An ordinal approach to the empirical analysis of games with monotone best responses

Natalia Lazzati, John K.-H. Quah, and Koji Shirai*

November 21, 2021

Abstract: We develop a nonparametric and ordinal approach for testing pure strategy Nash equilibrium play in games with monotone best responses, such as those with strategic complements/substitutes. The approach makes minimal assumptions on unobserved heterogeneity, requires no parametric assumptions on payoff functions, and no restriction on equilibrium selection from multiple equilibria. The approach can also be extended in order to make inferences and predictions. Both model-testing and inference can be implemented by a tractable computation procedure based on column generation. To illustrate how our approach works, we include an application to an IO entry game.

Keywords: revealed preference; monotone comparative statics; single crossing differences; supermodular games; revealed monotonicity axiom; nonparametric statistical tests

JEL classification numbers: C1, C6, C7, D4, L1

* Email addresses: nlazzati@ucsc.edu; john.quah@jhu.edu; kshirai1985@kwansei.ac.jp
For helpful discussions and comments, the authors are grateful to S. Berry, J. Fox, K. Hirano, T. Hoshino, A. Kajii, Y. Kitamura, B. Kline, E. Krasnokutskaya, C. Manski, W. Newey, J. Stoye, B. Strulovici, S. Takahashi, Y. Takahashi, and especially X. Tang. Various versions of this project have been presented to audiences at the following events and we are grateful for their comments: seminars at University of Arizona, Johns Hopkins, Kyoto, Louvain (CORE), New York University, Rice University, Simon Fraser University, the National University of Singapore, Northwestern University, University of Paris (Dauphine), Queensland, Shanghai University of Finance and Economics, University of Southern California, Singapore Management University, Stanford, UC Davis, UC San Diego, the Canadian Economic Theory Conference (Vancouver, 2017), the Conference on Econometrics for Incomplete Models (CeMMAP and Northwestern, 2018), 13th Greater New York Metropolitan Area Econometrics Colloquium (Princeton University, 2018), and the Econometric Society North American Summer Meeting (UC Davis, 2018). Koji Shirai gratefully acknowledges financial support from the Japan Society for Promotion of Science (Grant-in-Aid for JSPS Fellows) and the hospitality of Johns Hopkins University during his visit in the 2019-2020 academic year.

1 Introduction

Economic analysis is often concerned with the direction of effect of an exogenous or strategic variable on an agent's decision: would a consumer buy less of a good if its price falls? would a firm follow suit if a rival raises its price? is someone more likely to join a demonstration if it is known that more people are participating? The theory of monotone comparative statics identifies properties on payoff functions, such as the single crossing property (see Milgrom and Shannon (1994)), that are necessary and sufficient for optimal choices to be increasing or decreasing with respect to exogenous variables. The empirically relevant followup question is the following: what kind of observed choice behavior would be necessary and sufficient for the recovery of payoff functions obeying the single crossing property? The key contribution of this paper is to answer this revealed preference question and to explain how it forms the basis of an econometric approach to test the single crossing property and exploit it for making inferences. Our results are applicable to individual decision-making problems where we are interested in the complementarity between a decision taken by the agent and other observed variables, and also to multi-agent decision problems modelled as games with strategic complementarity (or more generally as complete information games with pure strategy Nash equilibria and monotone strategic effects).

One obvious and important area of application of our results is to the study of entry games (as in Bresnahan and Reiss (1990), Berry (1992), and Ciliberto and Tamer (2009)) and other games that arise in the empirical IO literature. In the papers cited, firms' entry decisions are modeled as games of complete information, where each firm's decision on whether or not to enter a given market is a best response to the entry decisions taken by other firms in that market. The payoff functions are assumed to depend on observable variables in a specific parametric fashion while the unobserved component is additively separable. The unobserved component is heterogenous across markets and belong to a known class of distributions. Entry decisions by firms across many markets are observed, from which one could then estimate firms' payoff functions. A major issue in this work concerns the effects of strategic interaction and market characteristics: does the entry of another firm encourage or deter entry? does an exogenous variable such as market size encourage or deter the entry for a particular firm? Obviously, these questions are empirically important in themselves, but imposing sign restrictions on these effects could also facilitate estimation procedures.[1]

In this context, our method allows us to test whether firms are playing pure strategy Nash equilibria, subject to single crossing restrictions on its payoff functions. For example, we can test the hypothesis that a firm's entry into a market is encouraged when the market is large and discouraged when another firm is also entering. Our method works without imposing any parametric assumptions on payoff functions, without restricting the distribution of unobserved heterogeneity to particular families, and without assumptions on equilibrium selection. To pass our test means that, with sufficiently high probability, the data is a sample drawn from a population of markets where firms with the hypothesized payoff functions play pure strategy Nash equilibria. The test recovers a distribution on firms' payoff functions that satisfy the single crossing property and agrees with the observations. Thus, when a data set passes the test, we can also form set estimates on the proportion of firms with payoff functions belonging to a particular type and make out-of-sample predictions on equilibrium behavior.

While we write of recovering "payoff functions", what we are really recovering are a player's preference over different actions, conditional on covariates and the actions of other players; this is as it should be, because in an environment where only pure strategy Nash equilibria are played, the information recovered from the data *has to be* just ordinal. The specific preference property we test (or when making inferences, assume) – the singe crossing property – is also an ordinal property.

We see our revealed preference method as providing a useful tool that could complement existing, mainly parametric, estimation strategies. For example, a model containing only single crossing restrictions that passes our test will provide motivation for a more specific version in which the impact of different factors enter parametrically and with sign restrictions.[2] On the other hand, a nonparametric version of the model that does not perform well in our test will raise questions about the validity of the model specification or the suitability of the data.

---

[1] This information could be used to build a mapping from specific moments of the data to the identified set of relevant parameters. For instance, in two-player games the sign of the strategic interaction parameters allows us to identify outcomes that could occur *only as* a unique equilibrium; it follows that the probabilities of these outcomes (conditional on various observable variables) do not depend on any equilibrium selection mechanism and can be nicely related to payoff relevant parameters. (See, e.g., Tamer (2003) and Kline and Tamer (2016).) In general, economically grounded shape restrictions improve both the identification and estimation of nonparametric econometric models. Shape restrictions can reduce the size of the identified set of relevant parameters (see, e.g., Matzkin (2007)) and allows for the more efficient use of small sample data sets (see, e.g., Beresteanu (2005, 2007)).

[2] See footnote 1 on the advantages of imposing sign restrictions a priori.

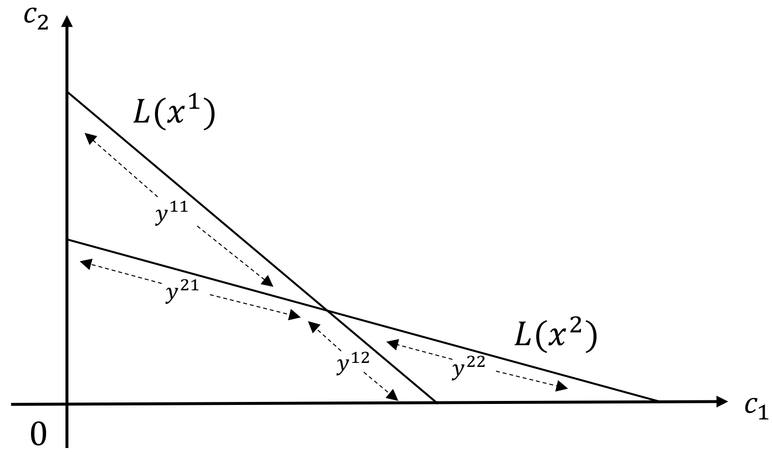

Figure 1: Dividing budget lines into 'patches'

1.1 Our approach

In Section 2, we explain our econometric approach in the context of a simple entry model. At this point, we motivate what we do by highlighting the connection with another model where the context and underlying economic theory is different, but where the econometric problem has the same structure. We have in mind the random utility model of consumer demand analyzed by Kitamura and Stoye (2018) (building on McFadden (2005) and McFadden and Richter (1991)). Suppose the econometrician observes the demand of a population of consumers for two goods at two observations ($t = 1, 2$). All consumers at observation t have the same income and face the same prices (which we denote by x^t) and thus have the same budget line $L(x^t)$, as depicted in Figure 1. Since they have heterogenous preferences, they will choose different bundles on the budget line. Graphically, their decisions appear as a cloud on each line. The question is the following: are the demand distributions at these two budgets consistent with all consumers in the population being utility-maximizers? Kitamura and Stoye answer this question using a two-step procedure.[3]

Step 1 consists of identifying those types of consumer behavior which are rational, in the sense of being consistent with utility-maximization. In this case there are four possible 'behavioral types,' of which three are rational. In Figure 1, each budget line has been divided into two 'patches.'[4] The three rational types of consumers are the following: those who choose a bundle from patch y^{11} on $L(x^1)$ and a bundle from patches y^{21} on $L(x^2)$ (behavioral type 1), those who choose from y^{11} on

[3] In their setup there could be multiple goods and multiple budget sets.
[4] The term patch follows Kitamura and Stoye (2018).

$L(x^1)$ and y^{22} on $L(x^2)$ (behavioral type 2) and those who choose from y^{12} on $L(x^1)$ and y^{22} on $L(x^2)$ (behavioral type 3). *All* rational consumers can be classified into one of those three behavioral types according to the choices they make on each budget line. The fourth behavioral type – those who choose a bundle from y^{12} on $L(x^1)$ and a bundle from y^{21} on $L(x^2)$ – is not rational.

Having identified the three rational behavioral types, **Step 2** of the test consists of answering a *dis*aggregation question: can the population be decomposed into just the three rational behavioral types? Suppose the proportion of behavioral type $b\,(=1,2,3)$ in the population is τ^b. Assuming that the distributions observed on $L(x^1)$ and $L(x^2)$ are the true population distributions, Step 2 involves solving the following set of linear equations:

$$\tau^1+\tau^2=\mathrm{P}(y^{11}\mid x^1);\quad \tau^2+\tau^3=\mathrm{P}(y^{21}\mid x^2);\quad \tau^1=\mathrm{P}(y^{21}\mid x^2);\quad \tau^3=\mathrm{P}(y^{12}\mid x^1), \tag{1}$$

where $\mathrm{P}(y^{tk}\mid x^t)$ consists of the proportion in region y^{tk} on the budget line $L(x^t)$ (so obviously $\mathrm{P}(y^{11}\mid x^1)+\mathrm{P}(y^{12}\mid x^1)=1$ and $\mathrm{P}(y^{21}\mid x^2)+\mathrm{P}(y^{22}\mid x^2)=1$). The right side of each of these equations is observable by the econometrician, so the issue is whether there are nonnegative numbers τ^1, τ^2 and τ^3 adding up to 1 that solve (1).[5] The data is consistent with utility-maximization if and only if a solution exists. Kitamura and Stoye (2018) formulates and implements an econometric version of this two-step test (i.e., one that takes into account of sampling issues, etc.).

Our econometric approach broadly follows this two-step procedure. In our context, the population consists of, not consumers, but *groups of players*. Players in different groups have different preferences, which lead to a distribution over joint actions (analogous to demand bundles in the consumer problem). Even though players' strategy sets are not changing, variation in the data that enables model testing is obtained through changes in exogenous variables (in other words, covariates) applicable to the whole population of groups (and observable to the econometrician).

For a player in a typical group playing the game, the player's preference over actions will be altered by changes in the exogenous variables and in the actions of other players; however, crucially, *the single-crossing property imposes restrictions on how the preference over actions will change.* The property requires that if a player prefers a higher action to a lower one, then this preference will be maintained if the exogenous variables increase or if other players raise their actions. These restrictions at the player level in turn restrict the way the pure strategy Nash equilibria of a group

[5] In this simple example, it is quite clear that a solution exists if and only if $\mathrm{P}(y^{12}\mid x^1)+\mathrm{P}(y^{21}\mid x^2)\leqslant 1$.

of players will vary with exogenous variables.[6] A 'behavioral type' in our model associates a group's joint-action to each realization of the exogenous variables (just as a behavioral type in the consumer model associates a patch to each budget line). **Step 1** in our paper consists of identifying those behavioral types that are compatible with the single-crossing property.

Notice that in the consumer problem, Step 1 is conceptually straightforward, since it is known that the strong axiom of revealed preference characterizes rational behavioral types. The theoretical contribution of our paper lies precisely in supplying the analog to this axiom in the context of games, i.e., to find a property on a behavioral type which is necessary and sufficient for it to arise as pure strategy Nash equilibria in games where players have payoff functions that satisfy the single-crossing property. We call the condition we find the *revealed monotonicity* (RM) axiom. This is the axiom we use to check if a given behavioral type is permitted by our hypothesis.

Having identified the permissible behavioral types using the RM axiom, **Step 2** involves finding a distribution with support *among these types* such that the resulting distribution over joint actions at different exogenous variables agree with the observations; formally one has to check if a solution exists to a set of linear equations similar to (1). The data is compatible with the hypothesized single-crossing properties if and only if a solution exists. In the econometric implementation of Step 2 of the analysis, we rely on the techniques developed by Kitamura and Stoye (2018).

A straightforward implementation of Step 1 may be impossible if the number of behavioral types satisfying the RM axiom (or the strong axiom in the Kitamura-Stoye model) is too large to be completely listed. Smeulders et al. (2021) proposes a *column generation* method to deal with this difficulty. Instead of listing all the types, an algorithm progressively searches among the relevant behavioral types in order to build up a set that is sufficient to rationalize the data. This is the approach we use in our empirical implementation. We also provide an extension of this method so that it can be used for certain problems of inference and not just for model testing.

1.2 Organization of paper

In Section 2 we provide an outline of how our approach works in the context of an entry game. Section 3 generalizes the example of Section 2, with test for the single crossing property in games

[6] The standard theory of games with strategic complementarity (Milgrom and Roberts (1990); Vives (1990)) guarantee that pure strategy Nash equilibria exist.

explained and justified in Section 3.2. If a data set passes the test, it is possible to extend our approach for the purpose of inference and out-of-sample predictions; this is discussed in Section 3.3. In Section 4 we explain the column generation method and provide an extension, so that it can also be used for problems of inference and not just for model testing. Lastly, in Section 5, to illustrate our approach, we carry out an empirical analysis of entry decisions made by airlines, using the data collected by Kline and Tamer (2016).

2 Motivating example

The seminal papers of Bresnahan and Reiss (1990) and Berry (1992) have given rise to a large empirical literature modelling oligopoly entry decisions. We will use this IO model to illustrate the type of questions we are interested in and the approach we propose to address them.

To simplify the exposition, we assume there are only two firms. Let $y_i \in \{N, E\}$ be the action set of firm i, where E means that the firm enters the market and N that it stays out. We indicate by y_{-i} the choice of the other firm, and let x_i be a real-valued, finite-dimensional vector of exogenous profit shifters (that might be market- or firm-specific) observed by the firms and also by the researcher. It is often assumed that profits have a linear functional form; for example, in Ciliberto and Tamer (2009) profits are given by, for $i = 1, 2$,

$$\pi_i\left(y_i, y_{-i}, x_i, \varepsilon_i\right) = \begin{cases} \alpha_i' x_i + \delta_i \mathbf{1}_{y_{-i}} + \varepsilon_i & \text{if} \quad y_i = E \\ 0 & \text{if} \quad y_i = N, \end{cases} \tag{2}$$

where $\mathbf{1}_E = 1$ and $\mathbf{1}_N = 0$. In this specification, the entry of firm j alters the profit of firm i by δ_i, and $(\varepsilon_1, \varepsilon_2)$ are profit shifters observed by the firms but not by the researcher. The aim of the econometrician is to estimate α_i and δ_i (and hence π_i), based on the observed entry decisions and profit shifters collected from a large cross-section of markets.

To be more specific, suppose these two firms interact with each other across many markets. Each market displays a set of observable profit shifters (or covariates) (x_1, x_2). Further assume that, for each value of the observable shifters, the econometrician observes the distribution of entry decisions. This is illustrated in Table 1, where $\mathrm{P}(E, N \mid x_1, x_2)$ denotes the fraction of those markets where Firm 1 enters and Firm 2 stays out, among those where the observable profit shifters take the value (x_1, x_2). (Obviously, the four entries in the table should add up to 1.)

<table>
<tr><td colspan="2" rowspan="2">(x_1, x_2)</td><td colspan="2">Firm 2</td></tr>
<tr><td>N</td><td>E</td></tr>
<tr><td rowspan="2">Firm 1</td><td>N</td><td>$\mathrm{P}(N, N \mid x_1, x_2)$</td><td>$\mathrm{P}(N, E \mid x_1, x_2)$</td></tr>
<tr><td>E</td><td>$\mathrm{P}(E, N \mid x_1, x_2)$</td><td>$\mathrm{P}(E, E \mid x_1, x_2)$</td></tr>
</table>

Table 1: Distribution of strategy profiles at (x_1, x_2)

Suppose that data of this form is collected at different values of (x_1, x_2). Then it would be possible to estimate α_i and δ_i under two main assumptions: (1) at each market, firms play a pure strategy Nash equilibrium of a complete information entry game and (2) the distribution of $(\varepsilon_1, \varepsilon_2)$ is independent of (x_1, x_2) and belongs to a specific family. For a recent study of how these estimates could be obtained from data in this format, see Kline and Tamer (2016).

A major focus of attention of the empirical literature is whether the entry decision of the other firm and/or the movement of different profit shifters tend to encourage or discourage a firm's entry. In (2), this manifests itself in the signs of α_i and δ_i. It is on precisely this issue that our paper makes a contribution: we develop a technique that allows us to test hypotheses about the *direction of impact* of different variables on a firm's entry decision, without imposing a parametric form on its payoff function and without requiring the unobservable profit shifters $(\varepsilon_1, \varepsilon_2)$ to (i) belong to any distribution family or (ii) to influence payoffs in an additively separable way. In our approach, there is no a priori restriction on the form of the profit functions or on the form of the unobserved heterogeneity over those functions, though we do maintain the assumption that the distribution of unobserved heterogeneity is independent of the observable profit shifters. At the same time, our approach shares the desirable features of the parametric procedure: the profit functions of the two players can be correlated, we allow for multiple pure strategy Nash equilibria, and we are agnostic about the selection rule among these equilibria. In the event that a data set is consistent with our hypotheses, our approach leads to the (set) estimation of the primitives of the model that would generate the observations as equilibrium outcomes. For example, we can provide bounds on the fraction of markets where (say) Firm 1 is nonstrategic, in the sense that its decision of whether or not to enter a market is independent of Firm 2's entry decision.

$x_2 = (0,0)$		Firm 2	
		N	E
Firm 1	N	3/12	3/12
	E	4/12	2/12

$x_2 = (0,1)$		Firm 2	
		N	E
Firm 1	N	1/12	5/12
	E	3/12	3/12

$x_2 = (1,0)$		Firm 2	
		N	E
Firm 1	N	2/12	4/12
	E	2/12	4/12

Table 2: Distribution of strategy profiles

2.1 Single crossing condition

To explain our approach in greater detail, suppose we wish to test the hypothesis that firm i's entry into a market is (1) encouraged when the profit shifter takes higher values and (2) discouraged when the other firm chooses to enter. The theoretical literature on monotone comparative statics (in particular, Milgrom and Shannon (1994)) tells us that this is captured by the following simple condition, which is a version of the *single crossing property* on the firm i's payoff function Π_i:

$$\Pi_i\left(E, y'_{-i}, x'_i\right) > \Pi_i\left(N, y'_{-i}, x'_i\right) \Longrightarrow \Pi_i\left(E, y''_{-i}, x''_i\right) > \Pi_i\left(N, y''_{-i}, x''_i\right) \tag{3}$$

for all $(-\mathbf{1}_{y''_{-i}}, x''_i) > (-\mathbf{1}_{y'_{-i}}, x'_i)$; that is, if firm i prefers entering a market to staying out when the other firm is also entering, then this preference is preserved if there is an increase in x_i or if the other firm decides not to enter. In the case where Π_i has the form given by (2), i.e., where $\Pi(y_i, y_{-i}, x_i) = \pi_i(y_i, y_{-i}, x_i, \varepsilon_i)$ for some ε_i, the single crossing property holds if $\alpha_i > 0$ and $\delta_i < 0$, but it is clear that this parametric form is *not* necessary for single-crossing to hold.[7]

We assume that the econometrician observes data of the type depicted in Table 1, i.e., the distribution over action profiles, $\mathrm{P}(\cdot \mid x_1, x_2)$, at certain realized values of (x_1, x_2). We shall ignore small sample issues for now and assume that this is the true distribution from a population of firm pairs. The heterogeneity in unobserved market characteristics leads to heterogeneity in the payoff function profiles (among firm pairs). This in turn leads to different pure strategy Nash equilibria being played by the two firms at different markets. The hypothesis being tested is that the distribution of (the random variable representing) the payoff function profiles of the two firms has the following property:

[7] Readers familiar with the single crossing property will notice that our definition ignores indifferences. We are assuming throughout this paper that preferences between actions are strict.

$x_2 = (0,0)$		Firm 2	
		N	E
Firm 1	N	0, 0	**0, 1**
	E	**1, 0**	-1, -1

$x_2 = (0,1)$		Firm 2	
		N	E
Firm 1	N	0, 0	**0, 1**
	E	1, 0	-1, 1

$x_2 = (1,0)$		Firm 2	
		N	E
Firm 1	N	0, 0	**0, 1**
	E	1, 0	-1, 1

Table 3: Example of payoffs satisfying the single crossing condition

a particular realization of the payoff function profile, (Π_1, Π_2)*, is in the distribution's support only if both* Π_1 *and* Π_2 *satisfy (3).*

This hypothesis is the natural nonparametric counterpart to the hypothesis that $\alpha_i > 0$ and $\delta_i < 0$ for $i = 1, 2$ in the linear representation (2); indeed, if payoff functions have that form and the heterogeneity in payoff function profiles is given by a distribution on $(\varepsilon_1, \varepsilon_2)$ which has $\mathbb{R}^2$ as its support, then the induced distribution over payoff function profiles will have a support confined to functions satisfying single-crossing condition (3) if and only if $\alpha_i > 0$ and $\delta_i < 0$, for $i = 1, 2$.

We first observe that this nonparametric hypothesis has testable implications. Consider an increase in the observable profit shifters from (x_1', x_2') to (x_1'', x_2''); then, at any particular realization Π_1 of Firm 1's payoff function, if it prefers to enter when the other firm enters at (x_1', x_2'), then the single-crossing condition guarantees that it will continue to prefer entry at (x_1'', x_2''). The same argument applies to Firm 2, and so we conclude that if (E, E) is the Nash equilibrium at (x_1', x_2') for a given realized profit function profile (Π_1, Π_2), then it will be the *unique* Nash equilibrium at (x_1'', x_2'') for this realized profile. Aggregating across all profiles, we establish that

$$\mathrm{P}\left(E, E \mid x_1'', x_2''\right) \geqslant \mathrm{P}\left(E, E \mid x_1', x_2'\right),$$

provided the distribution of profit function profiles is independent of the observable profit shifters. This inequality constitutes an observable restriction on the data but it is not the only restriction imposed by our hypothesis. The tightest possible restriction is obtained by checking whether the data set can be *rationalized*; loosely speaking, this involves finding a distribution over payoff functions for firms 1 and 2 obeying the single crossing condition, such that the equilibria they induce, along with some equilibrium selection rule, will generate the data observed.

2.2 Rationalizing a data set

To get a flavor of what testing for rationalizability involves, consider the data set depicted in Table 2, which specifies the distribution of joint actions by two firms at three distinct values of x_2. (In this example, x_2 is two-dimensional and x_1 is fixed throughout.) Given a particular realization (Π_1, Π_2), the firms will choose an action profile (either (E, E), (E, N), (N, E) or (N, N)) at each realization of x_2, and as x_2 takes different values the action profile of the two firms may change. We shall refer to each of these 'paths' or 'transitions' of the action profile as x_2 changes as a *behavioral type*. Notice that even though firms' profit functions may be heterogenous in infinitely many ways, its manifestation in behavior must be *finite*, since there are only finitely many possible actions and finite variation in the covariates (x_1, x_2).

To be precise, there are in total $4^3 = 64$ behavioral types and even then not all are consistent with pure strategy Nash equilibrium play with payoff functions obeying the single crossing property. For example, as we have already explained, a behavioral type where (E, E) is played at $x_2 = (0, 0)$ and (N, N) at $x_2 = (0, 1)$ is not compatible with single crossing. On the other hand, a behavioral type where (N, E) is played at all three values of x_2 can be justified with single crossing profit functions. To see this, suppose that for Firm 1 entry is profitable if and only if the other firm stays out; for Firm 2, entry is always profitable at $x_2 = (0, 1)$ or $(1, 0)$ but is not profitable at $x_2 = (0, 0)$ if Firm 1 also enters. The payoffs of the two firms are depicted in Table 3; it is straightforward to check that they satisfy the single crossing property (3) and justifies the behavioral type. Furthermore, since (E, N) is another equilibrium at $x_2 = (0, 0)$, the behavioral type where (E, N) is played at $x_2 = (0, 0)$ and (N, E) played at $x_2 = (1, 0)$ and $x_2 = (0, 1)$ is also compatible with single crossing.

Ascertaining if a data set can be rationalized involves a two-step procedure. In **Step 1**, we must identify all those behavioral types that are consistent with the single-crossing property in the sense that the action profile (y_1, y_2) at each value of the observed profit shifter (x_1, x_2) could be generated as pure strategy Nash equilibrium play from profit functions obeying the single-crossing property. This is a revealed preference problem and to carry out this step in general requires a revealed preference theorem analogous to Richter's or Afriat's Theorem (in the context of consumer demand). In **Step 2**, we check whether we can find weights on these behavioral types that account for the observed distribution of action profiles. Note that Step 2 reflects our assumption that the

Type	Weight	$x_2 = (0,0)$				$x_2 = (0,1)$				$x_2 = (1,0)$			
		Action profiles				Action profiles				Action profiles			
		N, N	N, E	E, N	E, E	N, N	N, E	E, N	E, E	N, N	N, E	E, N	E, E
1	1/12			1/12					1/12			1/12	
2	2/12	2/12					2/12			2/12			
3	2/12			2/12				2/12					2/12
4	1/12			1/12				1/12				1/12	
5	1/12	1/12				1/12					1/12		
6	2/12				2/12				2/12				2/12
7	3/12		3/12				3/12				3/12		
Sum	1	3/12	3/12	4/12	2/12	1/12	5/12	3/12	3/12	2/12	4/12	2/12	4/12

Table 4: Distribution of types rationalizing data in Table 2

distribution of profit functions is invariant across profit shifters (x_1, x_2).[8]

We claim that the data set depicted in Table 2 can be rationalized. To understand why, we list in Table 4 seven possible behavioral types for a pair of Firms 1 and 2. In this simple case, it is straightforward to check that each of these behavioral types is consistent with the single crossing property. For example, we have already justified behavioral type 7 where (N, E) is played at all values of x_2. When these behavioral types are represented in the population with the weights indicated in Table 4, they generate the distribution of entry decisions observed in Table 2. (Compare the entries in Table 2 with the last row of Table 4.)

Lastly, it is worth noting that our model can accommodate behavior which is disallowed by the parametric specification. Indeed, the data in Table 2 *cannot* be explained by profit functions of the form (2), in which Firm 2's profit upon entry takes the form of

$$\pi_2(E, y_1, x_{21}, x_{22}, \varepsilon_2) = \alpha_{21} x_{21} + \alpha_{22} x_{22} + \delta_{21} \mathbf{1}_{y_1} + \varepsilon_2, \tag{4}$$

where $(\alpha_{21}, \alpha_{22}) > 0$ and $\delta_{21} < 0$.[9] The essential reason for this is the following: when π_2 has the form (4), whether the boost to profits of an increase in x_{21} is greater or smaller than that obtained from the same increase to x_{22} depends on whether α_{21} is bigger or smaller than α_{22} and

[8] As we mentioned earlier, this exogeneity restriction is imposed in much of the literature on estimation.

[9] We are grateful to Aureo De Paula for suggesting that we construct an example with this specific feature.

is *independent of the realization of* ε_2. So it excludes the case where the realization of ε_2 influences the relative benefit of higher x_{21} versus higher x_{22}. To see why the parametric model cannot explain the data in Table 2, suppose instead that it does. Then

$$\mathrm{P}(E,E|x_1,(1,0))-\mathrm{P}(E,E|x_1,(0,0)) = \mu\left(\{\varepsilon_1 : \pi_1(E,E,x_1,\varepsilon_1) \geqslant 0\} \times \{\varepsilon_2 : -\delta_{21} \geqslant \varepsilon_2 \geqslant -\alpha_{21}-\delta_{21}\}\right),$$

where μ is the probability measure on the space of $(\varepsilon_1, \varepsilon_2)$; similarly,

$$\mathrm{P}(E,E|x_1,(0,1))-\mathrm{P}(E,E|x_1,(0,0)) = \mu\left(\{\varepsilon_1 : \pi_1(E,E,x_1,\varepsilon_1) \geqslant 0\} \times \{\varepsilon_2 : -\delta_{21} \geqslant \varepsilon_2 \geqslant -\alpha_{22}-\delta_{21}\}\right).$$

Since the former equals 2/12 while the latter equals 1/12, we conclude that $\alpha_{22} < \alpha_{21}$. However,

$$\frac{1}{12} = \mathrm{P}(N,N|x_1,(0,0)) - \mathrm{P}(N,N|x_1,(1,0)) = \mu\left(\{\varepsilon_1 : \pi_1(E,N,x_1,\varepsilon_1) \leqslant 0\} \times \{\varepsilon_2 : 0 \geqslant \varepsilon_2 \geqslant -\alpha_{21}\}\right)$$

and

$$\frac{2}{12} = \mathrm{P}(N,N|x_1,(0,0)) - \mathrm{P}(N,N|x_1,(0,1)) = \mu\left(\{\varepsilon_1 : \pi_1(E,N,x_1,\varepsilon_1) \leqslant 0\} \times \{\varepsilon_2 : 0 \geqslant \varepsilon_2 \geqslant -\alpha_{22}\}\right)$$

which tells us that $\alpha_{22} > \alpha_{21}$. So we obtain a contradiction.

3 Main results

3.1 Model, data, and empirical hypothesis

3.1.1 The model

The population is composed of many different groups, with the agents within each group playing a game among themselves. The set of agents in each group is denoted by $\mathcal{N} = \{1, 2, ..., n\}$. Agent i in the group chooses an action y_i from a *finite* set Y_i. We assume that Y_i is a *product set*, which means in particular that the agent's action can be multi-dimensional. Formally, $Y_i = \times_{k=1}^{K(i)} Y_{ik}$, where $Y_{ik} \subseteq \mathbb{R}$. The payoff of this agent over different actions depends on the actions of the other agents in its group $\mathbf{y}_{-i} = (y_j : j \in \mathcal{N}, j \neq i) \in \mathbf{Y}_{-i} = \times_{j \in \mathcal{N}, j \neq i} Y_j$ and on the value of an m_i-dimensional covariate $x_i \in X_i = \times_{m=1}^{M(i)} X_{im}$, where X_{im} is a subset of $\mathbb{R}$. Note that we allow for the covariates to take infinitely many distinct values. Thus the payoff of agent i is given by a function[10]

$$\Pi_i : Y_i \times \mathbf{Y}_{-i} \times X_i \to \mathbb{R}.$$

[10] In certain settings, it is natural for some covariate to simultaneously influence the payoffs of different players. There is no difficulty with introducing the same covariate to X_i (as an added dimension) for every player i. An example of this is found in Section 5.

We denote the set of best replies at $(\mathbf{y}_{-i}, x_i)$ by

$$\mathrm{BR}_i(\mathbf{y}_{-i}, x_i) = \mathrm{argmax}_{y_i \in Y_i} \Pi_i \left(y_i, \mathbf{y}_{-i}, x_i\right).$$

Throughout this paper, we assume that agents have *strict* preferences over actions, so that $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i)$ has a unique value. We use $\mathbf{\Pi} = (\Pi_i : i \in \mathcal{N})$, $\mathbf{y} = (y_i : i \in \mathcal{N})$, and $\mathbf{x} = (x_i : i \in \mathcal{N})$ to indicate a profile of payoff functions, actions, and covariate values, respectively.

A pair of payoff function and covariate profiles $(\mathbf{\Pi}, \mathbf{x})$ induces a game of complete information $\mathrm{G}(\mathbf{\Pi}, \mathbf{x})$. We denote the set of pure strategy Nash equilibria of this game by

$$\mathrm{NE}(\mathbf{\Pi}, \mathbf{x}) = \{\mathbf{y}^* \in \mathbf{Y} : y_i^* = \mathrm{BR}_i(\mathbf{y}_{-i}^*, x_i) \text{ for all } i \in \mathcal{N}\}$$

where $\mathbf{Y} = \times_{i \in \mathcal{N}} Y_i$ is the set of possible strategy profiles.

We are interested in the econometric analysis of games where payoff functions obey single crossing conditions (Milgrom and Shannon, 1994).

Single-Crossing Differences. The payoff function Π_i has *single-crossing differences in* $(y_i; (\mathbf{y}_{-i}, x_i))$ if the following holds:[11]

for every $y_i'' > y_i'$ and $(\mathbf{y}_{-i}'', x_i'') > (\mathbf{y}_{-i}', x_i')$,

$$\Pi_i\left(y_i'', \mathbf{y}_{-i}', x_i'\right) > \Pi_i\left(y_i', \mathbf{y}_{-i}', x_i'\right) \Longrightarrow \Pi_i\left(y_i'', \mathbf{y}_{-i}'', x_i''\right) > \Pi_i\left(y_i', \mathbf{y}_{-i}'', x_i''\right). \quad (5)$$

This condition states that if it is advantageous for agent i to choose a higher action y_i'' over a lower one y_i', then it remains advantageous to do so when other players raise their actions and/or covariates take higher values. It is the nonparametric analog to the property captured by the signs of α_i and δ_i in the linear payoff form (2) (of the previous section).

In the case where i's action space Y_i is one-dimensional, the condition (5) is sufficient to guarantee that $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i)$ is increasing in $(\mathbf{y}_{-i}, x_i)$ (see the Basic Theorem stated below). However, the condition is *not* sufficient when the action is multi-dimensional. In that case, Milgrom and Shannon (1994) show that, in addition to (5), $\Pi_i(y_i, \mathbf{y}_{-i}, x_i)$ must also be a quasisupermodular function of y_i for all $(\mathbf{y}_{-i}, x_i)$.[12] When the action space Y_i is a product set, it is straightforward to show that

[11] The term used in Milgrom and Shannon (1994) is single-crossing property and not single-crossing differences. The latter term follows Milgrom (2004) and seems more descriptive since the single-crossing condition is imposed on the difference of the payoff function at two values.

[12] Let A be a lattice. A function $F : A \to \mathbb{R}$ is quasisupermodular if $F(a' \vee a'') - F(a'') > (\geqslant) 0$ whenever $F(a' \wedge a'') - F(a') > (\geqslant) 0$.

the combination of quasisupermodularity and condition (5) is equivalent to the following stronger version of single crossing differences:

$$\begin{aligned}&\text{for every nonempty set } J \subset \{1,2,\ldots,K(i)\},\ y''_{iJ} > y'_{iJ} \text{ and } (y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i) > (y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i),\\ &\Pi_i\left(y''_{iJ}, y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i\right) > \Pi_i\left(y'_{iJ}, y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i\right)\\ &\qquad \Longrightarrow \Pi_i\left(y''_{iJ}, y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i\right) > \Pi_i\left(y'_{iJ}, y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i\right). \qquad (6)\end{aligned}$$

(Note that y_{iJ} and $y_{i(-J)}$ denote the subvectors on J and its complement respectively that together constitute y_i.) In other words, if over some subset of dimensions J, the agent prefers a higher action y''_{iJ} to a lower one y'_{iJ}, keeping fixed the actions on the other dimensions and the covariates, then that preference is maintained if actions on the other dimensions and/or the covariates are raised. Condition (6) is stronger than (5) since the latter is the special case where $J = \{1,2,\ldots,K(i)\}$.

An intuitive sufficient condition for a generalized version of single-crossing differences (6) is the *increasing differences* property:

$$\begin{aligned}&\text{for every nonempty set } J \subset \{1,2,\ldots,K(i)\} \text{ and } y''_{iJ} > y'_{iJ}\\ &\Pi_i\left(y''_{iJ}, y_{i(-J)}, \mathbf{y}_{-i}, x_i\right) - \Pi_i\left(y'_{iJ}, y_{i(-J)}, \mathbf{y}_{-i}, x_i\right) \text{ is weakly increasing in } (y_{i(-J)}, \mathbf{y}_{-i}, x_i). \qquad (7)\end{aligned}$$

This condition says that the marginal payoff to player i from increasing his action in any subset of dimensions J weakly increases with his actions in other dimensions, the actions of other players, and the covariates. (In the context of continuous domains, (7) is equivalent to the familiar requirement on nonnegative cross derivatives.[13]) As a further special case, suppose that $\Pi_i(y_i, \mathbf{y}_{-i}, x_i)$ is additively separable in y_i, in the sense that there are functions $\Pi_{ik} : Y_{ik} \times \mathbf{Y}_{-i} \times X_i \to \mathbb{R}$ such that

$$\Pi_i(y_i, \mathbf{y}_{-i}, x_i) = \sum_{k=1}^{K(i)} \Pi_{ik}(y_{ik}, \mathbf{y}_{-i}, x_i). \qquad (8)$$

Then for every nonempty set $J \subset \{1,2,\ldots,K(i)\}$ and $y''_{iJ} > y'_{iJ}$,

$$\Pi_i\left(y''_{iJ}, y_{i(-J)}, \mathbf{y}_{-i}, x_i\right) - \Pi_i\left(y'_{iJ}, y_{i(-J)}, \mathbf{y}_{-i}, x_i\right) = \sum_{k \in J}\left[\Pi_{ik}(y''_{ik}, \mathbf{y}_{-i}, x_i) - \Pi_{ik}(y'_{ik}, \mathbf{y}_{-i}, x_i)\right],$$

and condition (7) will be satisfied if each Π_{ik} has increasing differences in $(y_{ik}; (\mathbf{y}_{-i}, x_i))$, i.e., $\Pi_{ik}(y''_{ik}, \mathbf{y}_{-i}, x_i) - \Pi_{ik}(y'_{ik}, \mathbf{y}_{-i}, x_i)$ is increasing in $(\mathbf{y}_{-i}, x_i)$ for all $y''_{ik} > y'_{ik}$.

[13] Precisely, for all $k \in \{1,2,\ldots,K(i)\}$, $\frac{\partial^2 \Pi_i}{\partial y_{ik} \partial z}(y_i, \mathbf{y}_{-i}, x_i) \geqslant 0$ for $z = y_{i\ell}$ (with $\ell \neq k$), $z = y_{j\bar{k}}$ (for player $j \neq i$ and $\bar{k} \in \{1,2,\ldots,K(j)\}$), and $z = x_{im}$ (for $m \in \{1,2,\ldots,M(i)\}$) (see Topkis (1998)).

We denote by $\mathcal{SC}$ the set of payoff function profiles $\mathbf{\Pi} = (\Pi_i)_{i \in \mathcal{N}}$ where each Π_i satisfies (6). Sometimes we shall abuse terminology and refer to a particular payoff function Π_i as 'satisfying single crossing differences' or 'belonging to $\mathcal{SC}$'; that simply means that (6) holds for Π_i.

Note that single crossing differences is not just a nonparametric property, in fact it is an *ordinal* property since any strictly increasing transformation of a function that obeys single crossing differences will also obey single crossing differences. Furthermore, since a player's best responses are pinned down by a player's *preference over actions*,

$$\mathrm{NE}(\mathbf{\Pi}, \mathbf{x}) = \mathrm{NE}(\tilde{\mathbf{\Pi}}, \mathbf{x})$$

whenever $\tilde{\mathbf{\Pi}} = (\tilde{\Pi}_i)_{i \in \mathcal{N}}$ is a strictly increasing transformation of $\mathbf{\Pi} = (\Pi_i)_{i \in \mathcal{N}}$, in the sense that $\tilde{\Pi}_i$ is a strictly increasing transformation of Π_i for all i.[14]

The property of single-crossing differences has two key implications which are central to our study. (See Milgrom and Roberts (1990), Milgrom and Shannon (1994), and Vives (1990).)

Basic Theorem. *If $\mathbf{\Pi} \in \mathcal{SC}$, the family of games $\{G(\mathbf{\Pi}, \mathbf{x}) : \mathbf{x} \in \mathbf{X}\}$ has the following properties:*

(i) $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i)$ *is increasing in* $(\mathbf{y}_{-i}, x_i)$ *for each* $i \in \mathcal{N}$ *and*

(ii) $\mathrm{NE}(\mathbf{\Pi}, \mathbf{x})$ *is non-empty.*[15]

This result says that single crossing differences (in the sense of (5) when players' actions are one-dimensional and in the generalized sense of (6) when players' actions are multi-dimensional) guarantees that $\mathrm{G}(\mathbf{\Pi}, \mathbf{x})$ is a game of *strategic complements*, in the sense that a player optimally increases his action when other players raise theirs.[16] These games have pure strategy Nash equilibria. Furthermore, the best response of each player also increases with the (exogenous) covariate, and an increase in those covariates raises the set of Nash equilibria in a certain sense.

The entry game we studied in Section 2 is intuitively a game of *strategic substitutes* since a firm that prefers E to N when the other chooses E will continue to do so when the other firm chooses

[14] $\tilde{\Pi}_i$ is a strictly increasing transformation of Π_i if, for every $(\mathbf{y}_{-i}, x_i)$, there is a strictly increasing function h such that $\tilde{\Pi}_i(y_i, \mathbf{y}_{-i}, x_i) = h(\Pi_i(y_i, \mathbf{y}_{-i}, x_i))$ for all $y_i \in Y_i$. (Note that h is allowed vary with $(\mathbf{y}_{-i}, x_i)$.) It is straightforward to check that if Π_i satisfies (5), so does $\tilde{\Pi}_i$, and the analogous claim holds for (6). Finally, note that the increasing differences property (7) is *not* an ordinal property.

[15] It is also known that $\mathrm{NE}(\mathbf{\Pi}, \mathbf{x})$ has a smallest and a largest element and that they increase with $\mathbf{x}$; however, this property is of limited use in our setting since we make no assumptions on equilibrium selection.

[16] There is also a sense in which single-crossing differences in *necessary* for monotone optimal solutions; see Milgrom and Shannon (1994).

N. (In the parametric form (2) this is captured by requiring δ_1, $\delta_2 \leqslant 0$.) However, it is well-known and also easy to see that a two-player game of strategic substitutes can be formally treated as a game of strategic complements by reversing the order on the actions of one of the players. Thus it follows from the Basic Theorem that the two-player entry game in our motivating example will always have pure strategy Nash equilibria, so long as it is a game of strategic substitutes. This is the theoretical foundation in all empirical studies of entry games where agents are assumed to play pure strategy Nash equilibria.

3.1.2 Structure of the data

The econometrician has access to a large cross section of different groups. We assume that the set of observed covariates is finite and denote it by

$$\mathbf{X}^{\text{data}} = \left\{ \mathbf{x}^t : t \in \mathcal{T} = \{1, 2, ..., T\} \right\}.$$

We allow this set to be a *strict* subset of all the possible covariate values (which is $\mathbf{X}$). For each value of $\mathbf{x}^t \in \mathbf{X}^{\text{data}}$, the econometrician learns the joint distribution of actions by all agents. Thus the data set can be expressed as a set of conditional distributions

$$\mathcal{P} = \left\{ \mathrm{P}(\cdot \mid \mathbf{x}^t) : t \in \mathcal{T} \right\}$$

where $\mathrm{P}(\mathbf{y} \mid \mathbf{x}^t)$ is the fraction of groups in the population with covariate values $\mathbf{x}^t$ that select action profile $\mathbf{y}$. In this section, we assume these conditional distributions are the true population distributions; small sample issues are addressed in the empirical application in Section 5.

3.1.3 Empirical hypothesis

We offer a test to check whether a set of conditional distributions $\mathcal{P}$ is consistent with the hypothesis that all agents have payoff functions that satisfy single-crossing differences. The test assumes that the observed heterogeneity in joint actions across different groups in the population arises from heterogeneity in payoff functions and heterogeneity in equilibrium selection rules among pure strategy Nash equilibria (both of which are not directly observed by the econometrician).

Random payoff functions. To capture preference heterogeneity among the population of groups, we assume that the profile of payoff functions in each group, $\mathbf{\Pi}$, is random and distributed according to $\mathrm{P}_{\mathbf{\Pi}}$. (Notice that we are abusing notation by using $\mathbf{\Pi}$ to denote both the random variable and a particular realization.) By specifying a *joint* distribution on the payoff functions, we allow for the possibility that group formation is dependent on players' payoff functions. Put another way, the payoff function of a particular player j can depend on the payoff function of the player i to which player j is grouped. Let $\mathrm{P}_{\mathbf{\Pi}}|_{\mathbf{x}}$ be the distribution of payoff function profiles conditional on the realized values of the covariates $\mathbf{x}$; we assume that this does not depend on $\mathbf{x}$.

Assumption (Conditional Independence) $\mathrm{P}_{\mathbf{\Pi}}|_{\mathbf{x}} = \mathrm{P}_{\mathbf{\Pi}}$ for all $\mathbf{x}$ in the data set.

In the parametric entry model considered in Section 2, where profit functions are specified by (2), payoff heterogeneity is captured by the unobservables $(\varepsilon_1, \varepsilon_2)$. Each realization of $(\varepsilon_1, \varepsilon_2)$ induces a (deterministic) pair of profit functions (π_1, π_2), and a distribution on $(\varepsilon_1, \varepsilon_2)$ leads to a joint distribution on profit function pairs (π_1, π_2). If (as is typically assumed) the distribution of $(\varepsilon_1, \varepsilon_2)$ is independent of the distribution of the covariates (x_1, x_2), then the distribution on (π_1, π_2) will obey conditional independence in the sense defined. In our specification, we capture unobservable heterogeneity by directly specifying a distribution on the payoff function profiles $\mathbf{\Pi} = (\Pi_i)_{i \in \mathcal{N}}$ and assuming that this distribution is independent of the covariates.

Equilibrium selection rule. Given $\mathbf{x}$ and a particular realization $\mathbf{\Pi}$ in $\mathcal{SC}$, we denote the set of pure strategy Nash equilibria by $\mathrm{NE}(\mathbf{\Pi}, \mathbf{x})$. The Basic Theorem tells us that this set is nonempty, and even though we assume that best replies are single-valued, we cannot rule out the possibility of multiple equilibria. We denote the *equilibrium selection rule* by $\lambda\left(\mathbf{y} \mid \mathbf{\Pi}, \mathbf{x}^t\right)$; this refers to the fraction of groups in the population with payoff functions $\mathbf{\Pi}$ and covariates $\mathbf{x}^t$ that select the action profile $\mathbf{y}$. We assume $\lambda\left(\mathbf{y} \mid \mathbf{\Pi}, \mathbf{x}^t\right) = 0$ for all $\mathbf{y} \notin \mathrm{NE}(\mathbf{\Pi}, \mathbf{x}^t)$ and $\sum_{\mathbf{y} \in \mathbf{Y}} \lambda(\mathbf{y} \mid \mathbf{\Pi}, \mathbf{x}^t) = 1$.

The following definition formalizes our hypothesis.

Definition 1. $\mathcal{P}$ is single-crossing rationalizable or $\mathcal{SC}$-rationalizable if there exists a distribution $\mathrm{P}_{\mathbf{\Pi}}$ with support on $\mathcal{SC}$ and an equilibrium selection mechanism $\lambda\left(\cdot \mid \mathbf{\Pi}, \mathbf{x}\right)$ such that

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}^t) = \int \lambda\left(\mathbf{y} \mid \mathbf{\Pi}, \mathbf{x}^t\right) d\mathrm{P}_{\mathbf{\Pi}} \text{ for all } \mathbf{y} \in \mathbf{Y} \text{ and all } t \in \mathcal{T}. \tag{9}$$

In other words, $\mathcal{P}$ is $\mathcal{SC}$-rationalizable if we can find a distribution among payoff function profiles in $\mathcal{SC}$ and an equilibrium selection rule that could account for the observed distribution of joint actions at each $\mathbf{x} \in \mathbf{X}^{\mathrm{data}}$. Notice that this definition embodies the Conditional Independence assumption: if it were not assumed, the correct definition of $\mathcal{SC}$-rationalizability would require $\mathrm{P}(\mathbf{y} \mid \mathbf{x}^t) = \int \lambda(\mathbf{y} \mid \mathbf{\Pi}, t)\, d\mathrm{P}_{\mathbf{\Pi}}|_{\mathbf{x}^t}$, but the assumption allows us to replace $\mathrm{P}_{\mathbf{\Pi}}|_{\mathbf{x}}$ with $\mathrm{P}_{\mathbf{\Pi}}$. If we allowed the distribution of payoff functions to vary freely with the covariates, then *any* $\mathcal{P}$ is $\mathcal{SC}$-rationalizable: it is the combination of Conditional Independence and the restriction to payoff functions obeying single-crossing differences that leads to testable implications for $\mathcal{SC}$-rationalizability. Lastly, observe that $\mathcal{SC}$-rationalizability is an ordinal concept since both the set $\mathcal{SC}$ and the set of pure strategy Nash equilibria of a game depend only on players' preferences over actions.[17]

3.2 The test

As we illustrated in Section 2, the test to check whether $\mathcal{P}$ is $\mathcal{SC}$-rationalizable is a two-step procedure. In Step 1, the econometrician finds all behavioral types that are consistent with the hypothesis and in Step 2, weights on these behavioral types must be found that could account for the observed conditional distributions in $\mathcal{P}$.

STEP 1: Single-crossing rationalizable behavioral types

A *behavioral type* associates a profile of actions $\mathbf{y}^t$ to each observation t. Formally, it is a function BT: $\mathbf{X}^{\text{data}} \to \mathbf{Y}$, where $\mathrm{BT}(\mathbf{x}^t) = \mathbf{y}^t$. Abusing notation a little, we shall also write BT$= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}}$, which emphasizes that $\mathbf{y}^t$ is the action profile when the covariates are $\mathbf{x}^t$. The aim of this section is to characterize all behavioral types that could be generated as Nash equilibrium behavior from a set of players with payoff functions $\mathbf{\Pi} \in \mathcal{SC}$.

Definition 2. A behavioral type BT$= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}}$ is single-crossing rationalizable or $\mathcal{SC}$-rationalizable if there exists a profile of payoff functions $\mathbf{\Pi}$ in $\mathcal{SC}$ such that

$$\mathrm{BT}\left(\mathbf{x}^t\right) = \mathbf{y}^t \in \mathrm{NE}\left(\mathbf{\Pi}, \mathbf{x}^t\right) \text{ for all } t \in \mathcal{T}.$$

Note that our use of the term single-crossing rationalizable or $\mathcal{SC}$-rationalizable to describe a behavioral type, which we had previously used to describe a data set (see Definition 1), is appropriate

[17] For example, suppose $\mathcal{P}$ can be $\mathcal{SC}$-rationalized by an equilibrium selection rule λ and a distribution that (say) gives a weight of 3/4 on $\mathbf{\Pi}'' \in \mathcal{SC}$ and 1/4 on $\mathbf{\Pi}' \in \mathcal{SC}$. Then $\mathcal{P}$ can also be rationalized by a distribution giving weight 3/4 to $\tilde{\mathbf{\Pi}}'' \in \mathcal{SC}$ and 1/4 to $\tilde{\mathbf{\Pi}}' \in \mathcal{SC}$, where $\tilde{\mathbf{\Pi}}''$ and $\tilde{\mathbf{\Pi}}'$ are strictly increasing transformations of $\tilde{\mathbf{\Pi}}''$ and $\tilde{\mathbf{\Pi}}'$ respectively, and any equilibrium selection rule $\tilde{\lambda}$ such that $\tilde{\lambda}(\mathbf{y} \mid \tilde{\mathbf{\Pi}}'', \mathbf{x}^t) = \tilde{\lambda}(\mathbf{y} \mid \mathbf{\Pi}'', \mathbf{x}^t)$ and $\tilde{\lambda}(\mathbf{y} \mid \tilde{\mathbf{\Pi}}', \mathbf{x}^t) = \tilde{\lambda}(\mathbf{y} \mid \mathbf{\Pi}', \mathbf{x}^t)$ for all $\mathbf{y} \in \mathbf{Y}$ and $t \in \mathcal{T}$.

since one could think of a behavioral type BT as an imaginary data set in which we fix a group of agents and observe their joint actions across a finite number of covariate values.

Revealed Monotonicity (RM) Axiom. A behavioral type BT$= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t\in\mathcal{T}}$ obeys the *revealed monotonicity (RM) axiom* if, for every player $i \in \mathcal{N}$,

$$(\mathbf{y}^s_{-i}, x^s_i) \geqslant (\mathbf{y}^t_{-i}, x^t_i) \Longrightarrow y^s_i \geqslant y^t_i \text{ for any } s \text{ and } t \in \mathcal{T}. \tag{10}$$

This axiom imposes a monotonicity restriction on BT in the sense that it requires player i to take a weakly higher action at t compared to s whenever all other players are choosing higher actions and the covariate values are also higher. The following theorem states that this axiom fully characterizes $\mathcal{SC}$-rationalizability for a behavioral type.

Theorem 1. BT$= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t\in\mathcal{T}}$ *is $\mathcal{SC}$-rationalizable if and only if it satisfies the RM axiom.*

We can think of Theorem 1 as a revealed preference counterpart to the Basic Theorem. Whereas that theorem tells us that whenever $\mathbf{\Pi} \in \mathcal{SC}$, players have monotone best response functions, this result says that one could rationalize the observed action profiles in a given BT with some $\Pi \in \mathcal{SC}$, so long as it displays no violations of monotonicity. Theorem 1 gives the econometrician, through the RM axiom, a simple way of checking whether or not a given behavioral type is $\mathcal{SC}$-rationalizable.

Observe that the RM axiom is *necessary* for $\mathcal{SC}$-rationalizability. Indeed, suppose BT$= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t\in\mathcal{T}}$ is $\mathcal{SC}$-rationalizable and for some player i, we have $(\mathbf{y}^s_{-i}, x^s_i) \geqslant (\mathbf{y}^t_{-i}, x^t_i)$. Then there is a payoff function Π_i for player i that obeys single-crossing differences and under which $y^{\tilde{t}}_i$ is the best response to $(\mathbf{y}^{\tilde{t}}_{-i}, x^{\tilde{t}}_i)$ for all $\tilde{t} \in \mathcal{T}$, and (10) follows immediately from the Basic Theorem (i).

The proof of the *sufficiency* of the RM axiom is in the Appendix. In the proof we show that whenever (10) holds for a player i, then we can explicitly construct a single-crossing payoff function Π_i for that player such that y^t_i is the best response to $(\mathbf{y}^t_{-i}, x^t_i)$. (And, obviously, if this is repeated for every i then $\mathbf{x}^t$ is a Nash equilibrium of $G(\mathbf{\Pi}, \mathbf{x}^t)$ for all $t \in \mathcal{T}$.) The Π_i that we construct is additive in the players' own actions (i.e., has the form (8)) and satisfies the stronger increasing differences property (see (7)).[18]

[18] Note that this phenomenon is not altogether surprising and something similar is found in Afriat's Theorem: the generalized axiom of revealed preference (GARP) is necessary so long as the consumer has a locally nonsatiated preference, but when a data set obeys GARP then it can then be rationalized by a utility function with stronger properties: continuous, increasing, and concave.

Lastly, note that while the RM axiom forbids one player from taking a strictly lower action when other players are taking higher actions, it does not impose a joint monotonicity condition on the action profile: the axiom does *not* require that if $\mathbf{x}^2 > \mathbf{x}^1$, then $\mathrm{BT}(\mathbf{x}^2) = \mathbf{y}^2 > \mathbf{y}^1 = \mathrm{BT}(\mathbf{x}^1)$. The axiom allows for $\mathrm{BT}(\mathbf{x}^2) = \mathbf{y}^2 < \mathbf{y}^1 = \mathrm{BT}(\mathbf{x}^1)$ since it corresponds to the case where there are two ranked Nash equilibria (at both $\mathbf{x}^1$ and $\mathbf{x}^2$), with players' jointly playing the lower equilibrium at the higher covariate value.

STEP 2: Finding weights on behavioral types

Since the set of possible action profiles $\mathbf{Y}$ and the set of observations $\mathcal{T}$ are finite, the set of all possible behavioral types is also a finite set. Specifically, there are $|\mathbf{Y}|^T$ behavioral types, since for each of the T observed covariates $\mathbf{x}^t$ there are $|\mathbf{Y}|$ joint choices that a group can make. Some of these behavioral types are $\mathcal{SC}$-rationalizable while others are not. Theorem 1 allows the econometrician to check whether any given behavioral type is $\mathcal{SC}$-rationalizable by checking if the RM axiom holds. Ignoring computational issues for now (we shall address them in the next section), let us assume that the entire set of $\mathcal{SC}$-rationalizable behavioral types is worked out. We denote this set by $\mathcal{B}$.

Suppose $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, with the distribution $\mathrm{P}_{\mathbf{\Pi}}$ and the equilibrium selection rule λ. Let $\mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi})$ denote the probability that BT is the observed behavioral type, conditional on $\mathbf{\Pi}$ being the realized payoff profile. Then $\mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi}) = \times_{t=1}^{T} \lambda\left(\mathrm{BT}(\mathbf{x}^t) \mid \mathbf{\Pi}, \mathbf{x}^t\right)$. Let

$$\tau^{\mathrm{BT}} = \int \mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi})\, d\mathrm{P}_{\mathbf{\Pi}}. \tag{11}$$

By Theorem 1, if BT violates the RM axiom, then $\mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi}) = 0$ for all $\mathbf{\Pi} \in \mathcal{SC}$. Therefore, for $\mathbf{\Pi} \in \mathcal{SC}$, we have $\sum_{\mathrm{BT} \in \mathcal{B}} \mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi}) = 1$. Since the support of $P_{\mathbf{\Pi}}$ lies in $\mathcal{SC}$, we obtain

$$\sum\nolimits_{\mathrm{BT} \in \mathcal{B}} \tau^{\mathrm{BT}} = \int_{\mathcal{SC}} d\mathrm{P}_{\mathbf{\Pi}} = 1.$$

Furthermore, since $\lambda(\mathbf{y} \mid \mathbf{\Pi}, \mathbf{x}^t) = \sum_{\{\mathrm{BT} \in \mathcal{B}: \mathrm{BT}(\mathbf{x}^t) = \mathbf{y}\}} \mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi})$, we obtain from (9) that

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}^t) = \sum\nolimits_{\{\mathrm{BT} \in \mathcal{B}:\ \mathrm{BT}(\mathbf{x}^t) = \mathbf{y}\}} \tau^{\mathrm{BT}} \text{ for all } \mathbf{y} \in \mathbf{Y} \text{ and } t \in \mathcal{T}. \tag{12}$$

We have shown that when $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, then we can decompose the population of groups according to each group's behavioral type. The fraction of the population which takes a joint action $\mathbf{y}$ at $\mathbf{x}^t$ consists of precisely those behavioral types BT with $\mathrm{BT}(\mathbf{x}^t) = \mathbf{y}$.

This result admits a converse. Suppose there is a distribution $\tau = (\tau^{\mathrm{BT}})_{\mathrm{BT}\in\mathcal{B}}$ on $\mathcal{B}$ such that (12) holds. By definition, there is some $\mathbf{\Pi} \in \mathcal{SC}$ that rationalizes BT for each BT$\in \mathcal{B}$. By taking increasing transformations if necessary, we can guarantee that distinct behavioral types in $\mathcal{B}$ are rationalized by distinct payoff function profiles in $\mathcal{SC}$. We denote the profile that rationalizes BT by $\mathbf{\Pi}^{\mathrm{BT}}$. Then $\mathcal{P}$ is $\mathcal{SC}$-rationalizable with a distribution $\mathrm{P}_{\mathbf{\Pi}}$ that assigns probability τ^{BT} to $\mathbf{\Pi}^{\mathrm{BT}} \in \mathcal{SC}$ and an equilibrium selection rule λ where $\lambda\left(\mathbf{y} \mid \mathbf{\Pi}^{\mathrm{BT}}, \mathbf{x}^t\right) = 1$ if $\mathbf{y} = \mathrm{BT}(\mathbf{x}^t)$ and $\lambda\left(\mathbf{y} \mid \mathbf{\Pi}^{\mathrm{BT}}, \mathbf{x}^t\right) = 0$ if $\mathbf{y} \neq \mathrm{BT}(\mathbf{x}^t)$; in other words, all groups in the population with payoff profile $\mathbf{\Pi}^{\mathrm{BT}}$ will play $\mathrm{BT}(\mathbf{x}^t)$ at each t. The next result summarizes our observations.

Theorem 2. *$\mathcal{P} = \{\mathrm{P}(\cdot \mid \mathbf{x}^t)\}_{t\in\mathcal{T}}$ is $\mathcal{SC}$-rationalizable if and only if there exists a distribution $\tau = (\tau^{\mathrm{BT}})_{\mathrm{BT}\in\mathcal{B}}$ on $\mathcal{B}$ (the set of behavioral types obeying the RM axiom) such that (12) holds.*

Theorem 2 provides the final step needed to establish the $\mathcal{SC}$-rationalizability of $\mathcal{P}$: the econometrician must check if there is a distribution on $\mathcal{B}$ that solves (12), which in turn involves finding a positive solution to a set of equations linear in the unknowns τ^{BT} for all BT$\in \mathcal{B}$.

3.3 Inference and predictions with single-crossing

Having ascertained that a data set $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, we can go on to analyze the behavioral types that explain the data. To be specific, imagine that we are interested in the role played by a subset of types $\mathcal{B}^* \subset \mathcal{B}$. Since the set of distributions τ that solves system (12) is a convex set with typically more than one element, the total weight attributed to $\mathcal{B}^*$ cannot be uniquely identified. Nevertheless, it makes sense to find the greatest weight on $\mathcal{B}^*$ that is consistent with the $\mathcal{SC}$-rationalizability of $\mathcal{P}$, by solving the following linear program:

$$\max \textstyle\sum_{\mathrm{BT}\in\mathcal{B}^*} \tau^{\mathrm{BT}} \quad \text{subject to the distribution } \tau \text{ solving (12).} \tag{13}$$

We give two cases where this is useful, both of which are empirically implemented in Section 5.[19]

Application 1. Estimating the role of strategic interaction

While our model allows for the possibility that each player reacts strategically to other players in the game, it is conceivable that the data could be explained more simply, without appealing to strategic effects for one or more players in the game. This question can be explored using our techniques.

[19] Of course, there could be situations where we are interested in the least possible weight on $\mathcal{B}^*$ (obtained by solving (13) with the max operator replaced by min) or the entire interval of possible weights on $\mathcal{B}^*$.

To be specific, suppose we wish to check whether it is possible to regard a subgroup $\mathcal{N}'$ of the players as nonstrategic. This can be formalized by letting $\mathcal{B}^*$ (contained in $\mathcal{B}$) be the set of behavioral types that are $\mathcal{SC}$-rationalizable without strategic interactions for the players in $\mathcal{N}'$, i.e., $\{(\mathbf{y}^t, \mathbf{x}^t)\}_{t\in\mathcal{T}} \in \mathcal{B}^*$ if, for every agent $i \in \mathcal{N}'$, there is Π_i that depends *only on* y_i *and* x_i, such that $y_i^t = \text{argmax}_{y_i \in Y_i} \Pi_i(y_i, x_i^t)$ for all t. In other words, player i's behavior can be explained simply as a best response to x_i, without reference to the actions of other players. Notice also that the types in $\mathcal{B}^*$ can be characterized by a version of the RM axiom: for each $i \in \mathcal{N}'$, we require $y_i^s \geqslant y_i^t$ whenever $x_i^s \geqslant x_i^t$, for all s, $t \in \mathcal{T}$. This characterization allows us to construct $\mathcal{B}^*$, and thus we can solve (13). If the solution is 1, we conclude that $\mathcal{P}$ can be $\mathcal{SC}$-rationalized without requiring the players in $\mathcal{N}'$ to be strategic; on the other hand, if the solution is strictly below 1, then we *must* incorporate strategic interactions among these players to $\mathcal{SC}$-rationalize $\mathcal{P}$.

Application 2. Probability bounds for Nash equilibrium actions

Given a strategy profile $\overline{\mathbf{y}}$ and covariate $\overline{\mathbf{x}}$, we pose the following question: among all the possible $\mathcal{SC}$-rationalizations of $\mathcal{P}$, what is the greatest possible fraction of groups which have $\overline{\mathbf{y}}$ as a pure strategy Nash equilibrium at $\overline{\mathbf{x}}$? In the case where $\overline{\mathbf{x}} \in \mathbf{X} \backslash \mathbf{X}^{\text{data}}$, the answer to this question provides information on how the game would be played at an hitherto unobserved covariate value; this may be relevant (for example) to a policy-maker who could manipulate these covariates. But the question is also interesting if $\overline{\mathbf{x}} \in \mathbf{X}^{\text{data}}$.

To see why, notice that there is a distinction between $\text{P}(\overline{\mathbf{y}} \mid \overline{\mathbf{x}})$, the observed fraction of groups in the population that play $\overline{\mathbf{y}}$ at $\overline{\mathbf{x}}$, and the fraction of groups for which $\overline{\mathbf{y}}$ is a Nash equilibrium. The former is typically smaller than the latter because some groups who play strategy profiles other than $\overline{\mathbf{y}}$ *may also have* $\overline{\mathbf{y}}$ *as a Nash equilibrium.* As a very simple example of this phenomenon, suppose we observe the distribution of action profiles at an entry game with two firms at a single covariate value $\overline{\mathbf{x}}$. Assuming that preferences obey single crossing differences (in the sense of (3)) and are strict, if (E, E) or (N, N) is played by a pair of firms, then it has to be their unique equilibrium, but any pair that plays (E, N) may also have (N, E) as another (albeit unselected) equilibrium. Thus if $\text{P}(E, N \mid \overline{\mathbf{x}})$ and $\text{P}(N, E \mid \overline{\mathbf{x}})$ are the observed probabilities of action profiles (E, N) and (N, E) respectively, then the probability that (E, N) (similarly, (N, E)) is a Nash equilibrium profile at $\mathbf{x} = \overline{\mathbf{x}}$ is no greater than $\text{P}(E, N \mid \overline{\mathbf{x}}) + \text{P}(N, E \mid \overline{\mathbf{x}})$.

The distinction between $\mathrm{P}(\bar{\mathbf{y}} \mid \bar{\mathbf{x}})$ and the greatest possible weight on those groups which have $\bar{\mathbf{y}}$ as a Nash equilibrium at $\mathbf{x} = \bar{\mathbf{x}}$ is relevant because, if the gap is small, then we are sure that changing the equilibrium selection scheme *cannot* significantly increase the frequency with which $\bar{\mathbf{y}}$ is played. This means (for example) that a policy maker who wants $\bar{\mathbf{y}}$ to be played more often must alter payoffs in some way and it is not possible to simply convince players to coordinate on a different equilibrium. An earlier analysis of questions of this type can be found in Aradillas-Lopez (2011), which focuses on a different class of games.[20]

To answer our question, we let $\mathcal{B}^*$ be the set of behavioral types with the property that $\mathrm{BT} \in \mathcal{B}^*$ if there is $\mathbf{\Pi} \in \mathcal{SC}$ that rationalizes BT and $\bar{\mathbf{y}} \in \mathrm{NE}(\mathbf{\Pi}, \bar{\mathbf{x}})$. An application of Theorem 1 tells us that $\mathrm{BT} = \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}} \in \mathcal{B}^*$ if and only if $\{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}} \cup \{(\mathbf{y}^0, \mathbf{x}^0)\}$ obeys RM axiom, where $\{(\mathbf{y}^0, \mathbf{x}^0)\}$ is a notional observation added to the list, with $\mathbf{y}^0 = \bar{\mathbf{y}}$ and $\mathbf{x}^0 = \bar{\mathbf{x}}$; its inclusion guarantees that $\bar{\mathbf{y}}$ is a Nash equilibrium at $\bar{\mathbf{x}}$. Once $\mathcal{B}^*$ is determined, we can find the value of (13).[21] Suppose the value is β; then Theorem 1 guarantees that there is an $\mathcal{SC}$-rationalization of $\mathcal{P}$ where the probability that $\bar{\mathbf{y}}$ is a Nash equilibrium at $\bar{\mathbf{x}}$ is precisely β and no higher value is possible (or, in other words, β is a *tight* upper bound and not just an upper bound).

Other applications

In Application 2, we focused on a single action profile $\bar{\mathbf{y}}$ and asked how we could estimate its incidence at a given covariate value $\bar{\mathbf{x}}$. More generally, we could determine the *distributions over action profiles* that could arise at $\bar{\mathbf{x}}$, assuming consistency with $\mathcal{SC}$-rationalizability. This is explained in the Online Appendix A1, where we also show that when we compare the set of predicted distributions at $\bar{\mathbf{x}}$, with the set of predicted distributions at another covariate value $\hat{\mathbf{x}}$, with $\bar{\mathbf{x}} > \hat{\mathbf{x}}$, then the set at $\bar{\mathbf{x}}$ dominates the set at $\hat{\mathbf{x}}$ in a sense that generalizes first order stochastic dominance.

[20] Although the games considered in Aradillas-Lopez (2011) are referred to in that paper as games with strategic substitutes/complements, those properties are defined with respect to the monotonicity of the payoff functions in opponents' action, rather than the monotonicity of best responses. Thus the class of games treated in that paper is different from the one we consider. Nonetheless, the key idea behind that paper is captured by our simple example: under various assumptions on payoff functions, the probability that a given strategy profile $\mathbf{y}$ is a Nash equilibrium when $\mathbf{x} = \bar{\mathbf{x}}$ is not simply bounded above by 1, because there are some profiles $\mathbf{y}'$ that *cannot coexist* with $\mathbf{y}$ as Nash equilibria of a group when $\mathbf{x} = \bar{\mathbf{x}}$. In our model, we also make use of the structure we impose on how payoff functions vary *across* covariates, which enables us to say that a group/behavioral type that plays some profile $\mathbf{y}'$ at a different covariate value $\hat{\mathbf{x}}$ cannot play $\mathbf{y}$ at $\bar{\mathbf{x}}$ (see the example in footnote 27). These structural restrictions allow us to form a nontrivial upper bound on how often $\mathbf{y}$ is a Nash equilibria when $\mathbf{x} = \bar{\mathbf{x}}$.

[21] In the case where $\bar{\mathbf{x}} \in \mathbf{X}^{\text{data}}$, we should not confuse $\mathcal{B}^*$ with $\mathcal{B}^{**}$, the set of $\mathcal{SC}$-rationalizable types that actually play $\bar{\mathbf{y}}$ at $\mathbf{x}^s$; formally, $\mathrm{BT} \in \mathcal{B}^{**}$ if it is in $\mathcal{B}$ and $\mathrm{BT}(\bar{\mathbf{x}}) = \bar{\mathbf{y}}$. Obviously, $\mathcal{B}^{**} \subset \mathcal{B}^*$.

Another interesting problem is the estimation of the probability that a player i prefers a given action y_i'' to another action y_i', fixing the actions of other players in the game at $\overline{\mathbf{y}}_{-i}$ and the covariate value at $\overline{\mathbf{x}}$. Obviously, this probability is bounded below by $\mathrm{P}((y_i'', \overline{\mathbf{y}}_{-i}) \mid \overline{\mathbf{x}})$, which is observable, but this is not a tight lower bound because there could be other player i types who prefer y_i'' to y_i', even if y_i'' is not the most preferred strategy. In the Online Appendix A1, we formulate the relevant linear programs that would allow us to obtain tight bounds on this probability.

3.4 Variations on the baseline model

As we have already pointed out, our results could be applied to two-player games of strategic substitutes, since these games could be thought of as games of strategic complements, once we reverse the order of the strategy on one of the two players. More generally, our results could be applied to any class of games where the hypotheses involve pure strategy equilibrium play and monotone best responses for one or more players in the game (whether increasing or decreasing with respect to other players' actions). For example, in a three-player game one could test if player 1's best response is increasing with player 2's action and decreasing in player 3's; player 2's action is decreasing in the action of the other two players; etc. We are leaving aside here the issue of whether or not these games in general have pure strategy Nash equilibria; that may or may not be the case depending on the structure of strategic substitutes and complements, but so long as the modeler is convinced that such a hypothesis is reasonable and would like to test it with our method, the test itself is no more or less complicated than testing the case of strategic complements.

In aggregative games, a player's payoff is affected by other players' actions only through an aggregate value. The payoff of player i depends on the player's action y_i (which to keep our discussion simple we shall assume is one-dimensional), on $\sum_{j \neq i} y_j$ and on x_i, and thus could be written as $\Pi_i(y_i, \sum_{j \neq i} y_j, x_i)$. If Π_i has single-crossing differences in $(y_i; (\sum_{j \neq i} y_j, x_i))$ then the Basic Theorem guarantees that player i's best response is increasing in $(\sum_{j \neq i} y_j, x_i)$ so that, in particular, we have a game of strategic complements. On the other hand, if Π_i has single-crossing differences in $(y_i; (-\sum_{j \neq i} y_j, x_i))$ then player i's best response is increasing in $(-\sum_{j \neq i} y_j, x_i)$ and we have a game of strategic substitutes.[22] We can adapt our two-step procedure in an obvious way to check

[22] Even though games of strategic substitutes do not in general have pure strategy Nash equilibria, aggregative games with strategic substitutes do have pure strategy Nash equilibria (see Dubey, Haimanko and Zapechelnyuk (2006) and also Jensen (2010)).

for rationalizability subject to (aggregate) single crossing restrictions on the payoff functions, as well as make inferences and predictions along the lines indicated in Section 3.3.

3.5 Related nonparametric results

Our econometric modelling is close in spirit, though not in its specifics with the nonparametric random utility models studied in Tebaldi, Torgovitsky, and Yang (2020), Deb et al. (2018), Kitamura and Stoye (2018), Hoderlein and Stoye (2014), Manski (2007), McFadden (2005), McFadden and Richter (1991), and Marschak (1960). Manski (2007) includes a discussion of why this approach may be a useful complement to parametric methods. As far as we know, our paper is the first to exploit this nonparametric approach to study games.

In some nonparametric random utility models (such as those studied by Kitamura and Stoye (2018) and Manski (2007)) data generation is achieved through variation in budget sets. In this paper, we have avoided introducing variability in strategy sets, in order not to burden the reader with too many model features and also because our econometric application in Section 5 does not have such variation. In Lazzati, Quah, and Shirai (2018), we develop results on rationalizability, inference, etc. to data sets in which players' strategy sets are allowed to vary across observations (in addition to changing covariates); however, those results do not generalize Theorem 1 because it requires the actions of each player to be one-dimensional. (See also Carvajal (2004) for a related result in the context of one-dimensional action spaces.) Our proof of that result uses an approach significantly different from the one taken to proof Theorem 1.

There are a number of results in the monotone comparative statics literature that are related to ours. Topkis (1998, Theorem 2.8.9) considers a correspondence φ mapping elements of a totally ordered parameter set $\mathcal{R}$ to compact sublattices of $\mathbb{R}^\ell$. He shows that this correspondence is increasing in the strong set order (as the parameter increases) if and only if it can be exactly rationalized by a payoff function that is supermodular in the choice variable and has increasing difference between the choice variable and the parameter.[23] This is related to Theorem 1 since in that result we characterize the case where y_i^t can be rationalized by payoff function Π_i obeying (6); in fact, the rationalizing function constructed in the proof has the stronger property (7), which is

[23] By 'exactly rationalized' we mean that there is a rationalizing payoff function f such that $\operatorname{argmax}_{y \in \mathbb{R}^\ell} f(y, r) = \varphi(r)$ for all $r \in \mathcal{R}$.

precisely the property characterized in Topkis' result. We also assume throughout that preferences over actions are strict; this corresponds, in Topkis' setup, to assuming that $\varphi(r)$ is a singleton, for all $r \in \mathcal{R}$. However, Theorem 1 is not a special case of Topkis' result for two related reasons. Recall that we allow the agent i's choice to be observed only at *some* values of $\mathbf{Y}_{-i} \times X_i$, rather than the entire set; notice also that $\mathbf{Y}_{-i} \times X_i$ is a partially ordered set since there will be at least one other player in the game (choosing $\mathbf{y}_{-i}$), in addition to the covariate (x_i). In Topkis' notation, we are considering the case where $\mathcal{R}$ is a partially (rather than totally) ordered set and $\varphi(r)$ is specified for some (and not necessarily all) values of $r \in \mathcal{R}$.

Echenique and Komunjer (2009) consider a structural model where there could be multiple outcomes (which could be optimal choices made by an agent or equilibrium outcomes). They show that if there is a monotone relationship between the exogenous and dependent variables in the structural function, then there will be observable restrictions on the tail quantiles of the dependent variable. The issue of rationalizability is not addressed.

Apesteguia, Ballester, and Lu (2017) characterize a random utility model where the distribution of actions on any feasible set is generated by a set of preferences that are totally ordered by single crossing differences. Note that this model is distinct from ours even in the special case where the population consists of single agents making optimal choices (rather than groups of agents playing games within each group). We rationalize those situations where the actions of agents in the population are increasing with some observable covariates; this is neither necessary nor sufficient to guarantee that agents in the population have preferences that are totally ordered by single crossing differences *with each other*.

4 Column Generation

When implementing the test procedures outlined in the previous sections, we face two hurdles. The first is that we have assumed that we observe the true population distribution of strategy profiles, when actually the data would consist of a large sample. To deal with this issue, we simply follow the approach in Kitamura and Stoye (2018) (for model testing) and Deb et al. (2020) (for inference problems) and we make no methodological contribution.

The second issue has to do with finding a cleverer way of implementing the test when carrying out Step 1 (as stated) is impossible because the set $\mathcal{B}$ is too large to be completely listed. In this section we outline the *column generation method* that was recently proposed by Smeulders et al. (2021) to deal with essentially the same computational problem in the random consumer demand model of Kitamura and Stoye (2018). Loosely speaking, the column generation method first tests a stricter version of the model corresponding to a strict subset of $\mathcal{B}$ which is completely known. Then it progressively enlarges the set by including more behavioral types from $\mathcal{B}$, up to the point where further additions will not improve the model's ability at explaining the given data.

Smeulders et al. (2021) focused on the problem of model testing but did not consider how column generation could be used in problems of inference (of the type discussed in Section 3.3 and in Deb et al. (2020)). We make a methodological contribution here by showing how column generation can also be used to solve inference problems.

In order to focus our discussion on column generation, we shall maintain the assumption that we observe the true distribution. In any empirical implementation (such as the one in Section 5), column generation will have to be combined with the econometric methods in Kitamura and Stoye (2018) and Deb et al. (2020); the Online Appendix A2 gives an account of how they work together.

4.1 Column generation for model testing

We begin by re-formulating the test given in Theorem 2 in matrix notation. Since the data set $\mathcal{P} = \{\mathrm{P}(\cdot \mid \mathbf{x}^t)\}_{t\in\mathcal{T}}$ consists of T observations and there are $|\mathbf{Y}|$ strategy profiles in total, each behavioral type $\mathrm{BT} : \mathbf{X}^{data} \to \mathbf{Y}$ may be written as a vector $\mathbf{b} = (b_{(\mathbf{y},t)})_{\mathbf{Y}\times T}$ of length $|\mathbf{Y}|T$, with $b_{(\mathbf{y},t)} = 1$ if $\mathrm{BT}(\mathbf{x}^t) = \mathbf{y}$ and $b_{(\mathbf{y},t)} = 0$ otherwise. Plainly for any $\mathbf{b} \in \{0,1\}^{|\mathbf{Y}|\times T}$, there is some behavioral type corresponding to it, if and only if $\sum_{\mathbf{y}\in\mathbf{Y}} b_{\mathbf{y},t} = 1$ at each $t \in \mathcal{T}$. Similarly, the data set $\mathcal{P}$ can be succinctly captured by the column vector $\mathbf{p}$, where the $(\mathbf{y},t)$-th entry of $\mathbf{p}$ is $\mathrm{P}(\mathbf{y}|\mathbf{x}^t)$.

Let $\overline{\mathcal{B}}$ be the set of all logically possible behavioral types. In what follows, abusing notation, $\overline{\mathcal{B}}$ and its subset may also denote the set of binary vectors corresponding to their elements. Contained within $\overline{\mathcal{B}}$ is $\mathcal{B}$, the set of behavioral types that satisfy the RM axiom; associated with $\mathcal{B}$ is the matrix $\mathbf{B}$, with each column vector in this matrix representing a behavioral type in $\mathcal{B}$. Theorem 2 states that $\mathcal{P}$ is $\mathcal{SC}$-rationalizable if and only if there is $\tau \in \Delta^{\mathcal{B}}$, the set of distributions on $\mathcal{B}$, that solves $\mathbf{B}\tau = \mathbf{p}$. ($\Delta^{\mathcal{B}}$ could be thought of as elements of the standard $(|\mathcal{B}| - 1)$-simplex.) In other words, $\mathcal{P}$ is $\mathcal{SC}$-rationalizable if and only if

$$J := \min_{\tau \in \Delta^{\mathcal{B}}} (\mathbf{p} - \mathbf{B}\tau) \cdot (\mathbf{p} - \mathbf{B}\tau)$$

is equal to zero. Notice that $\{\mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}}\}$ gives all the (possible) data sets that could be generated by the behavioral types in $\mathcal{B}$, as we vary the distribution over these types; J is (the square of) the Euclidean distance between $\{\mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}}\}$ and the actual data set $\mathbf{p}$.

Let $\mathcal{B}_0$ be a subset of $\mathcal{B}$ and denote the corresponding matrix by $\mathbf{B}_0$. Then we can solve for

$$J_0 := \min_{\tau \in \Delta^{\mathcal{B}_0}} (\mathbf{p} - \mathbf{B}_0\tau) \cdot (\mathbf{p} - \mathbf{B}_0\tau)$$

J_0 gives the distance between $\mathbf{p}$ and the set of possible data sets obtained as we vary the distribution over $\mathcal{B}_0$. We say that $\mathcal{B}_0$ is *improvable* if $J_0 > J$ and we say that a behavioral type $\widehat{\mathbf{b}} \in \mathcal{B}$ *improves* $\mathcal{B}_0$ if, when $\widehat{\mathbf{b}}$ is added to $\mathcal{B}_0$, the new value of J_0 is *strictly* lower. The following result provides us with a way of checking whether J_0 is improvable.

Proposition 1. *If $\mathcal{B}_0 \subset \mathcal{B}$ is improvable, then there is $\widehat{\mathbf{b}} \in \mathcal{B}$ such that*

$$(\mathbf{p} - \eta_0) \cdot (\widehat{\mathbf{b}} - \eta_0) > 0, \tag{14}$$

where $\tau_0 = \arg\min_{\tau \in \Delta^{\mathcal{B}_0}} (\mathbf{p} - \mathbf{B}_0\tau) \cdot (\mathbf{p} - \mathbf{B}_0\tau)$ and $\eta_0 = \mathbf{B}_0\tau_0$. Conversely, if there is $\widehat{\mathbf{b}} \in \mathcal{B}$ such that (14) holds, then $\widehat{\mathbf{b}}$ improves $\mathcal{B}_0$.

This result states that if $\mathcal{B}_0$ is improvable, then there is $\widehat{\mathbf{b}} \in \mathcal{B}$ that, when included in $\mathcal{B}_0$ will shrink the distance between (the new) $\mathcal{B}_0$ and $\mathbf{p}$. Furthermore, (14) provides a criterion for finding $\widehat{\mathbf{b}}$. Notice that this result provides us with an *algorithm to solve for J via column generation.* First, we begin with a nonempty $\mathcal{B}_0$ that is easy to list and calculate J_0. Then J_0 is improvable if and only if we can find $\widehat{\mathbf{b}} \in \mathcal{B}$ that satisfies (14). If such a $\widehat{\mathbf{b}}$ can be found, then we can add it to $\mathcal{B}_0$ and recalculate J_0 and η_0. Then we can try to find another element in $\mathcal{B}$ that improves on $\mathcal{B}_0$ via (14); if one exists we can repeat the process. At each stage of the algorithm either no improving $\widehat{\mathbf{b}}$ can be found, in which case we can conclude that $\mathcal{B}_0$ is not improvable, or one can be found, which leads to an improvement. Since $\mathcal{B}$ is finite, this algorithm must terminate, and at the end we can be sure we have found $\mathcal{B}_0$ such that $J_0 = J$.

This algorithm is essentially the same as the one formulated by Smeulders et al. (2021). Notice that the effectiveness of this approach depends on the ease with which we could find $\widehat{\mathbf{b}}$ that solves

(14) or ascertain that it does not exist, which in turn depends on how it is to characterize $\mathcal{B}$. In our application, $\mathcal{B}$ can be characterized as solutions to a linear equation.

PROPOSITION 2. *We can construct a matrix C and a column vector θ, both with nonnegative integer entries, such that for any $\mathbf{b} \in \overline{\mathcal{B}}$, we have $\mathbf{b} \in \mathcal{B}$ if and only if $C\mathbf{b} \leqslant \theta$.*

The formulae for C and θ are found in our proof of this proposition in the Appendix. Given this result, there is $\widehat{\mathbf{b}} \in \mathcal{B}$ that solves (14) if and only if the value of the problem

$$\max\,(\mathbf{p} - \eta_0) \cdot (\mathbf{b} - \eta_0) \text{ subject to } \mathbf{b} \in \overline{\mathcal{B}} \text{ and } C\mathbf{b} \leqslant \theta. \tag{15}$$

is strictly positive. If it is, we can add $\widehat{\mathbf{b}}$ to $\mathcal{B}_0$ and then repeat the process. More details on the column generation procedure can be found in Online Appendix A2.1.

Column generation and testing on an empirical distribution. Suppose that a data set $\mathbf{p}$ is an empirical distribution with sample size N, rather than the true population distribution. In such a case, even if there is no distribution τ that solves $\mathbf{B}\tau = \mathbf{p}$, we have to consider if the violation of $\mathcal{SC}$-rationalizability is statistically significant. Letting

$$J_N := \min_{\tau \in \Delta^{\mathcal{B}}} N(\mathbf{p} - \mathbf{B}\tau) \cdot (\mathbf{p} - \mathbf{B}\tau), \tag{16}$$

it is reasonable to conclude that we cannot reject the hypothesis that the population distribution is $\mathcal{SC}$-rationalizable if J_N is sufficiently close to 0. Kitamura and Stoye (2018) propose a procedure for constructing a bootstrap sample $\widehat{\mathbf{p}}^{(r)}$ for $r = 1, 2, \ldots, R$, calculating the corresponding test statistic $J_N^{(r)}$, and then using these to calculate the p-value

$$p = \#\{J_N^{(r)} > J_N\}/R. \tag{17}$$

The column generation method could be used both in calculating J_N and in calculating $J_N^{(r)}$.[24] Furthermore, as noted by Smeulders et al. (2021), given the p-value's formula (17), there is no need to calculate $J_N^{(r)}$ precisely: we only need to ascertain if it is smaller or larger than J_N. This means that when implementing the column generation procedure to calculate the value of $J_N^{(r)}$, we can terminate the calculation once the tentative value of $J_N^{(r)}$ drops below J_N;, which significantly decreases the computation time in our application (discussed in Section 5). More details on the Kitamura and Stoye (2018) statistical test can be found in the Online Appendix A2.1.

[24] The application of column generation for calculating $J_N^{(r)}$ is somewhat more involved than that for calculating J_N because of the introduction of a 'tuning parameter' in the former. See Online Appendix A2.1 for the details.

4.2 Column generation for inference

Our starting point is an $\mathcal{SC}$-rationalizable data set $\mathcal{P}$, and we are interested in finding the greatest possible weight one could assign to the types in $\mathcal{B}^*$ (contained in $\mathcal{B}$) in any rationalization of $\mathcal{P}$ (in other words, the solution to problem (13)). This can be obtained by checking, for a given $\beta > 0$, whether there is a distribution $\tau = (\tau^{\mathbf{b}})_{\mathbf{b}\in\mathcal{B}}$ on $\mathcal{B}$ that solves

$$\mathbf{p} = \mathbf{B}\tau \quad \text{subject to } \textstyle\sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta; \tag{18}$$

provided this problem can be solved, the maximum weight on $\mathcal{B}^*$ can be obtained by a binary search over different values of β.[25] However, as in the case with model testing, the problem of finding $\tau \in \Delta^{\mathcal{B}}$ that solves (18) cannot be approached directly if $\mathcal{B}$ or $\mathcal{B}^*$ is too large to be fully listed.

Clearly, there is a distribution that solves (18) if and only if

$$J^{\beta} := \min\,(\mathbf{p} - \mathbf{B}\tau)\cdot(\mathbf{p} - \mathbf{B}\tau) \text{ subject to } \tau \in \Delta^{\mathcal{B}} \text{ and } \textstyle\sum_{\mathbf{b}\in\mathbf{B}^*} \tau^{\mathbf{b}} \geqslant \beta$$

equals zero. Let $\mathcal{B}_0$ be a subset of $\mathcal{B}$ such that $\mathcal{B}_0 \cap \mathcal{B}^*$ is nonempty, and let $\Delta^{\mathcal{B}_0}$ be the set of distributions on $\mathcal{B}_0$. Then

$$J_0^{\beta} := \min\,(\mathbf{p} - \mathbf{B}_0\tau)\cdot(\mathbf{p} - \mathbf{B}_0\tau) \text{ subject to } \tau \in \Delta^{\mathcal{B}_0} \text{ and } \textstyle\sum_{\mathbf{b}\in\mathcal{B}^*\cap\mathcal{B}_0} \tau^{\mathbf{b}} \geqslant \beta \tag{19}$$

is well-defined and gives the distance between the data set $\mathbf{p}$ and

$$\left\{\mathbf{B}_0\tau :\ \tau \in \Delta^{\mathcal{B}_0} \text{ and } \textstyle\sum_{\mathbf{b}\in\mathcal{B}^*\cap\mathcal{B}_0} \tau^{\mathbf{b}} \geqslant \beta\right\}.$$

We say that $\mathcal{B}_0$ *is improvable given problem* (18) if $J_0^{\beta} > J^{\beta}$. A pair of types $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$, with $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ and $\widehat{\mathbf{b}} \in \mathcal{B}$ is said to *improve* $\mathcal{B}_0$ if the new $\mathcal{B}_0$ after the inclusion of $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$ gives rise to a strictly lower value of J_0^{β}. The next result says that the improvability of $\mathcal{B}_0$ can be characterized by the presence of an improving pair.

PROPOSITION 3. *If the set $\mathcal{B}_0 \subset \mathcal{B}$ is improvable given problem (18), then there is a pair of types $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$, with $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ and $\widehat{\mathbf{b}} \in \mathcal{B}$ such that*

$$(\mathbf{p} - \eta_0)\cdot\left(\beta\widehat{\mathbf{b}}^* + (1-\beta)\widehat{\mathbf{b}} - \eta_0\right) > 0, \tag{20}$$

[25] There could be occasions where we are interested in finding the minimum weight on a subset of behavioral types $\mathcal{B}^*$ rather than the maximum. Our procedure is potentially applicable to those problems as well, since it is always possible to convert a minimum weight problem to one of finding the greatest weight on the set $\mathcal{B}\backslash\mathcal{B}^*$.

where τ_0 *is the distribution that achieves* J_0^β *and* $\eta_0 = \mathbf{B}_0\tau_0$*. Conversely, suppose there is* $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ *and* $\widehat{\mathbf{b}} \in \mathcal{B}$ *such that (20) holds; then* $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$ *improves* $\mathcal{B}_0$ *given problem (18).*

Clearly, this result has the same structure as Proposition 1, and just as that proposition provides us with the basis for using column generation to calculate J, so Proposition 3 justifies the use of a column generation algorithm to calculate J^β. Indeed, we begin with $\mathcal{B}_0$ and then try to find a pair $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$ that satisfies (20). If such a pair can be found, the proposition tells us that it improves $\mathcal{B}_0$. We then re-calculate J_0^β after including the pair in $\mathcal{B}_0$. Based on the new $\mathcal{B}_0$, we again search for a pair of types that satisfies (20). We repeat the process until no improving pair can be found, at which point Proposition 3 tells us that $J_0^\beta = J^\beta$.

The effectiveness of this algorithm hinges in part on how easy it is to find a pair $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ and $\widehat{\mathbf{b}} \in \mathcal{B}$ that satisfies (20). For this purpose it would be helpful if, like $\mathcal{B}$ (see Proposition 2), $\mathcal{B}^*$ can also be characterized by a system of integer linear inequalities. That is indeed the case for the two applications considered in Section 3.3 as shown in the Online Appendix A2.3.

As explained at the end of Section 4.1, when $\mathbf{p}$ is a sample rather than the true population distribution, we need a statistical procedure to check whether it is (statistically) $\mathcal{SC}$-rationalizable. And, in this case, we also need a statistical procedure to estimate the greatest weight on $\mathcal{B}^*$ (formally, the value of the solution to (13)); Deb et al. (2020) develops this procedure, building on Kitamura and Stoye (2018). The column generation method we developed for inference can be applied to help with the required calculations. We provide the details in the Online Appendix A2.2.

5 Empirical illustration

We apply our results in the preceding sections to an entry game using a data set taken from Kline and Tamer (2016). The data set contains the entry decisions of airlines in 7,882 markets, where a market is defined as a trip between two airports irrespective of intermediate stops. Airline firms are divided into two categories: LCC (low cost carriers) and OA (other airlines).[26] In Kline and Tamer's analysis (and in ours) the two categories are treated as two firms. Thus, in each market, the two firms, LCC and OA, can either both enter a market, both stay out, or one could enter with

[26] The data were collected from the second quarter of the 2010 Airline Origin and Destination Survey (DB1B). The low cost carriers are AirTran, Allegiant Air, Frontier, JetBlue, Midwest Air, Southwest, Spirit, Sun Country, USA3000, and Virgin America. A firm that is not a low cost carrier is, by definition, an 'other airline.'

the other staying out.

This data set also contains information on two covariates: market presence (MP) and market size (MS). Market presence is a market- and airline-specific variable. For each airline and for each airport, one counts the number of markets that the airline serves from that airport and divide it by the total number of markets served from that airport by any airline; the market presence variable for a given market and airline is the average of these ratios at endpoints of that market/trip. The construction and inclusion of this covariate is not novel and follows Berry (1992). Since the airlines are aggregated into two firms, the market presence variable is also aggregated: the market presence for LCC (resp. OA) is the maximum among the actual airlines in LCC category (resp. OA category). The second covariate, market size, is a market-specific variable (shared by all airlines in that market) and is defined as the population at endpoints of the corresponding trip. Lastly, Kline and Tamer (2016) discretize these variables, where each of them takes value 1 if the variable is higher than its median value and 0 otherwise. Thus, in our data set, there are three binary covariates, MP_{LCC}, MP_{OA}, and MS, and markets are partitioned into eight types according to realizations of them. Formally, $\mathbf{X} = \mathbf{X}^{\text{data}} = \{0,1\}^3$. Note that, in this case, MS simultaneously influences the payoffs of both LCC and OA, and we may consider $x_{LCC} = (\mathrm{MP}_{LCC}, \mathrm{MS})$ and $x_{OA} = (\mathrm{MP}_{OA}, \mathrm{MS})$.

Observations in the data set are summarized in Table 5. Ignoring small sample issues for the time being, notice that it has the same form as the cross sectional data set $\mathcal{P}$ we studied in Section 2. It consists of eight blocks, with the markets in each block sharing the same covariates. For example, there are 1,271 markets with $(\mathrm{MP}_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS}) = (0,0,0)$, of which around 30% are not served by either airline and about 68% are served only by airlines in the OA category (an action profile is written as $(y_{LCC}, y_{OA}) \in \{E,N\} \times \{E,N\}$). The entries in Table 5 seem 'reasonable,' in the sense that it appears as though a firm's entry is encouraged whenever its market presence is large or the market size is large, and it is deterred by the entry of the other firm. For example, going from $(0,0,0)$ to $(1,0,0)$ (so the market presence of LCC has increased), both $\mathrm{P}(N,N)$ and $\mathrm{P}(N,E)$ fall, while $\mathrm{P}(E,N)$ and $\mathrm{P}(E,E)$ both increase.

Testing $\mathcal{SC}$-rationalizability. Our hypothesis here is that, in each market, two firms (LCC and OA) are playing a pure strategy Nash equilibrium in a game of strategic substitutes with monotone effects from covariates. The payoff function of LCC $\Pi_{LCC}(y_{LCC}, y_{OA}, \mathrm{MP}_{LCC}, \mathrm{MS})$ is required

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (0,0,0)$ 1271 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.304	0.682	0.006	0.009

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (0,1,0)$ 763 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.190	0.785	0.003	0.022

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (1,0,0)$ 1125 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.194	0.367	0.253	0.186

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (1,1,0)$ 782 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.122	0.542	0.050	0.286

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (0,0,1)$ 869 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.159	0.823	0.001	0.017

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (0,1,1)$ 1039 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.078	0.889	0.000	0.033

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (1,0,1)$ 677 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.106	0.326	0.306	0.261

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (1,1,1)$ 1356 markets			
$\text{P}(N,N)$	$\text{P}(N,E)$	$\text{P}(E,N)$	$\text{P}(E,E)$
0.055	0.501	0.021	0.423

Table 5: Distribution of entry decisions across each realization of covariates

to obey single-crossing differences in $(y_{LCC}; (-y_{OA}, \text{MP}_{LCC}, \text{MS}))$, while the payoff function of OA, $\Pi_{OA}(y_{OA}, y_{LCC}, \text{MP}_{OA}, \text{MS})$ is required to obey single-crossing differences in $(y_{OA}; (-y_{LCC}, \text{MP}_{OA}, \text{MS}))$. By Basic Theorem in Section 3.1, this ensures that a firm's entry is discouraged by the opponent's entry and enhanced by increase in own covariates. The data set is supposed to arise from a population of those firms, with unobserved heterogeneity generating a distribution of realizations of payoff functions $\mathbf{\Pi} = (\Pi_{LCC}, \Pi_{OA})$, which we denote by $\text{P}_{\mathbf{\Pi}}$. We assume Conditional Independence (see Section 3.1), which guarantees that $\text{P}_{\mathbf{\Pi}}$ is independent of the realization of $(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$.

It is not difficult to see that the empirical distribution displayed in Table 5 is not exactly $\mathcal{SC}$-rationalizable. Notice from the data that

$$\text{P}\left(N, N|1,1,0\right) + \text{P}\left(E, N|1,1,0\right) = 17.14\% < 19\% = \text{P}\left(N, N|0,1,0\right).$$

This is not compatible with $\mathcal{SC}$-rationalizability because any pair of firms with single-crossing payoff functions that select (N, N) at $(0, 1, 0)$ would either select (N, N) or (E, N) at $(1, 1, 0)$. However, this 'failure' turns out to be not statistically significant. Employing Kitamura and Stoye's (2018) statistical test outlined at the end of Section 4.1 (with a sample of 2000 generated via the bootstrap procedure), we find that the p-value defined in (17) is equal to 0.138, and hence, our hypothesis is not rejected at 5% (or 10%) significance level. In this application there are four possible strategy profiles and eight covariate values, so the number of logically possible behavioral types is 4^8. We employ the column generation method to calculate J_N and $J_N^{(r)}$. Using our R code on a desktop computer with Apple M1 processor and 16 GB RAM, the p-value was calculated in 2.590 minutes.

Significance of strategic interaction. Having established that the data set is (statistically) $\mathcal{SC}$-rationalizable we can now go on to explore its properties. In particular, we can assess the extent to which strategic interaction plays a role in explaining the data, in the sense discussed in Section 3.3 (Application 1), by considering the following sub-classes of $\mathcal{SC}$-rationalizable behavioral types: (i) those that can be rationalized with the LCC firm having a payoff function that is independent of the actions of OA; (ii) those that can be rationalized with the OA firm having a payoff function that is independent of the actions of LCC; and (iii) those that can be rationalized with both firms having payoff functions that are independent of the other firm's action. We find that, in order to $\mathcal{SC}$-rationalize the data, the greatest possible weights on these three sub-classes of behavioral types are (i) 0.902, (ii) 0.777, and (iii) 0.777 (within 5% significance level). Since these weights are all strictly less than 1, we conclude that any $\mathcal{SC}$-rationalization of the data *requires* strategic behavior for both LCC and OA firms. These upper bounds are based on the statistical procedure in Deb et al. (2020), with the computations carried out using the column generation method for inference (as set out in Section 4.2); the computation time (for all three cases together) was 8.512 seconds. The Online Appendix A2.2 provides more information on our statistical and computational procedures.

Probability bounds for equilibrium actions. Under our hypothesis, the action profiles (N, N) and (E, E) can only be played as *unique* equilibrium for each realization of $\mathbf{x} = (\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$. On the other hand, when (N, E) is played, it is possible that (E, N) is also a Nash equilibrium of the game. For this reason, the probability that (E, N) is a Nash equilibrium of the game can be strictly higher than the observed frequency with which this profile is played, even after accounting for sampling error (and by an analogous argument the same is true of (N, E)). Section 3.3 (Application 2) sets out (in principle) how we can estimate the greatest weight on those types in the population which have (E, N) as a Nash equilibrium, at a given covariate value; in our calculations, we use the the column generation method, as explained in Section 4.2.

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$	$(0,0,0)$		$(0,1,0)$		$(1,0,0)$		$(1,1,0)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
Observed Prob.	0.682	0.006	0.785	0.003	0.367	0.253	0.542	0.050
$\max \Pr[\mathbf{y} \in \mathbf{NE}(\mathbf{\Pi}, \mathbf{x})]$	0.699	0.535	0.816	0.496	0.496	0.637	0.551	0.551
$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$	$(0,0,1)$		$(0,1,1)$		$(1,0,1)$		$(1,1,1)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
Observed Prob.	0.832	0.001	0.910	0.000	0.326	0.306	0.501	0.021
$\max \Pr[\mathbf{y} \in \mathbf{NE}(\mathbf{\Pi}, \mathbf{x})]$	0.832	0.605	0.910	0.480	0.480	0.652	0.520	0.496

Table 6: Probability bounds for equilibrium action profiles

Table 6 reports that at $(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (1,0,0)$, the greatest possible weight on those behavioral types that may have (N,E) as a Nash equilibrium of the game is 0.496: this includes types which are already playing (N,E) (with observed frequency 0.367) as well those types which are playing (E,N) but may have (N,E) as an alternative Nash equilibrium.[27] Thus, even if we allow for equilibrium selection rules to change, the frequency with which (N,E) is played at $(1,0,0)$ will not exceed 0.496. Notice that, in general, $\max \Pr[\mathbf{y} \in \mathbf{NE}(\mathbf{\Pi}, \mathbf{x})]$ is closer to the observed frequency in the case where $\mathbf{y} = (N,E)$, while the same gap in the case of $\mathbf{y} = (E,N)$ is considerably bigger. For $\mathbf{y} = (N,E)$, the calculation of $\max \Pr[\mathbf{y} \in \mathbf{NE}(\mathbf{\Pi}, \mathbf{x})]$ for all possible $\mathbf{x}$ took around 25 seconds in total using the column generation method (and similarly for $\mathbf{y} = (E,N)$).

Further tests. The tests that we have done so far do not really put the column generation method through its paces because the problem is quite small. Indeed, the total number of possible behavioral types ($4^8 = 65,536$) is just about small enough to be completely listed; one could then find all the SC-rationalizable behavioral types using the RM axiom (of which there are 482) and avoid using the column generation altogether.

To check the performance of the column generation method in a 'larger' model, we repeat our analysis with a finer division of the covariates. (A fuller discussion is found in Online Appendix

[27] But it is *not* the case that every $\mathcal{SC}$-behaviorial type with (E,N) as a Nash equilibrium at $(1,0,0)$ must also have (N,E) as a Nash equilibrium at $(1,0,0)$. For example, if the behavioral type chooses (E,N) at $(1,0,0)$ and (E,E) at $(1,1,0)$, then (N,E) cannot be a Nash equilibrium at $(1,0,0)$. On the other hand, there are $\mathcal{SC}$-behavioral types that choose (E,N) at $(1,0,0)$ and (N,E) at $(1,1,0)$; in these cases, (N,E) may be a Nash equilibrium at $(1,0,0)$. The latter types are the ones included in the estimated weight, while the former types are excluded.

A3.) Instead of aggregating covariates into binary variables, we let each of MP_{LCC}, MP_{OA} and MS take four possible values using quantiles: each variable takes value $k-1$, if it is in the k-th quartile. In this way, all markets are partitioned into $4^3 = 64$ covariate values and there is a distribution of entry decisions at each of them. In this environment, the total number of possible behavioral types is enormous (4^{64}) and the same is true of the number of $\mathcal{SC}$-rationalizable behavioral types.[28] While a direct approach is no longer possible, the column generation method still works, with the test of $\mathcal{SC}$-rationalizability finishing in 19.051 minutes, including the bootstrap procedure.[29] In this case, we find that the hypothesis is rejected (with p-value, defined by (17), being equal to 0). We also consider the case where only the MS (market size) variable takes four values, while MP_{LCC} and MP_{OA} remain binary. In this case, the data set *is* SC-rationalizable, with p-value equal to 0.412. As in the case when all covariates are binary, we can carry out further analyses. We check if strategic interaction is crucial to rationalizing the data (it is) and also calculate the probability bounds for equilibrium actions. In all cases, computations finish within reasonable time.

Appendix

Proof of Theorem 1. It remains for us to show that a behavioral type BT is $\mathcal{SC}$-rationalizable if it satisfies the RM axiom. For each agent i, we need to find a payoff function $\Pi_i : Y_i \times \mathbf{Y}_{-i} \times X_i \to \mathbb{R}$ that obeys single-crossing differences in the sense of (6) such that y_i^t uniquely maximizes $\Pi_i(\cdot, \mathbf{y}_{-i}^t, x_i^t)$, for all $t \in \mathcal{T}$. In fact, we shall explicitly construct a payoff function of the form

$$\Pi_i(y_i, \mathbf{y}_{-i}^t, x_i^t) = \sum_{k=1}^{K(i)} \Pi_{ik}(y_{ik}, \mathbf{y}_{-i}^t, x_i^t),$$

with $\Pi_{ik}(y_{ik}, \mathbf{y}_{-i}^t, x_i^t)$ having increasing differences (for $k = 1, \ldots, K(i)$).[30] As argued in Section 3.1.1, the function Π_i will then satisfy increasing differences and hence single-crossing differences. In what follows, for the sake of notational simplicity, let $\mathbf{Z} = \mathbf{Y}_{-i} \times X_i$ and denote the typical element of $\mathbf{Z}$ by $\mathbf{z}$ and $\mathbf{y}_{-i}^t \times x_i^t$ by $\mathbf{z}^t$.

[28] It is straightforward to see that any behavioral type where either (E, N) or (N, E) is played at a covariate obeys the RM axiom. Hence, there are at least 2^{64} ($\approx 3.1 \times 10^{19}$) types obeying the RM axiom.

[29] One may compare this to the test where there are $2^3 = 8$ covariate values (as opposed to $4^3 = 64$), which finished in 2.590 minutes.

[30] Recall that $Y_i = \times_{k=1}^{K(i)} Y_{ik}$, for $Y_{ik} \subset \mathbb{R}$.

We now explain the construction of $\Pi_{ik}(\cdot, \mathbf{z})$. First, we find a family of single-peaked functions, $f_{ik} : Y_{ik} \times \mathcal{T} \to \mathbb{R}$, with the following two properties: (i) $f_{ik}(y^t_{ik}, t) > f_{ik}(a, t)$ for all $a \neq y^t_{ik}$ and $a \in Y_{ik}$; and (ii) if $y^s_{ik} = y^t_{ik}$ then $f_{ik}(\cdot, s) = f_{ik}(\cdot, t)$ and if $y^s_{ik} > y^t_{ik}$, then $f_{ik}(a'', s) - f_{ik}(a', s) > f_{ik}(a'', t) - f(a', t)$ for all $a'' > a'$ in Y_{ik}. This can be obtained, for example, by letting

$$f_{ik}(a, t) = -(a - y^t_{ik})^2.$$

Then, for each $\mathbf{z} \in \mathbf{Z}$, let $T(\mathbf{z}) = \{t \in \mathcal{T} : \mathbf{z}^t \geqslant \mathbf{z}\} \cup \{\hat{t}\}$, where $\hat{t}$ is any observation that satisfies $y^{\hat{t}}_{ik} \geqslant y^t_{ik}$ for all $t \in \mathcal{T}$. Since it contains $\hat{t}$ at least, $T(\mathbf{z})$ is nonempty. Choose $\tilde{t}(\mathbf{z}) \in T(\mathbf{z})$ such that $y^{\tilde{t}(\mathbf{z})}_{ik} \leqslant y^t_{ik}$ for all $t \in T(\mathbf{z})$, and define $\Pi_{ik}(\cdot, \mathbf{z}) = f_{ik}(\cdot, \tilde{t}(\mathbf{z}))$. Although there may be more than one candidate for $\tilde{t}(\mathbf{z})$, by property (ii) of f_{ik}, the value of Π_{ik} is not affected by the choice.

We claim that $\Pi_{ik}(\cdot, \mathbf{z})$ defined above obeys two key properties that are analogous to properties (i) and (ii). Firstly, notice that at any $\mathbf{z} \in \mathbf{Z}$ such that $\mathbf{z} = \mathbf{z}^s$ for some observation s, we have $s \in T(\mathbf{z}^s)$. Furthermore, by the RM axiom, for any $t \in T(\mathbf{z}^s)$, we have $y^t_{ik} \geqslant y^s_{ik}$. It follows that $\Pi_{ik}(\cdot, \mathbf{z}^s) = f_{ik}(\cdot, s)$, and so $\text{argmax}_{a \in I_{ik}} \Pi_{ik}(a, \mathbf{z}^s) = y^s_{ik}$. Secondly, Π_{ik} has increasing differences in $(a; \mathbf{z})$. Suppose at $\mathbf{z}'$, we have $\Pi_{ik}(\cdot, \mathbf{z}') = f_{ik}(\cdot, t')$ and for $\mathbf{z}''$, we have $\Pi_{ik}(\cdot, \mathbf{z}'') = f_{ik}(\cdot, t'')$. If $\mathbf{z}'' > \mathbf{z}'$, then $T(\mathbf{z}'') \subseteq T(\mathbf{z}')$, and so $y^{t''}_{ik} \geqslant y^{t'}_{ik}$. By property (ii) of f_{ik}, we obtain

$$\Pi_{ik}(a'', \mathbf{z}'') - \Pi_{ik}(a', \mathbf{z}'') \geqslant \Pi_{ik}(a'', \mathbf{z}') - \Pi_{ik}(a', \mathbf{z}') \text{ for all } a'' > a'.$$

Then, letting $\Pi_i(y_i, \mathbf{z}) = \sum_{k=1}^{K(i)} \Pi_{ik}(y_{ik}, \mathbf{z})$, since $\text{argmax}_{y_{ik} \in Y_{ik}} \Pi_{ik}(y_{ik}, \mathbf{z}^t) = y^t_{ik}$ for all k and t, we obtain $\text{argmax}_{y_i \in Y_i} \Pi_i(y_i, \mathbf{z}^t) = y^t_i$ for all t. Lastly, with Y_i taking only finitely many values, we can always guarantee that $\Pi_i(\cdot, \mathbf{z})$ has *strict* preference over Y_i at every value of $\mathbf{Z}$ by perturbing f_{ik} if necessary. **QED**

The proof of Proposition 1 requires the following well-known lemma.

LEMMA 1. *Let V be a closed convex set in $\mathbb{R}^n$ and let $\mathbf{r} \in \mathbb{R}^n \backslash V$. Then there is a unique $\mathbf{v}^* \in V$ such that $\|\mathbf{r} - \mathbf{v}^*\| = \min_{\mathbf{v} \in V} \|\mathbf{r} - \mathbf{v}\|$. The point $\mathbf{v}^*$ is the unique point in V with the property that $(\mathbf{r} - \mathbf{v}^*) \cdot (\mathbf{v} - \mathbf{v}^*) \leqslant 0$ for all $\mathbf{v} \in V$.*

Proof of Proposition 1. If for all $\mathbf{b} \in \mathcal{B}$, we have $(\mathbf{p} - \eta_0) \cdot (\mathbf{b} - \eta_0) \leqslant 0$, then $(\mathbf{p} - \eta_0) \cdot (\mathbf{v} - \eta_0) \leqslant 0$ for all $\mathbf{v}$ in the convex hull of $\mathcal{B}$. This implies, by Lemma 1 that the distance between $\mathbf{p}$ and any $\mathbf{v}$ in the convex hull of $\mathcal{B}$ is again minimized η_0, which means that $\mathcal{B}_0$ is not improvable. Conversely, if there is $\widehat{\mathbf{b}}$ in $\mathcal{B}$ such that (14) holds, then, appealing to Lemma 1 again, we know that η_0 does *not* minimize the distance between $\mathbf{p}$ and the convex hull of $\mathcal{B}_0 \cup \{\widehat{\mathbf{b}}\}$ and $\widehat{\mathbf{b}}$ improves $\mathcal{B}_0$. **QED**

Proof of Proposition 2. For each $(\mathbf{y}, t)$, define $\mathcal{R}(\mathbf{y}, t) \subset \mathbf{Y} \times \mathcal{T}$ such that

$$\mathcal{R}(\mathbf{y}, t) = \{(\mathbf{y}', t') : \mathrm{BT}(\mathbf{x}^t) = \mathbf{y} \Longrightarrow \mathrm{BT}(\mathbf{x}^{t'}) \neq \mathbf{y}' \text{ for all BT} \in \mathcal{B}\}.$$

Recalling the definition of RM axiom (10), $(\mathbf{y}', t') \in \mathcal{R}(\mathbf{y}, t)$ holds if there exists some $i \in \mathcal{N}$ such that $y_i' < (>) y_i$ and $(y_{-i}', x_i^{t'}) \geqslant (\leqslant)(y_{-i}, x_i^t)$. Impose any linear ranking on the elements of $\mathbf{Y} \times \mathcal{T}$; we define $C = (c_{(\mathbf{y},t),(\mathbf{y}',t')})_{\mathbf{Y}\times\mathcal{T},\mathbf{Y}\times\mathcal{T}}$ to be a $|\mathbf{Y} \times \mathcal{T}| \times |\mathbf{Y} \times \mathcal{T}|$ matrix where $c_{(\mathbf{y},t),(\mathbf{y},t)} = |\mathcal{T}|$ and, if $(\mathbf{y}, t) \neq (\mathbf{y}', t')$, then $c_{(\mathbf{y},t),(\mathbf{y}',t')} = 1$ if $(\mathbf{y}', t') \in \mathcal{R}(\mathbf{y}, t)$ and zero otherwise. By setting $\theta = (|\mathcal{T}|, |\mathcal{T}|, \ldots, |\mathcal{T}|)$ (a column vector of length $|\mathbf{Y} \times \mathcal{T}|$), we claim that a behavioral type $\mathbf{b} \in \overline{\mathcal{B}}$ (thought of as a column vector) obeys RM axiom if and only if $C\mathbf{b} \leqslant \theta$. Indeed, since $\mathbf{b} \in \overline{\mathcal{B}}$, we have $\sum_{\mathbf{y}\in\mathbf{Y}} \mathbf{b}_{(\mathbf{y},t)} = 1$ for all $t \in \mathcal{T}$, which guarantees that $(C\mathbf{b})_{(\mathbf{y},t)} \leqslant \theta$ if $\mathbf{b}_{(\mathbf{y},t)} = 0$. Note that $(C\mathbf{b})_{(\mathbf{y},t)} \geqslant c_{(\mathbf{y},t),(\mathbf{y},t)} = \theta$ if $\mathbf{b}_{(\mathbf{y},t)} = 1$. If $\mathbf{b}$ satisfies the RM axiom holds, then $(C\mathbf{b})_{(\mathbf{y},t)} = \theta$ for all $(\mathbf{y}, t)$ with $\mathbf{b}_{(\mathbf{y},t)} = 1$; if $\mathbf{b}$ violates the RM axiom, then there is $(\tilde{\mathbf{y}}, \tilde{t})$ with $\mathbf{b}_{(\tilde{\mathbf{y}},\tilde{t})} = 1$ such that $(C\mathbf{b})_{(\tilde{\mathbf{y}},\tilde{t})} > \theta$. **QED**

We make crucial use of the following result in our proof of Proposition 3.

LEMMA 2. *Suppose that $\mathcal{B}' \subset \mathcal{B}$ where $\mathcal{B}' \cap \mathcal{B}^*$ is nonempty. Let $V(\mathcal{B}')$ be the set such that $\mathbf{v} \in V(\mathcal{B}')$ if $\mathbf{v} = \mathbf{B}'\tau$ and $\sum_{\mathbf{b}\in\mathcal{B}'\cap\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta$, where $\mathbf{B}'$ is a matrix representation of $\mathcal{B}'$. Then $V(\mathcal{B}')$ is the convex hull of vectors of the form $\beta\mathbf{b}^* + (1-\beta)\mathbf{b}$, where $\mathbf{b}^* \in \mathcal{B}' \cap \mathcal{B}^*$ and $\mathbf{b} \in \mathcal{B}'$.*

Proof. Clearly, the convex hull of those vectors is contained in $V(\mathcal{B}')$, so we need only show the other inclusion. Note that any $\mathbf{v} \in V(\mathcal{B}')$ can be written as $\beta(\sum_{q=1}^{\bar{q}} t_q \mathbf{b}_q^*) + (1-\beta)(\sum_{k=1}^{\bar{k}} s_k \mathbf{b}_k)$ where t_q, $s_k \geqslant 0$, $\sum_{q=1}^{\bar{q}} t_q = \sum_{k=1}^{\bar{k}} s_k = 1$, $\mathbf{b}_q^* \in \mathcal{B}^*$, and $\mathbf{b}_k \in \mathcal{B}'$. By breaking up the convex sums into smaller parts if necessary, we can, with no loss of generality, assume that $t_q = s_k$ and $\bar{q} = \bar{k}$. So then

$$\mathbf{v} = \beta\left(\sum_{q=1}^{\bar{q}} t_q \mathbf{b}_q^*\right) + (1-\beta)\left(\sum_{q=1}^{\bar{q}} t_q \mathbf{b}_k\right) = \sum_{q=1}^{\bar{q}} t_q \left[\beta\mathbf{b}_q^* + (1-\beta)\mathbf{b}_q\right],$$

which establishes our claim. **QED**

Proof of Proposition 3. Note that J_0^β is the distance between $\mathbf{p}$ and $V(\mathcal{B}_0)$ and this distance is achieved at $\eta_0 \in V(\mathcal{B}_0)$. If, for all $\beta\mathbf{b}^* + (1-\beta)\mathbf{b}$ where $\mathbf{b}^* \in \mathcal{B}^*$ and $\mathbf{b} \in \mathcal{B}$, we have

$$(\mathbf{p} - \eta_0) \cdot (\beta\mathbf{b}^* + (1-\beta)\mathbf{b} - \eta_0) \leqslant 0,$$

then $(\mathbf{p} - \eta_0) \cdot (\mathbf{v} - \eta_0) \leqslant 0$ for all $\mathbf{v} \in V(\mathcal{B})$, by Lemma 2. This in turn means, by Lemma 1, that $\mathcal{B}_0$ is not improvable given problem (18). Conversely, suppose that there is a pair of types $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$, with $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ and $\widehat{\mathbf{b}} \in \mathcal{B}$, such that (20) holds, then, by Lemma 1, η_0 does *not* minimize the distance between $\mathbf{p}$ and the convex hull of $V(\mathcal{B}_0)$ and $\beta\widehat{\mathbf{b}}^* + (1 - \beta)\widehat{\mathbf{b}}$. We conclude that $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$ improves $\mathcal{B}_0$ given problem (18). **QED**

References

Apesteguia, J., M. A. Ballester, and J. Lu (2017): Single-crossing random utility models. *Econometrica*, 85, 661–674.

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

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

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

Beresteanu, A. (2005): Nonparametric analysis of cost complementarities in the telecommunications industry. *RAND Journal of Economics*, 36, 870–889.

Beresteanu, A. (2007): Nonparametric estimation of regression functions under restrictions on partial derivatives. *http://www.pitt.edu/ arie/PDFs/shape.pdf.*

Carvajal, A. (2004): Testable restrictions of Nash equilibrium with continuous domains. *Discussion paper series, Royal Holloway College, University of London*, 2004-2026.

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

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2018): Revealed price preference: theory and empirical analysis. *arXiv:1801.02702*.

Echenique, F., and I. Komunjer (2009): Testing models with multiple equilibria by quantile methods. *Econometrica*, 77, 1281–1297.

Kitamura, Y., and J. Stoye (2018): Nonparametric analysis of random utility models. *Econometrica*, 86, 1883-1909.

Kline, B., and E. Tamer (2016): Bayesian inference in a class of partially identified models. *Quantative Economics*, 7, 329–366.

Hoderlein, S., and J. Stoye (2014): Revealed preferences in a heterogeneous population. *Review of Economics and Statistics*, 96, 197-213.

Lazzati, N., J. K.-H. Quah, and K. Shirai (2018) Nonparametric analysis of monotone choice. *Discussion Paper Series 184, School of Economics, Kwansei Gakuin University.*

Manski, C. (2007): Partial identification of counterfactual choice probabilities. *International Economic Review*, 48, 1393-1410.

Marschak, J. (1960): Binary choice constraints on random utility indicators. *Stanford Symposium on Mathematical Methods in the Social Sciences*, edited by K. Arrow. Stanford University Press.

Matzkin, R. (2007): Nonparametric identification. *Handbook of Econometrics, Vol. 6b*, edited by J. J. Heckman and E. E. Leamer, Elsevier Science.

McFadden, D.L. (2005): Revealed Stochastic Preference: a synthesis. *Economic Theory*, 26, 245-264.

McFadden, D.L., and K. Richter (1991): Stochastic rationality and revealed stochastic preference, in *Preferences, Uncertainty and Rationality*, ed. by J. Chipman, D. McFadden, and M.K. Richter. Boulder: Westview Press, pp. 161-186.

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

Milgrom, P., and C. Shannon (1994): Monotone comparative statics. *Econometrica*, 62, 157-180.

Milgrom, P. (2004): *Putting Auction Theory to Work*, Cambridge: Cambridge University Press.

Smeulders, B., Cherchye, L. and R. De Rock, B. (2021): Nonparametric analysis of random utility models: computational tools for statistical testing. *Econometrica*, 89(1), 437–455.

Tebaldi, P., A. Torgovitsky, and H. Yang (2020): Nonparametric estimates of demand in the California health insurance exchange. Mimeo.

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

Topkis, D. (1998): *Supermodularity and complementarity.* Princeton University Press.

Varian, H. (1982): The nonparametric approach to demand analysis. *Econometrica*, 50, 945-73.

Vives, X. (1990): Nash equilibria with strategic complementarities. *Journal of Mathematical Economics*, 19, 305-321.

Online Appendix

A1. More results on inference and predictions

This section contains results omitted from Section 3.3 of the main paper. In the first subsection, we explain how we can obtain a tight bound on the probability that an agent has a given ranking between a pair of actions. The second subsection expands on the discussion of Nash equilibrium predictions in Section 3.3 (Application 2) and also establishes that *the set of Nash equilibrium predictions increases with the covariate*, in a sense related to first order stochastic dominance.

Throughout we shall assume that that the data set $\mathcal{P} = \{P(\cdot \mid \mathbf{x}^t)\}_{t\in\mathcal{T}}$ is $\mathcal{SC}$-rationalizable. Recall (from Theorem 2) that $\mathcal{P}$ is $\mathcal{SC}$-rationalizable if and only if there exists a distribution $\tau = (\tau^{\mathrm{BT}})_{\mathrm{BT}\in\mathcal{B}}$ on $\mathcal{B}$ (the set of behavioral types obeying the RM axiom) such that

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}^t) = \sum\nolimits_{\{\mathrm{BT}\in\mathcal{B}:\ \mathrm{BT}(\mathbf{x}^t)=\mathbf{y}\}} \tau^{\mathrm{BT}} \text{ for all } \mathbf{y} \in \mathbf{Y} \text{ and } t \in \mathcal{T}. \tag{a.1}$$

A1.1 Predicting player preferences

We focus on a given player i and suppose that Y_i is one-dimensional. We are interested in estimating the proportion of groups in the population where agent i prefers some action y_i'' over another action y_i', when the covariate takes a specific value x_i^* and other players are playing a given profile of strategies $\mathbf{y}_{-i}^*$. We let

$$\mathbf{z}^* = (\mathbf{y}_{-i}^*, x_i^*).$$

In formal terms, we would like to estimate the probability of

$$S = \{\mathbf{\Pi} \in \mathcal{SC} : \Pi_i \text{ satisfies } \Pi_i(y_i'', \mathbf{z}^*) > \Pi_i(y_i', \mathbf{z}^*)\} \tag{a.2}$$

Suppose that there some observation t' for which $\mathbf{y}_{-i}^{t'} = \mathbf{y}_{-i}^*$ and $x_i^{t'} = x_i^*$. Then simply by assuming that all groups are playing Nash equilibria, we know that the probability of S must be weakly higher than $\mathrm{P}((y_i, \mathbf{y}_{-i}^*) \mid \mathbf{x}^{t'})$ (which is part of the data $\mathcal{P}$). But in fact we can obtain a sharper bound on the probability of S by exploiting the assumption that players have single crossing payoff functions.

Proposition A.1. *Suppose Y_i is one-dimensional and $\mathcal{P} = \{\mathrm{P}(\cdot \mid \mathbf{x}^t)\}_{t\in\mathcal{T}}$ is $\mathcal{SC}$-rationalizable by some distribution $\mathrm{P}_{\mathbf{\Pi}}$. Then for S defined by (a.2),*

$$\mathbf{m}(y_i'', y_i') \leqslant \int_S d\mathrm{P}_{\mathbf{\Pi}}$$

where $\mathbf{m}(y_i'', y_i')$ *is defined as follows:*

- *if* $y_i' < y_i''$ *then* $\mathbf{m}(y_i'', y_i') = \min \sum_{\mathrm{BT} \in \underline{\mathcal{B}}} \tau^{\mathrm{BT}}$ *subject to* τ *solving (a.1), with*

$$\underline{\mathcal{B}} = \{\mathrm{BT} \in \mathcal{B} : \text{ for some } t \in \mathcal{T},\ y_i^t = y_i'' \text{ and } (\mathbf{y}_{-i}^t, x_i^t) \leqslant \mathbf{z}^*\}; \tag{a.3}$$

- *if* $y_i' > y_i''$ *then* $\mathbf{m}(y_i'', y_i') = \min \sum_{\{q : \mathrm{BT}^q \in \overline{\mathcal{B}}\}} \tau^q$ *subject to* τ *solving (a.1), with*

$$\overline{\mathcal{B}} = \{\mathrm{BT} \in \mathcal{B} : \text{ for some } t \in \mathcal{T},\ y_i^t = y_i'' \text{ and } (\mathbf{y}_{-i}^t, x_i^t) \geqslant \mathbf{z}^*\}, \tag{a.4}$$

Proof. We consider the case where $y_i'' > y_i'$. Let

$$S' = \{\mathbf{\Pi} \in \mathcal{SC} : \mathbf{\Pi} \text{ rationalizes a behavioral type in } \underline{\mathcal{B}}\},$$

where $\underline{\mathcal{B}}$ is defined by (a.3). If $\mathbf{\Pi} \in S'$, then for some $t \in \mathcal{T}$, $\Pi_i(y_i'', \mathbf{y}_{-i}^t, x_i^t) > \Pi_i(y_i', \mathbf{y}_{-i}^t, x_i^t)$. Since $(\mathbf{y}_{-i}^t, x_i^t) \leqslant \mathbf{z}^*$ and Π_i satisfies single-crossing differences, $\Pi_i(y_i'', \mathbf{z}^*) > \Pi_i(y_i', \mathbf{z}^*)$. We conclude that

$$\begin{aligned}\sum\nolimits_{\mathrm{BT} \in \underline{\mathcal{B}}} \tau^{\mathrm{BT}} &= \sum\nolimits_{\mathrm{BT} \in \underline{\mathcal{B}}} \int \mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi})\, d\mathrm{P}_{\mathbf{\Pi}} \\ &= \int \sum\nolimits_{\mathrm{BT} \in \underline{\mathcal{B}}} \mathrm{P}(\mathrm{BT} \mid \mathbf{\Pi})\, d\mathrm{P}_{\mathbf{\Pi}} \leqslant \int_{S'} d\mathrm{P}_{\mathbf{\Pi}} \leqslant \int_S d\mathrm{P}_{\mathbf{\Pi}},\end{aligned}$$

where the first equality follows from (11) in the main paper, the penultimate inequality holds since $\sum_{\mathrm{BT} \in \underline{\mathcal{B}}} \mathrm{P}(\mathrm{BT}, \mathbf{\Pi})$ is less than 1 and equals zero if $\mathbf{\Pi} \notin S'$, and the final inequality is true since $S' \subset S$. Since τ must satisfy (a.1), a lower bound on $\sum_{\mathrm{BT} \in \underline{\mathcal{B}}} \tau^{\mathrm{BT}}$ is $\mathbf{m}(y_i'', y_i')$.

The proof of the case where $y_i'' < y_i'$ proceeds in analogous fashion. **QED**

Since there is typically more than one distribution $\mathrm{P}_{\mathbf{\Pi}}$ that $\mathcal{SC}$-rationalizes $\mathcal{P}$, the probability of S would typically only be set estimated. Proposition A.1 says that there is a uniform lower bound on the probability of S, which is $\mathbf{m}(y_i'', y_i')$. It follows immediately from this proposition that there is also a uniform upper bound on the probability of S, which is $1 - \mathbf{m}(y_i', y_i'')$ and thus we conclude that for any $\mathrm{P}_{\mathbf{\Pi}}$ that rationalizes $\mathcal{P}$,

$$\mathbf{m}(y_i'', y_i') \leqslant \int_S d\mathrm{P}_{\mathbf{\Pi}} \leqslant 1 - \mathbf{m}(y_i', y_i''). \tag{a.5}$$

We can calculate $\mathbf{m}(y_i'', y_i')$ and $\mathbf{m}(y_i', y_i'')$ from the data by solving the relevant linear program. The next result strengthens Proposition A.1 by showing that the bounds in (a.5) are tight.

PROPOSITION A.2. *There is a distribution $P_{\mathbf{\Pi}}$ with support on $\mathcal{SC}$ that rationalizes the data and satisfies*

$$\mathbf{m}(y_i'', y_i') = \int_S dP_{\mathbf{\Pi}}; \tag{a.6}$$

similarly, there is another distribution $P_{\mathbf{\Pi}}$ with support on $\mathcal{SC}$ that rationalizes the data and satisfies

$$\int_S dP_{\mathbf{\Pi}} = 1 - \mathbf{m}(y_i', y_i''). \tag{a.7}$$

Proof. Notice that (a.7) is equivalent to there being a distribution $\mathrm{P}_{\mathbf{\Pi}}$ with support on $\mathcal{SC}$ such that $\int_{\hat{S}} d\mathrm{P}_{\mathbf{\Pi}} = \mathbf{m}(y_i', y_i'')$ where

$$\hat{S} = \{\mathbf{\Pi} \in \mathcal{SC} : \Pi_i \text{ satisfies } \Pi_i(y_i'', \mathbf{z}^*) < \Pi_i(y_i', \mathbf{z}^*)\}.$$

Therefore, to prove (a.5) it suffices to establish (a.6).

We first consider the case where $y_i'' > y_i'$. Suppose that $\tau = \underline{\tau}$ solves $\min \sum_{BT \in \underline{\mathcal{B}}\}} \tau^{\mathrm{BT}}$ subject to τ satisfying (12), with $\underline{\mathcal{B}}$ given by (a.3), so that $\mathbf{m}(y_i'', y_i') = \sum_{BT \in \underline{\mathcal{B}}} \underline{\tau}^{\mathrm{BT}}$. We know from our proof of Theorem 2 (see the discussion immediately preceding the statement of the theorem in Section 3.2) that $\mathcal{P}$ can be rationalized by a distribution $\mathrm{P}^*_{\mathbf{\Pi}}$ that gives weight of τ^{BT} to a profile $\mathbf{\Pi}^{\mathrm{BT}} \in \mathcal{SC}$ that rationalizes BT; by taking strictly increasing transformations if necessary, we can guarantee that $\mathbf{\Pi}^{\mathrm{BT}} \neq \mathbf{\Pi}^{\mathrm{BT}'}$ for any BT$\neq$BT$'$. If BT$\in \underline{\mathcal{B}}$, then any $\mathbf{\Pi}^{\mathrm{BT}}$ that rationalizes BT will satisfy $\Pi_i^{\mathrm{BT}}(y_i'', \mathbf{z}^*) > \Pi_i^{\mathrm{BT}}(y_i', \mathbf{z}^*)$, so $\int_S d\mathrm{P}^*_{\mathbf{\Pi}} \geqslant \mathbf{m}(y_i'', y_i')$. To show that (a.6) holds for the distribution $\mathrm{P}^*_{\mathbf{\Pi}}$ it suffices to show that if BT$\notin \underline{\mathcal{B}}$ then there is $\mathbf{\Pi}^{\mathrm{BT}} \in \mathcal{SC}$ rationalizing BT such that Π_i^{BT} satisfies

$$\Pi_i^{\mathrm{BT}}(y_i'', \mathbf{z}^*) < \Pi_i^{\mathrm{BT}}(y_i', \mathbf{z}^*), \tag{a.8}$$

so that $\mathbf{\Pi}^{\mathrm{BT}} \notin S$. Suppose BT$= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}}$. Since BT is in $\mathcal{B}$ by assumption, we know there is Π_i obeying single-crossing differences such that, for all t, we have $\Pi_i^{\mathrm{BT}}(y_i^t, \mathbf{z}^t) > \Pi_i^{\mathrm{BT}}(y_i, \mathbf{z}^t)$ for all y_i in Y_i and not equal to y_i^t (where $\mathbf{z}^t$ denotes $(\mathbf{y}_{-i}^t, x_i^t)$). The issue is whether there is Π_i^{BT} that satisfies (a.8), in addition. The answer is 'Yes,' and we shall explicitly construct such a function Π_i^{BT}.

We denote the set $\mathbf{Y}_{-i} \times X_i$ by $\mathbf{Z}$ and a typical element $(\mathbf{y}_{-i}, x_i)$ by $\mathbf{z}$. We define a binary relation $\succ$ on $Y_i \times \mathbf{Z}$ in the following way: for any pair $(\bar{y}_i, \mathbf{z})$ and $(\hat{y}_i, \mathbf{z})$ with $\hat{y}_i < \bar{y}_i$, (i) if there is $\tilde{\mathbf{z}} \leqslant \mathbf{z}$ and some t such that $\bar{y}_i = y_i^t$ and $\tilde{\mathbf{z}} = \mathbf{z}^t$ then $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$; (ii) if there is $\tilde{\mathbf{z}} \geqslant \mathbf{z}$ and some t such that $\hat{y}_i = y_i^t$ and $\tilde{\mathbf{z}} = \mathbf{z}^t$, then $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$; (iii) if neither (i) nor (ii) holds then $\hat{y}_i \succ \bar{y}_i$.

We claim that $\succ$ has the following properties: (**P1**) $\succ$ rationalizes the data in the sense that, for all t, we have $(y_i^t, \mathbf{z}^t) \succ (y_i, \mathbf{z}^t)$ for $y_i \neq y_i^t$; (**P2**) $(y_i', \mathbf{z}^*) \succ (y_i'', \mathbf{z}^*)$; (**P3**) any two distinct elements $(\bar{y}_i, \mathbf{z})$ and $(\hat{y}_i, \mathbf{z})$ are strictly comparable, in the sense that either $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$ or $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$ must hold but not both; (**P4**) $\succ$ is transitive when restricted to the set $Y_i \times \{\mathbf{z}\}$, for any $\mathbf{z} \in \mathbf{Z}$. (**P5**) $\succ$ has the single-crossing property in the sense that if $(y_i^{\star\star}, \mathbf{z}) \succ (y_i^\star, \mathbf{z})$ for some $y_i^{\star\star} > y_i^\star$ then $(y_i^{\star\star}, \tilde{\mathbf{z}}) \succ (y_i^\star, \tilde{\mathbf{z}})$ for any $\tilde{\mathbf{z}} > \mathbf{z}$.

Assuming that these properties hold, it is clear that any function Π_i that represents $\succ$ (in the sense that $\Pi_i(y_i^{\star\star}, \mathbf{z}) > \Pi_i(y_i^\star, \mathbf{z})$ whenever $(y_i^{\star\star}, \mathbf{z}) \succ (y_i^\star, \mathbf{z})$) will be a payoff function that obeys single-crossing differences, rationalizes i's actions, and (because of (**P2**)) satisfy (a.8). Note that the existence of a representation for $\succ$ is clear since $\succ$ satisfies (**P3**) and (**P4**) and Y_i is a finite set.

(**P1**) follows from parts (i) and (ii) of the definition of $\succ$ and (**P5**) from part (i). Notice that it follows immediately from the definition of $\succ$ that either $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$ or $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$ must hold, for any $\hat{y}_i < \bar{y}_i$. Furthermore, since BT is $\mathcal{SC}$-rationalizable and thus obeys the RM axiom, they cannot hold simultaneously because conditions (i) and (ii) in the definition of $\succ$ cannot both be satisfied. Thus we have established (**P3**). Since $\text{BT}^q \notin \underline{\mathcal{B}}$, we know that for y_i'' and y_i', we cannot have $(y_i'', \mathbf{z}^*) \succ (y_i', \mathbf{z}^*)$ as a result of (i) holding. Therefore, we must have $(y_i', \mathbf{z}^*) \succ (y_i'', \mathbf{z}^*)$, which is (**P2**). It remains for us to show (**P4**). Suppose instead that transitivity is violated. Then there must be $y_i^\star$, $y_i^{\star\star}$, $y_i^{\star\star\star}$, and $\mathbf{z}$ such that $y_i^{\star\star} > y_i^\star$, $y_i^{\star\star\star}$ and $(y_i^\star, \mathbf{z}) \succ (y_i^{\star\star}, \mathbf{z}) \succ (y_i^{\star\star\star}, \mathbf{z})$. By definition, $(y_i^{\star\star}, \mathbf{z}) \succ (y_i^{\star\star\star}, \mathbf{z})$ can only occur if there is $\mathbf{z}' \leqslant \mathbf{z}$ and $t \in \mathcal{T}$ such that $\mathbf{z}' = \mathbf{z}^t$ and $y_i^{\star\star} = y_i^t$. But this also implies that $(y_i^{\star\star}, \mathbf{z}) \succ (y_i^\star, \mathbf{z})$, which means (by (**P3**)) that we cannot have $(y_i^\star, \mathbf{z}) \succ (y_i^{\star\star}, \mathbf{z})$.

To recap, we have shown that if $y_i'' > y_i'$ then the distribution $\mathrm{P}^*_{\mathbf{\Pi}}$ rationalizes the data and satisfies (a.6). It remains for us to prove the same result for $y_i'' < y_i'$. Using an analogous proof strategy, we need to show that for any BT$\notin \overline{\mathcal{B}}$, we can find $\mathbf{\Pi}^{\text{BT}} \in \mathcal{SC}$ rationalizing BT such that Π_i^{BT} satisfies (a.8) and so $\mathbf{\Pi}^q \notin S$. The proof proceeds by defining $\succ$ in the following way: for any pair $(\bar{y}_i, \mathbf{z})$ and $(\hat{y}_i, \mathbf{z})$ with $\hat{y}_i < \bar{y}_i$, (i) if there is $\tilde{\mathbf{z}} \leqslant \mathbf{z}$ and some t such that $\bar{y}_i = y_i^t$ and $\tilde{\mathbf{z}} = \mathbf{z}^t$ then $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$; (ii) if there is $\tilde{\mathbf{z}} \geqslant \mathbf{z}$ and some t such that $\hat{y}_i = y_i^t$ and $\tilde{\mathbf{z}} = \mathbf{z}^t$, then $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$; (iii) if neither (i) nor (ii) holds then $\bar{y}_i \succ \hat{y}_i$. In other words, the definition is the same as the one for the other case, except thar (iii) has been modified. One could check that (**P1**) to (**P5**) hold and, in particular, (the new version of) (iii) guarantees (**P2**) since we now assume $y_i'' < y_i'$. With these properties on $\succ$, there is a function Π_i that represents $\succ$ and it will be a payoff function that obeys single-crossing differences, rationalizes i's actions, and satisfies (a.8). **QED**

A1.2. Nash Equilibrium predictions

In Section 3.3 of the main paper we posed the following question: given a strategy profile $\overline{\mathbf{y}}$ and covariate $\overline{\mathbf{x}}$, what is the greatest possible fraction of groups which have $\overline{\mathbf{y}}$ as a pure strategy Nash equilibrium at $\overline{\mathbf{x}}$, among all the possible $\mathcal{SC}$-rationalizations of $\mathcal{P}$? In this section, we pose a more general question: what are the possible distributions of joint actions at the covariate value $\overline{\mathbf{x}}$? In formal terms, this amounts to identifying the *set* of conditional distributions $\mathrm{P}(\cdot \mid \overline{\mathbf{x}})$ such that the augmented stochastic data set $\mathcal{P} \cup \{\mathrm{P}(\cdot \mid \overline{\mathbf{x}})\}$ is still $\mathcal{SC}$-rationalizable. If $\overline{\mathbf{x}} \in \mathbf{X}\backslash\mathbf{X}^{\mathrm{data}}$ then we are making predictions of the possible distributions of joint action profiles at an out-of-sample covariate value but it makes sense to carry out this exercise even when $\overline{\mathbf{x}} = \mathbf{x}^s \in \mathbf{X}^{\mathrm{data}}$. In the latter case, we are identifying all the joint distributions that are possible at $\mathbf{x}^s$, allowing for all possible Nash equilibrium selection rules; the observed distribution $\mathrm{P}(\cdot \mid \mathbf{x}^s)$ will be one of the predicted distributions, but the set of predicted distributions will typically be larger. We now explain how we can find all the permissible distributions ovr action profiles at $\overline{\mathbf{x}}$.

Let $\mathrm{BT} : \{\overline{\mathbf{x}}\} \cup \mathbf{X}^{\mathrm{data}} \to \mathbf{Y}$ be a behavioral type defined on the enlarged domain $\{\overline{\mathbf{x}}\} \cup \mathbf{X}^{\mathrm{data}}$. Let $\widetilde{\mathcal{B}}$ be the set of all behavioral types defined on this domain that obey the RM axiom; obviously this set is finite. Applying Theorem 2, we know that $\mathcal{P} \cup \{\mathrm{P}(\cdot \mid \overline{\mathbf{x}})\}$ is $\mathcal{SC}$-rationalizable if and only if we can find a probability distribution $\tilde{\tau} = (\tilde{\tau}^{\mathrm{BT}})_{\mathrm{BT}\in\widetilde{\mathcal{B}}}$ over $\widetilde{\mathcal{B}}$ such that

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}^t) = \sum\nolimits_{\{\mathrm{BT}\in\widetilde{\mathcal{B}}\,:\,\mathrm{BT}(\mathbf{x}^t)=\mathbf{y}\}} \tilde{\tau}^{\mathrm{BT}} \text{ for each } \mathbf{y} \in \mathbf{Y} \text{ and } t \in \mathcal{T}, \text{ and} \qquad \text{(a.9)}$$

$$\mathrm{P}(\mathbf{y} \mid \overline{\mathbf{x}}) = \sum\nolimits_{\{\mathrm{BT}\in\widetilde{\mathcal{B}}\,:\,\mathrm{BT}(\overline{\mathbf{x}})=\mathbf{y}\}} \tilde{\tau}^{\mathrm{BT}} \text{ for each } \mathbf{y} \in \mathbf{Y}. \qquad \text{(a.10)}$$

Note that the left hand side of the equations in (a.9) are observations in $\mathcal{P}$, so those equations constitute conditions that $\tilde{\tau}$ has to satisfy. For any $\tilde{\tau}$ that satisfies those conditions, the resulting $\mathrm{P}(\cdot \mid \overline{\mathbf{x}})$ obtained from (a.10) is a predicted distribution at $\overline{\mathbf{x}}$. In other words, if we let $\mathbb{P}(\overline{\mathbf{x}})$ be the *set* of predicted distributions at $\overline{\mathbf{x}}$, then $\mathrm{P}(\cdot \mid \overline{\mathbf{x}})$ is in $\mathbb{P}(\overline{\mathbf{x}})$ if and only if there is $\tilde{\tau}$ that solves (a.9) and (a.10). Since the conditions are linear, $\mathbb{P}(\overline{\mathbf{x}})$ is a convex set and its properties can be found by further investigating the linear program.

The following result states that $\mathbb{P}(\overline{\mathbf{x}})$ is nonempty so long as $\mathcal{P}$ is $\mathcal{SC}$-rationalizable; in other words, that there *is* a solution to (a.9) and (a.10). This requires a short proof using the Basic Theorem. The result also tells us that $\mathbb{P}(\overline{\mathbf{x}})$ is, in a sense, increasing with respect to first order stochastic dominance.[1]

PROPOSITION A.3. *Suppose* $\mathcal{P} = \{\mathrm{P}(\cdot \mid \mathbf{x}^t)\}_{t\in\mathcal{T}}$ *is* $\mathcal{SC}$*-rationalizable. Then* $\mathbb{P}(\overline{\mathbf{x}})$ *is nonempty for any* $\overline{\mathbf{x}} \in \mathbf{X}$ *and has the following monotone property: if* $\hat{\mathbf{x}} > \overline{\mathbf{x}}$*, then for any* $\mathrm{P}(\cdot \mid \overline{\mathbf{x}}) \in \mathbb{P}(\overline{\mathbf{x}})$ *there is* $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \in \mathbb{P}(\hat{\mathbf{x}})$ *such that* $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \geqslant_{FSD} \mathrm{P}(\cdot \mid \overline{\mathbf{x}})$ *and for any* $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \in \mathbb{P}(\hat{\mathbf{x}})$ *there is* $\mathrm{P}(\cdot \mid \overline{\mathbf{x}}) \in \mathbb{P}(\overline{\mathbf{x}})$ *such that* $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \geqslant_{FSD} \mathrm{P}(\cdot \mid \overline{\mathbf{x}})$*.*

Proof. If $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, then we know from the proof of Theorem 2 that it can be rationalized by some distribution $\mathrm{P}_{\mathbf{\Pi}}$ with a finite support in $\mathcal{SC}$. For each $\mathbf{\Pi}$ in that support, the Basic Theorem tells us that $\mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})$ is nonempty. Choose $n(\mathbf{\Pi})$ in $\mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})$. Let $\pi(\mathbf{y}) = \{\mathbf{\Pi} \in \mathcal{SC} : n(\mathbf{\Pi}) = \mathbf{y}\}$. Then the distribution on $\mathbf{Y}$ where $\mathrm{P}(\mathbf{y} \mid \overline{\mathbf{x}}) = \int_{\pi(\mathbf{y})} d\mathrm{P}_{\mathbf{\Pi}}$ for all $\mathbf{y} \in \mathbf{Y}$ is in $\mathbb{P}(\overline{\mathbf{x}})$ and so $\mathbb{P}(\overline{\mathbf{x}})$ is nonempty.

We show that if $\mathrm{P}(\cdot \mid \overline{\mathbf{x}}) \in \mathbb{P}(\overline{\mathbf{x}})$, then there is $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \in \mathbb{P}(\hat{\mathbf{x}})$ such that $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \geqslant_{FSD} \mathrm{P}(\cdot \mid \overline{\mathbf{x}})$ if $\hat{\mathbf{x}} > \overline{\mathbf{x}}$. The (omitted) proof of the other case is similar. Since $\mathrm{P}(\cdot \mid \hat{\mathbf{x}}) \in \mathbb{P}(\hat{\mathbf{x}})$, there is a distribution $\mathrm{P}_{\mathbf{\Pi}}$ with a finite support in $\mathcal{SC}$ and an equilibrium selection rule $\bar{\lambda}(\cdot \mid \mathbf{\Pi}, \mathbf{x})$ (for $\mathbf{x} \in \mathbf{X}^{\text{data}}$ and $\mathbf{x} = \overline{\mathbf{x}}$) that rationalizes $\mathcal{P}$ and satisfies $\mathrm{P}(\mathbf{y} \mid \overline{\mathbf{x}}) = \int \bar{\lambda}(\mathbf{y} \mid \mathbf{\Pi}, \overline{\mathbf{x}})\, d\mathrm{P}_{\mathbf{\Pi}}$ for all $\mathbf{y} \in \mathbf{Y}$. Let $\hat{\lambda}$ be a new equilibrium selection rule where $\hat{\lambda}(\cdot \mid \mathbf{\Pi}, \mathbf{x}^t) = \bar{\lambda}(\cdot \mid \mathbf{\Pi}, \mathbf{x}^t)$ for $t \in \mathcal{T}$ and, in the case where $\mathbf{x} = \overline{\mathbf{x}}$, we define $\hat{\lambda}$ in the following manner: for each $\mathbf{y}'$ in $\mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})$ for which $\bar{\lambda}(\mathbf{y}' \mid \mathbf{\Pi}, \overline{\mathbf{x}}) > 0$, choose $\mathbf{y}''$ in $\mathrm{NE}(\mathbf{\Pi}, \hat{\mathbf{x}})$ such that $\mathbf{y}'' \geqslant \mathbf{y}'$ and set $\hat{\lambda}(\mathbf{y}'' \mid \mathbf{\Pi}, \hat{\mathbf{x}}) = \bar{\lambda}(\mathbf{y}' \mid \mathbf{\Pi}, \overline{\mathbf{x}})$. We know that $\mathbf{y}''$ exists because the set of pure strategy Nash equilibria of a game with strategic complements admits a largest element and a smallest element and both are increasing with $\mathbf{x}$ (see Milgrom and Roberts (1990)). For any $\mathbf{y} \in \mathbf{Y}$ not assigned a positive probability in this manner, set $\hat{\lambda}(\mathbf{y} \mid \mathbf{\Pi}, \hat{\mathbf{x}}) = 0$. In this way, the distribution given by $\mathrm{P}(\mathbf{y} \mid \hat{\mathbf{x}}) = \int \hat{\lambda}(\mathbf{y} \mid \mathbf{\Pi}, \hat{\mathbf{x}})\, d\mathrm{P}_{\mathbf{\Pi}}$ for all $\mathbf{y} \in \mathbf{Y}$ is in $\mathbb{P}(\hat{\mathbf{x}})$ and first order stochastically dominates $\mathrm{P}(\cdot \mid \overline{\mathbf{x}})$. **QED**

[1] For two distributions ν and θ on a Euclidean space, we say that ν first order stochastically dominates θ if $\int_C d\nu(y) \geqslant \int_C d\theta(y)$ for all measurable sets C that are upward comprehensive, i.e., if $y \in C$ then $z \in C$ for any $z \geqslant y$. It is known that this holds if and only if $\int f(y) d\nu(y) \geqslant \int f(y) d\theta(y)$ for all increasing real-valued functions f.

A2. Statistical and Computational Procedures

In this section we explain how the statistical procedures in Kitamura and Stoye (2018) and Deb et al. (2020) could be combined with the column generation method explained in Section 4 of the main paper to provide viable approaches to model testing, inference, and predictions.

A2.1. Model testing from an empirical distribution

For testing monotone rationalizability based on an empirical distribution $\mathbf{p}$, rather than simply checking if $\min_\tau(\mathbf{p}-\mathbf{B}\tau)\cdot(\mathbf{p}-\mathbf{B}\tau)$ is equal to 0, we need to evaluate the statistical significance of its value. Kitamura and Stoye (2018) sets out the following procedure. Let

$$J_N := \min_{\tau\in\Delta^{\mathcal{B}}} N(\mathbf{p}-\mathbf{B}\tau)\cdot(\mathbf{p}-\mathbf{B}\tau) \tag{a.11}$$

which we use as the test statistics. For some $\kappa_N > 0$ (drawn from a decreasing sequence κ_n tending to zero with $\sqrt{N}\kappa_N$ increasing and tending to infinity), let

$$\underline{\Delta}^{\mathcal{B}} = \left\{\tau \in \Delta^{\mathcal{B}} : \tau_{\mathbf{b}} > \kappa_N/|\mathcal{B}| \text{ for all } \mathbf{b}\in\mathcal{B}\right\}.$$

Let $\eta^* = \mathbf{B}\tau^*$, where τ^* is a solution of the problem

$$\min_{\tau\in\underline{\Delta}^{\mathcal{B}}} N(\mathbf{p}-\mathbf{B}\tau)\cdot(\mathbf{p}-\mathbf{B}\tau), \tag{a.12}$$

The tuning parameter κ_N is introduced in order to avoid possible discontinuity due to corner solutions, and the problem (a.12) requires that each behavioral type obeying RM axiom should take at least $\kappa_N/|\mathcal{B}|$ weight.[2] Based on η^*, we generate a bootstrap sample $\hat{\mathbf{p}}^{(r)}$ for $r = 1, 2, ..., R$ such that $\hat{\mathbf{p}}^{(r)} := (\mathbf{p}^{(r)} - \mathbf{p}) + \eta^*$, where $\mathbf{p}^{(r)}$ is obtained by the standard nonparametric bootstrap resampling. Then, for each $r = 1, 2, ..., R$, we calculate

$$J_N^{(r)} := \min_{\tau\in\underline{\Delta}^{\mathcal{B}}} N(\hat{\mathbf{p}}^{(r)}-\mathbf{B}\tau)\cdot(\hat{\mathbf{p}}^{(r)}-\mathbf{B}\tau), \tag{a.13}$$

and use the empirical distribution of $J_N^{(r)}$ to calculate p-value

$$p = \#\{J_N^{(r)} > J_N\}/R. \tag{a.14}$$

To carry out this test, we must first compute J_N. When $\mathcal{B}$ is too large to be completely identified, J_N can be obtained using the column generation method (following Smeulders et al. (2021)) outlined

[2] Recall that $\mathcal{B}$ is the set of all behavioral types obeying RM axiom.

in Section 4.1 of the main paper. We now explain it in greater detail. Let $\mathcal{B}_0$ be an arbitrary subset of $\mathcal{B}$ which we can completely list. In our implementation, we choose $\mathcal{B}_0$ to be the set of *constant behavioral types*, in which every player takes the same action regardless of opponents' actions and covariates. (Such a behavioral type obviously obeys the RM axiom.) Let $\mathbf{B}_0$ be the matrix where the column vectors are elements of $\mathcal{B}_0$ and calculate

$$J_{N,0} := \min_{\tau \in \Delta^{\mathcal{B}_0}} N(\mathbf{p} - \mathbf{B}_0\tau) \cdot (\mathbf{p} - \mathbf{B}_0\tau). \tag{a.15}$$

It is obvious that $J_{N,0} \geqslant J_N$, and hence, if $J_{N,0} = 0$, then we can immediately conclude that $J_N = 0$. When $J_{N,0} > 0$, we need to check if $\mathcal{B}_0$ is improvable. By Propositions 1 and 2 in the main paper, this can be done by checking whether

$$\max\,(\mathbf{p} - \eta_0) \cdot (\mathbf{b} - \eta_0) \text{ subject to } \mathbf{b} \in \overline{\mathcal{B}} \text{ and } C\mathbf{b} \leqslant \theta \tag{a.16}$$

is strictly positive. Recall that $\eta_0 = \mathbf{B}_0\tau_0$, where $\tau_0 = \arg\min_{\tau \in \Delta^{\mathcal{B}_0}} (\mathbf{p} - \mathbf{B}_0\tau) \cdot (\mathbf{p} - \mathbf{B}_0\tau)$. If $\mathcal{B}_0$ is not improvable, then we conclude that $J_{N,0}$ is in fact J_N. Otherwise, a behavioral type $\widehat{\mathbf{b}} \in \mathcal{B}$ that solves (a.16) improves $\mathcal{B}_0$. We then repeat the procedure by setting $\mathcal{B}_0 = \mathcal{B}_0 \cup \{\widehat{\mathbf{b}}\}$ until we get $J_{N,0} = J_N$. Algorithm 1 summarizes the procedure to calculate J_N.

<u>Algorithm 1</u>

1. Solve the minimization problem (a.15) to get $J_{N,0}$.
2. If $J_{N,0} = 0$ then
3.   Output: $J_N = J_{N,0}$
4. else
5.   Solve the maximization problem (a.16) to check if there exists a behavioral type $\mathbf{b} \in \mathcal{B}$ that improves $\mathcal{B}_0$.
6.   If there is no such a behavioral type, then
7.   Output: $J_N = J_{N,0}$.
8.   Else
9.   Update $\mathcal{B}_0$ by adding a behavioral type found in Line 5 to $\mathcal{B}_0$.
10.   Go to Line 1
11.   End if
12. End if

The next step in the Kitamura-Stoye procedure involves constructing an empirical distribution of $J_N^{(r)}$ (defined by (a.13)), but this cannot be carried out using the column generation method because the minimization problem involves putting positive weights on *all* behavioral types in $\mathcal{B}$

and $\mathcal{B}$ is precisely what we would like to avoid computing in full. We therefore modify the procedure for constructing an empirical distribution of the test statistic, following that in Smeulders et al. (2021, Lemma 3.2). Their procedure involves altering the set $\underline{\Delta}^{\mathcal{B}}$, so that positive weight is put only on a subset of types $\mathcal{B}'$, so that

$$\underline{\Delta}^{\mathcal{B}} = \left\{\tau \in \Delta^{\mathcal{B}} : \tau_{\mathbf{b}} > \kappa_N/|\mathcal{B}'| \text{ for all } \mathbf{b} \in \mathcal{B}'\right\}.$$

The requirement on $\mathcal{B}'$ is that it should contain a basis of the space spanned by $\mathcal{B}$, which then guarantees that the distributions in $\underline{\Delta}^{\mathcal{B}}$ are in the interior of the cone generated by $\mathcal{B}$. With this new definition of $\underline{\Delta}^{\mathcal{B}}$ we could then use the column generation method to calculate η^* and generate $J_N^{(r)}$, as given by (a.12) and (a.13).

We shall only explain how we can calculate (a.12) (since the procedure for calculating (a.13) is the same). We initiate the column generation process with $\mathcal{B}_0$, which is any set of behavioral types in $\mathcal{B}$ which contains $\mathcal{B}'$. The first step is similar to (a.15), except that the minimization is over a different set of distributions; specifically, we calculate

$$J_{N,0} := \min_{\tau \in \underline{\Delta}^{\mathcal{B}_0}} N(\mathbf{p} - \mathbf{B}_0\tau) \cdot (\mathbf{p} - \mathbf{B}_0\tau). \tag{a.17}$$

If $J_{N,0}$ is nonzero, we can check if $\mathcal{B}_0$ is improvable by checking whether the problem (a.16) has a strictly positive solution, where $\eta_0 = \mathbf{B}_0\tau_0$ and τ_0 is the distribution that solves (a.17). If $\mathcal{B}_0$ is not improvable then we know that $J_{N,0}$ solves (a.12); if it is improvable then we add the improving behavioral type to $\mathcal{B}_0$ and repeat the column generation process.[3]

In the empirical application discussed in Section 5, we employ some heuristics to reduce computation time. The main bottleneck in implementing column generation is solving the problem (a.16), which is needed for checking the existence of $\mathbf{b} \in \mathcal{B}$ that improves $J_{N,0}$. In fact, to improve $J_{N,0}$ it is not necessary to have the exact solution to (a.16). It suffices to find $\mathbf{b} \in \mathcal{B}$ such that $(\mathbf{p} - \eta_0) \cdot (\mathbf{b} - \eta_0) > 0$. In our program, we impose a time limit for solving (a.16) and use the best feasible solution found within it.[4] It is only when this solution satisfies $(\mathbf{p} - \eta_0) \cdot (\mathbf{b} - \eta_0) \leqslant 0$ that we solve the maximization problem exactly.

[3] While the *span* of $\mathcal{B}_0$ will include every element of $\mathcal{B}$, the *convex hull* of $\mathcal{B}_0$ need not coincide with the convex hull of $\mathcal{B}$. So it is possible for $\mathcal{B}_0$ to be improvable.

[4] Using Rglpk package on R, we give 0.2 seconds for solving (a.16) in each round of the algorithm.

As noted by Smeulders et al. (2021), the computation time can be further reduced by not always calculating the value of $J_N^{(r)}$ exactly. Indeed, by the definition of the p-value in (17), it suffices to determine whether each $J_N^{(r)}$ is larger or smaller than J_N. Thus we can terminate the algorithm for calculating $J_N^{(r)}$ once a tentative value of $J_{N,0}^{(r)}$ in the algorithm becomes lower than J_N.

The calculation of $J_N^{(r)}$ requires finding a subset of $\mathcal{B}$ that forms a basis for the space it spans. We obtain this set using the following method. First, recall that by Proposition 2 in the main paper, a behavioral type is in $\mathcal{B}$ if and only if it solves the integer programming problem $C\mathbf{b} \leqslant \theta$. Let $\mathcal{B}''$ be a linearly independent set of behavioral types in $\mathcal{B}$. We could check the existence of behavioral types in $\mathcal{B}$ which are linearly independent of the ones in $\mathcal{B}''$ by checking if there is $\mathbf{b} \in \mathcal{B}$ (equivalently, that solve $C\mathbf{b} \leqslant \theta$) and a real-valued vector $\mathbf{w}$ such that $\mathbf{B}'' \cdot (\mathbf{b} - \mathbf{B}''\mathbf{w}) = 0$ and $\mathbf{b} \neq \mathbf{B}''\mathbf{w}$, where $\mathbf{B}''$ refers to the matrix made out of the vectors in $\mathcal{B}''$. If such a behavioral type $\mathbf{b}$ can be found, then we add it to $\mathcal{B}''$ and repeat the procedure. This process will stop when there are no vectors in $\mathcal{B}$ which are linearly independent of the ones in $\mathcal{B}''$, at which point we obtain a basis for $\mathcal{B}$. Notice that while it is not always feasible to completely list the elements of $\mathcal{B}$, it is feasible to list a subset of $\mathcal{B}$ that forms a basis of the space spanned by $\mathcal{B}$, since the dimension of this space grows a lot more slowly with the number of actions and covariate values.[5]

In our empirical implementation, the starter set $\mathcal{B}_0$ for calculating J_N^1 is chosen to be a basis of $\mathcal{B}$. Following Smeulders et al. (2021), we choose, as the starter set for calculating J_N^2, the terminal value of $\mathcal{B}_0$ in the algorithm for calculating J_N^1. Thus the starter set for calculating J_N^2 is a superset of the one for calculating J_N^1. And we repeat this procedure for the rest of the bootstrap sample. Progressively enlarging the starter set in this way improves on the computation time, compared to leaving the starter set unchanged.

A2.2. Inference and prediction from an empirical distribution

This subsection provides a procedure for inference and prediction based on an empirical distribution. Based on an empirical distribution, our interest is the maximal possible weight on $\mathcal{B}^* \subset \mathcal{B}$ so that a data set is still rationalized, allowing for a given level of sampling error. Suppose that the p-value calculated in the preceding subsection is larger than some $\bar{p} \in [0, 1]$, i.e. $\#\{J_N^{(r)} > J_N\}/R \geqslant \bar{p}$.

[5] It is straightforward to check that the dimension of the space spanned by $\bar{\mathcal{B}}$ (the set of all logically possible behavioral types) is precisely $|\mathbf{Y}|T - |\mathbf{Y}| + 1$ and so obviously the dimension of the space spanned by $\mathcal{B}$ can be no higher. In fact, the span of $\mathcal{B}$ coincides with that of $\bar{\mathcal{B}}$, even though $\mathcal{B} \subset \bar{\mathcal{B}}$.

Letting $J_N^*(\bar{p}) = \max\left\{J : \#\{J_N^{(r)} > J\}/R \geqslant \bar{p}\right\}$, we would like to calculate

$$\max_{\tau\in\Delta^{\mathcal{B}}} \sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \text{ subject to } N(\mathbf{p}-\mathbf{B}\tau)\cdot(\mathbf{p}-\mathbf{B}\tau) \leqslant J_N^*(\bar{p}). \tag{a.18}$$

This corresponds to the supremum of the $100(1-\bar{p})\%$ confidence interval of $\sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}}$ in the sense of Deb et al. (2020). For a given $\beta > 0$, we can check whether there is a distribution $\tau = (\tau^{\mathbf{b}})_{\mathbf{b}\in\mathcal{B}}$ on $\mathcal{B}$ that solves

$$N(\mathbf{p}-\mathbf{B}\tau)\cdot(\mathbf{p}-\mathbf{B}\tau) \leqslant J_N^*(\bar{p}) \text{ subject to } \textstyle\sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} \geqslant \beta; \tag{a.19}$$

the solution to (a.18) can then be obtained by a binary search over different values of β. Thus, the crucial step is checking (a.19). Testing (a.19) is in turn equivalent to checking if $J_N^\beta \leqslant J_N^*(\bar{p})$, where

$$J_N^\beta := \min_{\tau\in\Delta^{\mathcal{B}}} N(\mathbf{p}-\mathbf{B}\tau)\cdot(\mathbf{p}-\mathbf{B}\tau) \text{ subject to } \textstyle\sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} \geqslant \beta;. \tag{a.20}$$

We can derive J_N^β without fully listing $\mathcal{B}$ or $\mathcal{B}^*$ by using column generation. For $\mathcal{B}_0 \subset \mathcal{B}$ with $\mathcal{B}^* \cap \mathcal{B}_0$ being nonempty, we first calculate

$$J_{N,0}^\beta := \min_{\tau\in\Delta^{\mathcal{B}_0}} N(\mathbf{p}-\mathbf{B}_0\tau)\cdot(\mathbf{p}-\mathbf{B}_0\tau) \text{ subject to } \sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} \geqslant \beta. \tag{a.21}$$

We say that $\mathcal{B}_0$ is *improvable given problem (a.19)*, if $J_{N,0}^\beta > J_N^\beta$. By Proposition 3 in the main paper, $\mathcal{B}_0$ is improvable given problem (a.19), if and only if there is a pair $\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\}$ with $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ and $\widehat{\mathbf{b}} \in \mathcal{B}$ such that

$$(\mathbf{p}-\eta_0)\cdot\left(\beta\widehat{\mathbf{b}}^* + (1-\beta)\widehat{\mathbf{b}} - \eta_0\right) > 0, \tag{a.22}$$

where τ_0 is the distribution that achieves J_0^β and $\eta_0 = \mathbf{B}_0\tau_0$. (It is clear that the proposition does not depend whether $\mathbf{p}$ is a population distribution or an empirical distribution.) By Proposition 2 in the main paper, we can construct a matrix C and a column vector θ so that $\mathbf{b} \in \mathcal{B}$ if and only if $C\mathbf{b} \leqslant \theta$. Suppose that, in addition, we can construct a matrix C^* and a column vector θ^* with integer entries so that, for any $\mathbf{b} \in \overline{\mathcal{B}}$, we have

$$\mathbf{b}^* \in \mathcal{B}^* \iff C^*\mathbf{b}^* \leqslant \theta^*. \tag{a.23}$$

Then a pair $\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\}$ obeying (a.22) exists if and only if there is a strictly positive solution to

$$\begin{aligned} &\max\,(\mathbf{p}-\eta_0)\cdot(\beta\mathbf{b}^* + (1-\beta)\mathbf{b} - \eta_0) \\ &\text{s.t. } \mathbf{b}, \mathbf{b}^* \in \overline{\mathcal{B}} \text{ and } \begin{pmatrix} C^* & O \\ O & C \end{pmatrix}\begin{pmatrix} \mathbf{b}^* \\ \mathbf{b} \end{pmatrix} \leqslant \begin{pmatrix} \theta^* \\ \theta \end{pmatrix}. \end{aligned} \tag{a.24}$$

As we show in the next subsection, every $\mathcal{B}^*$ in our empirical application has the matrix characterization as above. If there is a pair $\{\widehat{\mathbf{b}}, \widehat{\mathbf{b}}^*\}$ improving $\mathcal{B}_0$, then we update $\mathcal{B}_0$ by including the pair in $\mathcal{B}_0$ and recalculate $J_{N,0}^{\beta}$. Repeat this procedure until no improving pair can be found, at which we can conclude $J_{N,0}^{\beta} = J_N^{\beta}$.

Summarizing, we can use the following algorithm for testing whether $J_N^{\beta} \leqslant J_N^*(\bar{p})$.

<u>ALGORITHM 2</u>

1. Solve the minimization problem (a.20) to get $J_{N,0}^{\beta}$.
2. IF $J_{N,0}^{\beta} \leqslant J_N^*(\bar{p})$ THEN
3. OUTPUT: Yes, $J_N^{\beta} \leqslant J_N^*(\bar{p})$.
4. ELSE
5. Solve the maximization problem (a.24) to check if there exist $\mathbf{b}^*$ and $\mathbf{b}$ such that $(\mathbf{p} - \eta_0) \cdot (\beta \mathbf{b}^* + (1-\beta)\mathbf{b} - \eta_0) > 0$.
6. IF there is no such a pair of behavioral types obeying the inequality in Line 5 THEN
7. OUTPUT: No, $J_{N,0}^{\beta} > J_N^*(\bar{p})$.
8. ELSE
9. Update $\mathcal{B}_0$ by adding two behavioral types found in Line 5 to $\mathcal{B}_0$.
10. GO TO LINE 1
11. END IF
12. END IF

For any given $\beta \in [0, 1]$, the above algorithm checks if $J_N^{\beta} \leqslant J_N^*(\bar{p})$. By using a binary search, we can approximate the supremum of β for which this inequality holds. In our empirical analysis, we calculate the supremum of β up to $1/2^8$.

A2.3. Matrix representation of $\mathcal{B}^*$ used in the empirical application

We outline how we obtain the matrix characterization of $\mathcal{B}^*$ (in the sense of (a.23)), for the different versions of $\mathcal{B}^*$ used in the empirical analysis in Section 5 of the main paper.

Significance of strategic interaction. In this case, we use the following three versions of $\mathcal{B}^*$: (i) those behavioral types that can be rationalized with the LCC firm having a payoff function that is independent of the actions of OA; (ii) those behavioral types that can be rationalized with the OA firm having a payoff function that is independent of the actions of LCC; and (iii) those behavioral types that can be rationalized with both firms having payoff functions that are independent of the other firm's action. As explained in Section 3.3 (Application 1), these behavioral types can each be characterized by a version of RM axiom. For this reason, the matrix C^* and the column vector θ^*

that characterizes $\mathcal{B}^*$ (in the sense of (a.23)) can be constructed in essentially the same way as the construction of C and θ in the case of $\mathcal{B}$ (as provided in Proposition 2).

Probability bounds for equilibrium actions. In this case, for a given $\overline{\mathbf{y}} \in \mathbf{Y}$ and $\overline{\mathbf{x}} \in \mathbf{X}^{\text{data}}$, $\mathcal{B}^*$ is the set of behavioral types that *can* support $\overline{\mathbf{y}}$ as a Nash equilibrium action profile at $\mathbf{x} = \overline{\mathbf{x}}$ (see Section 3.3 (Application 2) in the main paper). As referred to in Section 3.3 (Application 2), BT($= \{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}}$) $\in \mathcal{B}^*$ if and only if $\{(\mathbf{y}^t, \mathbf{x}^t)\}_{t \in \mathcal{T}} \cup \{(\mathbf{y}^0, \mathbf{x}^0)\}$ obeys RM axiom, where $\{(\mathbf{y}^0, \mathbf{x}^0)\}$ is a notional observation added to the list, with $\mathbf{y}^0 = \overline{\mathbf{y}}$ and $\mathbf{x}^0 = \overline{\mathbf{x}}$. Then, similar to the characterization of C^* in Proposition 2, we can derive the set $\mathcal{R}(\overline{\mathbf{y}}, \overline{\mathbf{x}})$. Using this, define a vector $\zeta \in \{0,1\}^{|\mathbf{Y} \times \mathbf{X}^{\text{data}}|}$ such that for each $(\mathbf{y}, \mathbf{x}) \in \mathbf{Y} \times \mathbf{X}^{\text{data}}$, $\zeta_{(\mathbf{y},\mathbf{x})} = 1$ if and only if $(\mathbf{y}, \mathbf{x}) \in \mathcal{R}(\overline{\mathbf{y}}, \overline{\mathbf{x}})$. Then, define C^* as a $(|\mathbf{Y} \times \mathbf{X}^{\text{data}}| + 1) \times |\mathbf{Y} \times \mathbf{X}|$-matrix such that the first $|\mathbf{Y} \times \mathbf{X}^{\text{data}}| \times |\mathbf{Y} \times \mathbf{X}^{\text{data}}|$ part must be equal to the matrix C constructed in Proposition 2 in the main paper, and the additional $(|\mathbf{Y} \times \mathbf{X}^{\text{data}}| + 1)$-th row is equal to ζ defined above. Finally, define the vector θ^* such as $\theta^* = (\theta, 0)$, where θ is as defined in the proof of Proposition 2 in the main paper. Then, for a given behavioral type $\mathbf{b} \in \bar{\mathcal{B}}$, $C^*\mathbf{b} \leqslant \theta^*$ is a necessary and sufficient condition under which $\mathbf{b} \in \mathcal{B}^*$: it ensures that $C\mathbf{b} \leqslant \theta$, which is equivalent to $\mathbf{b}$ obeying RM axiom, and $\zeta \cdot \mathbf{b} \leqslant 0$, which is in turn equivalent to $\mathbf{b}$ not containing a behavior contradicting $(\overline{\mathbf{y}}, \overline{\mathbf{x}})$ under RM axiom.

A3. Finer divisions of covariates in the empirical application

In the data set used in Section 5 of the main paper, we followed Kline and Tamer (2016) in aggregating the market presence covariates (MP_{LCC} and MP_{OA}) and market size (MS) covariates into binary variables depending on whether the observed value is higher or lower than the median. This means that the covariates take $2^3 = 8$ possible values and the distribution of firms' entry decisions is observed at each of those eight values. As we pointed in the main paper, this does not fully exploit the column generation method which can handle bigger problems.

Our purpose here is to implement the analysis in Section 5 of the main paper again, after reconstructing the data set with a finer division of covariates. Specifically, we let each covariate take four possible values: the value is equal to $k - 1$, if observed value in raw data is in its k-th quartile ($k = 1, 2, 3, 4$). Thus, $\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS} \in \{0, 1, 2, 3\}$, and there are $4^3 = 64$ possible realizations of covariates. Each realization of the covariate values contains from 34 to 259 markets (the median is 120), and the empirical distribution of entry decisions of LCC and OA is calculated (see Figure A.1 in the end of this appendix).[6] We carried out a statistical test of $\mathcal{SC}$-rationalizability, according to the procedure set out in Section 5, with a bootstrap sample of size 2000. Contrary to the result in Section 5, the hypothesis of $\mathcal{SC}$-rationalizability is rejected in this case, with the p-value equal to 0. Thus the conclusion of $\mathcal{SC}$-rationalizability appears to be sensitive to the level of discretization used.

Our main objective is to explore the performance of the column generation method and on that score the news is good. On a desktop computer with Apple M1 processor and 16 GB RAM), the computation for this test took 19.051 minutes, including the 2000 calculations performed on the bootstrap sample of that size. This compares with a computation time of 2.590 minutes for testing the original model, where the covariates took 8 possible values (rather than 64).

It is worth emphasizing that this eightfold increase in the number of covariates leads to an enormous increase in the number of behavioral types. In the case of binary covariates, there are only $4^8 \approx 65,000$ logically possible behavioral types, with a direct check yielding 482 behavioral types that are $\mathcal{SC}$-rationalizable (equivalently, obeying the RM axiom). However, when each covariate can take four values, the number of possible behavioral types is equal to $4^{64} (> 10^{38})$ and at least $2^{64} (> 10^{19})$ of them obey RM axiom.[7] It is clear that there is no way one could completely list the SC-rationalizable types in this case. On the other hand, once we provide a starter set of behavioral types, our column generation algorithm has no difficulty finding the additional types needed to compute the test statistics.[8]

We also repeated the exercise with only MS is divided into four values and both MP_{LCC} and MP_{OA} remaining binary. In this case, each realization of covariates contains from 243 to 695 observed markets, with its median being equal to 503. We find that the data pass the statistical

[6] While the covariates can be divided even more finely, this would decrease the number of markets in each realization of the covariates to unacceptable levels. Indeed if we use a sextile division, some realizations of covariates will contain no markets at all.

[7] The logic behind these numbers are as follows. There are 64 possible realizations of covariates, and for each of them, there are four possible joint actions, (N, N), (N, E), (E, N), and (E, E). Thus, the total number of behavioral types is equal to 4^{64}. In addition, one can check that any behavioral type where either (E, N) or (N, E) is played at a covariate obeys the RM axiom. Hence, there are at least 2^{64} types obeying the RM axiom.

[8] We use different starter sets $\mathcal{B}_0$ for computing J_N and $J_N^{(r)}$. For calculating J_N, the starter set is the set of constant behavioral types (of which there are just four). For calculating $J_N^{(1)}$, the starter set is a basis of the span of $\mathcal{B}$, which has dimension 193. For calculating $J_N^{(2)}$ and so forth, the starter sets are bigger (see Section A2.1).

test for $\mathcal{SC}$-rationalizability, with a p-value being equal to 0.412. Given this, we can perform further investigations. Repeating the analysis in the main paper, we find that the existence of strategic interaction is crucial to explaining the data. The maximal weight on behavioral types consistent with the hypothesis that (i) a payoff function of LCC is independent of the action of OA, (ii) a payoff function of OA is independent of the action of LCC, and (iii) both firms have payoff functions independent of the opponent's action are respectively, 0.902, 0.762, and 0.753 at the 5% significance level. The probability bounds for equilibrium actions were also calculated. The computations finish within reasonable time (testing the model finishes in 3.730 minutes, while the inference calculations take 5.664 minutes in total).

References

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2018): Revealed price preference: theory and empirical analysis. *arXiv:1801.02702*.

Kitamura, Y., and J. Stoye (2018): Nonparametric analysis of random utility models. *Econometrica*, 86, 1883-1909.

Kline, B., and E. Tamer (2016): Bayesian inference in a class of partially identified models. *Quantative Economics*, 7, 329–366.

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

Smeulders, B., Cherchye, L. and R. De Rock, B. (2021): Nonparametric analysis of random utility models: computational tools for statistical testing. *Econometrica*, 89(1), 437–455.