Nonparametric analysis of monotone choice

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

April 18, 2018

Abstract: We develop a nonparametric approach to test for monotone behavior in optimizing agents and to make out-of-sample predictions. Our approach could be applied to simultaneous games with ordered actions, with agents playing pure strategy Nash equilibria or Bayesian Nash equilibria. We require no parametric assumptions on payoff functions nor distributional assumptions on the unobserved heterogeneity of agents. Multiplicity of optimal solutions (or equilibria) is not excluded, and we are agnostic about how they are selected. To illustrate how our approach works, we include an empirical application to an IO entry game.

Keywords: revealed preference; monotone comparative statics; single crossing differences; supermodular games; entry games

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

[*] Email addresses: nlazzati@ucsc.edu; john.quah@jhu.edu; kshirai1985@gmail.com
For helpful discussions and comments, the authors are grateful to S. Berry, A. Carvajal, A. De Paula, J. Fox, K. Hirano, A. Kajii, Y. Kitamura, B. Kline, E. Krasnokutskaya, C. Manski, J. Stoye, B. Strulovici, A. Sweeting, S. Takahashi, and Y. Takahashi. 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, 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, and conference presentations at the Cowles Conference on Heterogenous Agents and Microeconometrics (Yale, 2015), the SAET Conference (Cambridge (UK), 2015), the World Congress of the Econometric Society (Montreal, 2015), the SWET Conference (UC Riverside, 2016), the Conference on Econometrics and Models of Social Interactions (CeMMAP and Vanderbilt, 2017), and the Canadian Economic Theory Conference (Vancouver, 2017). Koji Shirai gratefully acknowledges financial support from the Japan Society for Promotion of Science (Grant-in-Aid for JSPS Fellows); he would also like to thank St Hugh's College (Oxford), the Oxford Economics Department, and the National University of Singapore Economics Department for their hospitality during his extended visits to these institutions when he was a JSPS fellow.

1 Introduction

This paper proposes a practical method for testing monotone effects in games of complete and incomplete information. Our method is nonparametric, based on restrictions arising from revealed preference, and builds on the theory of monotone comparative statics developed in Milgrom and Shannon (1994) and Quah and Strulovici (2009). These papers identify properties on payoff functions (such as single crossing conditions) 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 single crossing or other key properties? This is the question that we set out to answer, firstly in the context of a panel data set, where the choices of an agent subject to different exogenous variables are observed, and secondly, in the context of a cross sectional data set where the econometrician observes the distribution of actions of a population under different exogenous conditions.

An 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 entry decision in a market is a best response to decisions taken by other firms. The payoff functions are assumed to depend on observable variables in a specific parametric fashion while the unobserved component is additively separable. The latter 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]

[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

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 it is 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 Nash equilibria. Our test requires *rationalizability*, in the sense that we can recover a distribution on firms' payoff functions (strictly speaking on their preferences) that satisfy the single crossing restrictions we require 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 to make out-of-sample predictions on equilibrium behavior.

Our approach is also useful in the context of Bayesian games, such as those studied by Sweeting (2009) and De Paula and Tang (2012). Sweeting (2009) shows how we could estimate a Bayesian game subject to payoff functions which depend parametrically on observable variables and also on heterogeneous shocks (unobservable both to the econometrician and to other players). Our setup is very similar, except that we make no parametric assumptions. We show how to test whether a data set is rationalizable as a Bayesian Nash equilibrium, subject to single crossing restrictions on payoff functions. De Paula and Tang (2012) demonstrates how to recover the sign of interaction effects in Bayesian games, without making parametric assumptions on payoff functions, so this is close in spirit to what we do. However, De Paula and Tang's test *assumes* that firms are playing Bayesian Nash equilibria, while in our approach that assumption is part of the test, in the sense that we require the recovery of a complete set of model primitives which are consistent with the observations and with the single crossing assumptions.[2]

Broadly speaking, we see our revealed preference method as providing a useful tool that could

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]Since our test is more stringent, our data requirements are also (in a sense) more demanding than that in De Paula and Tang (2012). See Section 5.2 for a detailed discussion.

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.[3] On the other hand, a nonparametric version of the model that does not perform well in our test may raise questions about the validity of the model or the suitability of the data.

The paper is organized as follows. In Section 2 we provide an outline of how our approach works in the context of an entry game. Section 3 focuses on individual decision making, and begins with a quick survey of theoretical results on monotone comparative statics. Section 3.2 is the theoretical heart of the paper, where we characterize panel data sets (of individual choices) which are consistent with the single crossing property; this is achieved through a condition on the data set we call the *revealed monotonicity axiom*. The rest of the section then explains how to extend the test for the single crossing property to cross sectional data sets, in which we observe a distribution of actions from a population of agents with heterogenous payoff functions. This extension (from panel to cross sectional data) follows an approach that has been taken by other authors such as Manski (2007). Manski (2007) also discusses out of sample predictions, and our approach to this issue is, in its essentials, the same as his. In the case where the feasible action set available to all agents in the population is unchanged across observations, we show that the single crossing property can be characterized by an intuitive formulation involving first order stochastic dominance.

Section 4 extends the ideas of Section 3 to games, where we show how to test for pure strategy Nash equilibria. The use of these results is then illustrated in Section 6, where we carry out an empirical analysis of entry decisions made by airlines, using the data collected by Kline and Tamer (2016); we show how strategic substitutability in airlines' entry decisions can be tested using our method, and also explain how we can recover information on the distribution of payoff functions among airlines. More details on the empirical implementation can be found in the Online Appendix. Section 5 explains how the results of Section 3 can be applied to Bayesian games. It is worth noting that while the examples in this paper are largely taken from IO games, our overall approach can be applied to other decision- or game-theoretic settings; for example, the authors have applied it to study joint smoking behavior among married couples (see Lazzati, Quah, and Shirai (2018)).

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

(x_1, x_2)		Firm 2	
		N	E
Firm 1	N	$\mathrm{P}(N, N \mid x_1, x_2)$	$\mathrm{P}(N, E \mid x_1, x_2)$
	E	$\mathrm{P}(E, N \mid x_1, x_2)$	$\mathrm{P}(E, E \mid x_1, x_2)$

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

2 Motivating example

The seminal papers of Bresnahan and Reiss (1990) and Berry (1992) have given rise to a large literature in empirical IO and structural econometrics that model oligopoly entry. In these models there is a set of firms, $\mathcal{N} = \{1, 2, ..., n\}$ that may potentially operate in a given market. Firm i's action is denoted by $y_i \in \{E, N\}$, where E means that the firm enters and N that it stays out. The profit of firm i upon entry is determined by the entry decision of other firms and also by exogenous factors. We denote its profit by $\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i)$, where $\mathbf{y}_{-i}$ are the choices of the other firms, x_i is a real-valued, finite-dimensional vector of exogenous profit shifters (that might be market- or firm-specific) observed by all firms and also by the researcher, and ε_i are profit shifters observed by all firms but not by the researcher.

Typically, it is also assumed that Π_i has a linear functional form; for example, in Ciliberto and Tamer (2009) it is assumed that

$$\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i) = \begin{cases} \alpha_i' x_i + \sum_{j \neq i} \delta_{ij} \mathbf{1}_{y_j} + \varepsilon_i & \text{if} \quad y_i = E \\ 0 & \text{if} \quad y_i = N \end{cases} \tag{1}$$

where $\mathbf{1}_E = 1$ and $\mathbf{1}_N = 0$. Note that the entry of firm j alters the profit of firm i by δ_{ij}. The econometrician's first objective is to estimate α_i and δ_{ij} (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 there are just two firms, 1 and 2, both of which have payoff functions of the form given by (1), and which interact with each other across many markets. For some given value of the observable profit shifters (x_1, x_2), the econometrician observes the distribution of action profiles in a large sample of markets. This is illustrated in Table 1, where $\mathrm{P}(E, N \mid x_1, x_2)$ denotes the sample frequency of those markets where Firm 1 enters and Firm 2 stays out, when the observable profit shifters take the value (x_1, x_2). (Obviously, the four entries in the table should add up to

1.) 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_{ij}, under the following two assumptions: (1) at each market, each pair of firms is playing 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, with its parameters at least partially known. For a recent study of how these estimates could be obtained from data in this format, see Kline and Tamer (2016).

In the empirical study of entry games, a major focus of attention is whether the entry decisions of other firms and/or the movement of different profit shifters tend to encourage or discourage a firm's entry. (In the parametric form (1) this manifests itself in the signs of α_i and δ_{ij}.) 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. We also do not require the unobservable profit shifters $(\varepsilon_1, \varepsilon_2)$ to belong to any distribution family or to influence payoffs in an additively separable way, though we maintain the assumption that its distribution is independent of the observable profit shifters. In particular, we allow for ε_1 and ε_2 to be correlated with each other. In the event that a data set is consistent with our hypotheses, our procedure also leads to the (set) estimation of the primitives of the model that would generate the observations as pure strategy Nash equilibria.

2.1 Single crossing condition

To explain our approach in greater detail, suppose we wish to test the hypothesis that a firm'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 crucial observation to make here is that this hypothesis is precisely captured by the following simple condition, which is a version of the *single crossing property*, on the firm's payoff function:[4]

$$\Pi_i(E, y_j', x_i', \varepsilon_i) > \Pi_i(N, y_j', x_i', \varepsilon_i) \implies \Pi_i(E, y_j'', x_i'', \varepsilon_i) > \Pi_i(N, y_j'', x_i'', \varepsilon_i)$$

whenever $(-\mathbf{1}_{y_j''}, x_i'') > (-\mathbf{1}_{y_j'}, x_i')$. In other words, if firm i prefers entering a market to staying out when firm j is also entering, then this preference is preserved if there is an increase in x_i or if firm

[4]Throughout this section we shall assume that preferences are always strict. Indifferences are dealt with carefully in Section 3.

$x_2 = (0,0)$		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

j decides not to enter. In the case of (1), this holds if $\alpha_i > 0$ and $\delta_{ij} < 0$, but it is clear that such a parametric form is not necessary for the single crossing property to hold.

To gain some intuition on how our approach works, let us ignore small sample issues for the time being and suppose that we observe the true distribution of action profiles at some realized values of (x_1, x_2). We claim that the joint hypothesis that firms have payoff functions obeying the single crossing property and are playing pure strategy Nash equilibria is sufficient to impose restrictions on the data. Indeed, consider an increase in the observable profit shifters from (x_1', x_2') to (x_1'', x_2''); then, at any realization of ε_1, if firm 1 strictly prefers to enter when the other firm enters at (x_1', x_2'), the single crossing property guarantees that it will continue to prefer enter 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 pair of firms, it will continue to be the unique Nash equilibrium at (x_1'', x_2''). Since the distribution of $(\varepsilon_1, \varepsilon_2)$ is independent of the observable profit shifters, we conclude that[5]

$$\mathrm{P}(E, E | x_1'', x_2'') \geqslant \mathrm{P}(E, E | x_1', x_2').$$

This constitutes a restriction on the data that could be tested but it is certainly not the only restriction imposed by our hypothesis. The tightest possible restriction is obtained by checking whether the data set is *rationalizable*. This involves identifying those types of joint behavior (between the two firms) which are consistent with Nash equilibrium play and profit functions obeying the single crossing property, and then attributing weights to these types in a way that accounts for the observed distribution of action profiles.

[5]Strictly speaking, we need to assume here that the probability of a firm being exactly indifferent between entering and staying out is zero.

Path	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 3: Distribution of types rationalizing data in Table 2

2.2 Rationalizability

To get a flavor of what testing for rationalizability involves, suppose that we observe the distribution over action profiles at three values of (x_1, x_2), as depicted in Table 2. Note that x_1 is fixed throughout, while x_2 is two-dimensional and takes three values. We claim that these observations can be rationalized by our model. To understand why this is the case, we list in Table 3 seven possible *group paths* for a pair of Firms 1 and 2. For example, in group path 1, (E, N) is played by the two firms at $x_2 = (0, 0)$, with the firms switching to the Nash equilibrium (E, E) when $x_2 = (0, 1)$, and remaining at (E, N) when $x_2 = (1, 0)$. It is straightforward to check that each of these paths is compatible with single crossing, in the sense that one could find single crossing profit functions for each firm, so that for each realized value of x_2, the specified action profiles constitute a (not necessarily unique) Nash equilibrium.[6] Furthermore, when these paths are represented in the population with the weights indicated in Table 3, they rationalize the distribution of action profiles observed in Table 2. (Compare the entries in Table 2 with the last row of Table 3.)

More generally, the procedure for establishing the rationalizability of a data set (like that displayed in Table 2) consists of two steps. The first step consists of identifying *all* the group paths which are compatible with the single crossing condition and the second consists of determining if there is a distribution on these paths that explains the data. The second step is a computationally straightforward linear programming problem, but solving the first step is generally non-trivial.

[6]For example, the following preferences will lead to the pure strategy Nash equilibria in group path 1: firm 1 prefers E to N irrespective of the other firm's action and, irrespective of the action of firm 1, firm 2 prefers E to N if $x_2 = (0, 1)$, while it prefers N to E at other realizations of x_2.

What is required is an easy-to-check property on observed actions which *characterizes* payoff functions that obey the single crossing property: this is fundamentally a theoretical question and it is addressed in the Section 3 and 4.[7]

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 (1), in which Firm 2's profit upon entry will be

$$\Pi_2(E, y_1, x_{21}, x_{22}) = \alpha_{21}x_{21} + \alpha_{22}x_{22} + \delta_{21}\mathbf{1}_{y_1} + \varepsilon_2, \tag{2}$$

where $(\alpha_{21}, \alpha_{22}) > 0$ and $\delta_{21} < 0$.[8] The essential reason for this is the following: when Π_2 has the form (2), 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 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,

$$\frac{1}{12} = \mathrm{P}(N,N|x_1,(0,0)) - \mathrm{P}(N,N|x_1,(1,0)) = \mu\left(\{\varepsilon_1 : \Pi_1(E,N,x_1,\varepsilon_1) \leqslant 0\} \times \{\varepsilon_2 : 0 \geqslant \varepsilon_2 \geqslant -\alpha_{21}\}\right)$$

[7] The seven paths listed in Table 3 are only *some* of the possible group paths consistent with single crossing. For example, suppose firm 1 prefers E to N if and only if firm 2 chooses N, and firm 2 prefers E to N if and only if $x_2 = (0,1)$ *and* firm 1 chooses N. These preferences lead to two group paths, because there are two Nash equilibria at $x_2 = (0,1)$. One is Path 4, where (E,N) is always chosen. The other path involves (E,N) being played at $x_2 = (0,0)$ and $x_2 = (1,0)$, and (N,E) being played at $x_2 = (0,1)$; this path is not among those listed.

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

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 Revealed monotone choice

This section focuses on the single-agent model. Section 3.1 gives a quick survey of basic results on monotone comparative statics. Section 3.2 presents the theoretical result underpinning the whole paper: imagining a panel data set of choice behavior, it gives a necessary and sufficient condition under which it is generated by an agent choosing with a preference obeying the single crossing property. Section 3.3 explains how the results in Section 3.2 could be applied to repeated cross sections of choices made by a population of agents with such preferences.

3.1 Basic concepts and theory

Let $\mathcal{Y}$ be the set of all conceivable actions of a given agent. We assume it is either the real line $\mathbb{R}$ or a closed subset of $\mathbb{R}$ such as the natural numbers. The agent selects an action from a subset A of $\mathcal{Y}$. A set A is said to be *an interval of* $\mathcal{Y}$ if, for every $y', y'' \in A$ with $y' < y''$, we have that

$$y \in \mathcal{Y} \text{ and } y' < y < y'' \Longrightarrow y \in A.$$

Throughout this paper, we assume that observed feasible sets are *compact intervals*, where compactness is with respect to the Euclidean topology on $\mathbb{R}$. In this case, any compact interval A will have a largest element y'' and a smallest element y' so we may write it as $A = [y', y'']$. We denote the collection of compact interevals by $\mathbb{I}(\mathcal{Y})$.

The choice of the agent over different actions in a feasible set A is affected by a set of covariates $z \in \mathcal{Z}$. We assume $(\mathcal{Z}, \geqslant)$ is a partially ordered set. (For the sake of notational simplicity, we are using the same notation for the orders on $\mathcal{Y}$ and $\mathcal{Z}$ and for any other ordered sets; we do not anticipate any danger of confusion.) We refer to (z, A) as the agent's *environment*.

A binary relation $\succsim$ on $\mathcal{Y} \times \mathcal{Z}$ is a *preference* relation if, for every fixed z, it is reflexive, transitive, and complete. That is, $(y'', z) \gtrsim (y', z)$ means that y'' is weakly preferred to y' when the covariates are z.[9] The asymmetric part of $\gtrsim$ (the *strict* preference) is denoted by $\succ$. At the environment $(z, A) \in \mathcal{Z} \times \mathbb{I}(\mathcal{Y})$, the *best response* or *optimal choice* of the agent is the set

$$\mathrm{BR}(z, A) = \{y' \in A : (y', z) \gtrsim (y, z) \text{ for all } y \in A\}. \tag{3}$$

The preference $\gtrsim$ is *regular* if $\mathrm{BR}(z, A)$ is nonempty and compact for all $(z, A) \in \mathcal{Z} \times \mathbb{I}(\mathcal{Y})$. Regularity holds trivially in the important case where every bounded set of $\mathcal{Y}$ is finite (e.g., when $\mathcal{Y} \subset \mathbb{N}$) and, more generally, it holds if $\gtrsim$ is continuous at every z.

Oftentimes (and, indeed, in the previous section) it is convenient to think of the agent as having a *payoff function*, which is a real-valued map Π on $\mathcal{Y} \times \mathcal{Z}$. Clearly, this induces a preference $\gtrsim$ on $\mathcal{Y} \times \mathcal{Z}$, where $(y'', z) \gtrsim (y', z)$ if $\Pi(y'', z) \geqslant \Pi(y', z)$. Then $\mathrm{BR}(z, A) = \arg\max_{y \in A} \Pi(y, z)$ and we can speak of the payoff function being regular, etc., if the preference it induces has the corresponding property. Since observed choices only reveal ordinal information, it is appropriate that our discussion in this subsection and the next should focus on preferences. In Section 3.3, we shall revert to using payoff functions, so as to follow the convention in much of the econometric and structural IO literature.

The best response $\mathrm{BR}(z, A)$ is said to be *increasing in* z if, for every $z'' > z'$,

$$y'' \in \mathrm{BR}(z'', A) \text{ and } y' \in \mathrm{BR}(z', A) \Longrightarrow y'' \geqslant y'. \tag{4}$$

The preference $\gtrsim$ is said to obey *strict interval dominance (SID)* in $(y; z)$ if, for every $y'' > y'$ and $z'' > z'$,

$$(y'', z') \gtrsim (y, z') \text{ for all } y \in [y', y''] \Longrightarrow (y'', z'') \succ (y', z').$$

The following result is a straightforward adaptation of Theorem 1 in Quah and Strulovici (2009).

Basic Theorem. *Suppose an agent has a regular preference $\gtrsim$ on $\mathcal{Y} \times \mathcal{Z}$. Then $\mathrm{BR}(z, A)$ is increasing in $z \in \mathcal{Z}$ at every $A \in \mathbb{I}(\mathcal{Y})$ if and only if $\gtrsim$ obeys SID in $(y; z)$.*

[9] Note that this is not a standard definition of a 'preference' since we do not require it to be complete on $\mathcal{Y} \times \mathcal{Z}$. We *could* define it in the standard way but that does not seem like a meaningful thing to do given that, in our setting, even if the agent truly has a preference between (y, z) and (y', z') where $z \neq z'$, this will never be revealed since she never chooses between these two alternatives.

Readers familiar with the theory of monotone comparative statics will notice that our definition of monotonicity in (4) is stronger than the standard notion, which merely requires that $\mathrm{BR}(z'', A)$ dominates $\mathrm{BR}(z', A)$ in the *strong set order*.[10] This weaker notion of monotonicity can be characterized by preferences obeying *interval dominance* (rather than strict interval dominance), which can be defined as follows: for every $y'' > y'$ and $z'' > z'$,

$$(y'', z') \gtrsim (\succ)\, (y, z') \text{ for every } y \in [y', y''] \Longrightarrow (y'', z'') \gtrsim (\succ)\, (y', z'') \tag{5}$$

(see Theorem 1 in Quah and Strulovici (2009)). Throughout this paper we have chosen to work with a stronger notion of monotonicity; the weaker notion does not permit meaningful revealed preference analysis because it does not exclude the possibility that an agent is simply indifferent to all actions at every z. In this sense, our stronger assumption here is analogous to the assumption of local non-satiation made in Afriat's Theorem.[11] In the case where $\mathcal{Y}$ is discrete, it is quite natural to assume that the agent has a strict preference; in that context, there is no distinction between interval dominance and strict interval dominance.

The interval dominance order is Quah and Strulovici's (2009) generalization of the concept of *single crossing differences*, due to Milgrom and Shannon (1994).[12] Just as there is strict interval dominance, so there is a strict version of single crossing differences. A preference $\gtrsim$ obeys strict single crossing differences (SSCD) if, for every $y'' > y'$ and $z'' > z'$,

$$(y'', z') \gtrsim (y', z') \Longrightarrow (y'', z'') \succ (y', z''). \tag{6}$$

It is clear that SSCD implies SID, and hence the Basic Theorem applies if $\gtrsim$ obeys SSCD. In fact, it is known that the stronger SSCD property leads to a stronger conclusion: if $\gtrsim$ obeys SSCD then $\mathrm{BR}(z'', A)$ is increasing in z, for any nonempty set A, whether or not it is an interval.[13]

Consider the entry model in Section 2. For simplicity, let us assume the payoff shifter x_i for

[10]A set B'' dominates B' in this order if, for every $b'' \in B''$ and $b' \in B'$, $b'' \vee b' \in B''$ and $b'' \wedge b' \in B'$.

[11]It is clear that without such an assumption, any type of consumption data is rationalizable since one could simply suppose that the consumer is indifferent across all consumption bundles. For a statement and proof of Afriat's Theorem see Varian (1982).

[12]Milgrom and Shannon (1994) actually use the term 'single crossing property'; our reference to it as 'single crossing differences' follows Milgrom (2004).

[13]For a more detailed discussion of the connections, see Quah and Strulovici (2009).

player i is scalar, representing the size of the market, and that $\alpha_i > 0$ and $\delta_{ij} < 0$ for all $j \neq i$. In this case, firm i's payoff upon entry increases with x_i (market size) and diminishes as more firms enter the market; formally, $\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i)$ is increasing in $z_i = (-\mathbf{y}_{-i}, x_i)$, for every ε_i. This guarantees that at every possible realization of the unobserved payoff shifter ε_i, firm i's payoff function, $\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i)$, obeys SSCD in $(y_i; z_i)$. It follows from the Basic Theorem that if the firm chooses to enter the market at some $z_i = z_i'$ then it will also choose to enter when $z_i = z_i'' > z_i'$.

3.2 Revealed Monotonicity Axiom

We consider an observer who records the actions chosen by an agent under a finite set of environments. The data set can be denoted by

$$\mathcal{O} = \left\{ \left(y^t, z^t, A^t\right) : t \in \mathcal{T} = \{1, 2, ..., T\} \right\},$$

where at observation t, the agent chooses y^t in the environment (z^t, A^t). (Note that we allow for the agent to make different choices in the same environment, so it is possible for $(z^t, A^t) = (z^s, A^s)$ and yet $y^t \neq y^s$.)

Definition 1. *$\mathcal{O}$ is rationalizable if there is a regular preference $\succsim$ on $\mathcal{Y} \times \mathcal{Z}$ such that for each $t \in \mathcal{T}$, we have $(y^t, z^t) \succsim (y, z^t)$ for every $y \in A^t$.*

Our aim in this subsection is to characterize data sets that are rationalizable by preferences that obey SID in $(y; z)$. Our motivation is clear: if $\mathcal{O}$ is rationalizable by an SID preference then there is a regular preference that can both account for the observed behavior of the agent and guarantees that the optimal choice of the agent based on this preference is increasing in the covariates on any feasible action set in $\mathbb{I}(\mathcal{Y})$ (including environments *outside* the observations $\{(z^t, A^t)\}_{t \in \mathcal{T}}$).

Example 1. Consider a firm (for example, a power generator) producing a perishable good, whose production in each period depends on the spot price for its output and the forward contracts it has already signed. The observation at period t is (y^t, z^t, A^t), where $y^t \geqslant 0$ is the firm's output, $z^t \geqslant 0$ the spot price, and $A^t = [\bar{y}^t, K]$, where $\bar{y}^t$ is the amount the firm had already committed to supplying (at a price or prices which are not part of the observation) and K is the firm's capacity. Suppose that in each period t, the firm chooses $y \geqslant \bar{y}^t$ to maximize $z(y - \bar{y}^t) - C(y)$, where C is the cost of producing y. Then the data set $\mathcal{O} = \{(y^t, z^t, A^t)\}_{t\in\mathcal{T}}$ is rationalizable by a preference on output-price pairs $(y, z) \in \mathbb{R}_+ \times \mathbb{R}_+$ that obeys SID in $(y; z)$. Indeed, it is clear that we can choose the preference to be that induced by the payoff function $\Pi(y, z) = zy - C(y)$.

To determine whether a data set is rationalizable by an SID preference, we first define the revealed preference relations induced by $\mathcal{O}$. The *direct revealed preference* relation $\succsim^R$ is defined as follows: $(y'', z) \succsim^R (y', z)$ if $(y'', z) = (y^t, z^t)$ and $y' \in A^t$ for some $t \in \mathcal{T}$. The *indirect revealed preference* relation $\succsim^{RT}$ is the transitive closure of $\succsim^R$, i.e., $(y'', z) \succsim^{RT} (y', z)$ if there exists a finite sequence $\bar{y}_1, \bar{y}_2, ..., \bar{y}_k$ in $\mathcal{Y}$ such that

$$(y'', z) \succsim^R (\bar{y}_1, z) \succsim^R (\bar{y}_2, z) \succsim^R ... \succsim^R (\bar{y}_k, z) \succsim^R (y', z). \tag{7}$$

The motivation for this terminology is as follows. If the agent is optimizing according to some preference $\succsim$ and, at some environment (z, A) the agent selects y'' when $y' \in A$, then it must be the case that $(y'', z) \succsim (y', z)$. Furthermore, given that $\succsim$ is transitive, if $(y'', z) \succsim^{RT} (y', z)$ then $(y'', z) \succsim (y', z)$. We are now ready to introduce the axiom that characterizes rationalizability by an SID preference.

Definition 2. $\mathcal{O} = \{(y^t, z^t, A^t)\}_{t\in\mathcal{T}}$ *obeys the* Revealed Monotonicity (RM) axiom *if, for every* $s, t \in \mathcal{T}$,

$$z^t > z^s,\ y^t < y^s,\ \textit{and}\ (y^s, z^s) \succsim^{RT} (y^t, z^s) \Longrightarrow (y^t, z^t) \not\succsim^{RT} (y^s, z^t).$$

Remark: Suppose that $\mathcal{O}$ is such that $A^t = A \subseteq \mathcal{Y}$ for all $t \in \mathcal{T}$. Then it is clear that the RM axiom holds if and only if the optimal action is increasing in the covariates in the following sense: $y^t \geqslant y^s$ for any two observation $t, s \in \mathcal{T}$ such that $z^t > z^s$.

It is clear that the RM axiom is a non-vacuous restriction. So long as the dataset is finite, checking whether $\mathcal{O}$ obeys this property is a finite problem and indeed there are no computational difficulties, either theoretical or practical, associated with the implementation of this test. It helps with motivation at least to see why the RM axiom is a necessary condition.

Proposition 1. *Suppose* $\mathcal{O} = \{(y^t, z^t, A^t)\}_{t\in\mathcal{T}}$ *is rationalizable by a preference* $\succsim$ *that obeys SID in* $(y; z)$. *Then* $\mathcal{O}$ *obeys the RM axiom.*

Proof. We first establish that $\succsim^{RT}$ has what we call the *interval property*. In general, a binary relation $\mathcal{R}$ on $\mathcal{Y} \times \mathcal{Z}$ has this property if $(y'', z) \mathcal{R} (y', z)$ implies $(y'', z) \mathcal{R} (y, z)$ for any y between y'' and y', i.e., either $y' < y < y''$ or $y'' < y < y'$.

If $(y'', z) \succsim^R (y', z)$, then there is A^t such that $y'' = y^t$ and $y' \in A^t$. Since A^t is an interval, it is clear that $(y'', z) \succsim^R (y, z)$ for any y between y'' and y'. Now suppose $(y'', z) \succsim^{RT} (y', z)$, but $(y'', z) \not\succsim^R (y', z)$. Then, we have a sequence like (7). Suppose also that $y'' > y'$ and consider y such that $y'' > y > y'$. (The case where $y'' < y'$ can be handled in a similar way.) Letting $y_0 = y''$ and $y_{k+1} = y'$, we know that there exists at least one $0 \leqslant m \leqslant k$ such that $y_m \geqslant y \geqslant y_{m+1}$. Since $(y_m, z) \succsim^R (y_{m+1}, z)$, it must hold that $(y_m, z) \succsim^R (y, z)$. This in turn implies that $(y'', z) = (y_0, z) \succsim^{RT} (y, z)$, since $(y_0, z) \succsim^{RT} (y_m, z)$. So we have shown that $\succsim^{RT}$ has the interval property.

Suppose there are observations s and t such that $z^t > z^s$, $y^t < y^s$, and $(y^s, z^s) \succsim^{RT} (y^t, z^s)$ holds. The interval property guarantees that $(y^s, z^s) \succsim^{RT} (y, z^s)$ for all $y \in [y^t, y^s]$. Since $\mathcal{O}$ is rationalizable by an SID preference $\succsim$, we have $(y^s, z^s) \succsim (y, z^s)$ for all $y \in [y^t, y^s]$. The SID property on $\succsim$ guarantees that $(y^s, z^t) \succ (y^t, z^t)$, which means that $(y^s, z^t) \not\succ^{RT} (y^t, z^t)$. QED

Of course, our more substantial claim is that the RM axiom is also *sufficient* for rationalizability by an SID preference. In fact, an even stronger property is true: whenever a dataset obeys the RM axiom, then it is rationalizable by an SSCD (and not just SID) preference.[14]

THEOREM 1. *The following statements on $\mathcal{O} = \{(y^t, z^t, A^t)\}_{t \in \mathcal{T}}$ are equivalent:*

(a) *$\mathcal{O}$ is rationalizable by a preference that obeys SID in $(y; z)$.*

(b) *$\mathcal{O}$ obeys the RM axiom.*

(c) *$\mathcal{O}$ is rationalizable by a preference that obeys SSCD in $(y; z)$.*

As a very simple illustration of the use of this theorem, consider the following example.

EXAMPLE 1 (continued). Suppose there are just two observations, $(y^1, z^1, A^1 = [\bar{y}^1, K])$ and $(y^2, z^2, A^2 = [\bar{y}^2, K])$, with $z^2 > z^1$ and $\bar{y}^2 < \bar{y}^1$. In other words, the output price is higher and the

[14]This phenomenon, which may seem surprising, is not unknown to revealed preference analysis; for example, it is present in Afriat's Theorem. In that context, the data consist of observations of consumer's consumption bundles at different linear budget sets. If the agent is maximizing a locally non-satiated preference, then the data set must obey a property called the generalized axiom of revealed preference (GARP, for short); conversely, if a data set obeys GARP then it can be rationalized by a preference that is not just locally non-satiated but also obeys continuity, strong monotonicity, and convexity.

output commitment lower at the second observation. Suppose $\bar{y}^1 < y^2 < y^1$, then the RM axiom is violated since $(y^1, z^1) \succsim^R (y^2, z^1)$ and $(y^2, z^2) \succsim^R (y^1, z^2)$. On the other hand if $y^2 < \bar{y}^1 < y^1$, then the RM axiom holds, because the only revealed relation is $(y^2, z^2) \succsim^R (y^1, z^2)$ and thus these observations can be rationalized by an SID preference that obeys SID in $(y; z)$. Note that for any such preference $\succsim$, the SID property implies that $(y^2, z^1) \succ (y^1, z^1)$ but this does not contradict the optimality of y^1 in $[\bar{y}^1, K]$.

It would be natural to speculate given Theorem 1 that, if we allow A^t to be arbitrary subsets of $\mathcal{Y}$ (rather than intervals), then the RM axiom is necessary and sufficient for rationalizability with SSCD preferences. It is clear that the axiom will be necessary for rationalizability in this sense, but sufficiency does not hold.

Example 2. Let $\mathcal{Y} = \{u, v, w\}$ with $u < v < w$, and let $A^1 = \{u, w\}$, $A^2 = \{u, v\}$, and $A^3 = \{v, w\}$. Note that A^1 is not an interval of $\mathcal{Y}$. Suppose that $z^1 < z^2 < z^3$. Then $(w, z^1) \succsim^R (u, z^1)$, $(u, z^2) \succsim^R (v, z^2)$, and $(v, z^3) \succsim^R (w, z^3)$. The indirect revealed preference relation $\succsim^{RT}$ is equal to the direct revealed preference relation $\succsim^R$ in this example and, clearly, this set of three observations obeys the RM axiom. However, it cannot be rationalized by an SSCD preference. Suppose, instead, that an SSCD preference $\succsim$ rationalizes the data. Then, it must be that $(w, z^1) \succsim (u, z^1)$ and, by SSCD, $(w, z^2) \succ (u, z^2)$. In addition, we have $(u, z^2) \succsim (v, z^2)$ and so $(w, z^2) \succ (v, z^2)$. Since $\succsim$ obeys SSCD, we get $(w, z^3) \succ (v, z^3)$, which contradicts the direct revealed preference $(v, z^3) \succsim (w, z^3)$.

The RM axiom is an easy-to-understand property written in a form of no-cycling condition (which is reminiscent of GARP in Afriat's Theorem or the congruence axiom in Richter's Theorem), and the necessity of the axiom is relatively straightforward to show. Given the superficial familiarity, a reader could be forgiven for thinking that its sufficiency is also obvious. But there is more than what meets the eye in Theorem 1 and intuition can be misleading; indeed, any *correct* intuition will have to distinguish between arbitrary constraint sets and interval constraint sets because, as Example 2 demonstrates, the result is not true in the former case. The proof of the sufficiency of the RM axiom does proceed in a way which is vaguely familiar, in the sense that we extend $\succsim^{RT}$ further by relying on strict interval dominance, and then take the transitive closure of that extended revealed preference relation. While this seems like a natural approach to take, the issue is whether the resulting incomplete revealed preference relation is actually well-behaved enough to admit a completion that obeys SSCD or even SID, since the added requirements on the preference means that we cannot simply appeal to Szpilrajn's Theorem (or some other standard theorem) to complete the relation. The heart (and substantive part) of the proof lies in showing that the revealed preference relation *does* have the properties that allow for a completion that obeys SSCD. This in turn relies crucially on the observed constraint sets being intervals; indeed, Kukushkin, Quah, and Shirai (2016) provide an example of an incomplete preference ordering that obeys SSCD (which can be interpreted as arising from a data set with non-interval constraint sets), for which there is no completion that also obeys that property.

In the case where $\mathcal{Y}$ is finite, it is obvious that any SID preference on $\mathcal{Y} \times \mathcal{Z}$ can be represented by a payoff function $\Pi : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$, in the sense that $\Pi(y', z) \geqslant (>) \, \Pi(y, z)$ if $(y', z) \succsim (\succ) \, (y, z)$.[15] The next result asserts that such a representation generally exists.

PROPOSITION 2. *Suppose $\mathcal{Y}$ is a closed interval of $\mathbb{R}$ and that $\mathcal{O}$ obeys the RM axiom. Then $\mathcal{O}$ is rationalizable by a preference that obeys SSCD in $(y; z)$ and admits a payoff representation.*

Up to this point in our discussion, we have allowed for the possibility that an agent's preference is indifferent between two actions. It is sometimes convenient, especially in the case where $\mathcal{Y}$ consists of a finite set of actions, to rule out the possibility of indifference; in other words, for any (y, z) and (y', z) either $(y, z) \succ (y', z)$ or $(y', z) \succ (y, z)$. To characterize rationalizability by a strict SID preference would necessarily entail a strengthening of the RM axiom.

DEFINITION 3. *$\mathcal{O}$ obeys the* Strong Revealed Monotonicity (SRM) axiom *if, for every $s, t \in \mathcal{T}$,*

$$z^t \geqslant z^s,\ y^t < y^s, \ \textit{and}\ (y^s, z^s) \succsim^{RT} (y^t, z^s) \Longrightarrow (y^t, z^t) \not\succsim^{RT} (y^s, z^t).$$

Notice that this property strengthens the RM axiom by imposing the following additional condition: it excludes the possibility that $(y^s, z^s) \succsim^{RT} (y^t, z^s)$ and $(y^t, z^t) \succsim^{RT} (y^s, z^t)$ when $z^t = z^s$ and $y^t \neq y^s$. Clearly, this exclusion is needed for rationalizability by a strict preference. The following result is the analog to Theorem 1 and Proposition 2 for strict preferences.

[15] Note that there is a difference between our notion of 'representation' here and the textbook definition because $\succsim$ is not a complete binary relation on $\mathcal{X} \times \mathcal{Z}$: it need not compare elements with distinct z. So the representation requirement is also confined to elements (y', z') and (y, z) where $z' = z$.

THEOREM 2. *The following statements on $\mathcal{O} = \{(y^t, z^t, A^t)\}_{t\in\mathcal{T}}$ are equivalent:*

(a) *$\mathcal{O}$ is rationalizable by a strict preference that obeys SID in $(y; z)$.*

(b) *$\mathcal{O}$ obeys the SRM axiom.*

(c) *$\mathcal{O}$ is rationalizable by a strict preference that obeys SSCD in $(y; z)$.*

In the case where $\mathcal{Y}$ is a closed interval of $\mathbb{R}$, the strict SSCD preference in (c) *can be chosen to have a payoff representation.*

3.3 Cross-sectional data

So far we have assumed that the observer has access to panel data that gives the actions of the same agent across different environments. We now consider the case where data of this type is not available; instead, we only observe the distribution of actions taken by a population of agents, with possibly heterogeneous preferences, under different environments. It is possible to extend our revealed preference analysis to this stochastic setting, provided we assume that the distribution of preferences is the same in populations subject to different environments or, put another way, environments are assigned randomly. Throughout this section, we shall assume $\mathcal{Y}$ is finite, but bear in mind that the set of covariates can still be infinite.

The cross-sectional data consists of a finite set of environments and an associated distribution of choices for each of them. We can thus denote it as

$$\mathcal{P} = \left\{\mathrm{P}(\cdot|z^t, A^t) : t \in \mathcal{T} = \{1, 2, ..., T\}\right\}$$

where $\mathrm{P}(y|z^t, A^t)$ is the fraction of agents who choose action y in environment (z^t, A^t). It almost goes without saying that $\mathcal{P}$ is an idealized data set, in the sense that we assume that the distributions observed in $\mathcal{P}$ are true population distributions; in practice, we do not observe $\mathcal{P}$ but rather some sample which approximates $\mathcal{P}$. In this section, as well as the next two, we shall be abstracting from this specifically statistical issue and focus on idealized data sets. In the final, empirical section, finite sample issues will be addressed using the approach of Kitamura and Stoye (2016).

There is no loss of generality in assuming that all agents in the population have payoff functions (rather than just preferences), and so we shall present our results in that way, to make its connection with the empirical and econometric literature clearer. We introduce heterogeneity in the population via a term $\varepsilon \in \mathcal{E}$; one could think of ε simply as a parameterization of the payoff functions defined on $\mathcal{Y} \times \mathcal{Z}$. For a *type* ε, its payoff function is $\Pi(\cdot, \varepsilon) : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$. Each agent in the population knows its own type, but this is not directly observed by the econometrician. We denote a distribution on $\mathcal{E}$ by F. Since indifferences are allowed, the set $\mathrm{BR}(z, A, \varepsilon) = \arg\max_{y \in A} \Pi(y, z, \varepsilon)$ may be non-unique. A *selection rule* at $(\varepsilon, t) \in \mathcal{E} \times \mathcal{T}$ is a conditional probability on $\mathcal{Y}$ with support on $\mathrm{BR}(z^t, A^t, \varepsilon)$; specifically, $\lambda(y \mid \varepsilon, t)$ gives the probability that y is chosen by type ε at the observation t.

Definition 4. $\mathcal{P}$ is *stochastically rationalizable* (or, when there is no ambiguity, simply *rationalizable*) if there is $\mathcal{E}$, regular payoff functions $\Pi(\cdot, \varepsilon) : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$ for all $\varepsilon \in \mathcal{E}$, a selection mechanism λ, and a distribution F on $\mathcal{E}$ such that

$$\Pr\left(y|z^t, A^t\right) = \int \lambda(y \mid \varepsilon, t)\, \mathrm{dF}(\varepsilon) \text{ for each } t \in \mathcal{T}. \tag{8}$$

In other words, there is a distribution F on a family of payoff functions that accounts for the distribution of actions at each observed environment, assuming that agents choose from their best response set according to a selection rule.[16] This notion of rationalizability coincides with what Manski (2007) calls a *linear behavioral model.* The models of Marschak (1960) and McFadden and Richter (1991) are models of this type; see Manski (2007) for a discussion and other examples.

Our objective here is to formulate a test of rationalizability with payoff functions $\Pi(\cdot, \varepsilon)$ that obey SID in $(y; z)$. Before we do that we must introduce the notion of a *path*, which is a sequence of choices $y^* = \left(y^{*1}, y^{*2}, ..., y^{*T}\right)$ where $y^{*t} \in A^t$ for all $t \in \mathcal{T}$. A path y^* is said to obey the RM axiom if the induced data set $\{(y^{*t}, z^t, A^t)\}_{t \in \mathcal{T}}$ obeys the RM axiom. By Theorem 1, this guarantees that the induced data set can be rationalized by an SID (and indeed by an SSCD) preference. Since $\mathcal{Y}$ is finite, it is in principle possible to determine the entire set of paths that obey the RM axiom; we denote this set by $\mathcal{Y}^*$. The next result is a straightforward consequence of Theorem 1 and provides a way of testing whether or not a data set $\mathcal{P}$ is rationalizable by SID preferences.

Theorem 3. *$\mathcal{P}$ is rationalizable by payoff functions with the SID property in $(y; z)$ if and only if*

[16] Note that the selection mechanism can vary across observations, in the sense that $\lambda(y|\varepsilon, t)$ and $\lambda(y|\varepsilon, s)$ may not be equal, even when the environments are identical, i.e., $(y^t, A^t) = (y^s, A^s)$. In other words, our definition of rationalizability does not restrict which elements in the best response set are picked, and indeed allows for the possibility that the agent will choose differently when the same environment is repeated.

there exists a probability distribution Q *on* $\mathcal{Y}^*$ *such that*

$$\mathrm{P}(y|z^t, A^t) = \sum\nolimits_{y^* \in \mathcal{Y}^*} 1\left(y^{*t} = y\right) \mathrm{Q}\left(y^*\right) \quad \textit{for each } y \in A^t \textit{ and all } t \in \mathcal{T}. \tag{9}$$

When Q *exists, it is possible to choose* $\mathcal{E}$ *to be finite and the payoff functions to obey SSCD.*

Proof. Suppose $\mathcal{P}$ is rationalizable. Let $\mathrm{P}(y^*, \varepsilon)$ be the probability that an agent of type ε chooses the path y^* (in the sense that the agent chooses action y^{*t} at observation t). Note that $\mathrm{P}(y^*, \varepsilon) = \times_{t=1}^{T} \lambda(y^{*t}|\varepsilon, t)$. Since the agent ε has an SID payoff function, Theorem 1 guarantees that $\mathrm{P}(y^*, \varepsilon) > 0$ only if y^* obeys the RM axiom. Then the proportion of the population who choose the path y^* is $\int \mathrm{P}(y^*, \varepsilon)\mathrm{dF}(\varepsilon)$; if we set this as $\mathrm{Q}(y^*)$, it is clear that (9) holds.

Conversely, suppose Q exists that solves (9). We can list, in any particular order, the finite set of paths which have positive probability under Q. Let $\Pi(y, z, 1)$ be an SID payoff function that rationalizes the first path on the list, $\Pi(y, z, 2)$ an SID payoff function that rationalizes the second path, and so on. Then $\mathcal{E} = \{1, 2, \ldots\}$ is a finite set and let the distribution F assign a weight of $\mathrm{Q}(y^*_\varepsilon)$ to the type ε, where y^*_ε is the path rationalized by $\Pi(y, z, \varepsilon)$. Let $\lambda(y|\varepsilon, t) = 1$ if $y = y^{*t}_\varepsilon$, and let $\lambda(y|\varepsilon, t) = 0$ otherwise; in other words, the type ε chooses y^{*t}_ε at observation t with certainty. Then (9) guarantees that (10) holds. QED

This theorem sets out a procedure that allows us to determine whether a cross-sectional dataset is rationalizable by SID/SSCD payoff functions. First, we need to determine the set of all paths y^* that satisfy the RM axiom, and then we solve the linear equations given by (9). The implementability of this procedure in practice will depend crucially on the number of observed environments in the data and the ease with which we could work out the set $\mathcal{Y}^*$. Notice also that a solution to (9) is a distribution on $\mathcal{Y}^*$, and because (9) is a linear family of equations, *the collection of distributions on* $\mathcal{Y}^*$ *that solve this family form a convex set.* This set can be non-unique; when that occurs the distribution over payoff functions that rationalizes the data can also be non-unique. In other words, in this environment, the primitives of the model can typically only be partially identified.

In the special case where the feasible action sets are fixed across all observed environments, a path $y^* = (y^{*1}, y^{*2}, ..., y^{*T})$ obeys the RM axiom if and only if $y^{*t} \geqslant y^{*s}$ whenever $z^t > z^s$. In this case, it is clear that $\mathrm{P}(\cdot|z^t, A)$ will first order stochastically dominate $\mathrm{P}(\cdot|z^s, A)$ whenever

$z^t > z^s$, because every type in the population will be taking a weakly higher action. Less obviously, the converse is also true, so that monotonicity with respect to first order stochastic dominance *characterizes* rationalizability with SID preferences when the feasible action set is fixed.

THEOREM 4. *Suppose that $\mathcal{P}$ satisfies $A^t = A \subseteq \mathcal{Y}$ for all $t \in \mathcal{T}$. Then $\mathcal{P}$ is rationalizable by SID payoff functions if and only if* $\mathrm{P}(\cdot|z^t, A) \geqslant_{\mathrm{FSD}} \mathrm{P}(\cdot|z^s, A)$ *for all $s, t \in \mathcal{T}$ such that $z^t > z^s$.*

There are analogous versions of Theorems 3 and 4 for the case of rationalizability by strict SID preferences. In the case of the former, the result holds, provided we require Q to have its support on $\mathcal{Y}^{**}$, the set of paths obeying the SRM axiom. In the case of Theorem 4, the characterizing property requires, in addition, that $\mathrm{P}(\cdot|z^t, A) = \mathrm{P}(\cdot|z^s, A)$ for all s, $t \in \mathcal{T}$ such that $z^t = z^s$. We leave the reader to fill in the details.

Theorems 3 and 4 will be applicable in Section 5 (see, in particular, Theorem 8), where we study the rationalization of Bayesian Nash equilibria. In that context, a data set consists of observations of the distribution of actions of a player in different game environments, and the covariates will be (1) the player's observable characteristics and (2) some statistic of the actions of other players, such as their average action. (Thus the covariates will be at least two-dimensional.)

3.4 Related results

Topkis (1998, Theorem 2.8.9) considers a correspondence φ mapping elements of a totally ordered set (which can be interpreted as the set of covariates) to compact sublattices of $\mathbb{R}^\ell$. He shows that this correspondence is increasing in the strong set order 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 covariates. By 'exactly rationalized' we mean that the optimal choices at some value z of the covariate must coincide with (rather than simply contain) $\varphi(z)$. In the case where φ is a choice *function*, it is not hard to see that such a rationalization is possible even when the covariates form a partially (rather than totally) ordered set; this has been noted by Carvajal (2004) who applies this to a game setting.

Throughout this paper, we also permit the set of covariates to be partially rather than totally ordered; this is crucial in game theoretic applications, where the covariates will include the actions of other players in the game and thus will not generally be totally ordered. However, we confine ourselves to the case where actions are totally ordered (in essence, elements of $\mathbb{R}$ rather than $\mathbb{R}^\ell$), while allowing for observations of the choices made from different *subsets* of the set of all possible actions. Consequently, the observer may have partial information on the agent's ranking over different actions rather than simply the globally optimal action. In this respect, the problem we pose is more complicated than the one posed by Topkis, because the rationalizing preference we construct has to agree with this wider range of preference information (in addition to obeying single crossing differences).

Echenique and Komunjer (2009) considers a structural model where there could be multiple outcomes (which could be optimal choices made by an agent or equilibrium outcomes). It shows that a monotone relationship between the exogenous and dependent variables in the structural function leads to 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. The model we study in the previous subsection can also be interpreted as a random utility model; however, the rationalizability of a data set $\mathcal{P}$ does *not* require the population's preferences to be totally ordered by single crossing differences (i.e., in terms of our notation, we do not require $\Pi(y, z, \varepsilon)$ to have single crossing differences in $(y; \varepsilon)$). The single crossing property tested in the previous subsection is between actions and (observable) covariates, an issue which is not addressed in Apesteguia, Ballester, and Lu (2017).

There is a significant empirical literature on the endogenous adoption of (possibly) complementary technologies within the firm. Unlike our contribution, that literature is not always concerned with the recovery of payoff functions and, to the extent that it is, it uses parametric models (see, for example, Athey and Stern (1998)).

4 Monotone choice in games

This section extends the results in the previous section to analyze games of complete information. The first three subsections focus on games with strategic complements. The final subsection extends these ideas to games with strategic substitutes, etc.

4.1 Pure strategy Nash equilibrium

Let $\mathcal{N} = \{1, 2, ..., n\}$ be the set of agents in a complete information game, with $\mathcal{Y}_i$ (a closed subset of $\mathbb{R}$) being the set of all conceivable actions of agent i. Player i has a feasible action set A_i which is a compact interval of $\mathcal{Y}_i$, with the set of all such intervals denoted by $\mathbb{I}(\mathcal{Y}_i)$. Agent i's payoff over different feasible actions is affected by the actions of other players and also by an exogenous variable x_i, which is drawn from a partially ordered set $(\mathcal{X}_i, \geqslant)$. We refer to x_i as agent i's *observable characteristics* since it is observed by other players in the game and also by the econometrician. Note that if there are observable features of the game that affect *all* players, this could simply be folded into x_i for each i, since $\mathcal{X}_i$ can be multi-dimensional.

To connect this setup with the previous section, let $\mathcal{Z}_i = \mathbf{Y}_{-i} \times \mathcal{X}_i$ with $\mathbf{Y}_{-i} = \times_{j \neq i} \mathcal{Y}_j$. A typical element of $\mathcal{Z}_i$ is denoted by $z_i = (\mathbf{y}_{-i}, x_i)$ and $\mathcal{Z}_i$ is a partially ordered set if we endow it with the product order. Player i has a regular preference $\succsim_i$ on $\mathcal{Y}_i \times \mathcal{Z}_i$ (for all $i \in \mathcal{N}$). Given these preference, we denote by $\mathcal{G}(\mathbf{x}, \mathbf{A})$ the game arising when the joint feasible action set is $\mathbf{A} = \times_{i=1}^n A_i \in \mathbb{I}(\mathbf{Y})$ (where $\mathbb{I}(\mathbf{Y})$ denotes $\times_{i=1}^n \mathbb{I}(\mathcal{Y}_i)$) and the profile of observable characteristics is $\mathbf{x} \in \mathbf{X}$ (where $\mathbf{X}$ denotes $\times_{i=1}^n \mathcal{X}_i$). We refer to $(\mathbf{x}, \mathbf{A})$ as a *game environment*. We say that the family of games $\mathbb{G} = \{\mathcal{G}(\mathbf{x}, \mathbf{A})\}_{(\mathbf{x}, \mathbf{A}) \in \mathbf{X} \times \mathbb{I}(\mathbf{Y})}$ obeys *strategic complementarity* if, for every $\mathbf{A} \in \mathbb{I}(\mathbf{Y})$, the best response of each agent i (as given by (3)) is monotone in $z_i = (\mathbf{y}_{-i}, x_i)$. Games of strategic complementarity have a number of properties that make them particularly well-behaved; the most important for our purposes is that it always has a pure strategy Nash equilibrium (see Milgrom and Roberts, 1990). It follows immediately from the Basic Theorem that *the family of games* $\mathbb{G}$ *obeys strategic complementarity if and only if* $\succsim_i$ *is an SID preference for every agent* i.

Our objective is to develop revealed preference tests of the hypothesis that agents are playing games of strategic complementarity. We begin with the case where the observer has access to panel data, where the joint actions selected by the agents under a finite set of game environments are observed. Specifically, at observation t, the players choose an action profile $\mathbf{y}^t$ from the feasible action set $\mathbf{A}^t$, when the covariates are $\mathbf{x}^t$. Thus the data set is

$$\mathcal{O} = \left\{ \left(\mathbf{y}^t, \mathbf{x}^t, \mathbf{A}^t \right) : t \in \mathcal{T} = \{1, 2, ..., T\} \right\}.$$

Definition 5. *$\mathcal{O}$ is rationalizable as pure strategy Nash equilibria (PSNE) if, for each agent $i \in \mathcal{N}$,* *there is a regular preference $\succsim_i$ on $\mathcal{Y}_i \times \mathbf{Y}_{-i} \times \mathcal{X}_i$ such that, at each $t \in \mathcal{T}$, we have $(y_i^t, \mathbf{y}_{-i}^t, x_i^t) \succsim_i \left(y_i, \mathbf{y}_{-i}^t, x_i^t\right)$ for every $y_i \in A_i^t$.*

If $\mathcal{O}$ is rationalized as PSNE by a profile of preferences $\{\succsim_i\}_{i \in N}$ such that $\succsim_i$ obeys SID in $(y_i; z_i)$ where $z_i = (\mathbf{y}_{-i}, x_i)$ then the Basic Theorem guarantees that the resulting family of games $\mathbb{G}$ obeys strategic complementarity. Characterizing data sets with this rationalizability property is straightforward given the results of the previous section. By letting $z_i^t = (\mathbf{y}_{-i}^t, x_i^t)$, we can construct from $\mathcal{O}$ a 'personalized' dataset $\mathcal{O}_i = \{(y_i^t, z_i^t, A_i^t)\}_{t \in \mathcal{T}}$ for each agent i. The following result follows immediately from Theorem 1.

Theorem 5. *$\mathcal{O} = \{(\mathbf{y}^t, \mathbf{x}^t, \mathbf{A}^t)\}_{t \in \mathcal{T}}$ is rationalized as PSNE by $\{\succsim_i\}_{i \in \mathcal{N}}$, such that $\succsim_i$ obeys SID in $(y_i; z_i = (\mathbf{y}_{-i}, x_i))$ if and only if $\mathcal{O}_i$ obeys the RM axiom for each $i \in \mathcal{N}$.*

Example 3. Suppose that $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$, $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, and that $\mathcal{O}$ has two observations:
at observation 1, $y_1^1 = y_2^1 = 1$, $x_1^1 = x_2^1 = 0$, and $A_1^1 = A_2^1 = \{0, 1\}$; and
at observation 2, $y_1^2 = y_2^2 = 0$, $x_1^2 = x_2^2 = 1$, and $A_1^2 = A_2^2 = \{0, 1\}$.
In the case of a single agent and when the set of feasible actions is fixed across observations, the RM axiom is equivalent to the co-monotonicity of the observed action with the covariates. Clearly, co-monotonicity is violated here since the observed action profile *falls* from $(1, 1)$ to $(0, 0)$, as the covariates increase from $(0, 0)$ to $(1, 1)$. Nonetheless this data set *is* rationalizable as PSNE by SID preferences. By Theorem 5, it suffices to show that $\mathcal{O}_1$ and $\mathcal{O}_2$ obey the RM axiom. The two observations of $\mathcal{O}_1$ are the following:
at observation 1, $y_1^1 = 1$, $z_1^1 = (1, 0)$, with $A_1^1 = \{0, 1\}$; and
at observation 2, $y_1^2 = 0$, $z_1^2 = (0, 1)$, with $A_1^2 = \{0, 1\}$.
$\mathcal{O}_1$ trivially obeys the RM axiom (since z_1^1 and z_1^2 are not comparable), and similarly so does $\mathcal{O}_2$. More directly, it is straightforward to check that the following SID preferences rationalize the data: for agent 1, $(0, 0, x_1) \succ_1 (1, 0, x_1)$ and $(1, 1, x_1) \succ_1 (0, 1, x_1)$ for any x_1; in other words, agent 1 prefers coordinating on the same action with the other player irrespective of the exogenous variable. Similarly, for agent 2, $(0, 0, x_2) \succ_2 (1, 0, x_2)$ and $(1, 1, x_2) \succ_2 (0, 1, x_2)$. With these preferences, $(0, 0)$ and $(1, 1)$ are Nash equilibria when the covariates are (0,0) and also at (1,1).

4.2 Cross sectional data sets

We now extend our analysis to the case where we observe cross sectional data. Specifically, the data consists of a finite set of game environments and at each environment we observe the distribution of action profiles taken by a population of groups. Each group consists of n players, with a player in *role* i, for $i = 1, 2, \ldots, n$. As in Section 3.3, we now specialize to the case where $\mathcal{Y}_i$ is finite, for all i (while $\mathcal{X}_i$ may still be infinite). A typical data set can thus be denoted as

$$\mathcal{P} = \left\{ \mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t) : t \in \mathcal{T} = \{1, 2, ..., T\} \right\}$$

where $\mathrm{P}(\mathbf{y}|\mathbf{x}^t, \mathbf{A}^t)$ is the fraction of groups in the population who play the action profile $\mathbf{y}$ in game environment $(\mathbf{x}^t, \mathbf{A}^t)$.

Given that $\mathcal{Y}_i$ is finite, there is no loss of generality in assuming that an agent in role i has a payoff function (rather than just a preference). Heterogeneity among agents in this role is captured by a term $\varepsilon_i \in \mathcal{E}_i$, which parameterizes the payoff functions. For a *type* ε_i, we denote its payoff when it chooses action y_i, given that others are playing $\mathbf{y}_{-i}$, and the observable characteristics are x_i by $\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i)$. We refer to $\varepsilon = \times_{i=1}^n \varepsilon_i$ as a *group type* and denote the set of group types by $\mathbf{E} = \times_{i=1}^n \mathcal{E}_i$. Given a group type ε and a game environment $(\mathbf{x}, \mathbf{A})$, let $\mathrm{NE}(\mathbf{x}, \mathbf{A}, \varepsilon)$ be the corresponding set of PSNE; note that we are assuming that the players in a group know each other's type and are playing a game of complete information. A *selection rule* at $(\varepsilon, t) \in \mathbf{E} \times \mathcal{T}$ is a conditional probability on $\mathbf{Y}$ with support on $\mathrm{NE}(\mathbf{x}^t, \mathbf{A}^t, \varepsilon)$; specifically, $\lambda(\mathbf{y} \mid \varepsilon, t)$ gives the probability that the profile $\mathbf{y}$ is played by the group type ε at the observation t.

Definition 6. $\mathcal{P}$ is *stochastically rationalizable* (or simply *rationalizable*) if there is $\mathcal{E}_i$ and payoff functions $\Pi_i : (\cdot, \varepsilon_i) \to \mathbb{R}$ for each role i and type $\varepsilon_i \in \mathcal{E}_i$, a selection mechanism λ, and a distribution F on the set of group types $\mathbf{E} = \times_{i=1}^n \mathcal{E}_i$ such that

$$\Pr\left(\mathbf{y}|\mathbf{x}^t, \mathbf{A}^t\right) = \int \lambda(\mathbf{y} \mid \varepsilon, t)\, \mathrm{dF}(\varepsilon) \text{ for each } t \in \mathcal{T}. \tag{10}$$

In other words, the distribution of action profiles at each observation can be accounted for by a distribution F over group types (which does not vary across observations) and a selection rule over the PSNE of each group type. It is worth emphasizing that since we impose no restrictions at all on F, we are allowing for the possibility that groups are *not* formed randomly. For example, in two-player games, the probability that a role-1 player belongs to a particular type *can* depend on the type of role-2 player to which that role-1 player is paired.

We would like to formulate a test for the rationalizability of $\mathcal{P}$ with payoff functions that obey the SID property, in the sense that we require $\Pi_i(\cdot, \varepsilon_i)$ to obey SID in $(y_i; (\mathbf{y}_{-i}, x_i))$, for every role i and every type ε_i. The test provided by Theorem 3 for single-agent decision problems could be naturally extended to a game-theoretic setting.

We refer to a sequence of joint action profiles $\mathbf{y}^* = \left(\mathbf{y}^{*1}, \mathbf{y}^{*2}, ..., \mathbf{y}^{*T}\right)$, where $\mathbf{y}^{*t} \in \mathbf{A}^t$ for all $t \in \mathcal{T}$, as a *path*. Abusing terminology a little, we say that this path obeys the RM axiom if, for every i, the induced panel data set $\mathcal{O}_i = \{(y_i^t, z_i^t, A_i^t)\}_{t \in \mathcal{T}}$ (with $z_i^t = (\mathbf{y}_{-i}^t, x_i^t)$) obeys the RM axiom. We denote the set of paths that obey the RM axiom by $\mathbf{Y}^*$. Since $\mathbf{Y}$ is finite, it is in principle possible to find all the paths in $\mathbf{Y}^*$. Note that, by Theorem 5, if a path $\mathbf{y}^*$ obeys the RM axiom then $\{(\mathbf{y}^{*t}, \mathbf{x}^{*t}, \mathbf{A}^t)\}_{t \in \mathcal{T}}$ is rationalizable as PSNE with SID payoff functions. The next result is analogous to Theorem 3 and has a similar proof (which we shall omit).

THEOREM 6. *$\mathcal{P}$ is rationalizable with payoff functions that obey the SID property if and only if there exists a probability distribution* Q *on* $\mathbf{Y}^*$ *such that*

$$\mathrm{P}(\mathbf{y}|\mathbf{x}^t, \mathbf{A}^t) = \sum\nolimits_{\mathbf{y}^* \in \mathbf{Y}^*} 1\left(\mathbf{y}^{*t} = \mathbf{y}\right) \mathrm{Q}\left(\mathbf{y}^*\right) \quad \textit{for each } \mathbf{y} \in \mathbf{A}^t \textit{ and all } t \in \mathcal{T}. \tag{11}$$

When Q *exists, it is possible to choose the set* $\mathbf{E}$ *to be finite and the payoff functions to obey SSCD.*

There is an analogous version of Theorem 6 for the case of rationalizability by strict SID preferences: it is necessary and sufficient for Q to have its support on $\mathbf{Y}^{**}$, the set of paths that obey the SRM axiom, by which we mean that, for every i, the induced panel data set $\mathcal{O}_i = \{(y_i^t, z_i^t, A_i^t)\}_{t \in \mathcal{T}}$ (with $z_i^t = (\mathbf{y}_{-i}^t, x_i^t)$) obeys the SRM axiom.

In Section 2, we gave an example in Table 2 of a data set $\mathcal{P}$ with three observations and tested whether it is SID-rationalizable (with the order of Firm 2's strategies reversed). The paths listed in Table 3 obey the SRM axiom and a distribution Q that satisfies (11) is provided by the second column in that table.

In Section 6, we illustrate the use of Theorem 6 by applying it to data on the entry behavior of airlines. Readers who are keen to scrutinize the results of that application can skip to Section 6 immediately after reading Section 4.3. In fact there are many other potential empirical applications of our approach. In a related note, we apply Theorem 6 to test for the presence of spousal influence in smoking decisions (Lazzati, Quah, and Shirai, 2018).

4.3 Inference and out-of-sample predictions

Theorem 6 tells us that a data set $\mathcal{P}$ is rationalizable by payoff functions that obey SID if and only if there is a distribution Q over $\mathbf{Y}^*$ that solves (11). Beyond testing the model, we may wish to know more about the set Q, and thus about the types that form the population generating $\mathcal{P}$. To be specific, suppose we are interested in the proportion of the population with paths belonging to some set $\mathbf{Y}' \subset \mathbf{Y}^*$. (For examples of this type see the application in Section 6.) Since the set of distributions Q that solve (11) will typically be a non-singleton convex set, this proportion cannot be predicted uniquely. The largest (smallest) proportion which is data-consistent could be obtained by maximizing (minimizing) the linear objective $\sum_{\mathbf{y}^* \in \mathbf{Y}'} Q(\mathbf{y}^*)$ subject to (11). We could then conclude that the true proportion of the population with paths in $\mathbf{Y}'$ can lie anywhere between these two numbers, since the solution to (11) is convex.

Another natural exercise is to make out-of-sample predictions. Suppose $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)\}_{t \in \mathcal{T}}$ is rationalizable with SID payoff functions and we wish to predict the distribution of outcomes at some game environment $(\mathbf{x}^0, \mathbf{A}^0)$ (while maintaining the SID property on payoff functions). In other words, we would like to identify those distributions $\mathrm{P}(\cdot|\mathbf{x}^0, \mathbf{A}^0)$ (with their support on $\mathbf{A}^0$) such that the augmented stochastic data set $\mathcal{P} \cup \{\mathrm{P}(\cdot|\mathbf{y}^0, \mathbf{A}^0)\}$ is rationalizable by payoff functions that obey SID. Let $\mathbb{P}(\mathbf{x}^0, \mathbf{A}^0)$ be the *set* of these distributions; this set can be obtained using the following procedure.

Let $\phi : \mathbf{Y}^* \to \mathbf{A}^0$ be any map such that the path $(\phi(\mathbf{y}^*), \mathbf{y}^*)$, on the game environments $\{(\mathbf{x}^0, \mathbf{A}^0)\} \cup \{(\mathbf{x}^t, \mathbf{A}^t)\}_{t \in \mathcal{T}}$ obeys the RM axiom. It follows from Theorem 6 that $\mathrm{P}(\cdot|\mathbf{x}^0, \mathbf{A}^0)$ is in $\mathbb{P}(\mathbf{x}^0, \mathbf{A}^0)$ if and only if there is ϕ and a distribution Q on $\mathbf{Y}^*$ such that (11) holds and

$$\mathrm{P}(\mathbf{y}|\mathbf{x}^0, \mathbf{A}^0) = \sum\nolimits_{\mathbf{y}^* \in \mathbf{Y}^*} 1\left(\phi(\mathbf{y}^*) = \mathbf{y}\right) \mathrm{Q}\left(\mathbf{y}^*\right) \quad \text{for each } \mathbf{y} \in \mathbf{A}^0. \tag{12}$$

Note that $\mathrm{P}(\cdot|\mathbf{x}^0, \mathbf{A}^0)$ *exists* because both Q and ϕ exists, the former because $\mathcal{P}$ is rationalizable with SID payoff functions and the latter because any game of strategic complementarity has PSNE.[17]

4.4 Other classes of games

The tests we have outlined for games of strategic complementarity could potentially be applied to classes of games that require other monotone properties on a player's strategic choice. It could certainly 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 of one of the two players. It could also be applied to games of strategic substitutes more generally: instead of checking whether a player i has SID preferences in $(y_i; (\mathbf{y}_{-i}, x_i))$, we could check whether the player has SID preferences in $(y_i; (-\mathbf{y}_{-i}, x_i))$, which will guarantee that this player's action *decreases* with the actions of other players. There is, however, a caveat: while games with strategic complementarity always have pure strategy Nash equilibria, this is not true for games of strategic substitutes. So while it is possible to test whether a given set of observations is consistent with PSNE in the latter class of games, one may argue that the hypothesis itself is less plausible since there is no general PSNE existence result for these games. Furthermore, even if preferences could be found that rationalizes a data set as PSNE in games of strategic substitutes, these preferences may not guarantee the existence of PSNE outside the set of observed environments $\{(\mathbf{x}^t, \mathbf{A}^t)\}_{t\in\mathcal{T}}$; in other words, there may be difficulty in making out-of-sample PSNE predictions. The bottom line is that when we are testing a hypothesis that a set of observations consists of PSNE of games belonging to a particular class, it would be desirable for that class of games to have PSNE in general.

Another natural variation on the tests we have developed involves making stronger hypotheses on the way actions taken by other players impact a player's payoff. For example, in the entry model discussed in Section 2, we may hypothesize that firm i cares about the number of other firms that enter the market, irrespective of their identity. If the firm's payoff function has the linear form given by (1), this is equivalent to assuming that δ_{ij} is the same across all j. The analogous condition in our nonparametric setting is to require the SID property in $(y_i; (\sum_{j\neq i} y_i, x_i))$ (in the case of strategic complements) or the SID property in $(y_i; (-\sum_{j\neq i} y_i, x_i))$ (in the case of strategic

[17]Since $\mathbf{y}^*$ obeys the RM axiom, Theorem 5 guarantees that it admits a rationalization with SID preferences, and these preferences generate a nonempty set of PSNE in the game environment $(\mathbf{x}^0, \mathbf{A}^0)$ (Milgrom and Roberts, 1990). Let $\phi(\mathbf{y}^*)$ be any PSNE in this environment; then $(\phi(\mathbf{y}^*), \mathbf{y}^*)$ obeys the RM axiom (again by Theorem 5).

substitutes). Theorem 6 could be straightforwardly adapted to test such a hypothesis.

5 Bayesian Nash Equilibrium

In Section 4.2, we consider an idealized cross sectional data set of the form $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)\}_{t\in\mathcal{T}}$ and find necessary and sufficient conditions under which it could be rationalized by a population of groups playing complete-information games with strategic complementarity. In this section, we consider a data set of the same form and investigate its rationalizability as Bayesian Nash equilibria.

5.1 Payoffs and game structure

The players $\mathcal{N} = \{1, 2, \ldots, n\}$ play a simultaneous game. Each player i has a preference over its own actions that depends on its observable characteristic x_i and on its type ε_i. Unlike Section 4, ε_i is now *private information* and unknown to the other players. We assume that this preference is also affected by the player's belief on the actions of other players; we allow this belief to be non-deterministic, so it takes the form of a distribution ρ on $\mathbf{Y}_{-i}$. The impact of this belief on the agent's preference is captured by the expected value of some function $g_i : \mathbf{Y}_{-i} \times \mathcal{X}_i \to \mathbb{R}^{n_i}$; note that we have not excluded the possibility that the player's preference depends directly on $\mathbf{y}_{-i}$ since we can always set $g_i(\mathbf{y}_{-i}, x_i) = \mathbf{y}_{-i}$ but this formulation allows us to build in simpler forms of interaction, for example, by setting $g_i(\mathbf{y}_{-i}, x_i) = \sum_{j\neq i} y_j$. The function g_i is known to the players in the game and also to the econometrician; crucially, this means that, if ρ is the distribution on $\mathbf{y}_{-i}$ observed in the data set, then the expected value of g_i given ρ, i.e., $\mathbb{E}_\rho(g_i|x_i) = \int g_i(\mathbf{y}_{-i}, x_i)d\rho(\mathbf{y}_{-i})$, is also known to the econometrician.

Let $\mathcal{B}_i$ be the convex hull of the image of g_i. We assume that a player i of type $\varepsilon_i \in \mathcal{E}_i$ has a payoff function $\Pi_i(\cdot, \varepsilon_i) : \mathcal{Y}_i \times \mathcal{B}_i \times \mathcal{X}_i \to \mathbb{R}$ such that if the player's belief over $\mathbf{Y}_{-i}$ is given by ρ, then the payoff of action y_i when the observable characteristic is x_i is

$$\Pi_i(y_i, \mathbb{E}_\rho(g_i|x_i), x_i, \varepsilon_i). \tag{13}$$

In other words, the player's payoff can be captured by a function defined on the expected value of g_i. This is a nonstandard assumption: if $\Pi_i(y_i, g_i(\mathbf{y}_{-i}, x_i), x_i, \varepsilon_i)$ is the (ex post) payoff to player

i when other players are playing $\mathbf{y}_{-i}$ then the payoff when he has a belief ρ on $\mathbf{y}_{-i}$ is typically understood to be the expected value, which is $\int_{\mathbf{y}_{-i}\in\mathbf{Y}_{-i}} \Pi_i(y_i, g_i(\mathbf{y}_{-i}, x_i), x_i, \varepsilon_i)\, d\rho(\mathbf{y}_{-i})$ and not (13). However, many of the payoff functions assumed in the estimation of Bayesian games do have the property that the expected value equals (13).[18] This is obviously true for the payoff functions in the entry model discussed in Section 2, where $g_i(\mathbf{y}_{-i}, x_i) = \mathbf{y}_{-i}$ (see (1)). Another instance where this property obviously holds is the following payoff function

$$\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i) = a_i(y_i, x_i, \varepsilon_i) + \sum_{k=1}^{m_i} c_i^k(y_i, x_i, \varepsilon_i)\, g_i^k(\mathbf{y}_{-i}, x_i),$$

where a_i, c_i^k and g_i^k are real-valued functions; this is a generalization of the payoff function used in Aradillas-Lopez and Gandhi (2016) where it is assumed that

$$\Pi_i(y_i, \mathbf{y}_{-i}, x_i, \varepsilon_i) = a_i(y_i, \varepsilon_i) + c_i(y_i, \varepsilon_i)\, g_i(\mathbf{y}_{-i}, x_i),$$

with a_i, c_i and g_i being real-valued functions, and a_i and c_i independent of x_i.

We assume that the distribution of types for player i obeys the *independence property*, by which we mean that the distribution of ε_i can depend on some of player i's observable characteristics, but cannot depend on the characteristics of other players, or on their types. Specifically, we suppose that $\mathcal{X}_i = \tilde{\mathcal{X}}_i \times \bar{\mathcal{X}}_i$, with a typical element having the form $x_i = (\tilde{x}_i, \bar{x}_i)$ and allow the distribution of ε_i to depend on $\tilde{x}_i$, but to be independent of $(\bar{x}_i, \mathbf{x}_{-i}, \varepsilon_{-i})$. Note that we allow for the possibility that $\bar{x}_i$ is absent (the case where all of i's characteristics will affect the distribution of ε_i) and also the case where $\tilde{x}_i$ is absent (the case where none of i's characteristics affect the distribution of ε_i). That a player's private type is independent of the type of other players is widely assumed in the econometric literature on Bayesian games; in fact most estimation strategies also assume that ε_i is additively separable from the non-stochastic component of the payoff function and to belong to a *known* parametric family.

We denote the distribution of ε_i by $\mathrm{F}_i(\cdot, \tilde{x}_i) : \mathcal{E}_i \to \mathbb{R}$. The agent i chooses an action from its feasible action set A_i after observing ε_i, so its strategy is a map $\sigma_i : \mathcal{E}_i \to A_i$. Given the strategies

[18] An exception is Bajari, Hong, Krainer, and Nekipelov (2010), where the non-stochastic part of the payoff function is nonparametric and may not satisfy the property we require. In that model, however, the private type is additively separable and has a known distribution, requirements which we do not impose.

of other players, σ_{-i}, the distribution of their types $\mathrm{F}_{-i}(\cdot, \tilde{\mathbf{x}}_{-i})$, the payoff of an agent i of type ε_i is $\Pi_i(y_i, \mathbb{E}_{\mathrm{F}_{-i}}(g_i(\sigma_{-i})|x_i), x_i, \varepsilon_i)$ where

$$\mathbb{E}_{\mathrm{F}_{-i}}(g_i(\sigma_{-i})|x_i) = \int_{\mathbf{E}_{-i}} g_i([\sigma_j(\varepsilon_j)]_{j\neq i}\,,\, x_i) d\mathrm{F}_{-i}(\varepsilon_{-i}, \tilde{\mathbf{x}}_{-i}).$$

A profile of strategies $(\hat{\sigma}_i)_{i\in N}$ forms a *Bayesian Nash equilibrium* (BNE) in the game environment (x, A) if, whenever ε_i is in the support of $\mathrm{F}_i(\cdot, \tilde{x}_i)$,

$$\Pi_i(\hat{\sigma}(\varepsilon_i), \mathbb{E}_{\mathrm{F}_{-i}}(g_i(\hat{\sigma}_{-i})), x_i, \varepsilon_i) \geqslant \Pi_i(y_i, \mathbb{E}_{\mathrm{F}_{-i}}(g_i(\hat{\sigma}_{-i})), x_i, \varepsilon_i) \text{ for all } y_i \in A_i. \tag{14}$$

Suppose g_i is an increasing function and $\Pi_i(y_i, b_i, x_i, \varepsilon_i)$ has the SID property in $(y_i; b_i)$ (for all $i \in N$). Let us consider what happens when σ_j increases pointwise (for all $j \neq i$), in the sense that player j's action at every type ε_j is higher. Then b_i increases because g_i is increasing and this in turn leads to player i's optimal action at every type ε_i to increase, since Π_i has the SID property. In other words, player i's optimal strategy σ_i also increases pointwise. Thus the strategies of the players are strategic complements and the existence of a Bayesian Nash equilibium is guaranteed (Vives, 1990).

5.2 Rationalization

Definition 7. $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)\}_{t\in\mathcal{T}}$ is rationalizable as a BNE if, for each player i, there is a set of types $\mathcal{E}_i$, distributions $\mathrm{F}_i(\cdot, \tilde{x}_i^t)$ on $\mathcal{E}_i$, and payoff functions $\Pi_i(\cdot, \varepsilon_i) : \mathcal{Y}_i \times \mathcal{B}_i \times \mathcal{X}_i \to \mathbb{R}$ (for each type $\varepsilon_i \in \mathcal{E}_i$) such that in each game environment $(\mathbf{x}^t, \mathbf{A}^t)$, there is a BNE $(\hat{\sigma}_i^t)_{i\in N}$ that leads to $\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)$ as the distribution on the action profiles $\mathbf{A}^t$.

Our objective is to characterize those data sets which are rationalizable as BNE with payoff functions obeying SID. As an example of what we have in mind, suppose that the players in the game are radio stations, whose strategy consists of choosing the time during a particular hour at which to have a commercial break (as in Sweeting (2009)). For a given station, the optimal timing of a break depends on observable factors, such as the choices made by rival radio stations, but also on idiosyncratic causes unobserved by other stations; for example, on any particular day, a station must fit its commercial breaks around other pieces of programming. For this reason, we may think of radio stations as playing a game of incomplete information. Each observation t corresponds to a radio market (corresponding to particular geographically area), with $\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)$ being the joint distribution of commercial break times (at different stations) observed in that market.

Note that, in practice, obtaining the distribution $\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)$ could involve sampling the outcomes of the same game played many times. Our definition of rationalizability implicitly posits that the data collected to obtain $\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)$ are realizations from the *same* BNE; consequently, this feature is also part of the test we formulate. The estimation literature has addressed this issue in different ways. There are models where the primitives are such that multiplicity is excluded (for example, Seim (2006)), so the danger of sampling from different BNE of the same game does not arise; in other cases, such as Sweeting (2009), the estimation strategy assumes that a single equilibrium is played at each market, even if it allows for the possibility that different equilibria are played at different markets which have the same observable characteristics (and so players are playing the same game). Indeed, Sweeting (2009) emphasizes the usefulness for estimation of multiple equilibria. In terms of our notation, Sweeting (2009) allows for the possibility that $\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t) \neq \mathrm{P}(\cdot|\mathbf{x}^s, \mathbf{A}^s)$, even though $(\mathbf{x}^t, \mathbf{A}^t) = (\mathbf{x}^s, \mathbf{A}^s)$. Our definition and test of rationalizability allows for this possibility as well.

It may be implausible in certain applications to assume that the realized actions observed are from a single BNE. This corresponds, in our setup, to the case where the distribution $\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)$ is obtained from a mixture of several BNE of the same game. Our rationalization concept does not allow for that possibility, and we do not formulate a test for this case. Note that a distribution obtained from such a mixture is precisely what is considered in De Paula and Tang (2012).[19] However, their objective is not to rationalize a data set as BNE with SID preferences, but rather to determine the direction of interaction effects under the assumption that BNE are played; interestingly, the efficacy of their test hinges on the presence of multiple BNE.

5.3 BNE rationalizability with SID payoff functions

Let us suppose that $\mathcal{P}$ is rationalizable as a BNE. Then given our assumption that players' types are independent from each other, the distribution of players' actions must also be independent.

[19]See also the extension of Aradillas-Lopez and Gandhi (2016).

Consequently, for any $\mathbf{y}' \in \mathbf{Y}$ and at any observation t,

$$\mathrm{P}(\mathbf{y}'|\mathbf{x}^t, \mathbf{A}^t) = \times_{i=1}^{N} \mathrm{P}_i(y_i'|\mathbf{x}^t, \mathbf{A}^t),$$

where $\mathrm{P}_i(y_i'|\mathbf{x}^t, \mathbf{A}^t) = \sum_{y_{-i} \in \mathbf{Y}_{-i}} \mathrm{P}(y_i', \mathbf{y}_{-i}|\mathbf{x}^t, \mathbf{A}^t)$ is the probability of the agent i choosing action y_i'.

Notice also that, given the form of agent i's payoff function, the actions of other players only affect player i's payoff via the average value of g_i. The independence property guarantees that this value is independent of the realization of ε_i and given by

$$b_i^t = \mathbb{E}_{\mathrm{F}_{-i}}(g_i(\hat{\sigma}_{-i}^t)|x_i^t) = \sum_{\mathbf{y}_{-i} \in \mathbf{Y}_{-i}} g_i(\mathbf{y}_{-i}, x^t) \times \mathrm{P}_{-i}(\mathbf{y}_{-i}|\mathbf{x}^t, \mathbf{A}^t) \tag{15}$$

where $\mathrm{P}_{-i}(\mathbf{y}_{-i}|\mathbf{x}^t, \mathbf{A}^t) = \times_{j \neq i} \mathrm{P}_j(y_j|\mathbf{x}^t, \mathbf{A}^t)$ is the observed probability of $\mathbf{y}_{-i}$ at observation t. It follows from (15) that b_i^t is known to the econometrician. Somewhat abusing our notation, we shall now write $\mathrm{P}_i(y_i'|\mathbf{x}^t, \mathbf{A}^t)$, player i's observed probability of choosing action y_i', as $\mathrm{P}_i(y_i'|b_i^t, x_i^t, A_i^t)$, in order to highlight precisely those factors which have an impact on the player's action.

To rationalize $\mathcal{P}$ it is necessary and sufficient to rationalize the actions of each i, given the player's choice environment, as summarized by (b_i^t, x_i^t). In other words, we need to rationalize

$$\mathcal{P}_i = \{\mathrm{P}_i(\cdot|b_i^t, x_i^t, A_i^t) : t \in \mathcal{T}\} \tag{16}$$

in the sense of finding $\mathcal{E}_i$, distributions $\mathrm{F}_i(\cdot, \tilde{x}_i^t)$ on $\mathcal{E}_i$, payoff functions $\Pi_i(\cdot, \varepsilon_i) : \mathcal{Y}_i \times \mathcal{B}_i \times \mathcal{X}_i \to \mathbb{R}$ (for each type $\varepsilon_i \in \mathcal{E}_i$), and strategies $\hat{\sigma}_i^t : \mathcal{E}_i \to A_i^t$ at each observation t such that

(i) $\Pi_i(\hat{\sigma}_i^t(\varepsilon_i), b_i^t, x_i^t, \varepsilon_i) \geqslant \Pi_i(y_i, b_i^t, x_i^t, \varepsilon_i)$ for all $y_i \in A_i^t$ and ε_i in the support of $F_i(\cdot, \tilde{x}_i^t)$;

(ii) $\mathrm{P}_i(y_i|b_i^t, x_i^t, A_i^t) = \int_{\mathcal{E}_i} \mathbf{1}_{(\hat{\sigma}^t)_i^{-1}(y_i)} d\mathrm{F}_i(\varepsilon_i, \tilde{x}_i^t)$, where $(\hat{\sigma}^t)_i^{-1}(y_i) = \{\varepsilon_i \in \mathcal{E}_i : \hat{\sigma}_i^t(\varepsilon_i) = y_i\}$.

Condition (i) states that $\hat{\sigma}_i^t$ is an optimal response while (ii) guarantees that it generates a distribution over actions coinciding with the one observed.

Notice that the problem of rationalizing $\mathcal{P}_i$ with SID payoff functions is virtually identical to the one discussed in Section 3.3 and solved in Theorem 3, with the exception that while the distribution of types in that case is held fixed across all observations, we allow it to vary with $\tilde{x}_i$ in this case.

Since we can always choose the support of $\mathrm{F}_i(\cdot, \tilde{x}_i^t)$ and $\mathrm{F}_i(\cdot, \tilde{x}_i^s)$ to be non-overlapping whenever $\tilde{x}_i^t \neq \tilde{x}_i^s$, there are effectively no restrictions imposed by the rationalizability of $\mathcal{P}_i$ on $\mathrm{P}_i(\cdot|b_i^t, x_i^t, A_i^t)$ and $\mathrm{P}_i(\cdot|b_i^s, x_i^s, A_i^s)$ *jointly* when $\tilde{x}_i^t$ and $\tilde{x}_i^s$ are distinct. However, restrictions of the type imposed by Theorem 3 must hold (and are also sufficient) whenever $\tilde{x}_i^t = \tilde{x}_i^s$ since the rationalizability of $\mathcal{P}_i$ requires the distribution of types to be the same in these two case. To state this more formally, we partition $\mathcal{T}$ into $\mathcal{T}_{i1}$, $\mathcal{T}_{i2}$,..., $\mathcal{T}_{in(i)}$, separating observations which have distinct $\tilde{x}_i$. (Thus, for any two observations t and s in $\mathcal{T}_{ik}$, we have $\tilde{x}_i^t = \tilde{x}_i^s$.) Let

$$\mathcal{P}_{ik} = \{\mathrm{P}_i(\cdot|b_i^t, x_i^t, A_i^t) : t \in \mathcal{T}_{ik}\} \tag{17}$$

for each i and $k = 1, 2, \ldots, n(i)$. The following result summarizes our observations.

Theorem 7. *$\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)\}_{t \in \mathcal{T}}$ is rationalizable as a BNE with independent types and payoff functions of the form (13) obeying SID in $(y_i; b_i, \bar{x}_i)$ if and only if, for each $i \in N$ and $k = 1, 2, \ldots, n(i)$, the set $\mathcal{P}_{ik}$ (as defined by (17)) passes the test specified in Theorem 3.*

There are a number of features of this result that are worth remarking on. Firstly, suppose $\mathcal{X}_i = \bar{\mathcal{X}}_i$, which means that we require the distribution of ε_i to be completely independent of i's observable characteristics. In that case, there is no nontrivial partition of $\mathcal{P}_i$, and so the rationalizability condition in Theorem 7 is simply that $\mathcal{P}_i$ (as defined by (16)) passes the test specified in Theorem 3 (for each player i).

On the other hand, if we wish to allow the type distribution to be partially dependent on observable characteristics, then there will typically be a nontrivial partition of $\mathcal{P}_i$. But notice that for the test specified in Theorem 7 to be non-vacuous, it is crucial that there is some $\mathcal{P}_{ik}$ containing more than one element. In other words, there must be observations t and s where $\tilde{x}_i^t = \tilde{x}_i^s$, even as other aspects of the game environment differ between these observations. This can be thought of as a version of the familiar exclusion assumptions found extensively in the estimation literature.

Since the rationalizability of $\mathcal{P}$ imposes no relationship *between* $\mathrm{F}_i(\cdot, \tilde{x}_i^t)$ and $\mathrm{F}_i(\cdot, \tilde{x}_i^s)$ whenever $\tilde{x}_i^t \neq \tilde{x}_i^s$, it is always possible to choose their supports to be completely non-overlapping. Essentially because of this, whenever Π_i can be chosen to obey SID in $(y_i; (b_i, \bar{x}_i))$ then it can be chosen to obey SID in $(y_i; (b_i, \tilde{x}_i, \bar{x}_i))$; in other words, the latter (seemingly stronger) condition in fact imposes no further restrictions on the data. We leave the reader to fill in the details.

Lastly, in the case where the constraint sets for each player do not vary across observations, there is a simple characterization of rationalizability in terms of first order stochastic dominance (FSD). This follows immediately from Theorem 4.

THEOREM 8. *Suppose that $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x}^t, \mathbf{A}^t)\}_{t\in\mathcal{T}}$ satisfies $\mathbf{A}^t = \times_{i=1}^n A_i$ for all t. Then $\mathcal{P}$ is rationalizable as a BNE with independent types and payoff functions of the form (13) obeying SID in $(y_i; b_i, \bar{x}_i)$ if and only if, for each $i \in N$ and $s, t \in \mathcal{T}_{ik}$,*

$$(b_i^t, \bar{x}_i^t) > (b_i^s, \bar{x}_i^s) \Longrightarrow \mathrm{P}_i(\cdot|b_i^t, \bar{x}_i^t, A_i) \geqslant_{\mathrm{FSD}} \mathrm{P}_i(\cdot|b_i^s, \bar{x}_i^s, A_i). \tag{18}$$

EXAMPLE 4. Suppose we have at least two observations of a two player game where $g_i(y_j) = y_j$ for $i \neq j$. Each observation consists of a distribution of actions for player 1 and a distribution of actions for player 2. We assume that there are no changes either in the exogenous variables or in the constraint sets; in other words, this is the case emphasized by Sweeting (2009), of potentially distinct BNE of the same game. Suppose that $b_2^2 > b_2^1$; in other words, player 2's average action is higher in observation 2 than in observation 1. For $\mathcal{P}$ to be rationalizable, condition (18) requires the distribution of player 1's action in the second observation to first order stochastically dominate the distribution at the first observation. In particular, player 1's action has a higher mean in the second observation compared to the first, which in turn requires (via (18) again) that the distribution of player 2's action in the second observation should first order stochastically dominate that in the first. To sum up, for each player, the distribution of actions must be comparable across any two observations (with respect to FSD) and they must be co-monotone across players. When this occurs, the observations can be rationalized as distinct BNE of the same game, with each player i obeying SID in $(y_i; b_i)$.

5.4 Variations on Theorem 7

In Theorem 7, the payoff function of player i is required to obey SID in $(y_i; b_i, \bar{x}_i)$. It may be the case that the econometrician is interested in a weaker criterion: that payoff functions obey SID in $(y_i; b_i)$. In other words, the focus is on restricting interaction effects between players, while the effect of the exogenous variable x_i on i's payoff is allowed to be arbitrary. It is straightforward to check that this case can be covered by an analogous version of Theorem 7 in which $\mathcal{P}_i$ is partitioned differently: two observations s and t in the same partitioned subset $\mathcal{P}_{ik}$ must satisfy $x_i^s = x_i^t$.[20]

So far in this section we have assumed that we can identify a player in role i *across observations* and that his payoff function could differ from another player in role j. We could require instead that all players are ex ante identical, in the sense of having the same payoff function $\Pi(y_i, b_i, x_i, \varepsilon_i)$, though players in a given game could still be different since they may be associated with different observable characteristics. For the sake of simplicity, suppose also that the private type of each player is independently and identically distributed across players and completely independent of each player's observable characteristic. In this case, it is straightforward to check that the following modified version of Theorem 7 holds: *$\mathcal{P}$ is rationalizable as a BNE with independent types and some payoff function Π obeying SID in $(y_i; (b_i, \bar{x}_i))$ if and only if $\bigcup_{i=1}^n \mathcal{P}_i$ passes the test specified in Theorem 3.* In other words, instead of testing each $\mathcal{P}_i$ separately (as in Theorem 7) they must now be tested jointly. Lastly, note that we could pursue this logic further by allowing the number of (ex ante identical) players at each observed game to vary. For each player at each observation, the observable characteristics, the interaction term, the constraint set, and the distribution over the player's actions are known. BNE-rationalizability with a single payoff function obeying SID could be checked by subjecting all of these distributions (as a single collection) to the test specified in Theorem 3.

6 Empirical Illustration: rationalizing an entry game

We implement the results of Section 4 to study a complete information entry game. Our data set is taken from Kline and Tamer (2016), who use it to illustrate the application of their new econometric techniques. The data set they construct contains the entry decisions of airlines in 7882 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).[21] In Kline and Tamer's analysis (and in ours) the two categories are treated as two firms; thus in each market,

[20] Another way of saying the same thing is that $\bar{\mathcal{X}}_i$ no longer exists, and $\mathcal{X}_i = \widetilde{\mathcal{X}}_i$.

[21] 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 an LCC is by definition an OA.

Covariates = $(0,0,0)$	1271 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
30.37%	68.21%	0.55%	0.87%

Covariates = $(0,1,0)$	763 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
19%	78.51%	0.26%	2.23%

Covariates = $(1,0,0)$	1125 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
19.38%	36.71%	25.33%	18.58%

Covariates = $(1,1,0)$	782 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
12.15%	54.22%	4.99%	28.64%

Covariates = $(0,0,1)$	869 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
15.88%	82.28%	0.12%	1.73%

Covariates = $(0,1,1)$	1039 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
7.80%	88.93%	0%	3.27%

Covariates = $(1,0,1)$	677 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
10.64%	32.64%	30.58%	26.14%

Covariates = $(1,1,1)$	1356 markets		
P(N,N)	P(N,E)	P(E,N)	P(E,E)
5.53%	50.07%	2.14%	42.26%

Table 4: Distribution of action profiles across exogenous variables

the two players (LCC and OA) can either both enter, both stay out, or one could enter with the other staying out.

This data set also contains information on two explanatory variables: market presence (P) and market size (S). Market presence is a market- and airline-specific variable. For each airline and for each airport, they count the number of markets that 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 the two endpoints of the trip. The construction and inclusion of this explanatory variable is not novel and follows Berry (1992). Since the airlines are aggregated into two firms (LCC and OA), the market presence variable is also aggregated: the market presence for the LCC firm (resp. OA firm) is the maximum among the actual airlines in the LCC category (resp. OA category). The second explanatory variable is market size, which is a market-specific variable (shared by all airlines in that market) and is defined as the population at the endpoints of the trip. Lastly, Kline and Tamer (2016) discretize the market size and market presence variables to take two values: 1 if the variable is higher than its median value and 0 otherwise.

This data set is presented in Table 4. It consists of eight blocks, with the markets in each block sharing the same exogenous variables. For example, there are 1271 markets where $(P_{LCC}, P_{OA}, S) = (0,0,0)$, of which 30.37% are not served by either airline and 0.87% are served by both.

$(\varepsilon_{LCC}, \varepsilon_{OA})$		Firm OA			
		N		E	
Firm LCC	N	0	0	0	$\Pi_{OA}(N, E, P_{OA}, S, \varepsilon_{OA})$
	E	$\Pi_{LCC}(E, N, P_{LCC}, S, \varepsilon_{LCC})$	0	$\Pi_{LCC}(E, E, P_{LCC}, S, \varepsilon_{LCC})$	$\Pi_{OA}(E, E, P_{OA}, S, \varepsilon_{OA})$

Table 5: Payoff matrix in market $(\varepsilon_{LCC}, \varepsilon_{OA})$

6.1 Testing Rationalizability

Notice that the entries in Table 4 are mostly 'reasonable', in the sense that it appears as though a firm's entry is encouraged whenever its market presence is large or the market 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. Our results in Section 4 allow us to formulate and test this claim rigorously.

First, the notion of 'reasonable' behavior can be formalized by requiring LCC and OA to have payoff functions with the SID property. Specifically, let ε_{LCC} be a type of the LCC player and denote its payoff function by $\Pi_{LCC}(y_{LCC}, y_{OA}, P_{LCC}, S, \varepsilon_{LCC})$, where y_{LCC} is the choice variable (either E or N) and (y_{OA}, P_{LCC}, S) are the other variables that affect its payoff. We require $\Pi_{LCC}(\cdot, \varepsilon_{LCC})$ to obey SID in $(y_{LCC}; (-y_{OA}, P_{LCC}, S))$; in other words, if LCC prefers to enter a market when OA has entered, its market presence is small, or the market size is small, then (respectively) it will also enter if OA has not entered, its market presence is large, or when the market is large.[22] LCC's payoff if it stays out of a market is normalized at 0. Similarly, we require $\Pi_{OA}(y_{LCC}, y_{OA}, P_{LCC}, S, \varepsilon_{OA})$ to obey SID in $(y_{OA}; (-y_{LCC}, P_{OA}, S))$. At a typical market, the econometrician observes (P_{LCC}, P_{OA}, S) but not $(\varepsilon_{LCC}, \varepsilon_{OA})$; the two players are playing a game of complete information so they observe everything. Table 5 depicts the game played at a typical market between the two firms. The issue we wish to address is whether the data displayed in Table 4 is rationalizable as PSNE (in the stochastic sense defined in Definition 6), with both firms having payoff functions obeying SID.

At this point, it may be helpful to mention two features of our rationalizability criterion.

[1] To reiterate what we have stated more generally in Section 4, our notion of rationalizability is silent regarding the joint distribution of types $(\varepsilon_{LCC}, \varepsilon_{OA})$ and it also imposes no restriction on the equilibrium selection rule. In particular, we allow any sort of correlation between ε_{LCC} and

[22]For the SID property, we treat E as ranked above N (associated with values 1 and 0 respectively).

ε_{OA} and the equilibrium selection rule may depend on particular realizations of $(\varepsilon_{LCC}, \varepsilon_{OA})$. We do assume (as does Kline and Tamer (2016)) that the joint distribution of $(\varepsilon_{LCC}, \varepsilon_{OA})$ is independent of the exogenous variables (P_{LCC}, P_{OA}, S).

[2] Since only a firm's *preference* is recovered when a data set is rationalizable, we are free to choose the payoff function so long as it generates the same preference. It is easy to check that any payoff function for (say) LCC that obeys SID (which is an ordinal property) is ordinally equivalent to some (other) payoff function Π_{LCC} where $\Pi_{LCC}(N, y_{OA}, P_{LCC}, S, \varepsilon_{LCC}) = 0$ and $\Pi_{LCC}(E, y_{OA}, P_{LCC}, S, \varepsilon_{LCC})$ is *increasing* in $(-y_{OA}, P_{LCC}, S)$. In other words, whenever a data set is stochastically rationalizable, with firms having payoff functions that obey SID, then we can always choose the payoff functions so that the payoff of entering increases when the other firm stays out, the market presence is high, and the market size is big, and the payoff of staying out is normalized at 0.[23]

By Theorem 6, the rationalizability test involves working out all the paths that obey the SRM axiom and then finding weights on them that are consistent with the distribution over joint actions observed at each value of the exogenous variables.[24] In this data set, there are $4^8 = 65,536$ possible group paths, since for each of the eight possible covariate values, there are four joint choices that a pair of firms can make. The set of paths that obey the SRM axiom, $\mathbf{Y}^{**}$, contains 482 paths. A path $\mathbf{y}^* = (\mathbf{y}^{(0,0,0)}, \mathbf{y}^{(0,1,0)}, \ldots)$ specifies the joint action at each value of the exogenous variables. For example the following path obeys the SRM axiom: $\mathbf{y}^{(0,0,0)} = (N, N)$, $\mathbf{y}^{(0,1,0)} = (N, E)$, $\mathbf{y}^{(1,0,0)} = (E, N)$, $\mathbf{y}^{(1,1,0)} = (N, E)$, $\mathbf{y}^{(0,0,1)} = (N, E)$, $\mathbf{y}^{(1,0,1)} = (E, E)$, $\mathbf{y}^{(0,1,1)} = (N, E)$, and $\mathbf{y}^{(1,1,1)} = (E, E)$. Indeed, this is the path of PSNE that arises if the firms have the following SID preferences: LCC prefers entering if and only if (i) its market presence is large and the other firm is absent or (ii) its market presence is large and the market size is large; OA prefers entering if and only if (i) its market presence is large or (ii) the market size is large.

Suppose for now that the entries in Table 4 give the true distribution at each value of the exogenous variables/covariates. In that case, we can implement the test in Theorem 6 directly by checking whether there is a solution to the linear system (11). It turns out that there is no

[23]Note that this hinges on there being just two actions for each firm.

[24]We focus on preferences where a firm is never indifferent between entering and staying out. It follows that the paths must obey the SRM axiom rather than the RM axiom (see remarks following Theorem 6).

Covariates = $(0,0,0)$ 1271 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
30.06%	67.90%	0.86%	1.18%

Covariates = $(0,1,0)$ 763 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
18.29%	78.96%	0.71%	2.05%

Covariates = $(1,0,0)$ 1125 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
19.38%	36.71%	25.33%	18.58%

Covariates = $(1,1,0)$ 782 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
12.73%	53.64%	5.57%	28.06%

Covariates = $(0,0,1)$ 869 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
15.46%	81.86%	0.54%	2.15%

Covariates = $(0,1,1)$ 1039 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
7.86%	89.19%	0.26%	2.69%

Covariates = $(1,0,1)$ 677 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
10.64%	32.64%	30.58%	26.14%

Covariates = $(1,1,1)$ 1356 markets			
P(N,N)	P(N,E)	P(E,N)	P(E,E)
5.63%	49.98%	2.24%	42.16%

Table 6: Rationalizable outcome closest to Table 4

solution, so this data set is not exactly rationalizable. This is fairly surprising, since the number of unknowns (482) far exceeds the number of linear constraints, which suggests that the conditions are very permissive. In fact, there is at least one easy-to-understand reason why the dataset is not rationalizable. Notice from the data that $\text{P}(N,N|1,1,0)+\text{P}(E,N|1,1,0) = 17.14\% < 19\% = \text{P}(N,N|0,1,0)$. This is *not compatible with rationalizability* for the following reason: any pair of firms with SID preferences that select (N,N) at $(0,1,0)$ would either select (N,N) or (E,N) at $(1,1,0)$.

If we solve for the dataset that is rationalizable and closest (as measured by the sum of square deviations) to the one we actually observe, we get the distribution displayed in Table 6. Notice that in this case $\text{P}(N,N|1,1,0)+\text{P}(E,N|1,1,0) = 18.3\% > 18.29\% = \text{P}(N,N|0,1,0)$. If we compare the entries in Tables 4 and 6, we see immediately that they look quite close, which naturally makes us wonder whether the observed violation of rationalizability is in fact significant.

To address this issue, we adopt the approach recently proposed by Kitamura and Stoye (2016); they develop a method of evaluating the statistical significance of a data set violating a set of linear constraints that directly applies to our framework.[25] Roughly speaking, the test assumes that the closest compatible distribution displayed in the last table is the true distribution, and uses a bootstrap procedure to calculate the likelihood of getting a sample like the one we observe. (See the Online Appendix for a fuller description of the Kitamura-Stoye procedure and our implementation

[25]Kitamura and Stoye (2016) apply their test to the consumer utility-maximization problem.

Set Description	# of paths	Est. Prob.	Conf. Int.
Non-strategic LCC ($\mathbf{Y}^{**}_{LCC}$)	69	$[71, 88]$	$[65, 92]$
Non-strategic OA ($\mathbf{Y}^{**}_{OA}$)	69	$[59, 75]$	$[55, 78]$
Non-strategic LCC and OA ($\mathbf{Y}^{**}_{LCC} \cap \mathbf{Y}^{**}_{OA}$)	36	$[48, 75]$	$[44, 78]$

Table 7: Incidence of non-strategic payoffs

of all these results.) By applying their test, we find that the probability of getting our sample (or a more extreme one), assuming that our modelling restrictions are true, is 15% without a tuning parameter and it decreases as the value of the tuning parameter gets larger.[26] These probabilities correspond to the p-values for the null hypothesis that our modelling assumptions are true and they indicate that our hypothesis cannot be rejected for small values of the tuning parameter.[27]

6.2 Estimation

In testing the model, we also recover information on the distribution of paths, and thus on the distribution of firms' preferences, that could have generated the data. This information allows us to investigate a variety of interesting issues. To illustrate how this can be carried out, we ask the following basic question: are strategic effects necessary to explain the data, and if so, in what way?

For LCCs, their choice of action is affected by the exogenous variables (P_{LCC}, S), which can take four possible values, in addition to the action of the other carrier. We say that an LCC's preference at some fixed (P_{LCC}, S) is *non-strategic* if its optimal action is independent of the other player's action, i.e., the sign of $\Pi_{LCC}(E, y_{OA}, P_{LCC}, S, \varepsilon_{LCC})$ is independent of y_{OA}; on the other hand, its preference is *strategic* if the payoff switches from being positive to negative if the other player enters. Therefore, if LCC is non-strategic in a given market, then either E or N is a dominant strategy, while a strategic LCC chooses E if and only if the other firm stays out.

We identify, among the 482 paths that obey the SRM axiom, those paths which can be ratio-

[26]The Kitamura-Stoye test relies on a tuning parameter that solves a discontinuity issue arising from the possibility of boundary solutions. From an economic perspective, the tuning parameter modifies the initial hypothesis by forcing all consistent types to have a strictly positive probability in the linear model. The p-value estimates are sensitive to the tuning parameter.

[27]A more permissive version of our rationalizability test would allow for the possibility that firms may be indifferent between entering and not entering a market (see footnote 24). In that case, the rationalizability test involves solving (11), with the paths obeying the RM axiom (rather than the SRM axiom); there are 1,809 paths with this weaker property. The p-value increases to 36% without a tuning parameter (see the Online Appendix).

nalized by LCC preferences that are *always non-strategic*, in the sense that the carrier's preference is non-strategic at every realization of (P_{LCC}, S). (Note that LCC's decision can still depend on (P_{LCC}, S)). Let us denote this set of paths by $\mathbf{Y}^{**}_{LCC}$. It is worth bearing in mind that since rationalizing preferences are non-unique, there is a difference between $\mathbf{Y}^{**}_{LCC}$ and paths that *must* be rationalized by non-strategic LCC preferences, which will be a subset of $\mathbf{Y}^{**}_{LCC}$.[28] Since we are investigating whether strategic behavior is *necessary* for explaining the data, we focus our attention on $\mathbf{Y}^{**}_{LCC}$. Similarly, we can identify $\mathbf{Y}^{**}_{OA}$, those paths with the feature that it can be rationalized by OA being always non-strategic.

It turns out that there are 69 paths each in $\mathbf{Y}^{**}_{LCC}$ and in $\mathbf{Y}^{**}_{OA}$. (The path example given in the previous subsection is a member of $\mathbf{Y}^{**}_{OA}$ but not of $\mathbf{Y}^{**}_{LCC}$.[29]) The interval under 'Estimated Probability' in Table 7 gives the interval estimate on the probability of $\mathbf{Y}^{**}_{LCC}$, subject to the model generating the closest rationalizable distribution (i.e., the one depicted in Table 6).[30] The last column gives the confidence intervals on the probability of $\mathbf{Y}^{**}_{LCC}$; in essence, it gives the values on the probability of $\mathbf{Y}^{**}_{LCC}$ which, when imposed as an additional restriction on the model, would still allow the model to approximately rationalize the data (in Table 4) at the 5% level of significance;[31] this interval must, by construction contain the estimated probability. The first thing to notice is that the confidence interval on $\mathbf{Y}^{**}_{LCC}$ has an upper bound that is clearly below 1; to be specific, it is 0.92. This means that, notwithstanding the greater flexibility of our nonparametric model, a more parsimonious version of the model in which LCC is assumed to be always non-strategic will *not* be able to explain the data. Indeed, the probability of LCC being strategic (for some value of (P_{LCC}, S)) must be at least 0.08.[32] Formally, we may conclude that

$$\mathrm{P}\left(\{\varepsilon_{LCC} : \Pi_{LCC}(\cdot, P_{LCC}, S, \varepsilon_{LCC}) \text{ is strategic at some } (P_{LCC}, S)\}\right) \geqslant 0.08.$$

[28]For example, consider a path where OA always chooses N and LCC always chooses E. This path obeys the SRM axiom and is in $\mathbf{Y}^{**}_{LCC}$ since it can be rationalized by a preference where LCC always prefers to enter, but it is clear that there also exists a rationalization with a strategic LCC.

[29]The reader can check that there is no non-strategic LCC preference that rationalizes the path.

[30]This calculation is described at the start of Section 4.3.

[31]The procedure for calculating these intervals follow Deb, Kitamura, Quah, and Stoye (2018) and is explained more carefully in the Online Appendix.

[32]$\mathbf{Y}^{**}_{LCC}$ is the set of path that *can be* (not *must be*) rationalized by non-strategic LCC preferences. This means that the upper bound of the confidence interval on $\mathbf{Y}^{**}_{LCC}$ really is the upper bound on the probability of non-strategic LCCs, but the lower bound of that interval could be an overestimate of non-strategic behavior, because there are some paths in $\mathbf{Y}^{**}_{LCC}$ which can also be rationalized by strategic LCCs.

Set Description (subsets of $\hat{\mathbf{Y}}^{**}$)		# of paths	Est. Prob.	Conf. Int.
Non-strategic LCC for all (P_{LCC}, S)		46	$[71, 88]$	$[65, 92]$
Strategic LCC at $(P_{LCC}, S) = (1, 1)$ or $(1, 0)$		22	$[11, 28]$	$[8, 34]$
	Strategic LCC at $(P_{LCC}, S) = (1, 1)$	13	$[11, 28]$	$[8, 33]$
of which	Strategic LCC at $(P_{LCC}, S) = (1, 0)$	17	$[9, 16]$	$[6, 19]$
	Strategic LCC at $(P_{LCC}, S) = (1, 1)$ & $(1, 0)$	8	[0,16]	$[0, 19]$
Non-strategic OA for all (P_{OA}, S)		46	$[59, 75]$	$[55, 78]$
Strategic OA at $(P_{OA}, S) = (0, 1)$ or $(0, 0)$		22	$[24, 40]$	$[21, 45]$
	Strategic OA at $(P_{OA}, S) = (0, 1)$	17	$[24, 26]$	$[21, 29]$
of which	Strategic OA at $(P_{OA}, S) = (0, 0)$	13	$[13, 40]$	$[10, 45]$
	Strategic OA at $(P_{OA}, S) = (0, 1)$ & $(0, 0)$	8	$[10, 26]$	$[6, 29]$

Table 8: Composition of data-rationalizing types

Analogously, we surmise from Table 7 that for OA

$$\mathrm{P}\left(\{\varepsilon_{OA} : \Pi_{OA}(\cdot, P_{OA}, S, \varepsilon_{OA}) \text{ is strategic at some } (P_{OA}, S)\}\right) \geqslant 0.22.$$

It is worthwhile finding out precisely *where* is strategic behavior crucial in explaining the data. Indeed, we discover that it is possible to restrict ourselves to paths in the set $\hat{\mathbf{Y}}^{**}$, which consists of paths where the preference of LCC can be chosen to be non-strategic whenever $P_{LCC} = 0$ (i.e., at $(P_{LCC}, S) = (0, 0)$ and at $(P_{LCC}, S) = (0, 1)$) *and* the preference of OA can be chosen to be non-strategic whenever $P_{OA} = 1$ (i.e., at $(P_{OA}, S) = (1, 0)$ and at $(P_{OA}, S) = (1, 1)$).[33] There are 68 paths in $\hat{\mathbf{Y}}^{**}$, its estimated probability is 0.99, and the confidence interval is [0.96,1]. In other words, if we have a more restrictive model in which we require LCC to be non-strategic at $P_{LCC} = 0$ and OA to be non-strategic at $P_{OA} = 1$, such a model will still be consistent with the data.

Combining these observations, we conclude that to explain the data, it is crucial that there are markets where (I) LCCs are strategic when $P_{LCC} = 1$ and (II) OAs are strategic when $P_{OA} = 0$. This phenomenon is recorded in Table 8 which gives a dissection of the types that explain the data. All the paths reported in this table are drawn from $\hat{\mathbf{Y}}^{**}$. The first row gives the weight on the set of paths where LCC's preference can be chosen to be always non-strategic (i.e., at all four possible

[33]It is important to note that flipping the restrictions will not work. Indeed, the estimated weight of those paths which are rationalizable with LCC being non-strategic at $(P_{LCC}, S) = (1, 1)$ is 0.88 while the estimated weight of those paths which are rationalizable with OA being non-strategic at $(P_{OA}, S) = (0, 0)$ is 0.86, so both numbers are some distance away from 1.

values of (P_{LCC}, S)); there are 46 paths in this group and its weight is significantly less than 1. Explaining the data therefore requires the remaining paths in $\hat{\mathbf{Y}}^{**}$, which are those where LCC *must be* strategic, in the sense that any rationalization of the path will involve LCC being strategic at either $(P_{LCC}, S) = (1, 0)$ or at $(P_{LCC}, S) = (1, 1)$ (and possibly both); there are 22 paths in this group. We can explore its composition further and estimate the probability of those paths where LCC has to be strategic at (1,0), at (1,1), and at both (1,0) and (1,1). We learn, for example, that

$$\mathrm{P}(\{\varepsilon_{LCC} : \Pi_{LCC}(\cdot, P_{LCC}, S, \varepsilon_{LCC}) \text{ is strategic at } (P_{LCC}, S) = (1, 1)\}) \geqslant 0.08.$$ [34]

In other words, the probability that LCC is strategic in a market with observable characteristic $(P_{LCC}, S) = (1, 1)$ is at least 0.08. Similarly, for OA, we find that

$$\mathrm{P}(\{\varepsilon_{OA} : \Pi_{OA}(\cdot, P_{OA}, S, \varepsilon_{OA}) \text{ is strategic at } (P_{OA}, S) = (0, 1)\}) \geqslant 0.21.$$

Appendix I

Proof of Theorem 1

It suffices to show that (b) implies (c): if $\mathcal{O}$ obeys RM axiom, then it is rationalizable by a preference that obeys SSCD. Our proof involves first working out the (incomplete) revealed preference relations on $\mathcal{Y} \times \mathcal{Z}$ that *must* be satisfied by any SID preference that rationalizes the data and then constructing a rationalizing preference on $\mathcal{Y} \times \mathcal{Z}$ that completes those relations and obeys SSCD.

Given a data set $\mathcal{O} = \{(y^t, z^t, A^t)\}_{t \in \mathcal{T}}$, the *single-crossing extension* of the indirect revealed preference relation $\succsim^{RT}$ is another binary relation $\succ^{RTS}$ defined in the following way: (i) for $y'' > y'$, $(y'', z) \succ^{RTS} (y', z)$ if there is $z' < z$ such that $(y'', z') \succ^{RT} (y', z')$ and (ii) for $y'' < y'$, $(y'', z) \succ^{RTS} (y', z)$ if there is $z'' > z$ such that $(y'', z'') \succsim^{RT} (y', z'')$. Let $\succsim^{RTS}$ be the relation given by $\succsim^{RTS} = \succ^{RTS} \cup \succsim^{RT}$. It follows immediately from its definition that $\succsim^{RTS}$ also obeys SSCD, in the following

[34]Note that the upper bound on that confidence interval, 0.33, is *not* a cap on the probability that LCC is strategic at $(1, 0)$ because paths could be rationalized in more than one way, and there are paths that could be rationalized by one LCC preference that is strategic at (1,0) and another that is not. See the related observation in Footnote 32.

sense: if $y'' > y'$ and $z'' > z'$ or $y'' < y'$ and $z'' < z'$, then

$$(y'', z') \succsim^{RTS} (y', z') \Longrightarrow (y'', z'') \succ^{RTS} (y', z''). \tag{19}$$

In addition, let $\succsim^{RTST}$ be the transitive closure of $\succsim^{RTS}$, i.e., $(y'', z) \succsim^{RTST} (y', z)$ if there is a sequence $\bar{y}_1, \bar{y}_2, ..., \bar{y}_k$[35] such that

$$(y'', z) \succsim^{RTS} (\bar{y}_1, z) \succsim^{RTS} \cdots \succsim^{RTS} (\bar{y}_k, z) \succsim^{RTS} (y', z). \tag{20}$$

If there is one strict relation $\succ^{RTS}$ in this sequence, then we say that $(y'', z) \succ^{RTST} (y', z)$.[36]

LEMMA A1: *The relations $\succsim^{RTS}$, $\succ^{RTS}$, and $\succsim^{RTST}$ have the interval property.*

Proof. Let $y'' > y > y'$. (The case where $y'' < y < y'$ can be proved in a similar way.) If $(y'', z) \succsim^{RTS} (\succ^{RTS}) (y', z)$ holds, there exists some $z' \leqslant (<) z$ such that $(y'', z') \succsim^{RT} (y', z')$. By the interval property of $\succsim^{RT}$ (see the proof of Proposition 1), we obtain $(y'', z') \succsim^{RT} (y, z')$. Since $y'' > y$ and $z' \leqslant (<) z$, we have $(y'', z) \succsim^{RTS} (\succ^{RTS}) (y, z)$. So we have shown that $\succsim^{RTS}$ and $\succ^{RTS}$ have the interval property. Lastly, if $(y'', z) \succsim^{RTST} (y', z)$, there exists a sequence $\bar{y}_1, \bar{y}_2, ..., \bar{y}_k$ such that

$$(y'', z) \succsim^{RTS} (\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z) \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_k, z) \succsim^{RTS} (y', z).$$

Letting $\bar{y}_0 = y''$ and $\bar{y}_{k+1} = y'$, since $y'' > y > y'$, we can find some $0 \leqslant m \leqslant k$ such that $\bar{y}_m \geqslant y \geqslant \bar{y}_{m+1}$. By the interval property of $\succsim^{RTS}$, we obtain $(\bar{y}_m, z) \succsim^{RTS} (y, z)$. Thus $(y'', z) \succsim^{RTST} (y, z)$ since $(y'', z) \succsim^{RTST} (\bar{y}_m, z) \succsim^{RTS} (y, z)$. QED

The relevance of the binary relations $\succsim^{RTST}$ and $\succ^{RTST}$ flows from the following lemma, which says that any rationalizing preference for the agent must respect the ranking implied by them.

LEMMA A2:[37] *Suppose that the preference obeys SID and rationalizes $\mathcal{O}$. Then $\succsim$ extends $\succsim^{RTST}$*

[35]Note that subscripts here denote generic numbering. When we wish to refer to an action at a particular observation t, we write y^t.

[36]Note that $\succ^{RTS}$ is *not* the asymmetric part of $\succsim^{RTS}$ and $\succ^{RTST}$ is *not* the asymmetric part of $\succsim^{RTST}$.

[37]Strictly speaking this result is not needed for the proof of Theorem 1, but it provides the motivation for why we are focusing on $\succsim^{RTST}$ and $\succ^{RTST}$.

and $\succ^{RTST}$ *in the following sense:*

$$(y'', z) \succsim^{RTST} (\succ^{RTST})\, (y', z) \Longrightarrow (y'', z) \succsim (\succ)\, (y', z). \tag{21}$$

Proof. We assume $y'' > y'$. (The other case has a similar proof.) Since $\succsim$ is transitive, we need only show that $(y'', z) \succsim (\succ)\, (y', z)$ whenever $(y'', z) \succsim^{RTS} (\succ^{RTS})\, (y', z)$. If $(y'', z) \succsim^{RTS} (\succ^{RTS})\, (y', z)$ then there exists some $z' \leqslant (<)\, z$ such that $(y'', z') \succsim^{RT} (y', z')$. By the interval property of $\succsim^{RT}$, we obtain $(y'', z') \succsim^{RT} (y, z')$ for all $y \in [y', y'']$. Since $\succsim$ rationalizes $\mathcal{O}$, we also have $(y'', z') \succsim (y, z')$ for all $y \in [y', y'']$. By SID of $\succsim$, we obtain $(y'', z) \succsim (\succ)\, (y', z)$ for $z' \leqslant (<)\, z$. QED

The next result establishes a property of $\succsim^{RTST}$ useful for the proof of Theorem 1.

Lemma A3: *Suppose* $(y'', z) \succsim^{RTST} (y', z)$*; then there is a sequence* $\{\bar{y}_j\}_{j=1}^{k}$ *such that*

$$(y'', z) \succsim^{RTS} (\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z) \succsim^{RTS} \cdots \succsim^{RTS} (\bar{y}_k, z) \succsim^{RTS} (y', z), \tag{22}$$

$$\textit{with } y'' > \bar{y}_1 > \bar{y}_2 > \cdots > \bar{y}_k > y', \tag{23}$$

if $y'' > y'$ *and the inequality (23) reversed if* $y'' < y'$.

Proof. By the definition of $\succsim^{RTST}$, we know there is $\{\bar{y}_j\}_{j=1}^{k}$ such that (22) holds, so what we need to do is to show that $\{\bar{y}_j\}_{j=1}^{k}$ obeys (23) if $y'' > y'$. (The case where $y'' < y'$ has an analogous proof which we shall skip.) To do this, we choose a chain linking (y'', z) and (y', z) with the property that (writing $\bar{y}_0 = y''$ and $\bar{y}_{k+1} = y'$) $(\bar{y}_m, z) \not\succsim^{RTS} (\bar{y}_{m'}, z)$ for $m' > m + 1$; in other words, no link in the chain can be dropped. We claim that (23) must hold in this case. First we note that $\bar{y}_j > y'$ for all $j < k + 1$. If not, there is ℓ such that $\bar{y}_\ell \leqslant y' < \bar{y}_{\ell-1}$, with $(\bar{y}_{\ell-1}, z) \succsim^{RTS} (\bar{y}_\ell, z)$; since $\succsim^{RTS}$ has the interval property (Lemma A1), we obtain $(\bar{y}_{\ell-1}, z) \succsim^{RTS} (y', z)$ and the chain has been shortened. To show that $\bar{y}_j$ is decreasing, suppose instead that there is m such that $\bar{y}_{m+1} > \bar{y}_m$. Let $\bar{y}_{m+n}$ be the first time after $\bar{y}_{m+1}$ such that $\bar{y}_{m+n} \leqslant \bar{y}_m$. (This must occur since $\bar{y}_m > y'$.) Then we have $\bar{y}_{m+n} \leqslant \bar{y}_m < \bar{y}_{m+n-1}$. Since $(\bar{y}_{m+n-1}, z) \succsim^{RTS} (\bar{y}_{m+n}, z)$, the interval property of $\succsim^{RTS}$ guarantees that $(\bar{y}_{m+n-1}, z) \succsim^{RTS} (\bar{y}_m, z)$. Thus we obtain a cycle

$$(\bar{y}_m, z) \succsim^{RTS} (\bar{y}_{m+1}, z) \succsim^{RTS} \ldots \succsim^{RTS} (\bar{y}_{m+n-1}, z) \succsim^{RTS} (\bar{y}_m, z).$$

Since $\succsim^{RTS}$ is cyclically consistent, this chain cannot be related by $\succ^{RTS}$ and must be related by $\succsim^{RT}$. In particular, $(\bar{y}_{m+n-1}, z) \nsucc^{RTS} (\bar{y}_m, z)$ and thus $(\bar{y}_{m+n-1}, z) \nsucc^{RTS} (\bar{y}_{m+n}, z)$ (by the interval property of $\succ^{RTS}$). We conclude that $(\bar{y}_m, z) \succsim^{RT} (\bar{y}_{m+n}, z)$ and thus we can shorten (22) to

$$(y'', z) \succsim^{RTS} (\bar{y}_1, z) \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_m, z) \succsim^{RTS} (\bar{y}_{m+n}, z) \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_k, z) \succsim^{RTS} (y', z)$$

which contradicts our assumption that no link in the chain can be dropped. QED

It follows from Lemma A2 that in order for $\mathcal{O}$ to be monotone rationalizable, the binary relation $\succsim^{RTST}$ must have the following property: for any (y', z) and (y'', z) in $\mathcal{Y} \times \mathcal{Z}$,

$$(y', z) \succsim^{RTST} (y'', z) \Longrightarrow (y'', z) \nsucc^{RTST} (y', z). \tag{24}$$

If not, we obtain both $(y', z) \succsim (y'', z)$ and $(y'', z) \succ (y', z)$, which is impossible. The following lemma says that $\succsim^{RTST}$ obeys this property as well as SSCD whenever $\mathcal{O}$ obeys RM axiom.

LEMMA A4: *Suppose that $\mathcal{O}$ obeys RM axiom. Then, $\succsim^{RTST}$ obeys SSCD and property (24).*

Proof. We first prove that (24) holds. The statement (24) is equivalent to $\succsim^{RTS}$ being *cyclically consistent*, i.e.,

$$(\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z) \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_k, z) \Longrightarrow (\bar{y}_k, z) \nsucc^{RTS} (\bar{y}_1, z). \tag{25}$$

Cyclical consistency can in turn be equivalently re-formulated as the following:

$$(\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z) \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_k, z) \succsim^{RTS} (\bar{y}_1, z) \tag{26}$$

$$\Longrightarrow (\bar{y}_1, z) \nsucc^{RTS} (\bar{y}_2, z) \nsucc^{RTS} ... \nsucc^{RTS} (\bar{y}_k, z) \nsucc^{RTS} (\bar{y}_1, z)$$

Thus, whenever there is a cycle like (26), it *must* be the case that

$$(\bar{y}_1, z) \succsim^{RT} (\bar{y}_2, z) \succsim^{RT} ... \succsim^{RT} (\bar{y}_k, z) \succsim^{RT} (\bar{y}_1, z)$$

We prove (24) by induction on the length of the chain, k, on the left side of (25). Whenever

(25) holds for chains of length k or less (equivalently, whenever the cycles in (26) have length k or less), we say that $\succsim^{RTS}$ is k-consistent. For 2-consistency, we need to show that

$$(\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z) \Longrightarrow (\bar{y}_2, z) \not\succ^{RTS} (\bar{y}_1, z).$$

Suppose that $\bar{y}_1 > \bar{y}_2$; the case of $\bar{y}_1 < \bar{y}_2$ can be dealt with in a similar way. By definition, if $(\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z)$ then there is $z' \leqslant z$ such that $(\bar{y}_1, z') \succsim^{RT} (\bar{y}_2, z')$. On the other hand, if $(\bar{y}_2, z) \succ^{RTS} (\bar{y}_1, z)$, then there is $z'' > z$ such that $(\bar{y}_2, z'') \succsim^{RT} (\bar{y}_1, z'')$ and so we obtain a violation of RM axiom. Suppose that $\succsim^{RTS}$ is k-consistent for all $k < \bar{k}$. To show that $\bar{k}$-consistency holds, suppose the left side of (25) holds for $k = \bar{k}$ and $\bar{y}_1 < \bar{y}_{\bar{k}}$. Clearly, there must be $m < \bar{k}$ such that $\bar{y}_m < \bar{y}_{\bar{k}}$ and $\bar{y}_{m+1} \geqslant \bar{y}_{\bar{k}}$. We consider two cases: (A) $\bar{y}_m \geqslant \bar{y}_1$ and (B) $\bar{y}_m < \bar{y}_1$. In case (A), by the interval property of $\succsim^{RTS}$ (Lemma A1), we obtain $(\bar{y}_m, z) \succsim^{RTS} (\bar{y}_{\bar{k}}, z)$. By way of contradiction, suppose also that $(\bar{y}_{\bar{k}}, z) \succ^{RTS} (\bar{y}_1, z)$. Then the interval property of $\succ^{RTS}$ guarantees that $(\bar{y}_{\bar{k}}, z) \succ^{RTS} (\bar{y}_m, z)$ and so we obtain a violation of 2-consistency. For (B), since $(\bar{y}_m, z) \succsim^{RTS} (\bar{y}_{m+1}, z)$, the interval property guarantees that $(\bar{y}_m, z) \succsim^{RTS} (\bar{y}_1, z)$. So we obtain the cycle

$$(\bar{y}_1, z) \succsim^{RTS} (\bar{y}_2, z) \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_m, z) \succsim^{RTS} (\bar{y}_1, z) \tag{27}$$

which has length strictly lower than $\bar{k}$. By the induction hypothesis, we obtain

$$(\bar{y}_1, z) \not\succ^{RTS} (\bar{y}_2, z) \not\succ^{RTS} ... \not\succ^{RTS} (\bar{y}_m, z) \not\succ^{RTS} (\bar{y}_1, z)$$

and so we can replace each $\succsim^{RTS}$ in (27) by $\succsim^{RT}$. Furthermore, $(\bar{y}_m, z) \not\succ^{RTS} (\bar{y}_1, z)$ guarantees that $(y^m, z) \not\succ^{RTS} (\bar{y}_{m+1}, z)$, by the interval property of $\succ^{RTS}$. Therefore, $(\bar{y}_1, z) \succsim^{RT} (\bar{y}_{m+1}, z)$ and, by the interval property of $\succsim^{RT}$, we obtain $(\bar{y}_1, z) \succsim^{RT} (\bar{y}_{\bar{k}}, z)$. 2-consistency then ensures that $(\bar{y}_{\bar{k}}, z) \not\succ^{RTS} (\bar{y}_1, z)$. This completes the proof that (24) holds.

By definition, $\succsim^{RTST}$ obeys SSCD if whenever $y'' > y'$ and $z'' > z'$ *or* $y'' < y'$ and $z'' < z'$,

$$(y'', z') \succsim^{RTST} (y', z') \Longrightarrow (y'', z'') \succ^{RTST} (y', z'').$$

We shall concentrate on the case where $y'' > y'$; the other case has a similar proof. If $(y'', z') \succsim^{RTST}$

(y', z'), then by Lemma A3, there is $\bar{y}_j$ (for $j = 1, 2, ..., k$) such that

$$(y'', z') \succsim^{RTS} (\bar{y}_1, z') \succsim^{RTS} (\bar{y}_2, z') \succsim^{RTS} ... \succsim^{RTS} (\bar{y}_k, z') \succsim^{RTS} (y', z').$$

with $y'' > \bar{y}_1 > \bar{y}_2 > ... > \bar{y}_k > y'$. Since $\succsim^{RTS}$ obeys SSCD (see (19)), we obtain

$$(y'', z'') \succ^{RTS} (\bar{y}_1, z'') \succ^{RTS} (\bar{y}_2, z'') \succ^{RTS} ... \succ^{RTS} (\bar{y}_k, z'') \succ^{RTS} (y', z'')$$

and so $(y'', z'') \succ^{RTST} (y', z'')$. QED

Our final step consists of constructing the SSCD preference that rationalizes $\mathcal{O}$. Since $\succsim^R \subset \succsim^{RTST}$, it is clear that Lemma A2 has the converse: if there is a regular and SID preference $\succsim$ on $\mathcal{Y} \times \mathcal{Z}$ that obeys (21), then this preference rationalizes $\mathcal{O}$. This observation, together with Lemma A4, suggest that a reasonable way of constructing a rationalizing preference is to *begin with* $\succsim^{RTST}$ and $\succ^{RTST}$ and then complete these incomplete relations in a way that gives a preference with the required properties. This is precisely the approach we take. Define the binary relation $\succsim^*$ on $\mathcal{Y} \times \mathcal{Z}$ in the following manner:

$$\begin{aligned} (y'', z) \succsim^* (y', z) \text{ if } & (y'', z) \succsim^{RTST} (y', z) \\ & \text{or } (y'', z) \parallel^{RTST} (y', z) \text{ and } y' \geqslant y'', \end{aligned} \tag{28}$$

where $(y'', z)||^{RTST}(y', z)$ means neither $(y'', z) \succsim^{RTST} (y', z)$ nor $(y', z) \succsim^{RTST} (y'', z)$. The following result completes our argument that (b) implies (c) in Theorem 1.

LEMMA A5: *Suppose that $\mathcal{O}$ obeys RM axiom. The binary relation $\succsim^*$ is an SSCD preference that rationalizes $\mathcal{O}$. On every set $K \subset \mathcal{Y}$ that is compact in $\mathbb{R}$ and for every $z \in \mathcal{Z}$, $BR(z, K, \succsim^*)$ is nonempty and finite; in particular, $\succsim^*$ is a regular preference.*

Proof. We first show that $\succsim^*$ is a preference that rationalizes $\mathcal{O}$. Since $\succsim^{RTST} \subset \succsim^*$ by construction, $\succsim^*$ must rationalize $\mathcal{O}$. Furthermore, $\succsim^*$ is complete and reflexive by construction, so to demonstrate that it is a preference we need only show that it is transitive. Indeed, suppose

$$(a, z) \succsim^* (b, z) \succsim^* (c, z) \succsim^* (a, z). \tag{29}$$

There are essentially four possible cases we need to consider:

Case 1. None of the three elements are related by $\succsim^{RTST}$. Given the definition of $\succsim^*$, this means that $a < b < c < a$, which is impossible.

Case 2. $(a, z) \parallel^{RTST} (b, z)$, $(b, z) \parallel^{RTST} (c, z)$, and $(c, z) \succsim^{RTST} (a, z)$. Then (29) can only occur if $a < b < c$, but if this is the case, the interval property of $\succsim^{RTST}$ (Lemma A1) will imply that $(c, z) \succsim^{RTST} (b, z)$. So this case is impossible.

Case 3. $(a, z) \parallel^{RTST} (b, z)$, $(b, z) \succsim^{RTST} (c, z) \succsim^{RTST} (a, z)$. This is impossible because, by the transitivity of $\succsim^{RTST}$, we obtain $(b, z) \succsim^{RTST} (a, z)$.

Case 4. $(a, z) \succsim^{RTST} (b, z) \succsim^{RTST} (c, z) \succsim^{RTST} (a, z)$. By (24), this is only possible if

$$(a, z) \succsim^{RT} (b, z) \succsim^{RT} (c, z) \succsim^{RT} (a, z),$$

but then we also obtain, by the transitivity of $\succsim^{RT}$, $(a, z) \succsim^{RT} (c, z)$ and, hence, $(a, z) \succsim^* (c, z)$, which establishes the transitivity of $\succsim^*$.

To show that $\succsim^*$ obeys SSCD, let $y'' > y'$ and $z'' > z'$; then

$$\begin{aligned}(y'', z') \succsim^* (y', z') &\Longrightarrow (y'', z') \succsim^{RTST} (y', z') \\ &\Longrightarrow (y'', z'') \succ^{RTST} (y', z'') \\ &\Longrightarrow (y'', z'') \succ^* (y', z''),\end{aligned}$$

in which the first implication follows from the definition of $\succsim^*$, the second implication from the SSCD property of $\succsim^{RTST}$, and the third from the fact that $\succ^*$ contains $\succ^{RTST}$ (so $\succsim^*$ extends $\succsim^{RTST}$ in the sense of (21)). The last claim is true because if $(y'', z) \succ^{RTST} (y', z)$, then Lemma A4 says that $(y', z) \not\succsim^{RTST} (y'', z)$; thus $(y', z) \not\succsim^* (y'', z)$ and we obtain $(y'', z) \succ^* (y', z)$.

It remains for us to show that, for every $z \in \mathcal{Z}$, $\mathrm{BR}(z, K, \succsim^*)$ is nonempty and finite, where $K \subset \mathcal{Y}$ and K is compact in $\mathbb{R}$. If $K \not\ni y^t$ for every $t \in \mathcal{T}$, then it follows from the definition of $\succsim^*$ that $(m, z) \succsim^* (y, z)$, where $m = \min K$ and $y \in K$. In this case, m is the only maximiser of $\succsim^*$ in K. Suppose that $K \ni y^t$ for some t. Since there are a finite number of observations, we can find some $y^s \in K$ such that $(y^s, z) \succsim^* (y^t, z)$ for every $y^t \in K$. We claim that either m or y^s is optimal in K at the parameter value z, so that $\mathrm{BR}(z, K, \succsim^*)$ is nonempty and finite. Indeed, suppose there is $y \in K$ such that $(y, z) \succ^* (m, z)$. Then, since $m < y$, it must hold that $(y, z) \succ^{RTST} (m, z)$ and there is $\underline{t} \in \mathcal{T}$ such that $y = y^{\underline{t}}$, in which case we obtain $(y^s, z) \succsim^* (y^{\underline{t}}, z)$ by the definition of y^s. So for all $y \in K$, either $(m, z) \succsim^* (y, z)$ or $(y^s, z) \succsim^* (y, z)$. QED

Proof of Proposition 2

Let λ by a measure on $\mathcal{Y}$ with the following properties: (i) $\lambda(\mathcal{Y}) < \infty$; (ii) on any nonempty interval I of $\mathcal{Y}$, $\lambda(I) > 0$; (iii) $\lambda(\{y^t\}) > 0$ for all $t \in \mathcal{T}$. For any $(y, z) \in \mathcal{Y} \times \mathcal{Z}$, we define the set $L(y, z) = \{\bar{y} \in \mathcal{Y} : (y, z) \succsim^* (\bar{y}, z)\}$. This set is measurable since $\mathcal{O}$ is finite and $\mathbb{I}(\mathcal{Y})$ consists of compact intervals. Furthermore, λ is a finite measure (according to (i)), so $\lambda(L(y, z))$ is well-defined. We claim that $u(y, z) = \lambda(L(y, z))$ represents $\succsim^*$. It follows immediately from the definition that $u(y'', z) \geqslant u(y', z)$ if $(y'', z) \succsim^* (y', z)$. So we need only show that $u(y'', z) > u(y', z)$ if $(y'', z) \succ^* (y', z)$. Suppose there exists an observed action, y^s, such that $y^s \in L(y'', z)\backslash L(y', z)$; then $u(y'', z) > u(y', z)$ since $\lambda(\{y^s\}) > 0$ (by (iii)). If such an y^s does not exist, then, in particular, $y'' \notin \{y^t\}_{t \in \mathcal{T}}$. For $y'' \succsim^* y'$, it must be the case that $(y'', z) \parallel^{RTST} (y', z)$ and $y'' < y'$. We claim that there is a sufficiently small $\epsilon > 0$ such that $y'' + \epsilon < y'$ and for any $\bar{y} \in [y'', y'' + \epsilon]$, $(y, z) \parallel^{RTST} (y', z)$ and hence $(\bar{y}, z) \succ^* (y', z)$. If this is true, $[y'', y'' + \epsilon]$ is contained in $L(y'', z)\backslash L(y', z)$ and has positive measure (by (ii)), so again $u(y'', z) > u(y', z)$. It remains for us to show that $\epsilon > 0$ exists. If it does not exist, then there must be a sequence $y_n > y''$ and tending towards y'' such that $(y', z) \succsim^{RTST} (y_n, z)$ (since, with a finite data set, it is impossible for there to be a sequence y_n tending y'' such that $(y_n, z) \succsim^{RTST} (y', z)$). This leads to $(y', z) \succsim^{RTST} (y'', z)$, which is a contradiction.[38] QED

Proof of Theorem 2

It suffices to see that (b) implies (c). First, the SRM axiom guarantees that $\succsim^{RT}$ is antisymmetric. This in turn implies that $\succsim^{RTST}$ is antisymmetric. By Lemma A5, $\succsim^*$, as defined by (28) obeys SSCD and rationalizes $\mathcal{O}$. Lastly, it is clear from its definition that $(y'', z) \succsim^* (y', z)$ and $(y', z) \succsim^* (y'', z)$ only if $(y'', z) \succsim^{RTST} (y', z)$ and $(y', z) \succsim^{RTST} (y'', z)$, but the latter is impossible. QED

Proof of Theorem 4

[38] In general, if a sequence y_n tends to $y'' \in \mathcal{Y}$, and $(y', z) \succsim^{RTST} (y_n, z)$ for all n, then $(y', z) \succsim^{RTST} (y'', z)$. Analogous closure properties are true of $\succsim^{RT}$ and $\succsim^{RTS}$. It is straightforward to check that these properties follow from the finiteness of the data set and the compactness of the sets in $\mathbb{I}(\mathcal{Y})$.

For each $t \in \mathcal{T}$, we may regard $\mathrm{P}(\cdot|z^t, A)$ as an element of $\Delta = \left\{y \in \mathbb{R}_+^{|A|} : \sum_{k=1}^{|A|} y_k = 1\right\}$, and $\mathcal{P}$ as an element of Δ^T. We simplify our notation and denote $\mathrm{P}(\cdot|z^t, A)$ by $\mu^t \in \Delta$; a cross sectional data set $\mathcal{P}$ can be written as $\mathcal{P} = \{(\mu^t, z^t, A)\}_{t\in\mathcal{T}}$. We denote the set of paths obeying RM axiom by $\mathcal{Y}^*$; each path can also be regarded as an element of Δ^T. Let Δ^T_{RM} (contained in Δ^T) be the set of all possible $\{\mu^t\}_{t\in\mathcal{T}}$ such that $\{(\mu^t, z^t, A)\}_{t\in\mathcal{T}}$ is stochastically rationalizable. By Theorem 3, this set is the convex hull of $\mathcal{Y}^*$. Since $A^t = A$ for all $t \in \mathcal{T}$, $\mathcal{Y}^*$ consists precisely of those paths where a higher parameter leads to a weakly higher action; it follows immediately from this that Δ^T_{RM} is contained in Δ^T_{FSD}, the set of $\{(\mu^t, z^t, A)\}_{t\in\mathcal{T}}$ that obey first order stochastic dominance in the sense that $\mu^t \geqslant_{\mathrm{FSD}} \mu^s$ whenever $z^t > z^s$. Both Δ^T_{RM} and Δ^T_{FSD} are convex and compact sets in Δ^T. The Krein-Milman Theorem tells us that Δ^T_{FSD} is the convex hull of its extreme points; therefore, to show that $\Delta^T_{\mathrm{RM}} = \Delta^T_{\mathrm{FSD}}$ (as the theorem claims), we need only show that any extreme point of Δ^T_{FSD} is an element of $\mathcal{Y}^*$. Equivalently, we shall show the following: if $\{\mu^t\}_{t\in\mathcal{T}} \in \Delta^T_{\mathrm{FSD}}$ is not in $\mathcal{Y}^*$, then it is not an extreme point of Δ^T_{FSD}.

Suppose $\{\mu^t\}_{t\in\mathcal{T}} \in \Delta^T_{\mathrm{FSD}} \setminus \mathcal{Y}^*$ and for each $t \in \mathcal{T}$, let $m^t \in A$ be the median of μ^t, i.e., $m^t = \inf\left\{a : \sum_{a\leqslant y} \mu^t(y) \geqslant 0.5\right\}$. Let α^t be a distribution defined in the following manner: $\alpha^t(y) = 2\mu^t(y)$ if $y < m^t$; $\alpha^t(y) = 1 - 2\sum_{y<m^t} \mu^t(y)$ if $y = m^t$; $\alpha^t(y) = 0$ if $y > m^t$. We also define the distribution β^t: $\beta^t(y) = 0$ if $y < m^t$; $\beta^t(y) = 1 - 2\sum_{y>m^t} \mu^t(y)$ if $y = m^t$; and $\beta^t(y) = 2\mu^t(y)$ if $y > m^t$. Clearly, it holds that $\mu^t = 0.5\alpha^t + 0.5\beta^t$ for all t. Since $\{\mu^t\}_{t\in\mathcal{T}} \notin \mathcal{Y}^*$, there exists $t \in \mathcal{T}$ for which this convex combination is non-degenerate; therefore, $\{\mu^t\}_{t\in\mathcal{T}}$ is not an extreme point of Δ^T_{FSD} if $\{\alpha^t\}_{t\in\mathcal{T}}$ and $\{\beta^t\}_{t\in\mathcal{T}}$ are both in Δ^T_{FSD}. We only show this for $\{\alpha^t\}_{t\in\mathcal{T}}$ since the other case is similar. Suppose $z^t > z^s$ for some $s, t \in \mathcal{T}$. Since $\{\mu^t\}_{t\in\mathcal{T}}$ is in Δ^T_{FSD} it must hold that $m^s \leqslant m^t$. If $a < m^s \leqslant m^t$, it follows from $\{\mu^t\}_{t\in\mathcal{T}} \in \Delta^T_{\mathrm{FSD}}$ that

$$\sum_{y\leqslant a} \alpha^t(y) = 2\sum_{y\leqslant a} \mu^t(x) \leqslant 2\sum_{x\leqslant a} \mu^s(y) = \sum_{y\leqslant a} \alpha^s(y).$$

If $a \geqslant m^s$, then $\sum_{y\leqslant a} \alpha^t(y) \leqslant \sum_{y\leqslant a} \alpha^s(y) = 1$. We conclude that $\alpha^t \geqslant_{\mathrm{FSD}} \alpha^s$. QED

References

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

Aradillas-Lopez, A., and A. Gandhi (2016): Estimation of games with ordered actions: An application to chainstore entry. *Quantative Economics*, 7, 727–780.

Athey, S., and S. Stern (1998): An empirical framework for testing theories about complimentarity in organizational design. *NBER Working Paper*, No. 6600.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): Estimating static models of strategic interactions. *Journal of Business & Economics Statistics*, 28, 469 –482.

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*, 2004-26.

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.*

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

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

Kitamura, Y., and J. Stoye (2016): Nonparametric analysis of random utility models. *Cemmap Working Paper*, CWP27/16. (Conditionally accepted, *Econometrica*.)

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

Kukushkin, N., J. K.-H. Quah, and K. Shirai (2016): A counterexample on the completion of preferences with single crossing differences. *Munich Personal REPEc Archive 73760.*

Lazzati, N., J. K-H. Quah, and K. Shirai (2018): Strategic complementarity in spousal smoking behavior, *mimeo.*

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., 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.

Quah, J. K.-H., and B. Strulovici (2009): Comparative statics, informativeness, and the interval dominance order. *Econometrica*, 77, 1949-1999.

Seim, K. (2006): An empirical model of firm entry with endogenous producttype choices. *RAND Journal of Economics*, 37, 619–640.

Sweeting, A. (2009): The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *RAND Journal of Economics*, 40, 710–742.

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 (not part of main paper)

Data and testing procedures

We implement the results of Section 4 to a model of a binary entry game, with data from airline markets. The data is taken from Kline and Tamer (2016). The empirical question concerns the entry behavior of two kinds of firms: LCC (low cost carriers) and OA (other airlines). A firm that is not an LCC is by definition an OA. The dataset comes from the second quarter of the 2010 Airline Origin and Destination Survey (DB1B). It contains observations from 7882 markets that are defined as trips between two airports irrespective of intermediate stops.

In Kline and Tamer's analysis (and in ours) the two categories of firms are treated as two players. So the LCC player (respectively OA) enters if some LCC (OA) firm enters the market. Thus in each market, the two players (LCC and OA) can either both enter, both stay out, or one could enter with the other staying out. The data set also contains information on two explanatory variables: market presence (P) and market size (S). Market presence is a market- and airline-specific variable. For each airline and for each airport, they count the number of markets that 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 the two endpoints of the trip. Since the airlines are aggregated into two firms (LCC and OA), the market presence variable is also aggregated: the market presence for the LCC firm (resp. OA firm) is the maximum among the actual airlines in the LCC category (resp. OA category). The second explanatory variable is market size, which is a market-specific variable (shared by all airlines in that market) and is defined as the population at the endpoints of the trip. Lastly, Kline and Tamer discretize the market size and market presence variables to take two values: 1 if the variable is higher than its median value and 0 otherwise. Thus, for each market, the covariates are a triplet P_{LCC}, P_{OA}, S in the set product $\{0,1\} \times \{0,1\} \times \{0,1\}$. The distribution at each observable covariate is presented in Table 4 in the main paper.

Test and closest compatible distribution

Testing whether a data set is SID-rationalizable involves checking whether a linear system

$$Ax = B \tag{30}$$

has a positive solution x. We describe next all the components of this system.

Matrix A: This matrix is composed of 0s and 1s. Each column describes the behavior (in terms of choices) of a specific group path that satisfies the SRM axiom. Recall that a group path specifies the profile of choices that the group makes for each possible vector of covariate values (or treatments). Each row of A corresponds to one of the 32 possible combinations of (joint) entry choices and treatment values. Since there are 482 group paths that obey the SRM axiom, A is a 32×482-matrix. In Sheet "Consistent Paths" of the file "Matrices, Data, and Results.xlsx" (included with this submission as a separate file) we describe all possible group paths for the entry game application; all paths that satisfy the SRM axiom get number 1 in column SRM —SRM 1 and SRM 2 check the SRM axiom for firm LCC and OA, respectively. In Sheet "Matrix A" of "Matrices, Data, and Results.xlsx" we show how to construct matrix A in our application. We also considered the case where a player is allowed to be indifferent between entering and not entering a market. When indifferences are allowed, group paths must obey the RM (rather than SRM axiom) and the number of paths increases to 1,809, so that A is a $32 \times 1,809$-matrix. In these two sets of Sheets we write NI if we do not allow for indifferences and I if we allow for them.

Vector B: The size of this column vector is 32. It is composed of 8 conditional probability distributions. Each conditional distribution specifies the fraction of groups in the data that, for a given treatment, make each of the four possible joint choices. Sheet "Data" of "Matrices, Data, and Results.xlsx" describes all the information from the available data on the airlines that we use to construct vector B, and shows how to construct it.

Vector x: This vector represents a probability distribution over the set of SID-rationalizable group paths —whenever the system has a positive solution. In the entry game application, x has 482 entries when we do not allow for indifferences.

We implement our test by using Matlab. Specifically, we use the built-in function

$$x = \text{linprog}(lb, [], [], A, B, lb, [])$$

to check whether system (30) has a positive solution in x. In this specification, inputs A and B are described as above and lb corresponds to a column vector of zeros with the size of vector x. When no solution exists, Matlab reports that the primal solution appears infeasible.

For those data vectors B that do not pass this test, we use built-in function "lsqnonneg" in Matlab to find a positive vector $\widehat{x}$, with its components adding up to 1, that minimizes $(B - Ax)^{'} (B - Ax)$. We refer to $A\widehat{x}$ as the closest compatible distribution of choices. Sheet "Results" of "Matrices, Data, and Results.xlsx" describes the closest compatible vectors in columns "Closest" for the model that does not allow indifferences and allows them, respectively.

Small sample inference procedure

As Kitamura and Stoye (2016) explain, the null hypothesis is equivalent to

$$\mathbf{H}\text{:}\ \min\nolimits_{x \in \mathbb{R}_+^K} (B - Ax)' (B - Ax) = 0$$

where K is the number of group paths. (In the entry game without allowing for indifference, $K = 482$.) A natural sample counterpart of the objective function in $\mathbf{H}$ is given by

$$\left(\widehat{B} - Ax\right)' \left(\widehat{B} - Ax\right)$$

where $\widehat{B}$ estimates B by sample choice frequencies. Normalizing the latter by sample size N, we get

$$J_N = N \min\nolimits_{x \in \mathbb{R}_+^K} \left(\widehat{B} - Ax\right)' \left(\widehat{B} - Ax\right).$$

Let x^{**} be any solution to this problem. If $Ax^{**} = \widehat{B}$, so that the observed choices are compatible with our restrictions, then $J_N = 0$ and the null hypothesis cannot be rejected.

Kitamura and Stoye (2016) propose the following bootstrap algorithm to test $\mathbf{H}$:

(i) Obtain a vector x^* that solves

$$J_N = N \min_{[x - \tau_N 1_K] \in \mathbb{R}_+^K} \left(\widehat{B} - Ax\right)' \left(\widehat{B} - Ax\right)$$

and compute $\widehat{C}_{\tau_N} = Ax^*$. Here, 1_K is a vector of 1s of dimension K.

Broadly following the practice in Kitamura and Stoye (2016), we report results for $\tau_N = 0$ and for $\tau_N = 5.15 \times 10^{-7} = \sqrt{\ln(N)/N}/65,536$, where the denominator is the number of all possible paths (including those that violate the RM axiom) and $N = 7882$ (the sample size). (The tuning parameter τ_N plays the role of a similar tuning parameter in the moment selection approach; in Kitamura and Stoye (2016), it is required to satisfy two properties, namely, $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$.)

(ii) Calculate the boostrap estimators under the restriction

$$\widehat{B}_{\tau_N}^{(r)} = \widehat{B}^{(r)} - \widehat{B} + \widehat{C}_{\tau_N} \quad r = 1, ..., R$$

where $\widehat{C}_{\tau_N}$ derives from step (i) and $\widehat{B}^{(r)}$ is a re-sampled choice probability vector obtained via standard nonparametric boostrap. In addition, R is the number of boostrap replications. In our paper, we let $R = 2000$.

(iii) Calculate the boostrap test statistic by solving the following problem

$$J_N^{(r)}(\tau_N) = N \min_{[x - \tau_N 1_K] \in \mathbb{R}_+^K} \left(\widehat{B}_{\tau_N}^{(r)} - Ax\right)' \left(\widehat{B}_{\tau_N}^{(r)} - Ax\right)$$

for $r = 1, ..., R$.

(iv) Use the empirical distribution of $J_N^{(r)}(\tau_N)$, $r = 1, ..., R$, to obtain the critical value of J_N.

We repeat this procedure four times: for the model that does not allow indifferences (NI) and the one that allows them (I) and we implement the test under two specifications of the τ_N-parameter. For the airline application, we obtain the following p-values.

	NI	I
$\tau_N = 0$	0.150	0.307
$\tau_N = 5.15 \times 10^{-7}$	0.131	0.273

In Sheet "Results" of "Matrices, Data, and Results.xlsx" we expand on these findings. In particular, we also provide information regarding the closest compatible distribution for $\tau_N = 0$ (in column "Closest") and for $\tau_N = 5.15 \times 10^{-7}$ (in column "Closest τ").

Bounds for subset of types

Suppose the dataset is consistent with our modeling restrictions, i.e., it passes the test in the previous sub-section. This means that the set of types that our model allows can explain the observed choices. In this context, we might also want to recover, from the data, the relative importance of a specific subset of types allowed by the model. In general, this fraction can only be partially identified.

Recall that x represents a probability distribution over the set of all permissible paths. Let us define a vector ρ of equal length with x, such that the entry in ρ is 1 on all those types included in the subset of interest and 0 otherwise. For each x, the probability weight on the subset of interest is just $x'\rho$. Let $\widehat{C}$ be the closest compatible distribution of choice data (if the dataset passes the test, then $\widehat{C}$ is just $\widehat{B}$). We can construct bounds for $x'\rho$ by solving

$$\min_{x \in \mathbb{R}_+^K} \left\{ x'\rho : Ax = \widehat{C} \right\} \text{ and } \max_{x \in \mathbb{R}_+^K} \left\{ x'\rho : Ax = \widehat{C} \right\}.$$

The "Estimated Probability" in Tables 7 and 8 in the main paper is calculated in this way.

Finally, the confidence interval for the identified set can be constructed using a procedure developed in Deb, Kitamura, Quah, and Stoye (2017). Let $\theta \in [0, 1]$ be a specific weight for the group of interest. We can then use our previous test to verify whether this fraction is consistent with the empirical evidence by adding the restriction $x'\rho = \theta$ to the initial hypothesis $\mathbf{H}$. Bounds for the relevance of a specific subset of types can be recovered by implementing the previous test for all $\theta \in [0, 1]$ and including in the confidence interval all those values of θ for which the p-value is above 5%. This test is relatively easily to carry out and we implement it to obtain confidence intervals for various subsets of SID-rationalizable paths. The confidence intervals reported in Tables 7 and 8 of the main paper correspond to the case where the tuning parameter is set at zero but the results are robust to small modifications of the tuning parameter.