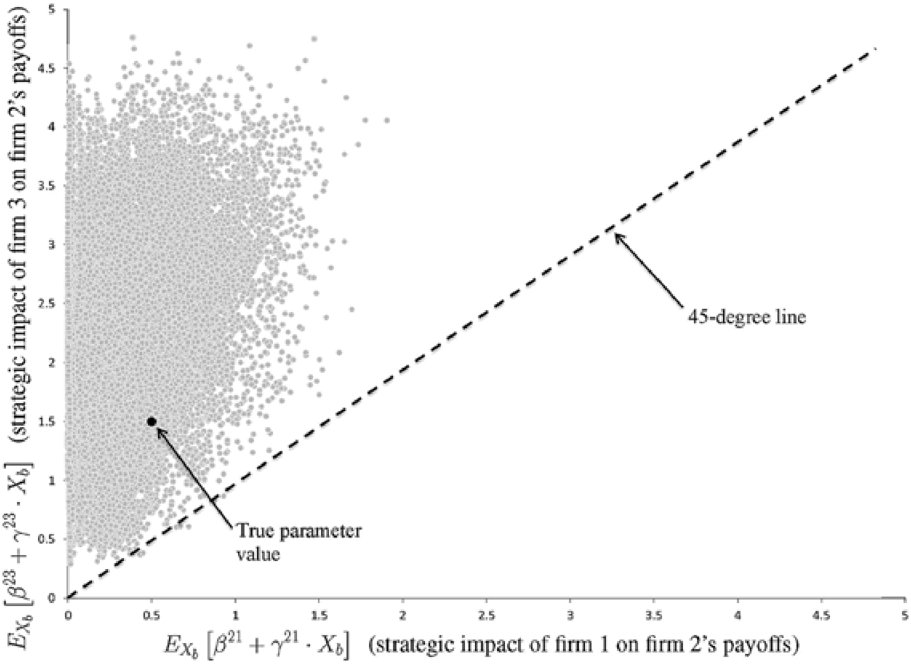

Figure 1. Relative effects of firms 1 and 3 on firm 2: 95% joint CS for $E_{X_b}[\beta^{21} + \gamma^{21} \cdot X_b]$ and $E_{X_b}[\beta^{23} + \gamma^{23} \cdot X_b]$.

Note that $x^* = 0.20$ given the parameters of our Monte Carlo design. We can use our 95% CS to construct a corresponding confidence interval (CI) for x^*. This is given by

$$95\% \text{ CI for } x^* : [0.003, 0.384].$$

The true value $x^* = 0.2$ is practically the midpoint of our CI.

Violations to our assumptions The Supplement continues our Monte Carlo analysis by describing the performance of our method under certain violations to our assumptions. Specifically, we introduce correlation in players' privately observed shocks and we also allow the strategic index to be a function of unobserved payoff shifters (and not only of X). For our designs, we find that our method still performs relatively well (i.e., the true parameter value is still contained in our CS with a frequency not too far from our target coverage probability) as long as the magnitude of correlation between these privately observed shocks remains relatively small, but that our results break down (i.e., the true parameter value is excluded from our CS with greater probability) as the magnitude of this correlation increases.

6. Application: Entry in the U.S. drugstore industry

Next we include an illustration of our results in the analysis of empirical data. One of the most important econometric applications of games has been the study of entry decisions by competing firms. Our model allows us to approach this problem by combining the usual extensive-margin enter/not enter dimension with an intensive-margin decision regarding the intensity of entry. In our application, this intensive margin is captured by the number of stores that a chain store decides to open in a market. The key advantage of taking the intensive margin into account is that it will give us a structural interpretation of the strategic index in terms of an underlying model of supply and demand (see Section 2.3). This stands in contrast to the "reduced form" profit function that dominates applied work on the binary entry margin. As we show below, the intensive margin provides new insights into the nature of competition in the market we study. It is important to note that our assumptions are compatible with the existence of fixed costs of entry and thus our model strictly nests the binary entry case (see Section 2.3).

Our application focuses on the U.S. retail drugstore industry, which we study because of three different considerations. First, it is an industry with three clearly identifiable main competitors: Walgreen's, CVS, and Rite Aid. According to *IBISWorld*, their market shares in 2011 were approximately 31%, 26%, and 12%, respectively.[11] Second, there has been a recent discussion among industry watchers of a takeover of Rite Aid by one of its competitors. This is a natural policy application for us since our approach can help us identify, for example, which is the closest competitor to Rite Aid. Third, we believe it is a case of an industry without an obvious, compelling demand side unobservable at the market level (i.e., an unexplained taste for health) that cannot be conditioned out with observables (such as the number of doctors in the market).

Naturally, entry takes place at different points in time but these dates are largely unobserved in our data set. Our justification for modeling this as a static game is the commonly made assumption that the choices observed are the realization of a long-run equilibrium whereby firms precommitted to their strategies before observing the strategies of others. According to this view, the fact that entry decisions take place at different points in time is incidental.

Throughout our exercise we identify these three players as

player 1: CVS, player 2: Rite Aid, player 3: Walgreens.

The term p denotes generically any one of the three players in the model, and q, r denote the two opponents of p. Let Y^p denote the number of stores opened by p in a market. Please note that we have abstracted away the competition effect these firms may face from supermarket and big-box stores (see Ellickson and Grieco (2013)), which is likely to be significant in items such as beauty products, personal care items, and over the counter medications, but less so in prescription medications and walk-in health services, although this may change in the future as Walmart and other supermarkets are added to increasingly more preferred pharmacy networks and expand their clinic services. We do this for simplicity of our empirical illustration and because we wanted to focus on the three closest competitors within the drugstore industry.

[11]Source: http://clients1.ibisworld.com/reports/us/industry/default.aspx?entid=1054.

Table 5. Summary statistics for Y^p.

	Y^1	Y^2	Y^3
Total†	7004	4318	7283
Mean	7.34	4.52	7.63
Stdev	21.95	15.57	23.88
25th Percentile	0	0	1
Median	1	0	1
75th Percentile	4	3	4
90th Percentile	16	10	17
95th Percentile	39	21	41
99th Percentile	112	71	106

Note: † Total number of stores in all markets. All other statistics shown are at the market level.

Table 6. Correlations observed for Y^1, Y^2, and Y^3.

	Y^1	Y^2	Y^3	Y^2+Y^3	Y^1+Y^3	Y^1+Y^2
Y^1	–	0.70	0.79	0.86	–	–
Y^2	0.70	–	0.49	–	0.62	–
Y^3	0.79	0.49	–	–	–	0.72

6.1 *Data overview*

6.1.1 *Units of observation* The decision variable Y_i^p denotes the total number of stores by p in market i *in the year 2011*. We define a market as a CBSA (core-based statistical area) in the continental United States. Metropolitan[12] CBSAs were split into the divisions determined by Office of Budget and Management and each division was considered a market. We exclude CBSAs with more than 5 million people because such large markets will likely consist of smaller submarkets. Our final sample consists of $N = 954$ observations.

6.1.2 *Choices and outcomes observed in the data* Table 5 summarizes some descriptive features of choices observed. It highlights the richness of the action space in this application. Table 6 shows the correlations observed across Y^1, Y^2, and Y^3. As we see there, a persistently positive association was observed across markets in the number of stores opened by each competitor. What is remarkable is that this pattern of positive correlation remains the same order of magnitude even after we condition on market observables such as market size aqnd so forth (we describe the market covariates in further depth in the next subsection).

[12]The Office of Budget and Management defines a CBSA as an area that consists of one or more counties and includes the counties containing the core urban area, as well as any adjacent counties that have a high degree of social and economic integration (as measured by commuting to work) with the urban core. Metropolitan CBSAs are those with a population of 50,000 or more. Under certain conditions, metropolitan CBSAs with 2.5 million people or more are split into divisions.

Table 7. Correlations if the game is reduced to binary choice.

	$\mathbb{1}\{Y^1 > 0\}$	$\mathbb{1}\{Y^2 > 0\}$	$\mathbb{1}\{Y^3 > 0\}$
$\mathbb{1}\{Y^1 > 0\}$	–	0.23	0.07
$\mathbb{1}\{Y^2 > 0\}$	0.23	–	0.04
$\mathbb{1}\{Y^3 > 0\}$	0.07	0.04	–

By their nature, the drugstores of each of these competitors provide the same type of services and can be rightly deemed, in general, as demand substitutes of each other. Given this observation and recalling the underlying Cournot model discussed in Section 2.3, basic economic theory would predict that, all else equal, more aggressive entry by a competitor would reduce a firm's marginal benefit to entry, leading us ex ante to consider entry decisions as strategic substitutes. Strategic substitution is assumed in numerous empirical applications of entry games (e.g., Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Berry (1992), Davis (2006)). Even though strategic substitutability is justified as the prediction of economic theory in our setting, the correlation pattern in Table 6 seems to fly in the face of it. This is especially true if we believe that there is no obvious, compelling demand side unobservable at the market level (i.e., an unexplained taste for medical drugs). Our framework can help us explore whether a model of strategic substitutes can produce this pattern of positive correlation in entry behavior. In our Monte Carlo experiments in Section 5, we learned that it is not hard to characterize multiplayer structural models, satisfying all our assumptions, where the underlying game is one of strategic substitutes and nevertheless all actions are pairwise correlated. See in particular the design described in footnote 8.

Ignoring the intensive-margin dimension of entry and focusing only on the binary choice decision of entry immediately obscures key features of the data. For example as Table 7 shows, it wipes out much of the positive association observed in the data.

By eliminating much of the positive association observed in the intensive margin, reconciling the data with an underlying game of strategic substitutes should be easier in a binary choice representation of the game compared to one that explicitly considers the intensive-margin decisions. A consequence of this would be that the inferential results for η^p in the latter case would be more precise. Our results will confirm this.

6.1.3 *Covariates included in X* Markets are defined as CBSAs with less than 5 million people. We included in our analysis the following market and player characteristics:

$$\text{POP} = \text{population}, \qquad \text{INC} = \text{average income per household},$$

$$\text{DENS} = \text{population density}, \qquad \text{AGE} = \text{median age in the population},$$

$$\text{BUS} = \text{total number of business establishments},$$

$$\text{DIST}^p = \text{distance to the nearest distribution center of } p,$$

TABLE 8. A statistical summary of covariates and market structure.

	Median Value in Markets With Zero Stores	Median Value in Markets With at Least One Store	Median Value in Markets With at Least Two Stores
POP	26,924	79,792	109,393
INC†	31,089	35,516	36,282
DENS‡	29.23	96.65	120.87
AGE	36.15	36.90	36.80
BUS§	621	1934	2607
DIST^1	381.87	151.19	124.78
DIST^2	669.65	186.09	150.02
DIST^3	280.93	141.94	124.14

Note: † Average income per household at the market level. Measured in 2000 U.S. dollars. ‡ Population per square mile. § Total number of business establishments in the market. All distances are measured in miles.

and we used

$$X = (\text{POP}, \text{INC}, \text{DENS}, \text{AGE}, \text{BUS}, \text{DIST}^1, \text{DIST}^2, \text{DIST}^3).$$

Population density was computed as the ratio of population/land area; X was treated as jointly continuously distributed.

Most of these covariates are fairly standard in empirical work. We do note that our inclusion of the number of business establishments (which we could empirically refine to be the number of retail establishments) is designed to control for supply side unobservables in a market. If it is just costly to locate a store in a market (because of say zoning restrictions), then this should affect the entry of stores in all industries, not just pharmacies. Table 8 presents summary statistics for our covariates and makes a comparison across different markets depending on the number of total stores. The table highlights the importance in particular of market size and density, as well as distance to distribution centers, as determinants of entry. To the extent that the covariates included in X may fail to fully control for unobservable market-level (demand or entry-cost) shocks and therefore leave some correlation in firms' unobserved payoff shocks, our results in Section 5 suggest that as long as the remaining correlation is relatively small, the properties of our inferential procedure can remain approximately valid in finite samples.

6.2 *Specifications for the strategic index* η^p

We refer generically to the three players as p, q, and r, and we consider specifications for the index of the form

$$\eta^p(y^{-p}; X|\theta^p) = (X'\theta^{p,q}) \cdot y^q + (X'\theta^{p,r}) \cdot y^r.$$

As we discussed above, we will maintain that actions are strategic substitutes. To this end we will choose the Θ such that strategic substitutability is ensured, that is,

$$X_i'\theta^{p,q} \geq 0, \qquad X_i'\theta^{p,r} \geq 0 \quad \forall i = 1, \dots, n\, \forall \theta \in \Theta.$$

We want to focus on simple specifications for the indices $X_i'\theta^{p,q}$ and $X_i'\theta^{p,r}$. Since θ can only be (partially) identified up to scale and location normalizations, these are also introduced in the parameter space in ways that will be described below.

Specification 1. Symmetry of opponents' strategic interaction effects First we study the special case where each p weighs the actions of his two opponents equally (a maintained, key assumption in De Paula and Tang (2012)) in every market. Given our assumptions, this is observationally equivalent to a strategic index of the form

$$\eta^p(y^{-p}; X|\theta^p) = \theta^p \cdot (y^q + y^r), \quad \text{where } \theta^p = 1.$$

In this case our inferential problem simply reduces to a specification test where we evaluate whether

$$\begin{aligned} &E[\mathbb{1}\{Y^p \leq y^p\} \cdot (Y^q + Y^r)|X = x] \\ &\quad \geq E[\mathbb{1}\{Y^p \leq y^p\}|X = x] \cdot E[(Y^q + Y^r)|X = x] \end{aligned} \tag{12}$$

for almost every (x, y^p) in our inferential range (to be described below).

Specification 2. Constant, possibly asymmetric relative strategic interaction effects Next we focus on the case where p may assign different weights to each opponent, but the relative effects remain constant across all markets. Letting $\theta^p = (\theta^{p,q}, \theta^{p,r})$, the strategic index is now of the form

$$\eta^p(y^{-p}; X|\theta^p) = \theta^{p,q} \cdot y^q + \theta^{p,r} \cdot y^r, \quad \text{where } \theta \geq 0\ \forall \theta \in \Theta. \tag{13}$$

We normalize Θ so that $\|\theta^p\| = 1$ for each p since our identified set is closed under nonnegative rescaling of θ^p (if θ satisfies (5), then so will $c \cdot \theta$ for any $c \geq 0$). This specification is of particular interest because strategic interaction effects have been typically modeled through constant coefficients in existing work that uses "reduced form" profit functions (e.g., Berry (1992), Tamer (2003) and many others).

Specification 3. A more flexible parameterization Here we allow for asymmetry and for covariate-dependent relative interaction effects. In our specification we express η^p solely as a function of market size (POP) and its distance to the nearest distribution center of each player ($\text{DIST}^1, \text{DIST}^2, \text{DIST}^3$). We wish to explore two conjectures through our parameterization:

(i) The difference in distance to the market ($\text{DIST}^p - \text{DIST}^q$) is a determinant of the strategic interaction effect of q on p. The basis for this effect is that if firm q's distribution center is located much closer than p's, then this will give q a cost side advantage relative to p in the market and thus make competition more intense with firm q's entry into the market.

(ii) Strategic interaction effects change with market size. One strand of the entry literature has modeled firm profits using "per capita" variable profits (see, e.g., Bresnahan and Reiss (1991a, 1991b)) , which would imply that the sensitivity of a firm's profits to another firm's entry is increasing with market size all else equal. However one can also imagine that larger markets offer more "room" for entry not just because there exist more people, but also because opportunities for market expansion relative to business stealing are larger, which would decrease the sensitivity of profit to a rival firm's entry in larger markets.

To explore both conjectures simultaneously we use the following parameterization of η^p. Denote $\theta^p = (\theta_1^p, \theta_2^p, \theta_3^p, \theta_4^p)'$ and $D^{p,q} \equiv \text{DIST}^p - \text{DIST}^q$ for every $p \neq q$. Define

$$\begin{aligned}
&\phi^{p,q}(X|\theta^p) \\
&\quad = \left(\theta_1^p + \theta_2^p \cdot \frac{1}{\text{POP}} + \theta_3^p \cdot \left(D^{p,q} - 200\right) \cdot \mathbb{1}\{D^{p,q} \geq 200\}\right. \\
&\qquad \left. + \theta_4^p \cdot \frac{\left(D^{p,q} - 200\right) \cdot \mathbb{1}\{D^{p,q} \geq 200\}}{\text{POP}}\right).
\end{aligned} \tag{14}$$

Population is measured in units of 500,000 inhabitants in (14). The strategic index for p is specified as

$$\eta^p\left(y^{-p}; X|\theta^p\right) = \phi^{p,q}\left(X|\theta^p\right) \cdot y^q + \phi^{p,r}\left(X|\theta^p\right) \cdot y^r. \tag{15}$$

Strategic substitutability is imposed by forcing the parameter space Θ to satisfy $\phi^{p,q}(X_i|\theta^p) \geq 0$ for each p, q and every market $i = 1, \ldots, n$. The individual signs of each coefficient were otherwise unrestricted. For the same reason given above we normalize $\|\theta^p\| = 1$ for $p = 1, 2, 3$ in our parameter space.

6.3 *Results*

Our target coverage probability is 95% throughout. Our parameter space Θ consisted of 1 million grid points with the scale normalization described above. An empty confidence set (CS) amounts to a rejection of the specification in question. The kernels and bandwidths used are described in detail in the Supplement. The kernel employed was bias-reducing of order 18, similar to the one used in Aradillas-López, Gandhi, and Quint (2013). Our bandwidths were of the form $h_n = c \cdot \widehat{\sigma}(X) \cdot n^{-\alpha_h}$ (we used individual bandwidths for each X, each proportional to $\widehat{\sigma}(X)$), $b_n = c_b \cdot \overline{\Omega} \cdot n^{-\alpha_b}$, and $\kappa_n = c_\kappa \cdot \overline{\Omega} \cdot \log(n)^{-1}$, where $\overline{\Omega} = \max_{\theta \in \Theta} |\widehat{\sigma}(\theta)|$. We chose these tuning parameters proportional to $\overline{\Omega}$ to ensure our procedure has a scale-invariant property. The choice of the constants c, c_b, c_κ, α_h, and α_b is described in the Supplement. For our sample size $n = 954$ the values of these tuning parameters were $h_n \approx 0.16 \cdot \widehat{\sigma}(X)$, $b_n \approx 10^{-5}$, and $\kappa_n \approx 10^{-7}$. The inference range used was

$$\mathcal{X} = \left\{x : \widehat{f}_X(x) \geq \widehat{f}_X^{(0.15)}, \text{POP} < 5 \text{ million}\right\},$$

where $\widehat{f}_X^{(0.15)}$ denotes the estimated 15th percentile of the density $\widehat{f}_X$. All of our main findings were robust to moderate changes in the tuning parameters used.

6.3.1 *Rejection of symmetry and of constant strategic interaction effects* Symmetry in the effects of opponents' actions on payoffs was rejected by our results. The value of our test statistic for testing (12) was 10.44, well above the critical value (1.645) for a 95% significance level. We conclude that if strategic substitutability is maintained across all markets, at least one player must assign different weights to the actions of his opponents in a subset of markets. Our results also rejected Specification 2 which assumed constant strategic effects. The smallest value of the test statistic across our parameter space Θ was 8.38, leading to an empty confidence set. Rejection of constant strategic effects is a relevant empirical finding because this is the type of specification used in the vast majority of existing parametric models. By Remark 2, rejecting any specification leads us to reject the assertion that the underlying game has a unique equilibrium w.p.1. In particular we reject the notion that there is no strategic interaction effect between the firms.

6.3.2 *Results for Specification 3* Our third specification produced a nonempty CS. Our first finding was a rejection of the assertion that $\theta^p = \theta^q$ for each $p \neq q$ (symmetry in parameters for all players). When we imposed this restriction we obtained an empty CS, with the smallest value of the test statistic being 2.01. Thus there is evidence of structural differences in payoff functions across these three players. We describe next the main features of the CS obtained.

6.3.3 *Evidence of asymmetric weights to opponents' strategies* Asymmetry of opponents' interaction effects is captured by the parameters θ_3^p and θ_4^p. Symmetry would hold for p in every market only if these parameters are jointly equal to zero. Figure 2 depicts the 95% joint CS for these parameters for each of the three players. As we can see, our results showed evidence of asymmetry for player 2 (Rite Aid).

We can study the asymmetry of strategic effects for specific markets. For example, Figure 3 depicts our confidence region for $\phi^{2,1}(X_i|\theta^2)$ (the effect of CVS on Rite Aid) and $\phi^{2,3}(X_i|\theta^2)$ (the effect of Walgreens on Rite Aid) corresponding to CBSA 29404 (Lake County–Kenosha County, IL–WI), where $\text{POP} = 820K$, $\text{DIST}^1 = 191$, $\text{DIST}^2 = 226$, and $\text{DIST}^3 = 21$. Our results show that, from the perspective of Rite Aid, the competition effect from Walgreens is stronger than the competition effect from CVS in that market.

We wanted to learn more about what the data revealed regarding the closeness of competition between rival firms. Since symmetry could only be rejected for Rite Aid we focused only on this firm. We say that the competition effect from CVS is stronger than that of Walgreens in market i if $\min_{\theta^2 \in \text{CS}_n(1-\alpha)}(\phi^{2,1}(X_i|\theta^2)) > \max_{\theta^2 \in \text{CS}_n(1-\alpha)}(\phi^{2,3}(X_i|\theta^2))$. The opposite would be true if the inequality holds with the superscripts 1 and 3 interchanged. We found that while the competition effect from CVS was stronger than that of Walgreens only in 9 markets, the opposite was true in 160 markets. Overall, our results provide evidence that Walgreens is a closer competitor to Rite Aid than CVS is. For policy purposes this closeness in competition could suggest that a merger between Rite Aid and Walgreens could potentially have a more significant anticompetitive effect than a merger between Rite Aid and CVS.[13]

[13] Rite Aid shares jumped sharply on March 14, 2012 following speculation from a Credit Suisse analyst about a potential merger with Walgreens (source: *New York Times*).

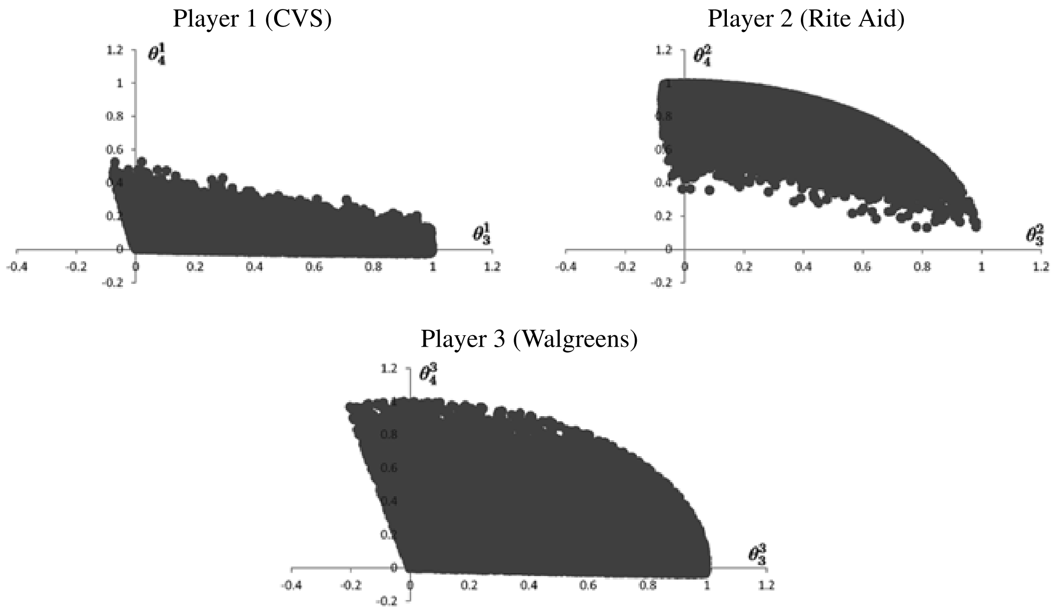

FIGURE 2. Asymmetry of strategic interaction effects; 95% joint CS for θ_3^p and θ_4^p.

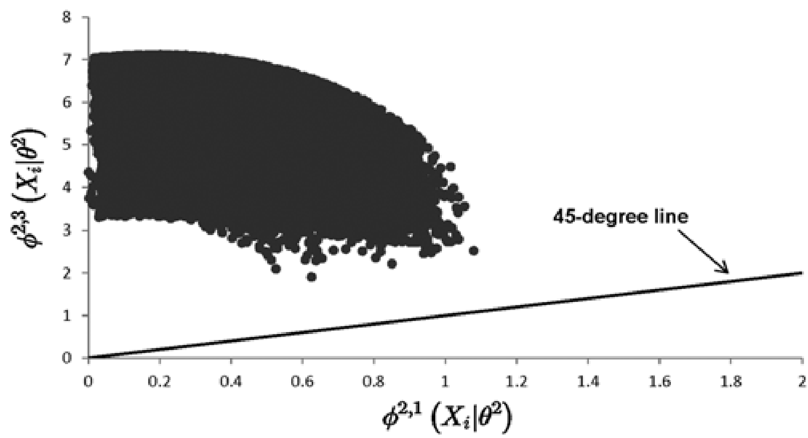

FIGURE 3. The CS for $\phi^{2,1}(X_i|\theta^2)$ and $\phi^{2,3}(X_i|\theta^2)$, for market $i =$ CBSA 29404 (Lake County–Kenosha County, IL–WI).

6.3.4 *Market size and strategic interaction* One of the goals of Specification 3 was to study the relationship between strategic interaction and market size. Positive signs for θ_2^p and θ_4^p would be consistent with interaction effects that decrease with the size of the market. Figure 4 depicts the 95% joint CS for these parameters for each firm. As we see there most of the values included in our CS for both coefficients are positive. Some negative values (except for θ_2^4) are included, but these are relatively small in absolute value.

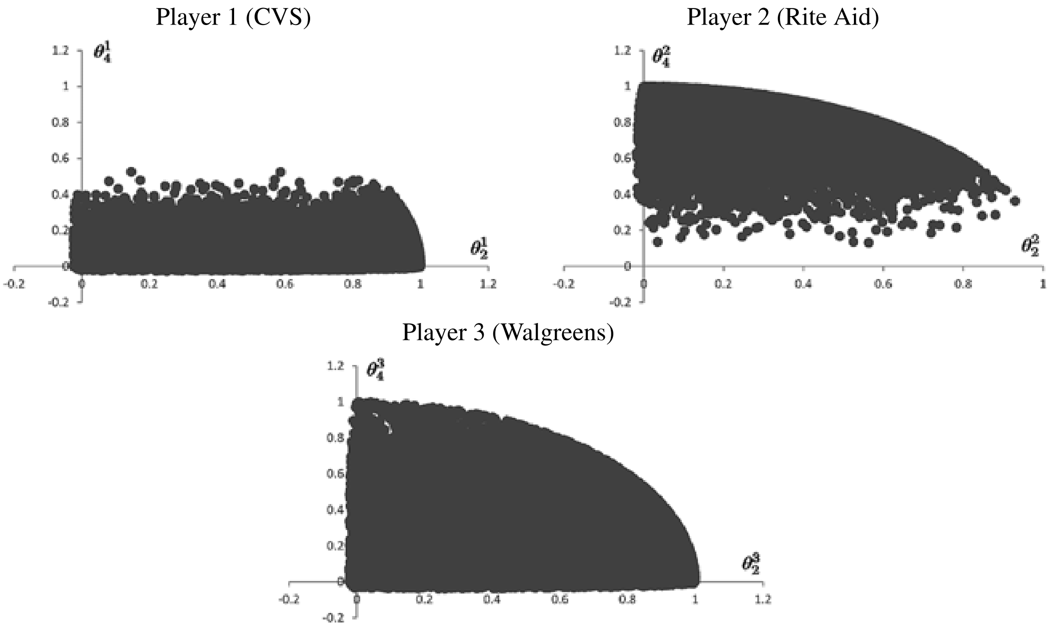

FIGURE 4. Market size and strategic interaction; 95% joint CS for θ_2^p and θ_4^p.

Let us focus on cases where relative distance is not significant (i.e., less than 200 miles) and the only determinant of strategic interaction is market size. In any such market the strategic coefficients are $\phi^{p,q}(X|\theta^p) = \theta_1^p + \theta_2^p \cdot \frac{1}{\text{POP}}$. Figure 5 shows how these strategic coefficients change with the size of the market. As we can see there, our results suggest that the strategic effect of opponents' strategies does not increase with market size and in fact could be less significant in larger markets.

6.3.5 *Evidence of multiple equilibria and nondegenerate equilibrium selection* By the arguments in Remark 2, the rejection of our first two specifications along with the parameter values that were rejected in our third specification are findings that are consistent with the existence of multiple equilibria in the underlying game and with an equilibrium selection mechanism that randomizes across these equilibria.

6.3.6 *Potential impact of unobserved market-level shocks* As our Monte Carlo results suggested (see in particular footnote 8), positive correlation between observed choices in a game that is assumed to be of strategic substitutes should not be automatically attributed to the presence of correlation in players' unobserved payoff shifters. In our empirical application we have tried to include multiple market-level relevant covariates in X. If there were to remain unobserved market-level heterogeneity that is publicly known to firms this would violate our independent private shocks assumption. However, consistent with the Monte Carlo findings included in the Supplement, our conjecture is that the asymptotic predictions in our model would remain approximately valid in the sample observed provided that the degree of correlation induced in firms' private shocks is relatively minor once we control for X. By the arguments in Remark 2, the

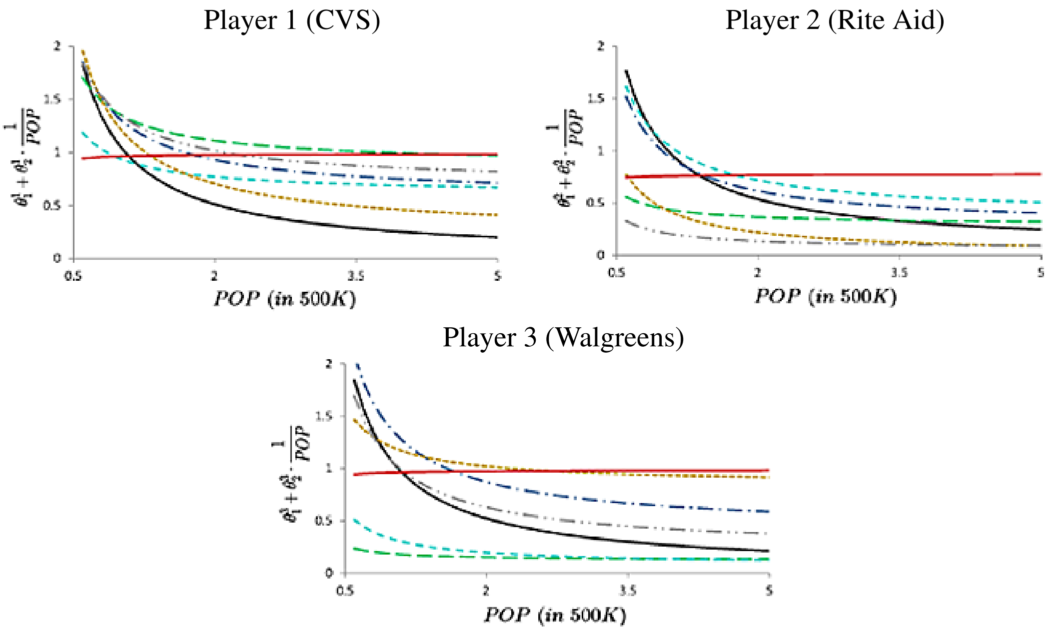

FIGURE 5. The $\theta_1^p + \theta_2^p \cdot \frac{1}{\text{POP}}$ for a range of POP values (measured in $500K$). The solid black line depicts the results for the largest value of θ_2^p in our CS. The solid horizontal line depicts the results for the smallest value of θ_2^p in our CS. Dotted lines correspond to five randomly drawn parameter values within our CS.

fact that Specification 3 was not rejected implies, in turn, that the independent private shocks assumption is not rejected by the data in this case. While this is not a definitive proof of the validity of this assumption, it does suggest that it is a reasonable approximation in this example.

6.4 *Results from modeling entry as a binary decision*

As Table 7 shows, much of the positive correlation in the intensive margin goes away when we look only at extensive margin decisions. This led us to conjecture that the range of models that would be consistent with strategic substitutes and with the choices observed would be larger if we limited attention to a binary choice representation of entry decisions. This intuition was confirmed by our methodology. While symmetry of weights to opponents (Specification 1) and constant relative interaction effects (Specification 2) were still rejected, modeling entry as a binary choice decision resulted in larger confidence sets in Specification 3. Furthermore, the closeness in competition between Rite Aid and Walgreens that our results uncovered was no longer evident. Specifically, as Figure 6 shows we now failed to reject that $\theta_3^p = \theta_4^p = 0$ for Rite Aid. Thus, we failed to reject that Rite Aid gives equal weights to both opponents across all markets. Hence, we conclude that key features of strategic interaction that are captured by intensive margin strategies are obscured if we focus attention solely on binary entry/no entry decisions.

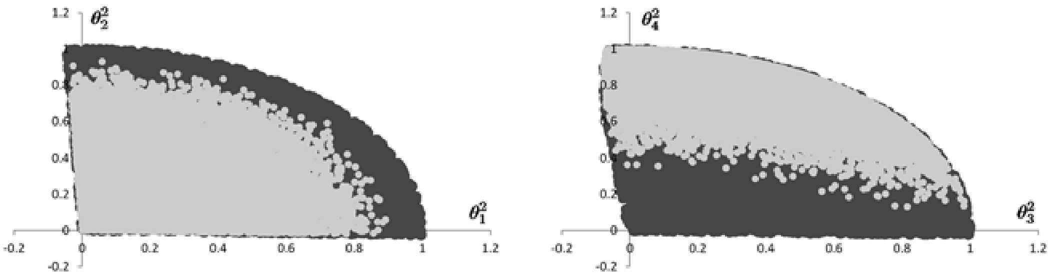

Figure 6. Comparing our previous results (shown in clear gray) with the CS from a binary choice entry model (shown in dark gray). Player 2 (Rite Aid).

7. Concluding remarks

We studied static games with very general strategy spaces. Making some general shape restriction assumptions on the underlying payoff functions we were able to characterize observable implications that allow us to do inference on the strategic interaction component that captures economically relevant features of strategic interaction. We showed how our assumptions can arise naturally in well known structural economic models. Our testable implications involve inequalities of nonlinear transformations of conditional moments. We introduced an econometric approach to do inference in this setting that is computationally easy to implement even in richly parameterized models with a large collection of conditioning covariates with a rich support. We described the asymptotic properties of our approach and we applied it to a model of entry in the pharmacy store industry where entry decisions are not merely binary choices but rather strategies about the number of stores that firms will open in a market. Our results uncovered economically relevant features of the underlying structural model such as a closeness in competition between two rivals: Rite Aid and Walgreens. While the very general assumptions we make about payoffs limit the extent to which we can perform policy analysis (such as constructing a formal measure of welfare or conducting detailed counterfactual analysis), the patterns of competition that can be revealed by our results can be policy-relevant in many cases. Furthermore, we can start with our assumptions and add more structure on the model, using our results to guide the specification of the strategic interaction component. If we choose to add more structure to payoffs, equilibrium selection, or any other component of the model, we can take our first-step confidence sets and refine them by keeping those elements that are consistent with the additional structure we may impose. In any case, our first-step results would guide the specification of any additional structure. In the case of our empirical application, for instance, we know that any model that assumes constant (and/or symmetric) strategic interaction effects would be misspecified. Even though inference in this paper focused on cases where the strategic index is parameterized (being nonparametric about all other components of the model), our identification result holds nonparametrically and thus can be the basis for nonparametric inference of the strategic interaction index. In that case, a fully nonparametric inferential approach such as sieves could be employed to approximate the strategic index, and the resulting asymptotic properties in the context of our conditional covariance inequalities would have to be characterized. Specific conjectures about the model (e.g., substitutability, symmetry, etc.) could be incorporated into the nonparametric estimator for the index. This is the subject of ongoing work in a more general context.

Appendix A: Proofs of our identification results

A.1 *Proof of Result 1*

Recall from (4) that

$$\begin{aligned}&\overline{\eta}^p_\sigma(X) \geq \overline{\eta'}^p_\sigma(X)\\&\quad\Longrightarrow\quad \overline{\nu}^p_\sigma\big(v;\xi^p\big) - \overline{\nu}^p_\sigma\big(u;\xi^p\big) \leq \overline{\nu}^p_{\sigma'}\big(v;\xi^p\big) - \overline{\nu}^p_{\sigma'}\big(u;\xi^p\big) \quad \forall u < v \in \mathcal{A}^p.\end{aligned}$$

Fix any $y^p \in \mathcal{A}^p$ and define the indicator function

$$\mathbb{I}^p_\sigma\big(y^p;\xi^p\big) = \max_{u \leq y^p}\Big(\min_{v \geq y^p+1}\big(\mathbb{1}\big\{\overline{\nu}^p_\sigma\big(v;\xi^p\big) - \overline{\nu}^p_\sigma\big(u;\xi^p\big) \leq 0\big\}\big)\Big).$$

By (4), we have

$$\overline{\eta}^p_\sigma(X) \geq \overline{\eta}^p_{\sigma'}(X) \quad\Longrightarrow\quad \mathbb{I}^p_\sigma\big(y^p;\xi^p\big) \geq \mathbb{I}^p_{\sigma'}\big(y^p;\xi^p\big).$$

Now suppose σ^{-p} and $\sigma^{-p'}$ are any pair of beliefs that produce unique expected-payoff-maximizing choices for p given the realization of ξ^p, and let $y^p_\sigma(\xi^p)$ and $y^p_{\sigma'}(\xi^p)$ denote the corresponding optimal choices. Then for any $y^p \in \mathcal{A}^p$,

$$\mathbb{1}\big\{y^p_\sigma\big(\xi^p\big) \leq y^p\big\} = \mathbb{I}^p_\sigma\big(y^p;\xi^p\big) \quad\text{and}\quad \mathbb{1}\big\{y^p_{\sigma'}\big(\xi^p\big) \leq y^p\big\} = \mathbb{I}^p_{\sigma'}\big(y^p;\xi^p\big).$$

Therefore, for any such pair of beliefs, if $\overline{\eta}^p_\sigma(X) \geq \overline{\eta}^p_{\sigma'}(X)$, then $\mathbb{1}\{y^p_\sigma(\xi^p) \leq y^p\} \geq \mathbb{1}\{y^p_{\sigma'}(\xi^p) \leq y^p\}$, which proves the statement in Result 1.

A.2 *Proof of Theorem 1*

Denote $\xi \equiv (\xi^p)^P_{p=1}$ and $\xi^{-p} \equiv (\xi^q)_{q \neq p}$. Given X, let J denote the number of BNE $\{\sigma_{*j}(X)\}_j$ that the selection mechanism $\mathcal{S}$ can choose with positive probability, and let $P^{\mathcal{S}}_j(X)$ denote the probability that $\mathcal{S}$ selects the jth BNE $(\sigma_{*j}(X))$, conditional on X. Our assumptions maintain that $\mathcal{S}$ concentrates on BNE that have a unique optimal choice. Denote it as $y^p_{\sigma_{*j}}(\xi^p)$ for the jth BNE. First, consider

$$E_{\xi^{-p}|X}\big[\eta^p\big(y^{-p}_{\sigma_{*j}}\big(\xi^{-p}\big);X\big)|X\big].$$

This is the expected value of η^p, conditional on X, in the jth BNE. By definition, this is equal to $\overline{\eta}^p_{\sigma_{*j}}(X)$, which was defined previously as

$$\overline{\eta}^p_{\sigma_{*j}}(X) = \sum_{y^{-p} \in \mathcal{A}^{-p}} \sigma^{-p}_{*j}\big(y^{-p}|X\big) \cdot \eta^p\big(y^{-p};X\big).$$

Now fix any $y^p \in \mathcal{A}^p$. By iterated expectations we have

$$E\big[\mathbb{1}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p}; X)|X\big] = \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot E_{\xi|X}\big[\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\} \cdot \eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p}); X)|X\big].$$

Assumption 3 (independent private shocks, i.e., $\xi^p \perp \xi^{-p}|X$) yields

$$\begin{aligned} E\big[\mathbb{1}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p}; X)|X\big] &= \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot E_{\xi^p|X}\big[\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\}|X\big] \cdot E_{\xi^{-p}|X}\big[\eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p}); X)|X\big] \\ &= \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot E_{\xi^p|X}\big[\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\}|X\big] \cdot \overline{\eta}^p_{\sigma_{*j}}(X). \end{aligned}$$

Therefore, by Assumption 3 we can express

$$E\big[\mathbb{1}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p}; X)|X\big] = E_{\xi^p|X}\left[\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\} \cdot \overline{\eta}^p_{\sigma_{*j}}(X)\Big|X\right]. \tag{A.1}$$

Next note that

$$\begin{aligned} E\big[\mathbb{1}\{Y^p \leq y^p\}|X\big] \cdot E\big[\eta^p(Y^{-p}; X)|X\big] &= \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot E_{\xi^p|X}\big[\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\}|X\big] \\ &\quad \times \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot E_{\xi^{-p}|X}\big[\eta^p(y^{-p}_{\sigma_{*j}}(\xi^{-p}); X)|X\big] \\ &= \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot E_{\xi^p|X}\big[\mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\}|X\big] \times \sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \overline{\eta}^p_{\sigma_{*j}}(X). \end{aligned} \tag{A.2}$$

Combining (A.1) and (A.2) we then have

$$\begin{aligned} &E\big[\mathbb{1}\{Y^p \leq y^p\} \cdot \eta^p(Y^{-p}; X)|X\big] - E\big[\mathbb{1}\{Y^p \leq y^p\}|X\big] \cdot E\big[\eta^p(Y^{-p}; X)|X\big] \\ &\quad = E_{\xi^p|X}\left[\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\} \cdot \overline{\eta}^p_{\sigma_{*j}}(X)\right. \\ &\qquad \left. - \left(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p) \leq y^p\}\right) \times \left(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X) \cdot \overline{\eta}^p_{\sigma_{*j}}(X)\right)\Big|X\right]. \end{aligned} \tag{A.3}$$

By Result 1, w.p.1 in (ξ^p) we have

$$\begin{aligned}&\sum_{j=1}^{J} P_j^{\mathcal{S}}(X)\cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p)\le y^p\}\cdot \overline{\eta}^p_{\sigma_{*j}}(X)\\&\quad-\left(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X)\cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p)\le y^p\}\right)\times\left(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X)\cdot \overline{\eta}^p_{\sigma_{*j}}(X)\right)\ge 0\\&\quad\forall y^p\in\mathcal{A}^p.\end{aligned}\tag{A.4}$$

To see why, simple algebra can be used to show that

$$\begin{aligned}&\sum_{j=1}^{J} P_j^{\mathcal{S}}(X)\cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p)\le y^p\}\cdot \overline{\eta}^p_{\sigma_{*j}}(X)\\&\qquad-\left(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X)\cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p)\le y^p\}\right)\times\left(\sum_{j=1}^{J} P_j^{\mathcal{S}}(X)\cdot \overline{\eta}^p_{\sigma_{*j}}(X)\right)\\&\quad=\sum_{\ell=1}^{J}\sum_{j=1}^{J} P_\ell^{\mathcal{S}}(X)P_j^{\mathcal{S}}(X)\cdot \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p)\le y^p\}\\&\qquad\cdot\left(1-\mathbb{1}\{y^p_{\sigma_{*\ell}}(\xi^p)\le y^p\}\right)\cdot\left(\overline{\eta}^p_{\sigma_{*j}}(X)-\overline{\eta}^p_{*\ell}(X)\right)\ge 0,\end{aligned}$$

where the last inequality follows from Result 1, which implies that, w.p.1 in ξ^p and $\forall y^p$,

$$\overline{\eta}^p_{\sigma_{*j}}(X)<\overline{\eta}^p_{*\ell}(X)\quad\Longrightarrow\quad \mathbb{1}\{y^p_{\sigma_{*j}}(\xi^p)\le y^p\}\le \mathbb{1}\{y^p_{\sigma_{*\ell}}(\xi^p)\le y^p\}.$$

From (A.3) and (A.4) it follows that, w.p.1 in X we must have

$$E\left[\mathbb{1}\{Y^p\le y^p\}\cdot\eta^p(Y^{-p};X)|X\right]\ge E\left[\mathbb{1}\{Y^p\le y^p\}|X\right]\cdot E\left[\eta^p(Y^{-p};X)|X\right]\quad\forall y^p.$$

This concludes the proof.

Appendix B: Econometric appendix

We focus on settings where the researcher observes an iid sample $((Y_i^p)_{p=1}^P, X_i)_{i=1}^n$, with[14] $((Y_i^p)_{p=1}^P, X_i)\sim F$. We assume that X can be split as $X=(X^c,X^d)$, where X^c have absolutely continuous distribution with respect to Lebesgue measure and X^d have a discrete distribution. We denote the dimension of X^c by q. We begin by describing the preliminary conditions needed for our construction.

[14] In the Supplement, we generalize our assumptions to a setting where $((Y_i^p)_{p=1}^P, X_i)_{i=1}^n$ is a triangular array.

B.1 *Specifying an "inference range"*

Let $\mathcal{X} \subset \operatorname{Supp}(X)$ denote a prespecified set such that

$$\mathcal{X} \cap \operatorname{Supp}(X^c) \subset \operatorname{int}(\operatorname{Supp}(X^c)).$$

We maintain the assumption that $f_X(x) \geq \underline{f} > 0$ for all $x \in \mathcal{X}$. Let[15] $\mathbb{I}_{\mathcal{X}}(x) = \mathbb{1}\{x \in \mathcal{X}\}$. Let

$$T^p_{\mathcal{X}}(\theta^p) = E_{Y^p,X}\big[\max\{\tau^p(Y^p|X;\theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X)\big]. \tag{B.1}$$

By construction, $T^p_{\mathcal{X}}(\theta^p) \geq 0$, and $T^p_{\mathcal{X}}(\theta^p) = 0$ if and only if $\Pr(\tau^p(Y^p|X;\theta^p) \leq 0|X \in \mathcal{X}) = 1$. We aggregate these one-sided expectations as

$$T_{\mathcal{X}}(\theta) = \sum_{p=1}^{P} T^p_{\mathcal{X}}(\theta^p).$$

Note that $T_{\mathcal{X}}(\theta) \geq 0$, and $T_{\mathcal{X}}(\theta) = 0$ if and only if $\Pr(\tau^p(Y^p|X;\theta^p) \leq 0|X \in \mathcal{X}) = 1$ for $p = 1, \dots, P$. The inference range $\mathcal{X}$ will be assumed to be such that the nonparametric estimators involved in our construction have appropriate asymptotic properties uniformly over it. Given our choice of $\mathcal{X}$, we focus attention on the superset of the identified set Θ^I:

$$\Theta^I_{\mathcal{X}} = \{\theta \in \Theta : T^p_{\mathcal{X}}(\theta^p) = 0 \text{ for } p = 1, \dots, P\}.$$

Note that $\Theta^I \subseteq \Theta^I_{\mathcal{X}}$, where $\Theta^I = \{\theta \in \Theta : \Pr(\tau^p(Y^p|X;\theta^p) \leq 0) = 1 \text{ for } p = 1, \dots, P\}$.

B.2 *Estimators involved in our construction*

We employ kernel-based nonparametric estimators. The term $K : \mathbb{R}^q \to \mathbb{R}$ will denote our kernel function. For a given $x \equiv (x^c, x^d)$ and $h > 0$ define

$$\mathcal{H}(X_i - x; h) = K\left(\frac{X^c_i - x^c}{h}\right) \cdot \mathbb{1}\{X^d_i - x^d = 0\}.$$

Let $h_n \longrightarrow 0$ be a nonnegative bandwidth sequence. For a given $x \equiv (x^c, x^d)$, y^p, and θ^p our estimators are of the form

$$\widehat{f}_X(x) = (nh^q_n)^{-1} \sum_{i=1}^{n} \mathcal{H}(X_i - x; h_n),$$

$$\widehat{F}_{Y^p}(y^p|x) = (nh^q_n \cdot \widehat{f}_X(x))^{-1} \sum_{i=1}^{n} \mathbb{1}\{Y^p_i \leq y^p\} \cdot \mathcal{H}(X_i - x; h_n),$$

$$\widehat{\lambda}^p(x;\theta^p) = (nh^q_n \cdot \widehat{f}_X(x))^{-1} \sum_{i=1}^{n} \eta^p(Y^{-p}_i; x|\theta^p) \cdot \mathcal{H}(X_i - x; h_n),$$

[15]The indicator function $\mathbb{I}_{\mathcal{X}}$ could be replaced with a smooth "trimming" function.

$$\widehat{\mu}^p\big(y^p|x;\theta^p\big)=\big(nh_n^q\cdot\widehat{f}_X(x)\big)^{-1}\sum_{i=1}^n\mathbb{1}\big\{Y_i^p\le y^p\big\}\cdot\eta^p\big(Y_i^{-p};x|\theta^p\big)\cdot\mathcal{H}(X_i-x;h_n),$$
$$\widehat{\tau}^p\big(y^p|x;\theta^p\big)=\widehat{F}_{Y^p}\big(y^p|x\big)\cdot\widehat{\lambda}^p\big(x;\theta^p\big)-\widehat{\mu}^p\big(x;\theta^p\big).$$

Our estimators for $T_{\mathcal{X}}^p(\theta^p)$ and $T_{\mathcal{X}}(\theta)$ are

$$\begin{aligned}\widehat{T}_{\mathcal{X}}^p\big(\theta^p\big)&=\frac{1}{n}\sum_{i=1}^n\widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big)\cdot\mathbb{1}\big\{\widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big)\ge -b_n\big\}\cdot\mathbb{I}_{\mathcal{X}}(X_i),\\ \widehat{T}_{\mathcal{X}}(\theta)&=\sum_{p=1}^P\widehat{T}_{\mathcal{X}}^p\big(\theta^p\big),\end{aligned}\tag{B.2}$$

where $b_n\longrightarrow 0$ is a nonnegative sequence whose properties will be described below.

B.3 *Basic assumptions*

Assumption B1 (Smoothness). *As before, express any $x\in\mathrm{Supp}(X)$ generically as $x\equiv(x^c,x^d)$ with x^c corresponding to the continuously distributed elements in X. Denote*

$$\mathcal{W}=\big\{(x,y)\in\mathrm{Supp}(X,Y):x\in\mathcal{X}\big\}.$$

Recall that we defined before

$$\begin{aligned}F_{Y^p}\big(y^p|x\big)&=E_{Y^p|X}\big[\mathbb{1}\big\{Y^p\le y^p\big\}|X=x\big],\\ \lambda^p\big(x;\theta^p\big)&=E_{Y^{-p}|X}\big[\eta^p\big(Y^{-p};x|\theta^p\big)|X=x\big],\\ \mu^p\big(y^p|x;\theta^p\big)&=E_{Y|X}\big[\mathbb{1}\big\{Y^p\le y^p\big\}\cdot\eta^p\big(Y^{-p};x|\theta^p\big)|X=x\big],\\ \tau^p\big(y^p|x;\theta^p\big)&=F_{Y^p}\big(y^p|x\big)\cdot\lambda^p\big(x;\theta^p\big)-\mu^p\big(y^p|x;\theta^p\big).\end{aligned}$$

For almost every $(x,y^p)\in\mathcal{W}$, $x'\in\mathcal{X}$, and every $\theta^p\in\Theta$, the following objects are M times differentiable with respect to x^c with bounded derivatives:

$$\begin{aligned}&F_{Y^p}\big(y^p|x\big),\qquad f_X(x),\qquad E_{Y^{-p}|X}\big[\eta^p\big(Y^{-p};x'|\theta^p\big)|X=x\big],\\ &E_{Y|X}\big[\mathbb{1}\big\{Y^p\le y^p\big\}\cdot\eta^p\big(Y^{-p};x'|\theta^p\big)|X=x\big].\end{aligned}$$

Now let

$$\begin{aligned}\gamma_p^{I}\big(y^p,x;\theta^p\big)&=E_{Y^p|X}\big[\mathbb{1}\big\{y^p\le Y^p\big\}\cdot\mathbb{1}\big\{\tau^p\big(Y^p|x;\theta^p\big)\ge 0\big\}|X=x\big],\\ \gamma_p^{II}\big(x;\theta^p\big)&=E_{Y^p|X}\big[F_{Y^p}\big(Y^p|x\big)\cdot\mathbb{1}\big\{\tau^p\big(Y^p|x;\theta^p\big)\ge 0\big\}|X=x\big],\\ \gamma_p^{III}\big(x;\theta^p\big)&=E_{Y^p|X}\big[\mu^p\big(Y^p|x;\theta^p\big)\cdot\mathbb{1}\big\{\tau^p\big(Y^p|x;\theta^p\big)\ge 0\big\}|X=x\big].\end{aligned}$$

For almost every $(x,y^p)\in\mathcal{W}$ and every $\theta^p\in\Theta$, the three objects defined above are M times differentiable with respect to x^c with bounded derivatives, and this is also satisfied by the *trimming function* $\mathbb{I}(x)$. *For given* y^p, x, *and* θ^p *define*

$$Q^p_{F_{Y^p}}\big(y^p|x\big) = F_{Y^p}\big(y^p|x\big)\cdot f_X(x), \qquad Q_{\lambda^p}\big(x;\theta^p\big) = \lambda^p\big(x;\theta^p\big)\cdot f_X(x),$$
$$Q_{\mu^p}\big(y^p|x;\theta^p\big) = \mu^p\big(y^p|x;\theta^p\big)\cdot f_X(x).$$

Then for some $\overline{Q} < \infty$,

$$\sup_{(x,y^p)\in\mathcal{W}} \big|Q_{F_{Y^p}}\big(y^p|x\big)\big| \le \overline{Q}, \qquad \sup_{x\in\mathcal{X},\theta^p\in\Theta} \big|Q_{\lambda^p}\big(x;\theta^p\big)\big| \le \overline{Q},$$
$$\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta} \big|Q_{\mu^p}\big(y^p|x;\theta^p\big)\big| \le \overline{Q}.$$

Note that the restrictions in Assumption B1 would likely rule out equilibrium selection rules that generate nonsmoothness. Following the analysis in Bajari et al. (2009), the smoothness conditions with respect to continuous state variables in the equilibrium selection mechanism require that the equilibrium paths not bifurcate for almost all values of the continuous state variable, or that a smooth path is chosen at the points of bifurcation.

Assumption B2 (Kernels and bandwidths). *Let M be as described in Assumption* B1. *We use a bias-reducing kernel K of order M with bounded support. The kernel is a function of bounded variation, is symmetric around zero, and satisfies* $\sup_{v\in\mathbb{R}^q}|K(v)| \le \overline{K} < \infty$. *The bandwidth sequences b_n and h_n are such that, for a small enough* $\epsilon_1 > 0$,

$$n^{1/2-\epsilon_1}\cdot h_n^q\cdot b_n \longrightarrow \infty, \qquad n^{1/2+\epsilon_1}\cdot b_n^2 \longrightarrow 0, \qquad n^{1/2+\epsilon_1}\cdot h_n^M \longrightarrow 0.$$

Focus on bandwidths of the type $h_n \propto n^{-\alpha_h}$ and $b_n \propto n^{-\alpha_b}$. Let $\overline{\epsilon} > 0$ be an arbitrarily small, but strictly positive constant, and let $\alpha_h = \frac{1}{2M} + \overline{\epsilon}$ and $\alpha_b = \frac{1}{4} + \overline{\epsilon}$. The conditions in Assumption B2 will be satisfied if

$$M \ge \left\lceil \frac{2\cdot q}{1 - 4\cdot\overline{\epsilon}(2+q)} \right\rceil.$$

For example, suppose $q = 8$ (as in our empirical application). Then we need $M \ge 17$. Recall that M is the number of derivatives assumed to exist in Assumption B1 and it also corresponds to the order of the kernel employed.

Our framework must allow for the existence of parameter values $\theta^p \in \Theta$ such that $\tau^p(Y^p|X;\theta^p)$ has a point mass at zero. While we allow for that, the following assumption restricts the way in which the distribution of $\tau^p(Y^p|X;\theta^p)$ approaches zero from the left. In essence the condition assumes that the density of $\tau^p(Y^p|X;\theta^p)$ is bounded in a neighborhood of the type $[-\overline{b}, 0)$, where $\overline{b} > 0$.

Assumption B3 (A regularity condition). *There exist constants* $\overline{b} > 0$ *and* $A > 0$ *such that, for each p and each* $\theta^p \in \Theta$,

$$\Pr\big(-b \le \tau^p\big(Y^p|X;\theta^p\big) < 0|X\in\mathcal{X}\big) \le b\cdot A \quad \forall 0 < b \le \overline{b}.$$

For a given θ^p, the functional $\tau^p(Y^p|X;\theta^p)$ is a random variable in its own right. Our setting allows for this random variable to have a point-mass at zero (which would occur if the inequalities are binding with positive probability at θ^p). Assumption B3 allows for this but simply requires that this random variable have a finite density in a neighborhood of the type $[-\overline{b}, 0)$ (that is, in a neighborhood to the left of zero), and that this be true uniformly over $\theta^p \in \Theta$.

Assumption B4 (Empirical process and manageability conditions). *For each p the following conditions are satisfied. Let*

$$\overline{\eta}^p(y^{-p}) = \sup_{x\in\mathcal{X},\theta^p\in\Theta} \left|\eta^p(y^{-p};x|\theta^p)\right|.$$

Then $E[\exp\{(\overline{\eta}^p(Y^{-p}))^2\cdot\epsilon\}] \leq C < \infty$ for some $\epsilon > 0$, that is, $(\overline{\eta}^p(Y^{-p}))^2$ possesses a moment generating function.

(i) *The classes of functions*

$$\begin{aligned}
\mathscr{F} &= \{f : f(y^{-p}) = \eta^p(y^{-p};x|\theta^p) \text{ for some } (x,\theta^p)\in\mathcal{X}\times\Theta\},\\
\mathscr{F}' &= \{f : f(x) = \lambda^p(x;\theta^p) \text{ for some } \theta^p\in\Theta\},\\
\mathscr{F}'' &= \{f : f(y^p,x) = \mu^p(y^p|x;\theta^p) \text{ for some } \theta^p\in\Theta\}
\end{aligned}$$

are Euclidean (see Definition 2.7 in Pakes and Pollard (1989)) with respect to envelopes $\overline{\eta}^p(\cdot)$, $\overline{F}'(\cdot)$, and $\overline{F}''(\cdot)$, respectively, where $\overline{\eta}^p(Y^{-p})$ satisfies the existence-of-moments conditions described above, and $\overline{F}'(\cdot)$ and $\overline{F}''(\cdot)$ satisfy $E[\overline{F}'(X)^2] < \infty$ and $E[\overline{F}''(Y^p,X)^2] < \infty$.

(ii) *Let $\overline{b} > 0$ be as described in Assumption B3. The class of functions*

$$\mathscr{G} = \{g : g(x,y) = \mathbb{1}\{-b \leq \tau^p(x,y;\theta^p) < 0\}\cdot\mathbb{I}_{\mathcal{X}}(x) \text{ for some } \theta^p\in\Theta, 0<b\leq\overline{b}\}$$

is Euclidean with respect to envelope 1.

Sufficient conditions for a class of functions to be Euclidean can be found,for example, in Nolan and Pollard (1987) and Pakes and Pollard (1989). Once a parametric family is chosen for η^p, those conditions can be used to verify part (i) of Assumption B4. In particular, η^p does not have to be smooth (or even continuous) to satisfy the Euclidean property. For part (ii) fix $b\in\mathbb{R}$ and let $\mathcal{N}(x,y;b)$ denote the number of points in Θ where $\tau(x,y;\theta^p) - b$ changes sign. Suppose $\sup_{(x,y)\in\mathcal{X}\times\mathcal{A}}\mathcal{N}(x,y;b) \leq \overline{\mathcal{N}} < \infty$ for all $0 < b \leq \overline{b}$. By Lemma 1 in Asparouhova et al. (2002) this ensures that the class of sets indexed by the indicator functions in part (ii) of our assumption is a Vapnik–Chervonenkis (VC) class of sets (see Definition 2.2 in Pakes and Pollard (1989)). The Euclidean property for said class of functions follows from here by the results in Pakes and Pollard (1989).

B.4 *Asymptotic properties of* $\widehat{T}^p_{\mathcal{X}}(\theta^p)$

The following theorem summarizes the key asymptotic properties of $\widehat{T}^p_{\mathcal{X}}(\theta^p)$ under our assumptions.

Theorem 2. *Let*

$$
\begin{aligned}
&\psi_U^p(Y, X; \theta^p) \\
&\quad = \big[\big(\gamma_p^I(Y^p, X; \theta^p) - \gamma_p^{II}(X; \theta^p)\big) \cdot \lambda^p(X; \theta^p) \\
&\qquad + \big(\eta^p(Y^{-p}; X|\theta^p) - \lambda^p(X; \theta^p)\big) \cdot \gamma_p^{II}(X; \theta^p) \\
&\qquad + \big(\gamma_p^I(Y^p, X; \theta^p) \cdot \eta^p(Y^{-p}; X|\theta^p) - \gamma_p^{III}(X; \theta^p)\big)\big] \cdot \mathbb{I}_{\mathcal{X}}(X)
\end{aligned}
$$

and

$$
\psi^p(Y_i, X_i; \theta^p) = \big(\max\{\tau^p(Y_i^p|X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) - T_{\mathcal{X}}^p(\theta^p)\big) + \psi_U^p(Y_i, X_i; \theta^p).
$$

If Assumptions B1–B4 *hold, then*

$$
\widehat{T}_{\mathcal{X}}^p(\theta^p) = T_{\mathcal{X}}^p(\theta^p) + \frac{1}{n}\sum_{i=1}^{n} \psi^p(Y_i, X_i; \theta^p) + \varepsilon_{p,n}(\theta^p),
$$

where

$$
\psi^p(Y_i, X_i; \theta^p) = \big(\max\{\tau^p(Y_i^p|X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) - T_{\mathcal{X}}^p(\theta^p)\big) + \psi_U^p(Y_i, X_i; \theta^p)
$$

and

$$
\sup_{\theta^p \in \Theta} \big|\varepsilon_{p,n}(\theta^p)\big| = O_p\big(n^{-1/2-\epsilon}\big) \quad \textit{for some } \epsilon > 0.
$$

The "influence function" ψ^p has two key properties:

(i) We have $E[\psi^p(Y_i, X_i; \theta^p)] = 0\ \forall \theta^p \in \Theta$.

(ii) We have $\psi^p(Y_i, X_i; \theta^p) = 0\ \forall \theta^p : \tau^p(Y^p|X; \theta^p) < 0$ w.p.1.

Property (ii) can be verified immediately by inspection. Property (i) can be verified using iterated expectations and we prove it in Appendix B.4.2 below. Let $\psi(Y_i, X_i; \theta) = \sum_{p=1}^{P} \psi^p(Y_i, X_i; \theta^p)$. By Theorem 2,

$$
\begin{aligned}
&\widehat{T}_{\mathcal{X}}(\theta) = T_{\mathcal{X}}(\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, X_i; \theta) + \varepsilon_n(\theta), \quad \text{where} \\
&\sup_{\theta \in \Theta} \big|\varepsilon_n(\theta)\big| = O_p\big(n^{-1/2-\epsilon}\big) \quad \text{for some } \epsilon > 0.
\end{aligned}
\tag{B.3}
$$

Additionally, the function $\psi(Y_i, X_i; \theta)$ is identified and has the two key properties:

(i) We have $E[\psi(Y_i, X_i; \theta)] = 0\ \forall \theta \in \Theta$.

(ii) Let

$$
\overline{\Theta}_{\mathcal{X}}^I = \big\{\theta \in \Theta : \tau^p\big(Y^p|X; \theta^p\big) < 0 \text{ w.p.1. } \forall p = 1, \dots, P\big\}.
$$

Then $\psi(Y_i, X_i; \theta) = 0$ w.p.1 $\forall \theta \in \overline{\Theta}_{\mathcal{X}}^I$.

B.4.1 *Proof of Theorem 2* A step-by-step, detailed proof is included in the Supplement. Here we present only a condensed version that summarizes the key steps. In Assumption B1 we described $\mathcal{W}$ as

$$\mathcal{W} = \{(x, y) \in \text{Supp}(X, Y) : x \in \mathcal{X}\},$$

where $\mathcal{X} \subset \text{Supp}(X)$ is a prespecified set such that $\mathcal{X} \cap \text{Supp}(X^c) \subset \text{int}(\text{Supp}(X^c))$. We maintain the assumption that $f_X(x) \geq \underline{f} > 0$ for all $x \in \mathcal{X}$. The proof is split into three main steps.

Step 1 In our first step, we show that under the assumptions of Theorem 2, there exist $D_1 > 0$, $D_2 > 0$, and $D_3 > 0$ such that

$$\begin{aligned}
&\Pr\Big(\sup_{(x,y^p)\in\mathcal{W},\theta^p\in\Theta} \big|\widehat{\tau}^p\big(y^p|x;\theta^p\big) - \tau^p\big(y^p|x;\theta^p\big)\big| \geq b_n\Big) \\
&\quad \leq D_1 \exp\{-\sqrt{n} h_n^q \big(D_2 \cdot b_n - D_3 \cdot h_n^M\big)\}.
\end{aligned}$$

Step 2 In the second step, we take the result from Step 1 and show that, combined with the assumptions of Theorem 2, it yields

$$\widehat{T}^p_{\mathcal{X}}\big(\theta^p\big) = \frac{1}{n}\sum_{i=1}^n \widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big) \cdot \mathbb{1}\{\tau^p\big(Y_i^p|X_i;\theta^p\big) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) + \varphi_n^p\big(\theta^p\big),$$

where

$$\sup_{\theta^p\in\Theta} \big|\varphi_n^p\big(\theta^p\big)\big| = O_p\big(n^{-1/2-\epsilon}\big) \quad \text{for some } \epsilon > 0.$$

Step 3 This is the last step in the proof. Note first that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big) \cdot \mathbb{1}\{\tau^p\big(Y_i^p|X_i;\theta^p\big) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\
&\quad = \frac{1}{n}\sum_{i=1}^n \max\{\tau^p\big(Y_i^p|X_i;\theta^p\big), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n \big(\widehat{\tau}^p\big(Y_i^p|X_i;\theta^p\big) - \tau^p\big(Y_i^p|X_i;\theta^p\big)\big) \\
&\qquad \cdot \mathbb{1}\{\tau^p\big(Y_i^p|X_i;\theta^p\big) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i).
\end{aligned} \tag{B.4}$$

The focus of Step 3 is on the properties of the second term. Take the objects defined in Assumption B1. For $h > 0$ and any x such that $f_X(x) > 0$ define

$$\begin{aligned}
&\psi_{F_{Y^p}}\big(Y_i^p, X_i, y^p, x; h\big) = \frac{\big(\mathbb{1}\big\{Y_i^p \le y^p\big\} - F_{Y^p}\big(y^p|x\big)\big)}{f_X(x)} \cdot \mathcal{H}(X_i - x; h),\\
&\psi_{\lambda^p}\big(Y_i^{-p}, X_i, x, \theta^p; h\big) = \frac{\big(\eta^p\big(Y_i^{-p}; x|\theta^p\big) - \lambda^p\big(x; \theta^p\big)\big)}{f_X(x)} \cdot \mathcal{H}(X_i - x; h),\\
&\psi_{\mu^p}\big(Y_i, X_i, y^p, x, \theta^p; h\big)\\
&\quad = \frac{\big(\mathbb{1}\big\{Y_i^p \le y^p\big\} \cdot \eta^p\big(Y_i^{-p}; x|\theta^p\big) - \mu^p\big(y^p|x; \theta^p\big)\big)}{f_X(x)} \cdot \mathcal{H}(X_i - x; h)
\end{aligned} \tag{B.5}$$

and define

$$\begin{aligned}
&\psi_{\tau^p}\big(Y_i, X_i, y^p, x, \theta^p; h\big)\\
&\quad = \lambda^p\big(x; \theta^p\big) \cdot \psi_{F_{Y^p}}\big(Y_i^p, X_i, y^p, x; h\big)\\
&\qquad + F_{Y^p}\big(y^p|x\big) \cdot \psi_{\lambda^p}\big(Y_i^{-p}, X_i, x, \theta^p; h\big) - \psi_{\mu^p}\big(Y_i, X_i, y^p, x, \theta^p; h\big)\\
&\quad = \big[\lambda^p\big(x; \theta^p\big) \cdot \big(\mathbb{1}\big\{Y_i^p \le y^p\big\} - F_{Y^p}\big(y^p|x\big)\big)\\
&\qquad + F_{Y^p}\big(y^p|x\big) \cdot \big(\eta^p\big(Y_i^p; x|\theta^p\big) - \lambda^p\big(x; \theta^p\big)\big)\\
&\qquad - \big(\mathbb{1}\big\{Y_i^p \le y^p\big\} \cdot \eta^p\big(Y_i^p; x|\theta^p\big) - \mu^p\big(y^p|x; \theta^p\big)\big)\big] \cdot \frac{\mathcal{H}(X_i - x; h)}{f_X(x)}.
\end{aligned} \tag{B.6}$$

From here, for any pair of observations i, j in $1, \dots, n$ and $h > 0$ denote

$$\begin{aligned}
&g_{\tau^p}\big(X_i, Y_i, X_j, Y_j; \theta^p, h\big)\\
&\quad = \frac{1}{h^q} \cdot \psi_{\tau^p}\big(Y_j, X_j, Y_i^p, X_i, \theta^p; h\big) \cdot \mathbb{1}\big\{\tau^p\big(Y_i^p|X_i; \theta^p\big) \ge 0\big\} \cdot \mathbb{I}_{\mathcal{X}}(X_i).
\end{aligned} \tag{B.7}$$

Using the results from Steps 1 and 2, we first show that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\tau}^p\big(Y_i^p|X_i; \theta^p\big) - \tau^p\big(Y_i^p|X_i; \theta^p\big)\big) \cdot \mathbb{1}\big\{\tau^p\big(Y_i^p|X_i; \theta^p\big) \ge 0\big\} \cdot \mathbb{I}_{\mathcal{X}}(X_i)\\
&\quad = \frac{1}{n^2}\sum_{j \ne i}\sum_{i=1}^{n} g_{\tau^p}\big(X_i, Y_i, X_j, Y_j; \theta^p, h_n\big) + \varrho_n^{p,1}\big(\theta^p\big), \quad \text{where}\\
&\sup_{\theta^p \in \Theta}\big|\varrho_n^{p,1}\big(\theta^p\big)\big| = O_p\big(n^{-1/2-\epsilon}\big) \quad \text{for some } \epsilon > 0.
\end{aligned} \tag{B.8}$$

From here we examine the Hoeffding decomposition (Serfling (1980)) of the U-statistic that appears in (B.8). Go back to the objects defined in Assumption B1 and let

$$\begin{aligned}
&\psi_U^p(Y, X; \theta^p) \\
&\quad = \big[\big(\gamma_p^I(Y^p, X; \theta^p) - \gamma_p^{II}(X; \theta^p)\big) \cdot \lambda^p(X; \theta^p) \\
&\qquad + \big(\eta^p(Y^{-p}; X|\theta^p) - \lambda^p(X; \theta^p)\big) \cdot \gamma_p^{II}(X; \theta^p) \\
&\qquad + \big(\gamma_p^I(Y^p, X; \theta^p) \cdot \eta^p(Y^{-p}; X|\theta^p) - \gamma_p^{III}(X; \theta^p)\big)\big] \cdot \mathbb{I}_{\mathcal{X}}(X).
\end{aligned} \tag{B.9}$$

We show that, under the assumptions of Theorem 2, the Hoeffding decomposition of the U-statistic in (B.8) yields

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\tau}^p(Y_i^p|X_i; \theta^p) - \tau^p(Y_i^p|X_i; \theta^p)\big) \cdot \mathbb{1}\big\{\tau^p(Y_i^p|X_i; \theta^p) \geq 0\big\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\psi_U^p(Y_i, X_i; \theta^p) + \vartheta_{p,n}(\theta^p),
\end{aligned}$$

where $\sup_{\theta^p \in \Theta} |\vartheta_{p,n}(\theta^p)| = O_p(n^{-1/2-\epsilon})$ for some $\epsilon > 0$. Combined with the result in Step 2 and with (B.4) we obtain

$$\begin{aligned}
&\widehat{T}_{\mathcal{X}}^p(\theta^p) = T_{\mathcal{X}}^p(\theta^p) + \frac{1}{n}\sum_{i=1}^{n}\psi^p(Y_i, X_i; \theta^p) + \varepsilon_{p,n}(\theta^p), \quad \text{where} \\
&\psi^p(Y_i, X_i; \theta^p) = \big(\max\{\tau^p(Y_i^p|X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) - T_{\mathcal{X}}^p(\theta^p)\big) \\
&\qquad\qquad\qquad\qquad + \psi_U^p(Y_i, X_i; \theta^p) \quad \text{and} \\
&\sup_{\theta^p \in \Theta}\big|\varepsilon_{p,n}(\theta^p)\big| = O_p\big(n^{-1/2-\epsilon}\big) \quad \text{for some } \epsilon > 0.
\end{aligned} \tag{B.10}$$

This concludes Step 3 and finishes the proof of Theorem 2.

B.4.2 *Two key properties of ψ^p* The "influence function" ψ^p has two key properties:

(i) We have $E[\psi^p(Y_i, X_i; \theta^p)] = 0 \ \forall \theta^p \in \Theta$.

(ii) We have $\psi^p(Y_i, X_i; \theta^p) = 0 \ \forall \theta^p : \tau^p(Y^p|X; \theta^p) < 0$ w.p.1.

Part (ii) is obvious by inspection. To see why (i) is true we can show how it holds for each one of the summands that comprise ψ^p. Note first that by definition,

$$E\big[\max\{\tau^p(Y_i^p|X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X) - T_{\mathcal{X}}^p(\theta^p)\big] = 0.$$

We will show how each of the three summands that comprise ψ_U^p has mean zero. We begin with the first term. Exchanging the order of integration, we have

$$\begin{aligned}
&E\big[\big(\gamma_p^I(Y_i^p, X_i; \theta^p) - \gamma_p^{II}(X_i; \theta^p)\big) \cdot \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&\quad = E_{X_i}\big[E_{Y_j|X_j}\big[E_{Y_i|X_i}\big[\big(\mathbb{1}\{Y_i^p \leq Y_j^p\} - F_{Y^p}(Y_j^p|X_i)\big)|X_i, Y_j, X_j\big]
\end{aligned}$$

$$
\begin{aligned}
&\quad \cdot \mathbb{1}\{\tau^p(Y_j^p|X_i;\theta^p) \geq 0\}|X_j = X_i, X_i\big] \times \lambda^p(X_i;\theta^p)\cdot \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&= E_{X_i}\big[E_{Y_j|X_j}\big[E_{Y_i|X_i}\big[\big(F_{Y^p}(Y_j^p|X_i) - F_{Y^p}(Y_j^p|X_i)\big)|X_i, Y_j, X_j\big] \\
&\quad \cdot \mathbb{1}\{\tau^p(Y_j^p|X_i;\theta^p) \geq 0\}|X_j = X_i, X_i\big] \times \lambda^p(X_i;\theta^p)\cdot \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&= 0.
\end{aligned}
$$

For the second term we have

$$
\begin{aligned}
&E\big[\big(\eta^p(Y_i^{-p};X_i|\theta^p) - \lambda^p(X_i;\theta^p)\big)\cdot \gamma_p^{II}(X_i;\theta^p)\cdot \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&\quad = E_{X_i}\big[\big(\lambda^p(X_i;\theta^p) - \lambda^p(X_i;\theta^p)\big)\cdot \gamma_p^{II}(X_i;\theta^p)\cdot \mathbb{I}_{\mathcal{X}}(X_i)\big] = 0,
\end{aligned}
$$

where we simply used the fact that $\lambda^p(X_i;\theta^p) = E_{Y^{-p}|X}[\eta^p(Y_i^{-p};X_i|\theta^p)|X_i]$. For the third term, by exchanging the order of integration we have

$$
\begin{aligned}
&E\big[\big(\gamma_p^I(Y_i^p, X_i;\theta^p)\cdot \eta^p(Y_i^{-p};X_i|\theta^p) - \gamma_p^{III}(X_i;\theta^p)\big)\cdot \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&\quad = E_{X_i}\big[E_{Y_j|X_j}\big[E_{Y_i|X_i}\big[\big(\mathbb{1}\{Y_i^p \leq Y_j^p\}\cdot \eta^p(Y_i^{-p};X_i|\theta^p) \\
&\qquad - \mu^p(Y_j|X_i;\theta^p)\big)|X_i, Y_j, X_j\big] \\
&\qquad \times \mathbb{1}\{\tau^p(Y_j^p|X_i;\theta^p) \geq 0\}|X_j = X_i, X_i\big] \times \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&\quad = E_{X_i}\big[E_{Y_j|X_j}\big[\big(\mu^p(Y_j|X_i;\theta^p) - \mu^p(Y_j|X_i;\theta^p)\big) \\
&\qquad \times \mathbb{1}\{\tau^p(Y_j^p|X_i;\theta^p) \geq 0\}|X_j = X_i, X_i\big] \times \mathbb{I}_{\mathcal{X}}(X_i)\big] \\
&\quad = 0.
\end{aligned}
$$

Combining these results we have $E[\psi^p(Y_i, X_i;\theta^p)] = 0\ \forall \theta^p \in \Theta$, as claimed.

The supplement

The accompanying Supplement to this paper includes a step-by-step proof of Theorem 2, a description of the construction of our confidence sets (CS) based on Theorem 2, an analysis of the uniform asymptotic properties of our CS, a description of the kernels and bandwidths used in our empirical application, a detailed discussion on identification and nontrivial CS, a discussion of identification when there exists correlation across players' unobserved payoff shocks (i.e., when our assumption of independent private shocks fails), and additional Monte Carlo experiment results when our assumptions are violated.

References

Andrews, D. and X. Shi (2011a), "Inference based on conditional moment inequality models." Working paper, Yale University. [729, 742]

Andrews, D. and X. Shi (2011b), "Nonparametric inference based on conditional moment inequalities." Working paper, Yale University. [729, 742]

Aradillas-Lopez, A. (2010), "Semiparametric estimation of a simultaneous game with incomplete information." *Journal of Econometrics*, 157 (2), 409–431. [727]

Aradillas-Lopez, A. (2011), "Nonparametric probability bounds for Nash equilibrium actions in a simultaneous discrete game." *Quantitative Economics*, 2, 135–171. [728, 729]

Aradillas-López, A., A. Gandhi, and D. Quint (2013), "Identification and inference in ascending auctions with correlated private values." *Econometrica*, 81 (2), 489–534. [760]

Armstrong, T. (2011a), "Asymptotically exact inference in conditional moment inequality models." Working paper, Stanford University. [729, 742]

Armstrong, T. (2011b), "Weighted ks statistics for inference on conditional moment inequalities." Working paper, Stanford University. [729, 742]

Asparouhova, E., R. Golanski, K. Kasprzyk, R. Sherman, and T. Asparouhov (2002), "Rank estimators for a transformation model." *Econometric Theory*, 18, 1099–1120. [772]

Bajari, P., H. Hong, J. Kreiner, and D. Nekipelov (2009), "Estimating static models of strategic interaction." *Journal of Business & Economic Statistics* (forthcoming). [728, 737, 771]

Bajari, P., H. Hong, and S. Ryan (2005), "Identification and estimation of discrete games of complete information." Working paper. [728]

Beresteanu, A., I. Molchanov, and F. Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79, 1785–1821. [728]

Berry, S. (1992), "Estimation of a model of entry in the airline industry." *Econometrica*, 60 (4), 889–917. [727, 729, 757, 759]

Berry, S. and E. Tamer (2006), "Identification in models of oligopoly entry." In *Advances in Econometrics. Theory and Applications, Ninth World Congress* (R. Blundell, W. Newey, and T. Persson, eds.), Econometric Society Monographs, Vol. 2, 46–85, Cambridge University Press, Cambridge. [727]

Bjorn, P. A. and Q. Vuong (1984), "Simultaneous equations models for dummy endogenous variables: A game theoretic formulation with an application to labor force participation." SSWP No. 537, Caltech. [727]

Bresnahan, T. and P. Reiss (1990), "Entry in monopoly markets." *Review of Economic Studies*, 57, 531–553. [738]

Bresnahan, T. and P. Reiss (1991a), "Empirical models of discrete games." *Journal of Econometrics*, 48, 57–81. [727, 757, 759]

Bresnahan, T. and P. Reiss (1991b), "Entry and competition in concentrated markets." *Journal of Political Economy*, 99, 977–1009. [727, 729, 757, 759]

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models." In *Handbook of Econometrics*, Vol. 6, Part B, 5549–5632, Chapter 76, Elsevier, Amsterdam. [746]

Chernozhukov, V., S. Lee, and A. Rosen (2011), "Intersection bounds: Estimation and inference." Cemmap working papers, CWP34/11, Institute for Fiscal Studies and Department of Economics, UCL. [729, 742]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77, 1791–1828. [728, 729]

Davis, P. (2006), "Estimation of quantity games in the presence of indivisibilities and heterogeneous firms." *Journal of Econometrics*, 134, 187–214. [727, 729, 757]

De Paula, A. and X. Tang (2012), "Inference of signs of interaction effects in simultaneous games with incomplete information." *Econometrica*, 80 (1), 143–172. [728, 736, 737, 739, 740, 759]

Ellickson, P. B. and P. L. Grieco (2013), "Wal-mart and the geography of grocery retailing." *Journal of Urban Economics*, 75 (0), 1–14. [755]

Galichon, A. and M. Henry (2011), "Set identification in models with multiple equilibria." *Review of Economic Studies*, 78 (4), 1264–1298. [728]

Gowrisankaran, G. and J. Krainer (2011), "Entry and pricing in a differentiated products industry: Evidence from the atm market." *RAND Journal of Economics*, 42 (1), 1–22. [728, 729]

Grieco, P. (2012), "Discrete choice games with flexible information structure: An application to local grocery markets." Working paper, Penn State University. [728]

Haile, P., A. Hortacsu, and G. Kosenok (2008), "On the empirical content of quantal response equilibrium." *American Economic Review*, 98 (1), 180–200. [737]

Harsanyi, J. (1973), "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points." *International Journal of Game Theory*, 2 (1), 1–23. [735]

Kline, B. and E. Tamer (2010), "Bounds for best response functions in binary games." Working paper, Northwestern University, Department of Economics. [728]

Lee, S., K. Song, and Y.-J. Whang (2011), "Testing functional inequalities." Cemmap working papers, CWP12/11, Institute for Fiscal Studies and Department of Economics, UCL. [745]

Lee, S., K. Song, and Y.-J. Whang (2013), "Testing functional inequalities." *Journal of Econometrics*, 172, 14–32. [729]

Lewbel, A. and X. Tang (2012), "Identification and estimation of games with incomplete information using excluded regressors." Report, Boston College. [728]

McKelvey, R. and T. Palfrey (1995), "Quantal response equilibria for normal-form games." *Games and Economic Behavior*, 10, 6–38. [737]

Milgrom, P. and R. Weber (1985), "Distributional strategies for games with incomplete information." *Mathematics of Operation Research*, 10 (4), 619–632. [737]

Morris, S. (2008), "Purification." In *The New Palgrave Dictionary of Economics*, 779–782, Palgrave Macmillan, London, United Kingdom. [735]

Nolan, D. and D. Pollard (1987), "U-processes: Rates of convergence." *The Annals of Statistics*, 15, 780–799. [772]

Pakes, A. and D. Pollard (1989), "Simulation and the asymptotics of optimization estimators." *Econometrica*, 57 (5), 1027–1057. [772]

Pesendorfer, M. and P. Schmidt-Dengler (2008), "Asymptotic least squares estimators for dynamic games." *Review of Economic Studies*, 75, 901–928. [727]

Ponomareva, M. and E. Tamer (2011), "Misspecification in moment inequality models: Back to moment equalities?" *The Econometrics Journal*, 14 (2), 186–203. [746]

Seim, K. (2006), "An empirical model of firm entry with endogenous product-type choices." *RAND Journal of Economics*, 37 (3), 619–640. [727, 729, 737]

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [776]

Sweeting, A. (2009), "The strategic timing of radio commercials: An empirical analysis using multiple equilibria." *RAND Journal of Economics*, 40 (4), 710–724. [727, 728, 737]

Tamer, E. T. (2003), "Incomplete bivariate discrete response model with multiple equilibria." *Review of Economic Studies*, 70, 147–165. [727, 738, 759]

Topkis, D. M. (1998), *Supermodularity and Complementarity*. Princeton University Press, Princeton, NJ. [732]

Vives, X. (1999), *Oligopoly Pricing*. MIT Press, Cambridge, MA. [732]

Co-editor Rosa L. Matzkin handled this manuscript.

Submitted January, 2014. Final version accepted January, 2016.