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

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

November 29, 2023

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

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

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

* Email addresses: nlazzati@ucsc.edu; john.quah@jhu.edu; kshirai1985@kwansei.ac.jp

For helpful discussions and comments, the authors are grateful to S. Berry, J. Fox, K. Hirano, T. Hoshino, A. Kajii, Y. Kitamura, B. Kline, E. Krasnokutskaya, C. Manski, W. Newey, T. Sekiguchi, J. Stoye, B. Strulovici, S. Takahashi, Y. Takahashi, and especially X. Tang. Various versions of this project have been presented to audiences at the following events and we are grateful for their comments: seminars at University of Arizona, Johns Hopkins, Kyoto, Louvain (CORE), New York University, Rice University, Simon Fraser University, the National University of Singapore, Northwestern University, University of Paris (Dauphine), Queensland, Shanghai University of Finance and Economics, University of Southern California, Singapore Management University, Stanford, UC Davis, UC San Diego, the Canadian Economic Theory Conference (Vancouver, 2017), the Conference on Econometrics for Incomplete Models (CeMMAP and Northwestern, 2018), 13th Greater New York Metropolitan Area Econometrics Colloquium (Princeton University, 2018), and the Econometric Society North American Summer Meeting (UC Davis, 2018). Koji Shirai gratefully acknowledges financial support from the Japan Society for Promotion of Science (KAKENHI 19K00155) and the hospitality of Johns Hopkins University during his visit in the 2019-2020 academic year.

1 Introduction

Economic analysis is often concerned with the effect of an exogenous or strategic variable on an agent's decision: would a consumer buy less of good A if the price of good B falls? would a firm follow its rival when the latter raises its price? is someone more likely to join a demonstration if it is known that more people are participating? The theory of monotone comparative statics identifies properties on payoff functions, such as the single crossing property (see Milgrom and Shannon (1994)), that are necessary and sufficient for optimal choices to be monotone with respect to opponents' strategies and exogenous variables. The empirically relevant follow-up question is the following: what kind of *observed* choice behavior would be necessary and sufficient for the recovery of payoff functions obeying the single crossing property? The key contribution of this paper is to answer this revealed preference question and to show that it forms the basis of an econometric analysis of games with strategic complements.

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

Our approach has as its starting point a data set of the same type as the papers cited above. With this data set we can test whether firms are playing pure strategy Nash equilibria (PSNE), subject to single crossing restrictions on its payoff functions. For example, we can test the hypothesis that a firm's entry into a market is encouraged when the market is large and discouraged when another firm is also entering. Our method works without imposing any parametric assumptions on payoff functions, without assuming that unobserved heterogeneity is additive or that its distribution belongs to a particular family, and without assumptions on equilibrium selection. By specifying a *joint* distribution on the payoff functions, we allow for correlation or other forms of dependence among firms' payoff functions, which is important in many settings (see Chen, Christensen and Tamer (2018)). To pass our test means that the hypothesis that the data are explained as PSNE by firms with payoff functions satisfying single crossing restrictions cannot be refuted.

At its most basic, our approach provides a way for researchers to test the general (nonparametric) features of a model, before the implementation of a more restrictive parametric model that could be used for inference and prediction. In some cases, the confirmation of monotone features which are part of our test could also facilitate estimation procedures.[1] Beyond this, since our test recovers the distributions on firms' payoff functions that satisfy single crossing restrictions and agree with the observations, the procedure can also be extended for the purposes of inference and prediction (when the data set passes the test).

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

Our econometric approach is similar to that in Kitamura and Stoye (2018) (henceforth KS).[2] This paper tests a random utility model of consumer demand. In the first step, it is assumed that the population distribution of consumer demand at a linear budget set B, which we denote by $\mathrm{P}(\cdot|B)$,

[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 Tamer (2003) and Kline and Tamer (2016)). Shape restrictions can also 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] Our approach is also close in spirit, though not in its specifics, with the nonparametric random utility models studied in Tebaldi, Torgovitsky, and Yang (2023), Deb, Kitamura, Quah and Stoye (2022), Apesteguia, Ballester and Lu (2017), Hoderlein and Stoye (2014), Manski (2007), McFadden (2005), McFadden and Richter (1991), and Marschak (1960). As far as we know, our paper is the first to exploit this nonparametric approach to study games. Note that Kitamura and Stoye's empirical approach (and hence our approach, too) is based on linear programming, which can be also found in earlier works, for example, such as Honoré and Tamer (2006) and Chernozhukov, Fernàndez-Val, Hahn and Newey (2013).

is known for a finite collection of budget sets $\mathcal{B}$. Then one could formulate necessary and sufficient conditions under which this idealized data set $\mathcal{P}_{\mathrm{KS}} = \{\mathrm{P}(\cdot|B)\}_{B\in\mathcal{B}}$ is generated by a population of utility-maximizing consumers, under the *conditional independence* assumption; this assumption requires the distribution of utility functions (which generates the distribution of demand) to be the same at each budget set $B \in \mathcal{B}$. The characterization of $\mathcal{P}_{\mathrm{KS}}$ in KS is facilitated by the well-known characterization of utility-maximizing demand behavior for a *single* consumer, known as the strong axiom of revealed preference (SARP). The second step in the KS approach is to show how the characterizing conditions on $\mathcal{P}_{\mathrm{KS}}$ could be statistically tested for an actual data set, with empirical frequencies at each budget set $B \in \mathcal{B}$.

The key observation in our paper is that the procedure implemented in KS could also be used for analyzing specific classes of games, provided certain obstacles are overcome. We have the same two-step approach as in KS. We assume that there is a large population of groups, with each group playing the same game. To begin, we assume that the population distribution over joint action profiles at a given vector of covariate values $\mathbf{x}$, which we denote by $\mathrm{P}(\cdot|\mathbf{x})$, is known for a finite set of covariate values $\widehat{\mathbf{X}}$.[3] (The set $\widehat{\mathbf{X}}$ takes the place of $\mathcal{B}$ in the KS model.) Then we can formulate necessary and sufficient conditions under which the idealized data set $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x})\}_{\mathbf{x}\in\widehat{\mathbf{X}}}$ is consistent with a population of groups made up of agents having payoff functions satisfy single crossing conditions and playing PSNE, under the assumption of conditional independence (which in this case means that the distribution of payoff function profiles across groups is the same at different $\mathbf{x} \in \widehat{\mathbf{X}}$). The second step in our approach shows how the characterizing conditions on $\mathcal{P}$ could be statistically tested on an actual data set with empirical frequencies over action profiles at different covariate values; for this second step, we simply follow the statistical procedure in KS.

Similar to KS, the characterization of $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x})\}_{\mathbf{x}\in\widehat{\mathbf{X}}}$ requires that we find necessary and sufficient conditions under which the joint actions from a *single* group at different covariate values are consistent with our hypothesis of PSNE play and payoff functions satisfying singe-crossing conditions (with respect to opponents' actions and covariates). Since, unlike KS, there is no ready-made characterization for this class of games, we need to develop it ourselves. We show that this

[3] Variation of feasible sets (as in KS) can be included in our analysis of games (see Lazzati, Quah and Shirai (2018)), but we have avoided it, in order not to burden the reader with too many model features and also because our empirical application does not have such variation. (See also Carvajal (2004) for a related result.)

hypothesis can be characterized by a property we call the revealed monotonicity (RM) axiom. This axiom plays the role of SARP in the KS model.

When the data set passes the test, our approach is in turn useful for making inference and prediction in the spirit of Deb, Kitamura, Quah and Stoye (2022), which deals with a version of the consumer model. For example, we can estimate the fraction of players who are effectively nonstrategic, in the sense that their actions depend only on covariate values and are independent of what other players do. We can also bound the proportion of groups which (at a given covariate vector) has a particular equilibrium profile as a PSNE (along the lines of the analysis in Aradillas-Lopez (2011)); note that this potentially differs from the observed fraction of groups playing that action profile, not just because of sampling variation, but also because a given action profile could be a non-chosen PSNE when there are multiple PSNE.

The procedure in KS is hard to implement when there is a large number of budget sets and Smeulders, Cherchye, and De Rock (2021) propose a column generation method to deal with this difficulty. This method is also applicable in our setting and is useful in easing the computational burden of our test when (for example) $\widehat{\mathbf{X}}$ is a big set. In our paper, we develop a new result on column generation that allows for this method to be used, not just for testing but also inference.

The rest of the paper is organized as follows. In Section 2, we provide an outline of how our procedure works in the context of an entry game and contrast it with a parametric approach. Section 3 presents our main results at the population level. We introduce the revealed monotonicity axiom and use it to characterize those distributions over joint actions that are consistent with our hypothesis; properties of the underlying distribution over payoff function profiles can also be recovered. Section 4 explains how the population-level analysis in Section 3 can be implemented on finite sample data. In this section we also introduce and extend the column generation method of Smuelders et al. (2021). To illustrate our approach, we carry out an empirical analysis of entry decisions made by airlines; this is found in Section 5.

2 Motivating example

There is a large empirical literature modelling oligopoly entry decisions. We shall use this model to illustrate the basic question we are interested in and the approach we propose to address this

$x_2 = (0,0)$		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 1: $\mathcal{P} = \{\mathrm{P}(\cdot \mid x_2 = (0,0)),\ \mathrm{P}(\cdot \mid x_2 = (0,1)),\ \mathrm{P}(\cdot \mid x_2 = (1,0))\}$

question. For simplicity, we treat the case of two firms. Let $y_i \in \{N, E\}$ be the action set of firm i, where E means that the firm enters the market and N that it stays out and let x_i be a real-valued, finite-dimensional vector of exogenous profit shifters (covariates) that affect firm i's profit and are observed by the other firm and the researcher.

We assume that there is a large population of markets, with each market consisting of a Firm 1 and a Firm 2 that make their entry decisions simultaneously. The designation of a player as Firm 1 or Firm 2 is made by the researcher and based on observable characteristics; for example, in Kline and Tamer (2016), one firm is 'Low-Cost Carrier' and the other firm 'Other Airlines' (see Section 5). There is a finite set of realized profit shifters which we denote by $\hat{\mathbf{X}}$. For each $(x_1, x_2) \in \hat{\mathbf{X}}$, we suppose that the population distribution of joint action profiles $\mathrm{P}(\cdot \mid x_1, x_2)$ is known to the researcher. The idealized data set can then be succinctly written as $\mathcal{P} = \{\mathrm{P}(\cdot \mid (x_1, x_2))_{(x_1,x_2)\in\hat{\mathbf{X}}}$. An example of $\mathcal{P}$ is given Table 1, where we assume there is only variation in x_2 and it takes three possible vector values; for example, the box in the left gives the value of $\mathrm{P}(\cdot \mid x_2 = (0,0))$ with (say) $\mathrm{P}((E,N) \mid x_2 = (0,0)) = 4/12$. We are interested in developing a procedure which allows us to identify those $\mathcal{P}$ which are compatible with our model of firm entry. Of course, in any empirical analysis these characterizing conditions on $\mathcal{P}$ would have to be statistically tested on an actual data set with sampling variation (as we explain in detail in Section 4). Confining our discussion to $\mathcal{P}$ at this stage allows us to focus on the more distinctive aspects of our analysis.

We now describe the model which (potentially) generates $\mathcal{P}$. We denote the payoff/profit of firms 1 and 2 by $\Pi_1(y_1, y_2, x_1)$ and $\Pi_2(y_1, y_2, x_2)$ respectively. We postulate that entry decisions are generated as pure strategy Nash equilibria (PSNE) of an entry game between Firms 1 and 2. We allow for multiple PSNE and impose no restriction on how firms select among these equilibria. There remains unobserved market heterogeneity even after conditioning on (x_1, x_2); this heterogeneity is captured by a joint distribution on (Π_1, Π_2), which in turn leads to a distribution over joint actions

$\mathrm{P}(\cdot \mid x_1, x_2)$. We assume that there is *conditional independence*, in the sense that the distribution over (Π_1, Π_2) does not vary with the realized value of (x_1, x_2).

Lastly, we postulate that the firms' profit functions satisfy single-crossing restrictions (see Milgrom and Shannon (1994)). In this context, it means that Firm 1's entry into the market is encouraged when the profit shifter x_1 takes higher values and is discouraged when Firm 2 chooses to enter. Formally, we require

$$\Pi_1 (E, y_2', x_1') > \Pi_1 (N, y_2', x_1') \Longrightarrow \Pi_1 (E, y_2'', x_1'') > \Pi_1 (N, y_2'', x_1'') \tag{1}$$

whenever $x_1'' \geqslant x_1'$ and either $y_2' = y_2''$ or $y_2' = E$ and $y_2'' = N$. (A similar requirement is imposed on Π_2.) For example, in Ciliberto and Tamer (2009),

$$\Pi_1 (y_1, y_2, x_1) = \begin{cases} \alpha_1' x_1 + \delta_1 \mathbf{1}_{y_2} + \varepsilon_1 & \text{if} \quad y_1 = E \\ 0 & \text{if} \quad y_1 = N, \end{cases} \tag{2}$$

where $\mathbf{1}_E = 1$ and $\mathbf{1}_N = 0$. In this specification, the entry of Firm 2 alters the profit of Firm 1 by δ_1 and unobserved heterogeneity in payoff functions is captured by ε_1, which enters the profit function additively. It is straightforward to check that our single crossing restrictions are satisfied if $\delta_1 < 0$ and $\alpha_1 > 0$. Note, however, that the converse is not true, i.e., there are distributions over payoff functions satisfying (1), that *cannot* be represented in the additive form given by (2), for any distribution on ε_1.

We say that $\mathcal{P}$ is consistent with the single-crossing model, or $\mathcal{SC}$*-rationalizable*, if there there is a joint distribution of payoff functions (Π_1, Π_2) that satisfy our single crossing conditions (1) such that the resulting distribution of PSNE (given some equilibrium selection rule) coincides with $\mathrm{P}(\cdot \mid x_1, x_2)$ for each $x \in \widehat{\mathbf{X}}$. We would like to answer the following question: what conditions on $\mathcal{P}$ characterize $\mathcal{SC}$-rationalizability? In other words, when presented with $\mathcal{P}$, how could we check if it is $\mathcal{SC}$-rationalizable?

We first observe that our model *does* have structural implications for $\mathcal{P}$. Consider an increase in the observable profit shifters from (x_1', x_2') to (x_1'', x_2''); then, at any particular realization Π_1 of Firm 1's payoff function, if it prefers to enter when the other firm enters at (x_1', x_2'), then the single-

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

Table 2: Distribution of types rationalizing data in Table 1

crossing condition guarantees that it will continue to prefer entry at (x_1'', x_2''). The same argument applies to Firm 2, and so we conclude that if (E, E) is the Nash equilibrium at (x_1', x_2') for a given realized profit function profile (Π_1, Π_2), then it will be the *unique* Nash equilibrium at (x_1'', x_2'') for this realized profile. Aggregating across all profiles, we establish that

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

provided conditional independence holds. This inequality constitutes a restriction on $\mathcal{P}$ but it is not the only restriction imposed by our model. We now sketch out the procedure for systematically checking whether $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, using $\mathcal{P}$ presented in Table 1 as an example.

Given a particular realization (Π_1, Π_2), the firms will choose an action profile (either (E, E), (E, N), (N, E) or (N, N)) at each realization of x_2, and as x_2 takes different values the action profile of the two firms may change. We shall refer to the map from x_2 to the action profile as a *group type*. Notice that even though firms' profit functions may be heterogenous in infinitely many ways, its manifestation in behavior must be *finite*, since there are only finitely many possible actions and the realized covariates (x_1, x_2) takes values in the finite set $\widehat{\mathbf{X}}$.

To be precise, there are in total $4^3 = 64$ group types, but not all are consistent with PSNE play and single-crossing payoff functions. For example, as we have already explained, a group type where (E, E) is played at $x_2 = (0, 0)$ and (N, N) at $x_2 = (0, 1)$ is not compatible with single crossing. On the other hand, it is quite clear a group type where (N, E) is played at all three values of x_2 can be justified with single crossing profit functions.

Ascertaining if $\mathcal{P}$ can be rationalized involves a two-step procedure. Firstly, we must identify all *single-crossing group types*, in the sense that the action profile (y_1, y_2) at each value of (x_1, x_2) could be generated as PSNE from payoff functions satisfying (1). This is do-able because we show in Section 3 that these group types are characterized by an easy-to-check condition called the *revealed monotonicity axiom*. Secondly, we have to check whether there are weights on these group types that could account for the observed distribution of action profiles; this involves solving a system of linear inequalities.

We claim that $\mathcal{P}$ depicted in Table 1 can be rationalized. To understand why, we list in Table 2 seven possible group types. One could check that each of these group types is consistent with the single crossing property. When these types are represented in the population with the weights indicated in Table 2, they generate the distribution of entry decisions observed in Table 1. (Compare the entries in Table 1 with the last row of Table 2.)

Lastly, we point out that while $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, it is *not* compatible with a model where profit functions have the form (2), so the latter specification does involve a loss of generality. Indeed, with this specification, Firm 2's profit upon entry is

$$\pi_2(E, y_1, x_{21}, x_{22}, \varepsilon_2) = \alpha_{21} x_{21} + \alpha_{22} x_{22} + \delta_{21} \mathbf{1}_{y_1} + \varepsilon_2, \tag{3}$$

where $(\alpha_{21}, \alpha_{22}) > 0$ and $\delta_{21} < 0$.[4] Whether the boost to profits of an increase in x_{21} is greater or smaller than that obtained from the same increase to x_{22} depends on whether α_{21} is bigger or smaller than α_{22} and 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 this parametric model cannot explain the data in Table 1, 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,

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

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

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

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

and

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

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

In Online Appendix A1, we provide a more elaborate discussion of the contrast between the observable restrictions imposed by a linear parametric model and our (more general) nonparametric model. In particular, using simulations based on an extended version of the above example, we show that the difference between the two models is also picked up at the sample level: the method of Kline and Tamer (2016) (correctly) finds that the data are inconsistent with the linear model, whereas our method (also correctly) finds that the data are consistent with the more general model.

3 $\mathcal{SC}$-rationalizable distributions

In this section, we consider a population of groups that play pure strategy Nash equilibria (PSNE) within each group. We characterize how the distribution of action profiles in this population will change with covariates when agents have best responses that are monotone with respect to both covariates and the actions of other agents in the group.

3.1 Games with single-crossing payoff functions

We assume that there is a population of groups, and for each group, we denote the set of agents by $\mathcal{N} = \{1, 2, ..., n\}$. Agent $i \in \mathcal{N}$ chooses an action y_i from an action space Y_i, which we assume is finite and totally ordered. We denote a joint action profile of the group by $\mathbf{y} \in \mathbf{Y} = \times_{i \in \mathcal{N}} Y_i$. For each $i \in \mathcal{N}$, there is an $M(i)$-dimensional covariate $x_i \in X_i = \times_{m=1}^{M(i)} X_{im}$, and the profile of $x_i = (x_{i1}, x_{i2}, ..., x_{iM(i)})$ across agents is denoted by $\mathbf{x} = (x_i : i \in \mathcal{N})$.

The payoff of each agent over different actions in Y_i depends on the actions of the other agents in its group $\mathbf{y}_{-i} = (y_j : j \in \mathcal{N}, j \neq i) \in \mathbf{Y}_{-i} = \times_{j \in \mathcal{N}, j \neq i} Y_j$ and $x_i \in X_i = \times_{m=1}^{M(i)} X_{im}$, where X_{im} is a subset of $\mathbb{R}$. Thus the payoff of agent i is given by a function $\Pi_i : Y_i \times \mathbf{Y}_{-i} \times X_i \to \mathbb{R}$. We use $\mathbf{\Pi} = (\Pi_i : i \in \mathcal{N})$ to indicate a profile of payoff functions.

A pair of payoff functions and covariate profiles $(\mathbf{\Pi}, \mathbf{x})$ induces a game of complete information $\mathrm{G}(\mathbf{\Pi}, \mathbf{x})$. In what follows, we let $\mathbf{X} \subset \times_{i \in \mathcal{N}} X_i$ denote the set of conceivable joint realizations of covariates $\mathbf{x}$. Since some covariates may be shared by multiple agents, $\mathbf{X}$ may not be equal to the direct product of X_i's (as in our empirical application in Section 5). We denote the best response of each player i at $(\mathbf{y}_{-i}, x_i)$ by $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i) = \operatorname{argmax}_{y_i \in Y_i} \Pi_i(y_i, \mathbf{y}_{-i}, x_i)$: throughout this paper, we assume that agents have *strict* preferences over actions, so that $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i)$ has a unique value.[5] The set of pure strategy Nash equilibria (PSNE) of this game is defined as

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

Importantly, even if the best response of every agent is single-valued, there could be multiple PSNE.

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

Definition 1. The payoff function Π_i has *single-crossing differences in* $(y_i; (\mathbf{y}_{-i}, x_i))$ if the following holds:[6] for every $y_i'' > y_i'$ and $(\mathbf{y}''_{-i}, x_i'') > (\mathbf{y}'_{-i}, x_i')$,

$$\Pi_i\left(y_i'', \mathbf{y}'_{-i}, x_i'\right) > \Pi_i\left(y_i', \mathbf{y}'_{-i}, x_i'\right) \Longrightarrow \Pi_i\left(y_i'', \mathbf{y}''_{-i}, x_i''\right) > \Pi_i\left(y_i', \mathbf{y}''_{-i}, x_i''\right). \quad (4)$$

For simplicity, we may refer to such a payoff function as a *single-crossing payoff function.*

This condition states that if it is advantageous for agent i to choose a higher action y_i'' over a lower one y_i', then it remains advantageous to do so when other players raise their actions and/or covariates take higher values. The focus of our analysis is the set of payoff profiles

$$\mathcal{SC} = \{\mathbf{\Pi} = (\Pi_i)_{i \in \mathcal{N}} : \Pi_i \text{ is strict and has single crossing differences in } (y_i; (\mathbf{y}_{-i}, x_i))\}.$$

Note that single-crossing differences is an *ordinal* property since any strictly increasing transfor-

[5] This assumption is needed for the revealed preference analysis to be meaningful, as, otherwise, we could justify any behavior by simply claiming that each agent is indifferent among all elements in the action space.

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

mation of a function that obeys single-crossing differences will also obey single-crossing differences. Furthermore, since a player's best responses are pinned down by a player's *preference over actions*, $\mathrm{NE}(\mathbf{\Pi}, \mathbf{x}) = \mathrm{NE}(\widetilde{\mathbf{\Pi}}, \mathbf{x})$ whenever $\widetilde{\mathbf{\Pi}} = (\widetilde{\Pi}_i)_{i \in \mathcal{N}}$ is a strictly increasing transformation of $\mathbf{\Pi} = (\Pi_i)_{i \in \mathcal{N}}$, in the sense that $\widetilde{\Pi}_i$ is a strictly increasing transformation of Π_i for all i.

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

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

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

(ii) $\mathrm{NE}(\mathbf{\Pi}, \mathbf{x})$ *is non-empty.*[7]

This result says that single crossing differences guarantees that $\mathrm{G}(\mathbf{\Pi}, \mathbf{x})$ is a game of *strategic complements*, in the sense that a player optimally increases his action when other players raise theirs,[8] and that these games have pure strategy Nash equilibria. Furthermore, the best response of each player also increases with the (exogenous) covariate.

Remark 1. By reversing the signs of opponents' actions and/or covariates, one can guarantee *decreasing* best responses with single-crossing payoff functions. However, the existence of PSNE is no longer ensured unless the game is a two-player game with strategic substitutes (as in Section 2) or an aggregative game. In the latter, each player's payoff has the form $\Pi_i(y_i, \sum_{j \neq i} y_j, x_i)$, and one can test single crossing differences either in $(y_i; (\sum_{j \neq i} y_j, x_i))$ (the case of strategic complements) or in $(y_i; (-\sum_{j \neq i} y_j, x_i))$ (the case of strategic substitutes); the existence of PSNE in these cases is guaranteed (see Dubey, Haimanko and Zapechelnyuk (2006) and Jensen (2010)).

3.2 Rationalizability

Since the population consists of many groups with heterogenous preferences, even at a given covariate value $\mathbf{x}$, different groups will take different joint actions. This generates a conditional distribution over joint actions $\mathbf{y} \in \mathbf{Y}$, which we denote by $\mathrm{P}(\cdot \mid \mathbf{x})$. Throughout this section, we

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

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

assume that $\mathrm{P}(\cdot \mid \mathbf{x})$ is known for $\mathbf{x}'s$ contained in some *finite* subset $\widehat{\mathbf{X}} \subset \mathbf{X}$. In applications, it may be the case that $\widehat{\mathbf{X}} = \mathbf{X}$ but it is also possible for $\widehat{\mathbf{X}} \subsetneq \mathbf{X}$. The set $\mathbf{X}$ may or may not be finite, but it is important for our results that $\widehat{\mathbf{X}}$ is finite.[9] We wish to consider the conditions under which a set of choice distributions

$$\mathcal{P} = \left\{ \mathrm{P}(\cdot \mid \mathbf{x}) : \mathbf{x} \in \widehat{\mathbf{X}} \right\}$$

is consistent with (in other words, generated by) pure strategy Nash equilibrium play in games with single-crossing payoff functions. Note that different choices across groups can arise not only from *heterogeneity in payoff functions* but also from *heterogeneity in equilibrium selection rules* among PSNE (both of which are not directly observed by the researcher).

Conditionally independent random payoff functions. To capture preference heterogeneity, we assume that the profile of payoff functions, $\mathbf{\Pi} = (\Pi_i)_{i \in \mathcal{N}}$, is random and distributed according to $\mathrm{P}_{\mathbf{\Pi}}$. (Notice that we are abusing notation by using $\mathbf{\Pi}$ to denote both the random variable and a particular realization.) By specifying a *joint* distribution on the payoff functions, we allow for correlation or any other type of dependence across the payoffs of the group members. In particular, we can be agnostic about correlations arising from group formation processes in the population (of the type observed in Chen et al. (2018), for example). Let $\mathrm{P}_{\mathbf{\Pi}}|_{\mathbf{x}}$ be the distribution of payoff function profiles conditional on the realized values of the covariates $\mathbf{x}$; we assume that $\mathrm{P}_{\mathbf{\Pi}}$ satisfies *conditional independence* in the sense that it does not depend on $\mathbf{x}$, i.e., $\mathrm{P}_{\mathbf{\Pi}}|_{\mathbf{x}} = \mathrm{P}_{\mathbf{\Pi}}$ for all $\mathbf{x} \in \widehat{\mathbf{X}}$.

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

We are now in position to spell out precisely what it means for a set of distributions $\mathcal{P}$ to be

[9] In Kitamura and Stoye (2018) and Deb et al (2022), the analog to $\widehat{\mathbf{X}}$ is the set of the price vectors at which the distribution of demand is known, whereas the analog to $\mathbf{X}$ is the set of all strictly positive price vectors; for essentially the same reasons, it is also important in their analyses that the former is finite. Possible ways of extending to the case where (what we call) $\widehat{\mathbf{X}}$ is infinite are discussed briefly in Kitamura and Stoye (2018, Section 8) and their observations are also potentially applicable here.

consistent with PSNE in games with single-crossing payoff functions.

Definition 2. A distribution $\mathrm{P}_{\mathbf{\Pi}}$, with support on $\mathcal{SC}$, *rationalizes* the set of choice distributions $\mathcal{P}$ if there is an equilibrium selection mechanism $\lambda\left(\cdot|\mathbf{\Pi},\mathbf{x}\right)$ such that

$$\mathrm{P}(\mathbf{y}|\mathbf{x}) = \int \lambda\left(\mathbf{y}|\mathbf{\Pi},\mathbf{x}\right) d\mathrm{P}_{\mathbf{\Pi}} \textit{ for all } \mathbf{y} \in \mathbf{Y} \textit{ and all } \mathbf{x} \in \widehat{\mathbf{X}}. \tag{5}$$

$\mathcal{P}$ is *single-crossing rationalizable* (or *$\mathcal{SC}$-rationalizable*) if it admits such a distribution $\mathrm{P}_{\mathbf{\Pi}}$; in other words, there is a distribution among payoff function profiles in $\mathcal{SC}$ and an equilibrium selection rule that could account for the observed distribution of joint actions at each $\mathbf{x} \in \widehat{\mathbf{X}}$.

Remark 2. $\mathcal{SC}$-rationalizability requires the domain of each agent's recovered payoff function to be $Y_i \times \mathbf{Y}_{-i} \times X_i$, rather than $Y_i \times \mathbf{Y}_{-i} \times \mathrm{proj}_i\,\widehat{\mathbf{X}}$, and single crossing differences must also be satisfied on the entire domain, i.e., even for actions that are available but not chosen and covariate values that are not part of $\widehat{\mathbf{X}}$.[10]

Remark 3. We adopt conditional independence in this paper, because it is a tractable and quite prevalent restriction in empirical work. There are ways to weaken or modify this condition. If the modeler has a specific belief about the way that $\mathbf{\Pi}$ depends on the covariate $\mathbf{x}$ other than conditional independence, then it may be possible to replace conditional independence with the new condition and develop a characterization for this modified notion of $\mathcal{SC}$-rationalizability (see Remark 4), along with a statistical procedure to test it. Alternatively, in applications where conditional independence may be suspect, but instrumental variables are available, one may develop control variables for which the distribution of the payoff function profiles (after conditioning on the control variables) is independent of $\mathbf{x}$; with this one could calculate endogeneity-corrected distributions on actions, to which our results are applicable (see Kitamura and Stoye (2018)).

3.3 The revealed monotonicity (RM) axiom

We now explain how $\mathcal{SC}$-rationalizable distributions may be characterized. The characterization has two parts and generalizes the procedure we used in the illustration in Section 2. Firstly,

[10] Note that $\mathrm{proj}_i\,\widehat{\mathbf{X}}$ is the projection of $\widehat{\mathbf{X}}$ to the set of possible values of x_i. That is, letting $\mathbf{x}_{-i}$ be a profile of covariates of agents other than i, $\mathrm{proj}_i\,\widehat{\mathbf{X}} = \{x_i : (x_i, \mathbf{x}_{-i}) \in \widehat{\mathbf{X}} \text{ for some } \mathbf{x}_{-i}\}$.

we characterize all group types that could be generated by payoff functions with single crossing functions. Secondly, we find weights on these types that could account for the distributions in $\mathcal{P}$.

I: Single-crossing rationalizable group types

A *group type* associates a profile of actions $\mathbf{y}$ to each covariate $\mathbf{x} \in \widehat{\mathbf{X}}$. Formally, it is a function B: $\widehat{\mathbf{X}} \to \mathbf{Y}$. For reasons which will be clear later, it is convenient to generalize the notion of group types to correspondences; thus a *generalized group type* is a map from $\widehat{\mathbf{X}}$ to a nonempty subset of $\mathbf{Y}$. We could interpret a generalized group type as a set of observations generated by a group of players with fixed preferences, where $\mathrm{B}(\mathbf{x})$ consists of the action profiles that are played at the game with covariate $\mathbf{x}$. We wish to characterize all group types where $\mathrm{B}(\mathbf{x})$ consists of PSNE and players have payoff functions $\mathbf{\Pi}$ in $\mathcal{SC}$.

Definition 3. A generalized group type $\mathrm{B} : \widehat{\mathbf{X}} \rightrightarrows \mathbf{Y}$ is a *single-crossing group type* if there exists a profile of payoff functions $\mathbf{\Pi}$ in $\mathcal{SC}$ such that $\mathrm{B}(\mathbf{x}) \subset \mathrm{NE}(\mathbf{\Pi}, \mathbf{x})$ for all $\mathbf{x} \in \widehat{\mathbf{X}}$.

The next definition provides the key observable feature of single-crossing group types.

Definition 4. A generalized group type $\mathrm{B} : \widehat{\mathbf{X}} \rightrightarrows \mathbf{Y}$ obeys the *revealed monotonicity (RM) axiom*, if for each $\mathbf{x}', \mathbf{x}'' \in \widehat{\mathbf{X}}$,

$$\mathbf{y}' \in \mathrm{B}(\mathbf{x}'), \mathbf{y}'' \in \mathrm{B}(\mathbf{x}''), \text{ and } (\mathbf{y}''_{-i}, x''_i) \geqslant (\mathbf{y}'_{-i}, x'_i) \implies y''_i \geqslant y'_i \text{ for each } i \in \mathcal{N}. \tag{6}$$

This axiom imposes a monotonicity restriction on B in the sense that it requires player i to take a weakly higher action whenever all other players are choosing higher actions and the covariate values are also higher. Note, however, that it does *not* require that if $\mathbf{x}'' \geqslant \mathbf{x}', \mathbf{y}'' \in \mathrm{B}(\mathbf{x}'')$, and $\mathbf{y}' \in \mathrm{B}(\mathbf{x}') \implies \mathbf{y}'' \geqslant \mathbf{y}'$. Indeed, the axiom even allows for $\mathbf{y}'' < \mathbf{y}'$, which corresponds to the case where there are two ranked Nash equilibria (at both $\mathbf{x}'$ and $\mathbf{x}''$), with players' jointly playing the lower equilibrium at the higher covariate value. The following theorem states that this axiom fully characterizes single-crossing group types.

Theorem 1. *The correspondence* $\mathrm{B} : \widehat{\mathbf{X}} \rightrightarrows \mathbf{Y}$ *is a single-crossing group type if and only if it satisfies the RM axiom.*

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

It is clear that the RM axiom is *necessary.* Indeed, suppose that $\mathrm{B}(\cdot)$ is a single-crossing group type and for some $\mathbf{y}'' \in \mathrm{B}(\mathbf{x}'')$ and $\mathbf{y}' \in \mathrm{B}(\mathbf{x}')$, it holds that $(\mathbf{y}''_{-i}, x''_i) \geqslant (\mathbf{y}'_{-i}, x'_i)$ for some $i \in \mathcal{N}$. Then, for this agent i, there is some single-crossing payoff function Π_i for which $y''_i = \mathrm{BR}_i(\mathbf{y}''_{-i}, x''_i)$ and $y'_i = \mathrm{BR}_i(\mathbf{y}'_{-i}, x'_i)$ and Basic Theorem (i) immediately implies that $y''_i \geqslant y'_i$. Thus the more substantial part of this theorem is the claim that the RM axiom is sufficient for a group type to be single-crossing. In the proof (see Appendix), we explicitly construct, for each $i \in \mathcal{N}$, a single-crossing payoff function Π_i that supports y_i as the best response to $(\mathbf{y}_{-i}, x_i)$ for every $\mathbf{y} \in \mathrm{B}(\mathbf{x})$. This payoff function is defined on $Y_i \times \mathbf{Y}_{-i} \times X_i$, rather than $Y_i \times \mathbf{Y}_{-i} \times \mathrm{proj}_i\, \widehat{\mathbf{X}}$, with single-crossing differences being satisfied on the entire domain. Indeed, the payoff functions we construct satisfy *increasing differences*[11] (rather than just single-crossing differences) and are also *single-peaked.*[12]

II: Finding weights on group types

Since the set of possible action profiles $\mathbf{Y}$ and $\widehat{\mathbf{X}}$ are finite, the set of all possible group types is also finite. We denote *the set of single-valued and single-crossing group types by $\mathcal{B}$.* The following result characterizes $\mathcal{SC}$-rationalizable choice distributions $\mathcal{P} = \{\mathrm{P}(\mathbf{y} \mid \mathbf{x}) : \mathbf{x} \in \widehat{\mathbf{X}}\}$ using the set of single-crossing group types $\mathcal{B}$ or, equivalently (by Theorem 1), those group types that obey the RM axiom.

Theorem 2. *$\mathcal{P}$ is $\mathcal{SC}$-rationalizable if and only if there exists a distribution $\tau = (\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$ on $\mathcal{B}$ such that the following holds:*

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \sum\nolimits_{\{\mathrm{B}\in\mathcal{B}:\ B(\mathbf{x})=\mathbf{y}\}} \tau^{\mathrm{B}} \ \textit{for all } \mathbf{y} \in \mathbf{Y} \textit{ and } \mathbf{x} \in \widehat{\mathbf{X}}. \tag{7}$$

[11] We say that $\Pi_i(y_i, \mathbf{y}_{-i}, x_i)$ has increasing differences in $(y_i; \mathbf{y}_{-i}, x_i)$, if $\Pi_i(y''_i, \mathbf{y}''_{-i}, x''_i) - \Pi_i(y'_i, \mathbf{y}''_{-i}, x''_i) \geqslant \Pi_i(y''_i, \mathbf{y}'_{-i}, x'_i) - \Pi_i(y'_i, \mathbf{y}'_{-i}, x'_i)$ for every $y''_i > y'_i$ and $(\mathbf{y}''_{-i}, x''_i) > (\mathbf{y}'_{-i}, x'_i)$. This implies single-crossing differences, but not vice versa. Note that ability to guarantee a stronger property under the RM axiom is not altogether surprising and something similar is found in Afriat's Theorem: the generalized axiom of revealed preference (GARP) is necessary so long as the consumer has a locally nonsatiated preference, but when a data set obeys GARP then it can then be rationalized by a utility function with stronger properties: continuous, increasing, and concave.

[12] This means that, for each $(\mathbf{y}_{-i}, \mathbf{x})$, there is $\bar{y}_i$ such that $\Pi_i(\bar{y}_i, \mathbf{y}_{-i}, \mathbf{x}) > \Pi_i(y_i, \mathbf{y}_{-i}, \mathbf{x})$ for all $y_i \neq \bar{y}_i$, with $\Pi_i(y_i, \mathbf{y}_{-i}, \mathbf{x})$ being strictly increasing in y_i for $y_i \leqslant \bar{y}_i$ and strictly decreasing in y_i for $\bar{y}_i \geqslant y_i$.

Proof of Theorem 2. The proof of the "if" part of this claim is particularly straightforward, since the characterization is *itself* an instance of $\mathcal{SC}$-rationalizability. Indeed, suppose there is a distribution $\tau = (\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$ on $\mathcal{B}$ such that (7) holds. By definition, there is some $\mathbf{\Pi} \in \mathcal{SC}$ that rationalizes B for each $\mathrm{B} \in \mathcal{B}$. By taking increasing transformations if necessary, we can guarantee that distinct group types in $\mathcal{B}$ are rationalized by distinct payoff function profiles in $\mathcal{SC}$. We denote the profile that rationalizes B by $\mathbf{\Pi}^{\mathrm{B}}$. Then $\mathcal{P}$ is $\mathcal{SC}$-rationalizable with a distribution $\mathrm{P}_{\mathbf{\Pi}}$ that assigns probability τ^{B} to $\mathbf{\Pi}^{\mathrm{B}} \in \mathcal{SC}$ and an equilibrium selection rule λ where $\lambda\left(\mathbf{y} \mid \mathbf{\Pi}^{\mathrm{B}}, \mathbf{x}\right) = 1$ if $\mathbf{y} = \mathrm{B}(\mathbf{x})$ and $\lambda\left(\mathbf{y} \mid \mathbf{\Pi}^{\mathrm{B}}, \mathbf{x}\right) = 0$ if $\mathbf{y} \neq \mathrm{B}(\mathbf{x})$; in other words, all groups in the population with payoff profile $\mathbf{\Pi}^{\mathrm{B}}$ will play $\mathrm{B}(\mathbf{x})$ at each $\mathbf{x} \in \widehat{\mathbf{X}}$.

Conversely, suppose $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, with the distribution $\mathrm{P}_{\mathbf{\Pi}}$ and the equilibrium selection rule λ. Let

$$p(\mathrm{B}, \mathbf{\Pi}) = \times_{\mathbf{x}\in\widehat{\mathbf{X}}} \lambda\left(\mathrm{B}(\mathbf{x}) \mid \mathbf{\Pi}, \mathbf{x}\right) \tag{8}$$

and let $\tau^{\mathrm{B}} = \int p(\mathrm{B}, \mathbf{\Pi})\, d\mathrm{P}_{\mathbf{\Pi}}$. If B is not a single crossing type, then $p(\mathrm{B}, \mathbf{\Pi}) = 0$ for all $\mathbf{\Pi} \in \mathcal{SC}$. Therefore, for $\mathbf{\Pi} \in \mathcal{SC}$,

$$\sum\nolimits_{\mathrm{B}\in\mathcal{B}} p(\mathrm{B}, \mathbf{\Pi}) = 1, \tag{9}$$

which guarantees that $\sum_{\mathrm{B}\in\mathcal{B}} \tau^{\mathrm{B}} = \int_{\mathcal{SC}} d\mathrm{P}_{\mathbf{\Pi}} = 1$ (since the support of $\mathrm{P}_{\mathbf{\Pi}}$ lies in $\mathcal{SC}$). Furthermore, it follows from (8) that $\lambda(\mathbf{y} \mid \mathbf{\Pi}, \mathbf{x}) = \sum_{\{\mathrm{B}\in\mathcal{B}:\mathrm{B}(\mathbf{x})=\mathbf{y}\}} p(\mathrm{B}, \mathbf{\Pi})$ and thus

$$\mathrm{P}(\mathbf{y}|\mathbf{x}) = \int \lambda\left(\mathbf{y}|\mathbf{\Pi}, \mathbf{x}\right) d\mathrm{P}_{\mathbf{\Pi}} = \sum\nolimits_{\{\mathrm{B}\in\mathcal{B}:\ \mathrm{B}(\mathbf{x})=\mathbf{y}\}} \int p(\mathrm{B}, \mathbf{\Pi})\, d\mathrm{P}_{\mathbf{\Pi}} = \sum\nolimits_{\{\mathrm{B}\in\mathcal{B}:\ \mathrm{B}(\mathbf{x})=\mathbf{y}\}} \tau^{\mathrm{B}}.$$

QED

Theorems 1 and 2 together provide us with a way of establishing the $\mathcal{SC}$-rationalizability of $\mathcal{P}$. First, we must identify the single-crossing group types, which, by Theorem 1, we can do via the RM axiom. Then Theorem 2 tells us that checking if $\mathcal{P}$ is SC-rationalizable boils down to finding a positive solution to a set of equations linear in the unknowns τ^{B} for all $\mathrm{B} \in \mathcal{B}$.[13]

Remark 4. It is part of the definition of $\mathcal{SC}$-rationalizability that the distribution of $\mathbf{\Pi} = (\Pi_i)_{i\in\mathcal{N}}$ is independent of $\mathbf{x}$. Suppose we drop this condition but still require all payoff functions to consist

[13] In some problems, it may not be computationally feasible to find all the elements of $\mathcal{B}$, but in those cases, one could still test for $\mathcal{SC}$-rationalizabiity by progressively enlarging the set of single-crossing types (see Section 4.1).

of single-crossing functions; then it is easy to see that the payoff functions and equilibrium selection rules will induce a distribution over group types in $\mathcal{B}$ at each $\mathbf{x}$, which we may denote by $(\tau_{\mathbf{x}}^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$, such that the following counterpart of (7) holds:

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \sum\nolimits_{\{\mathrm{B}\in\mathcal{B}:\ \mathrm{B}(\mathbf{x})=\mathbf{y}\}} \tau_{\mathbf{x}}^{\mathrm{B}} \text{ for all } \mathbf{y} \in \mathbf{Y} \text{ and } \mathbf{x} \in \widehat{\mathbf{X}}. \tag{10}$$

This condition is trivially true in the sense that one could always find $(\tau_{\mathbf{x}}^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$ such that it holds. Conditional independence imposes the additional requirement that $\tau_{\mathbf{x}'}^{\mathrm{B}} = \tau_{\mathbf{x}''}^{\mathrm{B}}$ for any $\mathbf{x}'$, $\mathbf{x}'' \in \widehat{\mathbf{X}}$ and this condition in combination with (10) is obviously equivalent to (7). One could imagine situations where the modeler has different views of how the distribution of $\mathbf{\Pi}$ (and hence the distribution of the associated group types) varies with $\mathbf{x}$, which may be more permissive than or different from conditional independence; these could be incorporated as further conditions on $\tau_{\mathbf{x}}^{\mathrm{B}}$ that could be tested in combination with (10). Obviously, such a test will remain a linear test if the added conditions are linear in $\tau_{\mathbf{x}}^{\mathrm{B}}$.

3.4 Recovering properties of a rationalizing distribution $\mathrm{P}_{\mathbf{\Pi}}$

When $\mathcal{P}$ is $\mathcal{SC}$-rationalizable, we are also able to extract information about this rationalization through the properties of $(\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$ that solve (7). In particular, let $\mathcal{SC}^*$ be a subset of single-crossing payoff functions (including all of its strictly increasing transformations) and let

$$\mathcal{B}^* = \{\mathrm{B} \in \mathcal{B} : \text{there is } \mathbf{\Pi} \in \mathcal{SC}^* \text{ that rationalizes B}\}\,. \tag{11}$$

By a straightforward adaptation of the proof of Theorem 2 (see Appendix), we can show that

$$\max\left\{\sum\nolimits_{\mathrm{B}\in\mathcal{B}^*} \tau^{\mathrm{B}} : (\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}} \text{ solves } (7)\right\} = \max\left\{\int_{\mathbf{\Pi}\in\mathcal{SC}^*} d\mathrm{P}_{\Pi} : \mathrm{P}_{\Pi} \text{ rationalizes } \mathcal{P}\right\} \tag{12}$$

Notice that the left hand side of this equation is straightforward to compute when $\mathcal{B}^*$ and $\mathcal{B}$ are known, since it simply involves solving a linear program. Thus we can find the greatest possible weight on a given set of payoff profiles, for any distribution that rationalizes $\mathcal{P}$.[14] We give two cases where this exercise is useful, both of which are empirically implemented in Section 5. Other examples can be found in the Online Appendix A3.

Application 1. Bounds on the role of strategic interaction

While our model allows for the possibility that each player reacts strategically to other players in the game, it is conceivable that the conditional choice distributions could be explained more simply, without appealing to strategic effects for one or more players in the game.

To be specific, suppose we wish to check whether it is possible to regard a subgroup $\mathcal{N}'$ of the players as nonstrategic. Let $\mathcal{SC}^*$ be the payoff profiles in $\mathcal{SC}$ such that Π_i does not depend on $\mathbf{y}_{-i}$ for every $i \in \mathcal{N}'$ and let $\mathcal{B}^*$ be its corresponding set of group types (as defined by (11)). The types in $\mathcal{B}^*$ can be characterized by a stricter version of the RM axiom: a group type is in $\mathcal{B}^*$ if and only if it obeys the RM axiom and, for each $i \in \mathcal{N}'$, we require that $\mathbf{y}'' \in \mathrm{B}(\mathbf{x}'')$, $\mathbf{y}' \in \mathrm{B}(\mathbf{x}')$, and $x_i'' \geqslant x_i' \implies y_i'' \geqslant y_i'$. With this characterization, we can construct $\mathcal{B}^*$. If we find that

$$\max\left\{\sum\nolimits_{\mathrm{B}\in\mathcal{B}^*} \tau^{\mathrm{B}} : (\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}} \text{ solves } (7)\right\} = 1,$$

we conclude (by (12)) that $\mathcal{P}$ can be $\mathcal{SC}$-rationalized without requiring the players in $\mathcal{N}'$ to be strategic; on the other hand, if the upper bound is strictly below 1, then we *must* incorporate strategic interactions among these players to $\mathcal{SC}$-rationalize $\mathcal{P}$.

Application 2. Probability bounds for Nash equilibrium profiles

Given a strategy profile $\overline{\mathbf{y}}$ and covariate $\overline{\mathbf{x}}$, we pose the following question: among all the possible $\mathcal{SC}$-rationalizations of $\mathcal{P}$, what is the greatest fraction of groups which could have $\overline{\mathbf{y}}$ as a pure strategy Nash equilibrium at $\overline{\mathbf{x}}$? Here, $\overline{\mathbf{x}} \in \mathbf{X}$ may or may not be an element of $\widehat{\mathbf{X}}$, and when $\overline{\mathbf{x}} \notin \widehat{\mathbf{X}}$, the answer to this question provides information on how the game would be played at an hitherto unobserved covariate value. However, the question is interesting even when $\overline{\mathbf{x}} \in \widehat{\mathbf{X}}$.

[14] To obtain $\min\left\{\int_{\Pi\in\mathcal{SC}^*} d\mathrm{P}_\Pi : \mathrm{P}_\Pi \text{ rationalizes } \mathcal{P}\right\}$, we use the similarly easy-to-prove identity

$$\min\left\{\sum\nolimits_{\mathrm{B}\in\mathcal{B}_0} \tau^{\mathrm{B}} : (\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}} \text{ solves } (7)\right\} = \min\left\{\int_{\Pi\in\mathcal{SC}^*} d\mathrm{P}_\Pi : \mathrm{P}_\Pi \text{ rationalizes } \mathcal{P}\right\}$$

where $\mathcal{B}_0 = \left\{\mathrm{B} \in \mathcal{B} : \mathrm{B} \text{ can only be rationalized by } \Pi \in \mathcal{SC}^*\right\}$.

To see why, notice that there is a distinction between $\mathrm{P}(\overline{\mathbf{y}} \mid \overline{\mathbf{x}})$, the observed fraction of groups in the population that play $\overline{\mathbf{y}}$ at $\overline{\mathbf{x}}$, and the fraction of groups for which $\overline{\mathbf{y}}$ is a Nash equilibrium. The former is typically smaller than the latter because some groups who play strategy profiles other than $\overline{\mathbf{y}}$ *may also have* $\overline{\mathbf{y}}$ *as a Nash equilibrium.*[15] The distinction between $\mathrm{P}(\overline{\mathbf{y}} \mid \overline{\mathbf{x}})$ and the greatest possible weight on those groups which have $\overline{\mathbf{y}}$ as a Nash equilibrium at $\mathbf{x} = \overline{\mathbf{x}}$ is relevant because, if the gap is small, then we are sure that changing the equilibrium selection scheme *cannot* significantly increase the frequency with which $\overline{\mathbf{y}}$ is played. This means (for example) that a policy maker who wants $\overline{\mathbf{y}}$ to be played more often must alter payoffs in some way and it is not possible to simply convince players to coordinate on a different equilibrium. An earlier analysis of questions of this type can be found in Aradillas-Lopez (2011), which focuses on a different class of games.[16]

To answer our question, let $\mathcal{SC}^* = \{\Pi \in \mathcal{SC} : \overline{\mathbf{y}} \in \mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})\}$ and let $\mathcal{B}^*$ be its corresponding set of group types. We can check whether B belongs to $\mathcal{B}^*$ by using the RM axiom. Indeed $\mathrm{B} \in \mathcal{B}^*$ if and only if the (possibly) multi-valued group type $\overline{B}$ defined as follows obeys the RM-axiom: $\overline{\mathrm{B}}(\overline{\mathbf{x}}) = \{\mathrm{B}(\overline{\mathbf{x}}), \overline{\mathbf{y}}\}$ and $\overline{\mathrm{B}}(\mathbf{x}) = \mathrm{B}(\mathbf{x})$ for every $\mathbf{x} \in \widehat{\mathbf{X}} \backslash \{\overline{\mathbf{x}}\}$. The proportion of the population which has $\overline{\mathbf{y}}$ as a PSNE cannot exceed $\max\left\{\sum_{\mathrm{B} \in \mathcal{B}^*} \tau^{\mathrm{B}} : (\tau^{\mathrm{B}})_{\mathrm{B} \in \mathcal{B}} \text{ solves } (7)\right\}$ and can equal this number. Notice that this value exceeds $\mathrm{P}(\overline{\mathbf{y}} \mid \overline{\mathbf{x}})$ since $\mathcal{B}^*$ contains the set $\{\mathrm{B} \in \mathcal{B} : \mathrm{B}(\overline{\mathbf{x}}) = \overline{\mathbf{y}}\}$ (see (7)).[17]

[15] For example, suppose we observe the distribution of action profiles at an entry game with two firms at a single covariate value $\overline{\mathbf{x}}$. Assuming that preferences obey single crossing differences (in the sense of (1)) and are strict, if (E, E) or (N, N) is played by a pair of firms, then it has to be their unique equilibrium, but any pair that plays (E, N) may also have (N, E) as another (albeit unselected) equilibrium. Thus if $\mathrm{P}(E, N \mid \overline{\mathbf{x}})$ and $\mathrm{P}(N, E \mid \overline{\mathbf{x}})$ are the observed probabilities of action profiles (E, N) and (N, E) respectively, then the probability that (E, N) (similarly, (N, E)) is a Nash equilibrium profile at $\mathbf{x} = \overline{\mathbf{x}}$ is no greater than $\mathrm{P}(E, N \mid \overline{\mathbf{x}}) + \mathrm{P}(N, E \mid \overline{\mathbf{x}})$.

[16] Although the games considered in Aradillas-Lopez (2011) are referred to in that paper as games with strategic substitutes/complements, those properties are defined with respect to the monotonicity of the payoff functions in opponents' action, rather than the monotonicity of best responses. Thus the class of games treated in that paper is different from the one we consider. Nonetheless, the key idea behind that paper is captured by our simple example: under various assumptions on payoff functions, the probability that a given strategy profile $\mathbf{y}$ is a Nash equilibrium when $\mathbf{x} = \overline{\mathbf{x}}$ is not simply bounded above by 1, because there are some profiles $\mathbf{y}'$ that *cannot coexist* with $\mathbf{y}$ as Nash equilibria of a group when $\mathbf{x} = \overline{\mathbf{x}}$. In our model, we also make use of the structure we impose on how payoff functions vary *across* covariates, which enables us to say that a group/group type that plays some profile $\mathbf{y}'$ at a different covariate value $\hat{\mathbf{x}}$ cannot play $\mathbf{y}$ at $\overline{\mathbf{x}}$ (see the example in footnote 24). These structural restrictions allow us to form a nontrivial upper bound on how often $\mathbf{y}$ is a Nash equilibria when $\mathbf{x} = \overline{\mathbf{x}}$.

[17] Our analysis here gives the most optimistic estimate on the possibility of switching the equilibrium action to $\overline{\mathbf{y}}$, in the sense that it assumes that every group type which *can* be rationalized by an element in $\mathcal{SC}^*$ actually does have a payoff function profile in $\mathcal{SC}^*$. We could also find the most conservative estimate of the proportion of the population that could switch to $\overline{\mathbf{y}}$ by changing equilibrium selection rules; this is explained in Online Appendix A6.2.

4 The statistical procedure

This section outlines the statistical procedure that implements the results in the previous section, which are based on population distributions. The test of $\mathcal{SC}$-rationalizability is explained in Section 4.1 and relies on the statistical hypothesis testing proposed by Kitamura and Stoye (2018). The efficient implementation of this test when $\mathcal{B}$ is large (and cannot be fully listed) uses the column generation approach proposed in Smeulders et al. (2021). Section 4.2 outlines the procedure (in essence provided by Deb et al. (2022)) to obtain confidence intervals for the weights on certain group types; the efficient implementation of this procedure requires a nontrivial extension of the column generation method in Smeulders et al. (2021) and we provide this in Proposition 3.

4.1 Statistical hypothesis testing

We begin with a matrix re-formulation of the characterization given in Theorem 2. Each generalized group type $\mathrm{B} : \widehat{\mathbf{X}} \rightrightarrows \mathbf{Y}$ can be represented as a vector $\mathbf{b} = (b_{\mathbf{y},\mathbf{x}})_{\mathbf{Y}\times\widehat{\mathbf{X}}}$ such that $b_{\mathbf{y},\mathbf{x}} = 1$ if $\mathbf{y} \in \mathrm{B}(\mathbf{x})$ and $b_{\mathbf{y},\mathbf{x}} = 0$ otherwise. Conversely, for any $\mathbf{b} \in \{0,1\}^{|\mathbf{Y}\times\widehat{\mathbf{X}}|}$ corresponds to a generalized group type, with a vector $\mathbf{b} \in \{0,1\}^{|\mathbf{Y}\times\widehat{\mathbf{X}}|}$ representing a single-valued) group type if and only if $\sum_{\mathbf{y}\in\mathbf{Y}} b_{\mathbf{y},\mathbf{x}} = 1$ at every $\mathbf{x} \in \widehat{\mathbf{X}}$. Similarly, since $\mathcal{P}$ consists of $|\widehat{\mathbf{X}}|$ distributions on $\mathbf{Y}$, it can be captured by the column vector $\mathbf{p} \in [0,1]^{|\mathbf{Y}\times\widehat{\mathbf{X}}|}$, where the $(\mathbf{y},\mathbf{x})$-th entry of $\mathbf{p}$ is $\mathrm{P}(\mathbf{y}|\mathbf{x})$ (and hence, $\sum_{\mathbf{y}\in\mathbf{Y}} p_{\mathbf{y},\mathbf{x}} = 1$ for each $\mathbf{x} \in \widehat{\mathbf{X}}$).

In what follows, we shall abuse notation and use $\mathcal{B}$ to denote both the set of group types obeying the RM axiom and also the vectors corresponding to those types. We denote by $\mathbf{B}$ the matrix where each column represents a group type in $\mathcal{B}$. Theorem 2 states that $\mathcal{P}$ is $\mathcal{SC}$-rationalizable if and only if there is $\tau \in \Delta^{\mathcal{B}}$, the set of distributions on $\mathcal{B}$, that solves $\mathbf{B}\tau = \mathbf{p}$. ($\Delta^{\mathcal{B}}$ could be thought of as elements of the standard $(|\mathcal{B}| - 1)$-simplex.) We would like to test if the data is consistent with the $\mathcal{SC}$-rationalizability of $\mathcal{P}$. Equivalently, letting $\mathbb{P}^{\mathrm{SC}} = \{\mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}}\}$ (i.e., the set of $\mathcal{SC}$-rationalizable distributions in vector form), our null hypothesis is

$$\min_{\eta\in\mathbb{P}^{\mathrm{SC}}} (\mathbf{p} - \eta) \cdot (\mathbf{p} - \eta) = 0. \tag{13}$$

The data set consists of $N_{\mathbf{x}}$ observations of the action profiles at each realization of $\mathbf{x} \in \widehat{\mathbf{X}}$. We assume that $N_{\mathbf{x}}/N \to \rho_{\mathbf{x}} \in (0,1)$ at each $\mathbf{x} \in \widehat{\mathbf{X}}$, as $N = \sum_{\mathbf{x}\in\widehat{\mathbf{X}}} N_{\mathbf{x}} \to \infty$. We denote the empirical distribution over action profiles by

$$\mathcal{Q} = \left\{ \mathrm{Q}(\cdot \mid \mathbf{x}) : \mathbf{x} \in \hat{\mathbf{X}} \right\},$$

and we estimate $\mathcal{P}$ by this sample analog. As in the case with $\mathcal{P}$, we can represent $\mathcal{Q}$ by a column vector $\mathbf{q} \in [\mathbf{0}, \mathbf{1}]^{|\mathbf{Y} \times \mathbf{X}|}$ where the $(\mathbf{y}, \mathbf{x})$-th entry is equal to $\mathrm{Q}(\mathbf{y}|\mathbf{x})$. The test statistic is

$$J_N := \min_{\eta \in \mathbb{P}^{\mathrm{SC}}} N(\mathbf{q} - \eta) \cdot (\mathbf{q} - \eta) = \min_{\tau \in \Delta^{\mathcal{B}}} N(\mathbf{q} - \mathbf{B}\tau) \cdot (\mathbf{q} - \mathbf{B}\tau). \tag{14}$$

As pointed out in Kitamura and Stoye (2018), care is needed to obtain the valid critical value. In particular, we cannot simply adopt a solution to the problem (14) as the bootstrap estimation for the empirical choice distribution, due to the possible discontinuity of the limiting distribution of J_N. Addressing this issue involves introducing a tuning parameter and considering the *tightened* problem defined below.

Choose $\mathcal{B}' \subset \mathcal{B}$ so that it contains a basis of the space spanned by $\mathcal{B}$, and define $\Delta^{\mathcal{B}}_{\kappa_N} = \left\{ \tau \in \Delta^{\mathcal{B}} : \tau_{\mathbf{b}} > \kappa_N / |\mathcal{B}'| \text{ for all } \mathbf{b} \in \mathcal{B}' \right\}$, with κ_N being selected so that $\kappa_N \downarrow 0$ and $\sqrt{N}\kappa_N \uparrow \infty$ as $N \to \infty$.[18] Letting $\mathbb{P}^{\mathrm{SC}}_{\kappa_N} = \{\mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}}_{\kappa_N}\}$, we adopt

$$\eta^* = \operatorname*{argmin}_{\eta \in \mathbb{P}^{\mathrm{SC}}_{\kappa_N}} N(\mathbf{q} - \eta) \cdot (\mathbf{q} - \eta) \tag{15}$$

as the bootstrap estimator of the empirical choice frequency. Compared to the problem (14), the feasible set in the minimization problem is tightened by the tuning parameter, with positive weights required of elements in $\mathcal{B}'$. We then generate a bootstrap sample $\mathbf{q}^{(r)}$ (for $r = 1, 2, \ldots, R$) using standard nonparametric bootstrap re-sampling from η^* and re-center this sample by setting $\hat{\mathbf{q}}^{(r)} := (\mathbf{q}^{(r)} - \mathbf{q}) + \eta^*$. With $\hat{\mathbf{q}}^{(r)}$ we can calculate the bootstrap test statistic

$$J_N^{(r)} := \min_{\eta \in \mathbb{P}^{\mathrm{SC}}_{\kappa_N}} N(\hat{\mathbf{q}}^{(r)} - \eta) \cdot (\hat{\mathbf{q}}^{(r)} - \eta) = \min_{\tau \in \Delta^{\mathcal{B}}_{\kappa_N}} N(\hat{\mathbf{q}}^{(r)} - \mathbf{B}\tau) \cdot (\hat{\mathbf{q}}^{(r)} - \mathbf{B}\tau). \tag{16}$$

With the empirical distribution of $J_N^{(r)}$ we can calculate the p-value $p = \#\{J_N^{(r)} > J_N\}/R$. The null

[18] In the original formulation by Kitamura and Stoye (2018), positive weights are required on all elements in $\mathcal{B}$, which is inconvenient when applying the column generation procedure described later in this subsection. We require positive weights only for the types in $\mathcal{B}'$. The validity of our approach (which is similar to that in Smeulders et al. (2021)) is shown in the Online Appendix A5.2.

hypothesis that $\mathbf{q}$ is a sample from some $\mathbf{p} \in \mathbb{P}^{\mathrm{SC}}$ is not rejected if the p-value is greater than the critical value.

A major hurdle in implementing the above test is that the computation of J_N and $J_N^{(r)}$ involves the enumeration of $\mathcal{B}$, which is often too large to compute. We cope with this problem by applying the *column generation* procedure in Smeulders et al. (2021). This procedure involves first testing a more stringent version of the model corresponding to a strict subset $\mathcal{B}_0$ of $\mathcal{B}$ which is completely known. For instance, we may choose the 'starter' set $\mathcal{B}_0$ to be the set of *constant types*, in which every player takes the same action regardless of opponents' actions and covariates; these group types obviously obey the RM axiom. Then the set $\mathcal{B}_0$ is progressively enlarged by including more group types from $\mathcal{B}$, up to the point where further additions will not improve the model's ability to explain the data.

To be precise, let $\mathbf{B}_0$ be the matrix where the column vectors are elements of $\mathcal{B}_0$. We can then calculate

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

Obviously, $J_{N,0} \geqslant J_N$ and we could check if it is possible to decrease $J_{N,0}$ by including some $\mathbf{b} \in \mathcal{B}$. We say that $\mathbf{b} \in \mathcal{B}$ *improves* $\mathcal{B}_0$ if, when $\mathbf{b}$ is included in $\mathcal{B}_0$, the new value of $J_{N,0}$ is strictly lower. The following result, which flows from the convex projection theorem, provides a necessary and sufficient condition for $\mathcal{B}_0$ to be improvable.

Proposition 1. *A set of group types $\mathcal{B}_0$ is improved by some $\mathbf{b} \in \mathcal{B}$, if and only if*

$$\max_{\mathbf{b} \in \mathcal{B}} (\mathbf{q} - \eta_0) \cdot (\mathbf{b} - \eta_0) > 0, \tag{18}$$

where $\eta_0 = \mathbf{B}_0\tau_0$ with $\tau_0 = \arg\min_{\tau \in \Delta^{\mathcal{B}_0}} (\mathbf{q} - \mathbf{B}_0\tau) \cdot (\mathbf{q} - \mathbf{B}_0\tau)$.

To solve the problem (18) without fully enumerating $\mathcal{B}$, we must find a computational efficient way of characterizing $\mathcal{B}$. Conveniently for us, the RM axiom — and hence the set $\mathcal{B}$ — can be characterized as solutions to an integer linear programming problem.

Proposition 2. *We can construct a matrix C and a column vector θ, both with nonnegative integer entries, such that for any $\mathbf{b} \in \{0,1\}^{|\mathbf{Y} \times \hat{\mathbf{X}}|}$, we have $\mathbf{b} \in \mathcal{B}$ if and only if $C\mathbf{b} \leqslant \theta$.*

The formulae for C and θ are found in our proof of this proposition in the Appendix. Combining this result with Proposition 1, $\mathcal{B}_0$ is improved by some $\mathbf{b} \in \mathcal{B}$, if and only if

$$\max\,(\mathbf{q} - \eta_0) \cdot (\mathbf{b} - \eta_0), \text{ subject to } \mathbf{b} \in \{0,1\}^{|\mathbf{Y} \times \widehat{\mathbf{X}}|} \text{ and } C\mathbf{b} \leqslant \theta \tag{19}$$

is strictly positive. If it is, we add this $\mathbf{b}$ to $\mathcal{B}_0$ and then repeat the process. In other words, we recalculate $J_{N,0}$ and η_0 based on the new $\mathcal{B}_0$, and try to find another element in $\mathcal{B}$ that improves on $\mathcal{B}_0$ by checking if (19) has a strictly positive solution. Since $\mathcal{B}$ is finite, this algorithm must terminate, and at the end we can be sure we have found $\mathcal{B}_0$ such that $J_{N,0} = J_N$.

The column generation procedure described above can be also applied to the computation of $J_N^{(r)}$ defined by (16). Since the constraint in the problem (16) requires positive weights on $\mathcal{B}'$, this set needs to be contained in the initial choice of $\mathcal{B}_0$.[19]

Remark 5. In the empirical application discussed in Section 5, we employ some procedures to reduce computation time. When $|\widehat{\mathbf{X}}|$ becomes large, obtaining the exact solution to (19) can be hard. In fact, to improve $J_{N,0}$, it suffices to find $\mathbf{b} \in \mathcal{B}$ such that $(\mathbf{q} - \eta_0) \cdot (\mathbf{b} - \eta_0) > 0$. In our program, we impose a time limit for solving (19) and use the best feasible solution found within that.[20] It is only when this solution satisfies $(\mathbf{q} - \eta_0) \cdot (\mathbf{b} - \eta_0) \leqslant 0$ that we solve the maximization problem exactly.

Remark 6. As noted by Smeulders et al. (2021), the computation time can be further reduced by not always calculating the value of $J_N^{(r)}$ exactly. Indeed, it suffices to determine whether each $J_N^{(r)}$ is larger or smaller than the critical value J_N. Thus we can terminate the procedure for calculating $J_N^{(r)}$ once a value of $J_{N,0}^{(r)}$ becomes lower than J_N.

4.2 Estimating weights on selected types

Suppose we have a data set that is consistent with $\mathcal{SC}$-rationalizability (in the sense that the null hypothesis (13) is not rejected) and would now like to form a confidence interval on $\sum_{\mathbf{b} \in \mathcal{B}^*} \tau^{\mathbf{b}}$, the

[19] Note that, even if $\mathcal{B}_0$ contains $\mathcal{B}'$, and $\mathcal{B}'$ contains a linear basis of $\mathcal{B}$, the *convex hull* of $\mathcal{B}_0$ need not coincide with the convex hull of $\mathcal{B}$ (though the linear hull of $\mathcal{B}_0$ of course coincides with the linear hull of $\mathcal{B}$). So it is still possible for $\mathcal{B}_0$ to be improvable.

[20] For example, using Rglpk package on R, our program gives 0.2 seconds for solving (19) at each repetition. This constraint tends to be binding when $\widehat{\mathbf{X}}$ is large, leading to a large integer programming problem.

total weight of a subset of single-crossing group types $\mathcal{B}^*$ (see Section 3.4). To do this, we follow the procedure in Deb et al. (2022). The problem of determining whether a given weight of $\mathcal{B}^*$ falls within the confidence interval can be determined by testing a suitably modified version of the null hypothesis (13), with $\mathbb{P}^{\text{SC}}$ replaced by a different set of distributions. To be specific, suppose we would like to find the upper bound of the confidence interval. For each $\beta \in (0,1)$, we let

$$\mathbb{P}^{SC}(\beta;\mathcal{B}^*) = \left\{ \mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}} \text{ and } \sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta \right\},$$

and test the null hypothesis

$$\min_{\eta\in\mathbb{P}^{\text{SC}}(\beta;\mathcal{B}^*)} (\mathbf{p}-\eta)\cdot(\mathbf{p}-\eta) = 0 \tag{20}$$

at some significance level $\bar{p}$. We then use binary search to obtain the maximal value of β under which the null hypothesis is not rejected; the resulting maximal value of β corresponds to the supremum of the $100(1-\bar{p})\%$ confidence interval of β.

For a given β, the test statistic is

$$\begin{aligned} J_N(\beta) &:= \min_{\eta\in\mathbb{P}^{\text{SC}}(\beta;\mathcal{B}^*)} N(\mathbf{q}-\eta)\cdot(\mathbf{q}-\eta) \\ &= \min_{\tau\in\Delta^{\mathcal{B}}} N(\mathbf{q}-\mathbf{B}\tau)\cdot(\mathbf{q}-\mathbf{B}\tau) \text{ subject to } \textstyle\sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta. \end{aligned} \tag{21}$$

Once again, it may not be possible to fully enumerate $\mathcal{B}$ or $\mathcal{B}^*$, and so a version of the column generation procedure outlined in the previous subsection is needed. This in turn requires an extension of Proposition 1 which we now explain.

Let $\mathcal{B}_0 \subset \mathcal{B}$ be such that $\mathcal{B}_0 \cap \mathcal{B}^* \neq \varnothing$, and let us calculate

$$J_{N,0}(\beta) = \min_{\tau\in\Delta^{\mathcal{B}_0}} N(\mathbf{q}-\mathbf{B}_0\tau)\cdot(\mathbf{q}-\mathbf{B}_0\tau) \;\; \text{s.t.} \sum_{\mathbf{b}\in(\mathcal{B}_0\cap\mathcal{B}^*)} \tau^{\mathbf{b}} \geqslant \beta.$$

We say that $\mathcal{B}_0$ is *improvable given problem (21)*, if $J_{N,0}(\beta) > J_N(\beta)$. The following proposition is the counterpart of Proposition 1 and provides a necessary and sufficient condition for a given $\mathcal{B}_0$ to be improvable.

Proposition 3. *If the set $\mathcal{B}_0 \subset \mathcal{B}$ is improvable given problem (21), then there is a pair of types $\{\mathbf{b}^*, \mathbf{b}\}$, with $\mathbf{b}^* \in \mathcal{B}^*$ and $\mathbf{b} \in \mathcal{B}$ such that*

$$(\mathbf{q}-\eta_0)\cdot(\beta\mathbf{b}^*+(1-\beta)\mathbf{b}-\eta_0)>0, \tag{22}$$

where $\eta_0=\mathbf{B}_0\tau_0$ *with* τ_0 *being the distribution that achieves* $J_{N,0}(\beta)$*. Conversely, suppose there is* $\mathbf{b}^*\in\mathcal{B}^*$ *and* $\mathbf{b}\in\mathcal{B}$ *such that (22) holds; then* $\{\mathbf{b}^*,\mathbf{b}\}$ *improves* $\mathcal{B}_0$ *given problem (21).*

We already know (from Proposition 2) that we can construct a matrix C and a column vector θ so that $\mathbf{b}\in\mathcal{B}$ if and only if $C\mathbf{b}\leqslant\theta$. Suppose that, in addition, we can construct a matrix C^* and a column vector θ^* with integer entries so that, for any $\mathbf{b}\in\{0,1\}^{|\mathbf{Y}\times\hat{\mathbf{X}}|}$, we have $\mathbf{b}^*\in\mathcal{B}^*\iff C^*\mathbf{b}^*\leqslant\theta^*$. Then a pair $\{\mathbf{b}^*,\mathbf{b}\}$ obeying (22) exists, if and only if the problem

$$\begin{aligned}&\max\,(\mathbf{q}-\eta_0)\cdot(\beta\mathbf{b}^*+(1-\beta)\mathbf{b}-\eta_0)\\&\text{s.t. }\mathbf{b},\,\mathbf{b}^*\in\{0,1\}^{|\mathbf{Y}\times\hat{\mathbf{X}}|}\text{ and }\begin{pmatrix}C^* & O\\ O & C\end{pmatrix}\begin{pmatrix}\mathbf{b}^*\\ \mathbf{b}\end{pmatrix}\leqslant\begin{pmatrix}\theta^*\\ \theta\end{pmatrix}\end{aligned} \tag{23}$$

has a positive optimal value. Note that every $\mathcal{B}^*$ in our empirical application has the matrix characterization described above (see Online Appendix A4 for the specific construction). If there is a pair $\{\mathbf{b},\mathbf{b}^*\}$ that improves $\mathcal{B}_0$, then we update $\mathcal{B}_0$ by including the pair in $\mathcal{B}_0$ and recalculate $J_{N,0}(\beta)$. Since $\mathcal{B}$ is finite, this process terminates and we obtain $J_{N,0}(\beta)=J_N(\beta)$.

To find the valid critical value, we need a suitable tightening that imposes strictly positive weights on a certain subset of group types. The tightening here must depend on β and its formulation is rather involved, and so we postpone this discussion to Online Appendix A5.3. That said, once we have constructed a suitably tightened subset of $\Delta^{\mathcal{B}}$ by some tuning parameter κ_N, the rest of the procedure is similar to that outlined in the preceding subsection. Denoting this subset by $\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*)$ and letting

$$\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*)=\left\{\mathbf{B}\tau:\tau\in\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*)\text{ and }\sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}}\geqslant\beta\right\}, \tag{24}$$

the bootstrap estimator and the recentered bootstrap samples can be obtained as in (15) – (16), after replacing the sets $\mathbb{P}^{\mathrm{SC}}_{\kappa_N}$ and $\Delta^{\mathcal{B}}_{\kappa_N}$ by, respectively, $\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*)$ and $\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*)$. Column generation can be also applied, in a way similar to its application for calculating $J_N(\beta)$, with the proviso that $\mathcal{B}_0$ must contain the group types on which the distributions in $\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*)$ give positive weight. The details of the procedure are provided in Online Appendix A5.3.

5 Empirical illustration

We apply our results in the preceding sections to an entry game using a data set taken from Kline and Tamer (2016). The data set contains the entry decisions of airlines in 7,882 markets, where a market is defined as a trip between two airports irrespective of intermediate stops. Airline firms are divided into two categories: LCC (low cost carriers) and OA (other airlines).[21] In Kline and Tamer's analysis (and in ours) the two categories are treated as two firms. Thus, in each market, the two firms, LCC and OA, can either both enter a market, both stay out, or one could enter with the other staying out.

This data set also contains information on two covariates: market presence (MP) and market size (MS). Market presence is a market- and airline-specific variable. For each airline and for each airport, one counts the number of markets that the airline serves from that airport and divide it by the total number of markets served from that airport by any airline; the market presence variable for a given market and airline is the average of these ratios at endpoints of that market/trip. The construction and inclusion of this covariate is not novel and follows Berry (1992). Since the airlines are aggregated into two firms, the market presence variable is also aggregated: the market presence for LCC (resp. OA) is the maximum among the actual airlines in the LCC category (resp. OA category). The second covariate, market size, is a market-specific variable (shared by all airlines in that market) and is defined as the population at endpoints of the corresponding trip.

Furthermore, Kline and Tamer (2016) discretize these variables, where each of them takes value 1 if the variable is higher than its median value and 0 otherwise. Thus, in our data set, there are three binary covariates, MP_{LCC}, MP_{OA}, and MS, and markets are partitioned into eight groups according to realizations of them. Formally, $\mathbf{X} = \{0, 1\}^3$, and in this case, it also holds that $\widehat{\mathbf{X}} = \mathbf{X}$. Note that MS simultaneously influences the payoffs of both LCC and OA, and hence the covariates affecting LCC's payoff can be written as $x_{LCC} = (\mathrm{MP}_{LCC}, \mathrm{MS})$ and, similarly, $x_{OA} = (\mathrm{MP}_{OA}, \mathrm{MS})$.

Observations in the data set can be used to calculate the empirical choice distributions, that we include in Table 3. It consists of eight blocks, with the markets in each block sharing the same covariates. For example, there are 1,271 markets with $(\mathrm{MP}_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS}) = (0, 0, 0)$, of which

[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 a low cost carrier is, by definition, an 'other airline.'

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

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

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

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

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

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

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

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

Table 3: Empirical distribution across each realization of covariates

around 30% are not served by either airline and about 68% are served only by airlines in the OA category (an action profile is written as $(y_{LCC}, y_{OA}) \in \{E, N\} \times \{E, N\}$). The entries in Table 3 seem 'reasonable', in the sense that it appears as though a firm's entry is encouraged whenever its market presence is large or the market size is large, and it is deterred by the entry of the other firm. For example, going from $(0,0,0)$ to $(1,0,0)$ (so the market presence of LCC has increased), both Q(N,N) and Q(N,E) fall, while Q(E,N) and Q(E,E) both increase.

Testing $\mathcal{SC}$-rationalizability. Our hypothesis is that, in each market, two firms (LCC and OA) are playing a pure strategy Nash equilibrium in a game of strategic substitutes with monotone effects from covariates. The payoff function of LCC, say, $\Pi_{LCC}(y_{LCC}, y_{OA}, \mathrm{MP}_{LCC}, \mathrm{MS})$, is required to obey single-crossing differences in $(y_{LCC}; (-y_{OA}, \mathrm{MP}_{LCC}, \mathrm{MS}))$, and similarly, the payoff function of OA, $\Pi_{OA}(y_{OA}, y_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS})$, is required to obey single-crossing differences in $(y_{OA}; (-y_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS}))$. This ensures that a firm's entry is discouraged by the opponent's entry and enhanced by an increase in own covariates. The data set is supposed to arise from a population of those firms, with unobserved heterogeneity generating a distribution of realizations of payoff functions $\mathbf{\Pi} = (\Pi_{LCC}, \Pi_{OA})$, which we denote by $\mathrm{P}_{\mathbf{\Pi}}$, and an equilibrium selection rule.

Employing the statistical test in Section 4.1, we find a p-value of 0.138, and hence, the hypothesis that the empirical choice frequencies are explained by our modeling restrictions cannot be refuted at 5% (or 10%) significance level. We choose the tuning parameter $\kappa_N = \sqrt{\log \underline{N}_{\mathbf{x}}/10^6 \underline{N}_{\mathbf{x}}}$, where

$\underline{N}_{\mathbf{x}} = \min_{\mathbf{x}\in\hat{\mathbf{X}}} N_{\mathbf{x}}$, and the number of bootstrap samples as $R = 2000$.[22] Note that having a p-value strictly less than 1 means that J_N defined in (14) is strictly positive, i.e., there is a strictly positive distance between our empirical distribution and $\mathbb{P}^{\mathrm{SC}}$, the set of (exactly) $\mathcal{SC}$-rationalizable distributions. Using our R code on a desktop computer with Apple M1 processor and 16 GB RAM, the p-value was calculated in less than 3 minutes.

In this setting, the set $\mathbf{X} = \hat{\mathbf{X}}$ has exactly eight elements, and hence the number of possible group types is $4^8 \approx 65,000$. In this small environment, it is in fact not difficult to check the RM axiom for each of these group types. Doing that, we find that only 482 types satisfy single-crossing (equivalently, satisfy the RM axiom). This gives a sense of the "empirical bite" of our test: the data set has to be explained by using a very small fraction (less than 1%) of all possible group types.

Significance of strategic interactions. Having established that the data set is (statistically) $\mathcal{SC}$-rationalizable we can now go on to explore its properties. In particular, we can assess the extent to which strategic interactions play a role in explaining the data, in the sense discussed in Section 3.4, by considering the sub-classes of single-crossing group types that correspond to: (i) the LCC firm having a payoff function that is independent of the actions of OA; (ii) the OA firm having a payoff function that is independent of the actions of LCC; and (iii) both firms having payoff functions that are independent of the other firm's action. Applying the procedure explained in Section 4.2, we find that the greatest possible weights on these three sub-classes of consistent group types are (i) 0.923, (ii) 0.790, and (iii) 0.789 (within 5% significance level). Since these weights are all strictly less than 1, we conclude that any $\mathcal{SC}$-rationalization of the data *requires* strategic behavior for both LCC and OA firms. The computation time for each case was about 27 minutes.

Probability bounds for equilibrium actions. Under our behavioral hypothesis, the action profiles (N, N) and (E, E) can only be played as the *unique* equilibrium at any realization of $\mathbf{x} = (\mathrm{MP}_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS})$. On the other hand, when (N, E) is played, it is possible that (E, N) is also a Nash equilibrium of the game. For this reason, the probability that (E, N) is a Nash equilibrium of the game can be strictly higher than the observed frequency with which this profile is played, even after accounting for sampling variability (the same goes for (N, E)). Applying the argument in Sections 3.4 and 4.2, we can recover the greatest possible weight on group types in the

[22]Recall that $N_{\mathbf{x}}$ is the number of observations with covariates $\mathbf{x}$. Our choice of κ_N largely follows Kitamura and Stoye (2018).

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$	$(0,0,0)$		$(0,1,0)$		$(1,0,0)$		$(1,1,0)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
$\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$	0.699	0.544	0.815	0.503	0.503	0.644	0.558	0.555
Observed Prob.	0.682	0.006	0.785	0.003	0.367	0.253	0.542	0.050
$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$	$(0,0,1)$		$(0,1,1)$		$(1,0,1)$		$(1,1,1)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
$\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$	0.841	0.616	0.913	0.496	0.485	0.661	0.523	0.497
Observed Prob.	0.832	0.001	0.910	0.000	0.326	0.306	0.501	0.021

Table 4: Probability bounds for equilibrium action profiles

population which have (E, N) as a Nash equilibrium at a given covariate value (and similarly for (N, E)). These are reported as $\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$ in Table 4.[23]

For example at $(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS}) = (1, 0, 0)$, the greatest possible weight on those group types that may have (N, E) as a Nash equilibrium of the game is 0.503: this includes types which are already playing (N, E) (with observed frequency 0.367) as well as types that are playing (E, N) but may have (N, E) as an alternative Nash equilibrium.[24] Thus, even if we allow for equilibrium selection rules to change, and (N, E) is chosen whenever it is a PSNE, the frequency with which (N, E) is played at $(1, 0, 0)$ will not exceed 0.503. Notice that, in general, $\max \Pr[\overline{\mathbf{y}} \in \text{NE}(\Pi, \overline{\mathbf{x}})]$ is closer to the observed frequency in the case where $\overline{\mathbf{y}} = (N, E)$, while the same gap in the case of $\overline{\mathbf{y}} = (E, N)$ is considerably bigger. For $\overline{\mathbf{y}} = (N, E)$ (and similarly for $\overline{\mathbf{y}} = (E, N)$), the calculation of $\max \Pr[\overline{\mathbf{y}} \in \text{NE}(\Pi, \overline{\mathbf{x}})]$ for all $\mathbf{x} \in \widehat{\mathbf{X}}$ took around 38 minutes.[25]

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

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

[23]To be precise, $\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$ is the upper limit of the confidence interval on $\sum_{\mathrm{B} \in \mathcal{B}^*} \tau^{\mathrm{B}}$ where $\mathcal{B}^*$ is defined as in Application 2 of Section 3.4.

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

[25]See Online Appendix A6.2 for an estimate of $\min \Pr[\overline{\mathbf{y}} \in \text{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$.

A6.1.) Instead of aggregating covariates into binary variables, we let each of MP_{LCC}, MP_{OA} and MS take four possible values using quantiles: each variable takes value $k-1$, if it is in the k-th quartile. In this way, all markets are partitioned into $4^3 = 64$ covariate values and there is a distribution of entry decisions at each of them. In this environment, the total number of possible group types is enormous (4^{64}) and the same is true of the number of $\mathcal{SC}$-rationalizable group types.[26] While a direct approach is no longer feasible, the column generation method still works, with the test of $\mathcal{SC}$-rationalizability finishing in around 3 minutes, including the bootstrap procedure. In this case, we find that the null hypothesis is rejected with p-value equal to 0.

The conflicting results cast doubts on the robustness of the model to explain choices of airline firms. In Online Appendix 6.1, we implement the tests for even finer discretizations. Naturally, the number of markets at each discrete value of the covariate falls as the discretization becomes finer and so we only use those covariate values that contain a certain minimal number of markets; in other words, we have a case where $\widehat{\mathbf{X}}$ is a strict subset of $\mathbf{X}$. Broadly speaking (see Online Appendix A6.1 for details), we find that the hypothesis is supported if the market presence variables remain binary, even with finer discretizations of market size. Finer discretizations of the market presence variables lead to rejection of the hypothesis, even when the market size variable remains binary.

The precise reasons for this failure are unclear and requires a more careful analysis. In terms of the model's explicit assumptions, the failure could be attributed to a failure of the single-crossing property on payoff functions, the failure of firms to play PSNE in each market, or the failure of the conditional independence assumption (especially with finer discretization of the market presence covariate). There could also be problems having to do with the model's basic structure, such as the particular way in which market presence is calculated or the modeling of interaction as a *two*-action, *two*-agent game (which ignores the possible presence of multiple carriers within the LCC or OA category in each market, the scale of their operations if they enter, or the fact that carriers may operate and interact in multiple markets and have more complicated payoff functions).[27]

It is worth noting that the uneven performance of the model with finer discretizations of the covariates is also detectable when we implement the procedure of Kline and Tamer (2016), which

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

[27] The last point is related to the literature on multimarket contact; see, for example, Evans and Kessides (1994).

assumes conditional independence but has a parametric specification (see Online Appendix A6.1). The two approaches yield consistent results. In particular, with binary covariates, the data pass both the Kline-Tamer test and our test; with finer covariates, whenever the model fails our test, it also fails the Kline-Tamer test (as it should since it tests a more restrictive model).

Appendix

Proof of Theorem 1. It remains for us to show that a generalized group type $\mathrm{B} : \widehat{\mathbf{X}} \rightrightarrows \mathbf{Y}$ is consistent if it satisfies the RM axiom. For each $i \in \mathcal{N} = \{1, 2, ..., n\}$, we need to find a payoff function $\Pi_i : Y_i \times \mathbf{Y}_{-i} \times X_i \to \mathbb{R}$ such that (a) Π_i obeys the single crossing differences in $(y_i; \mathbf{y}_{-i}, x_i)$, and (b) $\mathbf{y} \in \mathrm{B}(\mathbf{x}) \implies \mathbf{y} \in \mathrm{NE}(\mathbf{\Pi}, \mathbf{x})$, and (c) for each $(\mathbf{y}_{-i}, x_i)$, $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i)$ is a singleton (even when $x_i \notin \mathrm{proj}_i\, \widehat{\mathbf{X}}$).

For each $i \in \mathcal{N}$, let $\mathbf{Z}_i = \mathbf{Y}_{-i} \times X_i$. Since $\widehat{\mathbf{X}}$ and $\mathbf{Y}$ are finite sets, the graph of $\mathrm{B}(\mathbf{x})$, which is $\mathcal{G}(\mathrm{B}) := \left\{(\mathbf{y}, \mathbf{x}) : \mathbf{y} \in \mathrm{B}(\mathbf{x}) \text{ for some } \mathbf{x} \in \widehat{\mathbf{X}}\right\}$, is also a finite set. Hence, it can be written as $\mathcal{G}(\mathrm{B}) = \{(\mathbf{y}^t, \mathbf{x}^t) : \mathbf{y}^t \in \mathrm{B}(\mathbf{x}^t) \text{ for } t \in \mathcal{T}\}$ where $\mathcal{T} = \{1, 2, ..., T\}$ is a finite index set. Letting $\mathbf{z}_i^t := (\mathbf{y}_{-i}^t, x_i^t)$ for $t \in \mathcal{T}$, each $(\mathbf{y}^t, \mathbf{x}^t) \in \mathcal{G}(\mathrm{B})$ can be written as $(y_i^t, \mathbf{z}_i^t)$ for every $i \in \mathcal{N}$.

To obtain a payoff function Π_i defined on $Y_i \times \mathbf{Z}_i$ $(= Y_i \times \mathbf{Y}_{-i} \times X_i)$, we begin with a family of single-peaked functions, $f_i : Y_i \times \mathcal{T} \to \mathbb{R}$ satisfying the following properties: (i) $f_i(y_i^t, t) > f_i(a, t)$ for all $a \neq y_i^t$ and $a \in Y_i$; and (ii) if $y_i^s = y_i^t$ then $f_i(\cdot, s) = f_i(\cdot, t)$ and if $y_i^s > y_i^t$, then $f_i(a'', s) - f_i(a', s) > f_i(a'', t) - f(a', t)$ for all $a'' > a'$ in Y_i. This can be obtained, for example, by letting $f_i(a, t) = -(a - y_i^t)^2$. Then we define $\Pi_i : Y_i \times \mathbf{Z}_i \to \mathbb{R}$ as follows. For each $\mathbf{z} \in \mathbf{Z}_i$, let $T(\mathbf{z}) = \{t \in \mathcal{T} : \mathbf{z}^t \geqslant \mathbf{z}\} \cup \{\hat{t}\}$, where $\hat{t}$ is any index that satisfies $y_i^{\hat{t}} \geqslant y_i^t$ for all $t \in \mathcal{T}$. Since it contains $\hat{t}$ at least, $T(\mathbf{z})$ is nonempty. Choose $\tilde{t}(\mathbf{z}) \in T(\mathbf{z})$ such that $y_i^{\tilde{t}(\mathbf{z})} \leqslant y_i^t$ for all $t \in T(\mathbf{z})$, and define $\Pi_i(\cdot, \mathbf{z}) = f_i(\cdot, \tilde{t}(\mathbf{z}))$. Although there may be more than one candidate for $\tilde{t}(\mathbf{z})$, by property (ii) of f_i, the value of Π_i is not affected by the choice.

We claim that $\Pi_i(\cdot, \mathbf{z})$ defined above obeys properties (a) – (c). For property (a), suppose that at $\mathbf{z}'$, we have $\Pi_i(\cdot, \mathbf{z}') = f_i(\cdot, t')$ and for $\mathbf{z}''$, we have $\Pi_i(\cdot, \mathbf{z}'') = f_i(\cdot, t'')$. If $\mathbf{z}'' > \mathbf{z}'$, then $T(\mathbf{z}'') \subseteq T(\mathbf{z}')$, and so $y_i^{t''} \geqslant y_i^{t'}$. By property (ii) of f_i, we obtain

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

Thus Π_i satisfies *increasing-differences*, which means it satisfies single crossing differences. For property (b), notice that, at any $\mathbf{z}^s$, we have $s \in T(\mathbf{z}^s)$ and, by the RM axiom, $y_i^t \geqslant y_i^s$ for any $t \in T(\mathbf{z}^s)$. It follows that $\Pi_i(\cdot, \mathbf{z}^s) = f_i(\cdot, s)$, and so $\text{argmax}_{a \in Y_i} \Pi_i(a, \mathbf{z}^s) = y_i^s$. Lastly, property (c) flows from the single-peakedness of each $f_i(\cdot, t)$. **QED**

Proof of equation (12). If $\mathcal{P}$ is $\mathcal{SC}$-rationalizable and there is $(\tau^{\mathrm{B}})_{\mathrm{B} \in \mathcal{B}}$ that solves (7), then there is P_Π that rationalizes $\mathcal{P}$ such that $\int_{\Pi \in \mathcal{SC}^*} d\mathrm{P}_\Pi = \sum_{\mathrm{B} \in \mathcal{B}^*} \tau^{\mathrm{B}}$; this is clear from the proof of the "if" part of Theorem 2. Conversely, for any rationalization P_Π of $\mathcal{P}$, we claim there is $(\tau^{\mathrm{B}})_{\mathrm{B} \in \mathcal{B}}$ that solves (7) such that $\int_{\Pi \in \mathcal{SC}^*} d\mathrm{P}_\Pi \leqslant \sum_{\mathrm{B} \in \mathcal{B}^*} \tau^{\mathrm{B}}$. Indeed, construct $(\tau^B)_{\mathrm{B} \in \mathcal{B}}$ in the same way as in the proof of Theorem 2. Notice that if $\Pi \in \mathcal{SC}^*$, then $p(\mathrm{B}, \Pi) = 0$ for any $\mathrm{B} \notin \mathcal{B}^*$ and so it follows from (9) that $\sum_{\mathrm{B} \in \mathcal{B}^*} p(\mathrm{B}, \mathbf{\Pi}) = 1$ since $p(\mathrm{B}, \Pi) = 0$ for any $\mathrm{B} \notin \mathcal{B}^*$. This gives us (12). **QED**

The proof of Proposition 1 requires the following well-known result from convex analysis.

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

Proof of Proposition 1. If for all $\mathbf{b} \in \mathcal{B}$, we have $(\mathbf{p} - \eta_0) \cdot (\mathbf{b} - \eta_0) \leqslant 0$, then $(\mathbf{p} - \eta_0) \cdot (\mathbf{v} - \eta_0) \leqslant 0$ for all $\mathbf{v}$ in the convex hull of $\mathcal{B}$. This implies, by Lemma 1 that the distance between $\mathbf{p}$ and any $\mathbf{v}$ in the convex hull of $\mathcal{B}$ is again minimized at η_0, which means that $\mathcal{B}_0$ is not improvable. Conversely, if there is $\hat{\mathbf{b}}$ in $\mathcal{B}$ such that (18) holds, then, appealing to Lemma 1 again, we know that η_0 does *not* minimize the distance between $\mathbf{p}$ and the convex hull of $\mathcal{B}_0 \cup \{\hat{\mathbf{b}}\}$ and $\hat{\mathbf{b}}$ improves $\mathcal{B}_0$. **QED**

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

$$\mathcal{R}(\mathbf{y}, \mathbf{x}) = \{(\mathbf{y}', \mathbf{x}') : \mathbf{y} = \mathrm{B}(\mathbf{x}) \Longrightarrow \mathbf{y}' \neq \mathrm{B}(\mathbf{x}') \text{ for all } \mathrm{B} \in \mathcal{B}\}.$$

Recalling the definition of the RM axiom (6), $(\mathbf{y}', \mathbf{x}') \in \mathcal{R}(\mathbf{y}, \mathbf{x})$ holds if there exists some $i \in \mathcal{N}$ such that $y_i' < (>) y_i$ and $(y_{-i}', x_i') \geqslant (\leqslant)(y_{-i}, x_i)$. Impose any linear ranking on the elements of $\mathbf{Y} \times \hat{\mathbf{X}}$; we define $C = (c_{(\mathbf{y},\mathbf{x}),(\mathbf{y}',\mathbf{x}')})_{\mathbf{Y} \times \hat{\mathbf{X}}, \mathbf{Y} \times \hat{\mathbf{X}}}$ to be a $|\mathbf{Y} \times \hat{\mathbf{X}}| \times |\mathbf{Y} \times \hat{\mathbf{X}}|$ matrix where $c_{(\mathbf{y},\mathbf{x}),(\mathbf{y},\mathbf{x})} = |\hat{\mathbf{X}}|$ and, if $(\mathbf{y}, \mathbf{x}) \neq (\mathbf{y}', \mathbf{x}')$, then $c_{(\mathbf{y},\mathbf{x}),(\mathbf{y}',\mathbf{x}')} = \mathbf{1}[(\mathbf{y}', \mathbf{x}') \in \mathcal{R}(\mathbf{y}, \mathbf{x})]$. By setting $\theta = (|\hat{\mathbf{X}}|, |\hat{\mathbf{X}}|, \ldots, |\hat{\mathbf{X}}|)$ (a column vector of length $|\mathbf{Y} \times \hat{\mathbf{X}}|$), we claim that a single-valued group type $\mathbf{b}$ (thought of as a column vector) obeys RM axiom if and only if $C\mathbf{b} \leqslant \theta$. Indeed, since $\mathbf{b}$ is single-valued, we have $\sum_{\mathbf{y}\in\mathbf{Y}} \mathbf{b}_{(\mathbf{y},\mathbf{x})} = 1$ for all $\mathbf{x} \in \widehat{\mathbf{X}}$, which guarantees that $(C\mathbf{b})_{(\mathbf{y},\mathbf{x})} \leqslant |\widehat{\mathbf{X}}|$ if $\mathbf{b}_{(\mathbf{y},\mathbf{x})} = 0$. Note that $(C\mathbf{b})_{(\mathbf{y},\mathbf{x})} \geqslant c_{(\mathbf{y},\mathbf{x}),(\mathbf{y},\mathbf{x})} = |\widehat{\mathbf{X}}|$ if $\mathbf{b}_{(\mathbf{y},\mathbf{x})} = 1$. If $\mathbf{b}$ satisfies the RM axiom, then $(C\mathbf{b})_{(\mathbf{y},\mathbf{x})} = |\widehat{\mathbf{X}}|$ for all $(\mathbf{y},\mathbf{x})$ with $\mathbf{b}_{(\mathbf{y},\mathbf{x})} = 1$; if $\mathbf{b}$ violates the RM axiom, then there is $(\tilde{\mathbf{y}},\tilde{\mathbf{x}})$ with $\mathbf{b}_{(\tilde{\mathbf{y}},\tilde{\mathbf{x}})} = 1$ such that $(C\mathbf{b})_{(\tilde{\mathbf{y}},\tilde{\mathbf{x}})} > |\widehat{\mathbf{X}}|$. **QED**

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

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

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

$$\mathbf{v} = \beta\left(\sum_{l=1}^{\bar{l}} t_l \mathbf{b}_l^*\right) + (1-\beta)\left(\sum_{l=1}^{\bar{l}} t_l \mathbf{b}_l\right) = \sum_{l=1}^{\bar{l}} t_l \left[\beta\mathbf{b}_l^* + (1-\beta)\mathbf{b}_l\right],$$

which establishes our claim. **QED**

Proof of Proposition 3. Note that $J_{N,0}(\beta)$ is the distance between $\mathbf{q}$ and $V(\mathcal{B}_0)$ and this distance is achieved at $\eta_0 \in V(\mathcal{B}_0)$. If, for all $\beta\mathbf{b}^* + (1-\beta)\mathbf{b}$ where $\mathbf{b}^* \in \mathcal{B}^*$ and $\mathbf{b} \in \mathcal{B}$, we have

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

then $(\mathbf{q} - \eta_0) \cdot (\mathbf{v} - \eta_0) \leqslant 0$ for all $\mathbf{v} \in V(\mathcal{B})$, by Lemma 2. This in turn means, by Lemma 1, that $\mathcal{B}_0$ is not improvable given problem (21). Conversely, suppose that there is a pair of group types $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$, with $\widehat{\mathbf{b}}^* \in \mathcal{B}^*$ and $\widehat{\mathbf{b}} \in \mathcal{B}$, such that (22) holds, then, by Lemma 1, η_0 does *not* minimize the distance between $\mathbf{q}$ and the convex hull of $V(\mathcal{B}_0)$ and $\beta\widehat{\mathbf{b}}^* + (1-\beta)\widehat{\mathbf{b}}$. We conclude that $\left\{\widehat{\mathbf{b}}^*, \widehat{\mathbf{b}}\right\}$ improves $\mathcal{B}_0$ given problem (21). **QED**

References

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

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

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

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

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

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

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

Chen, X., T. M. Christensen, and E. Tamer (2018): Monte Carlo confidence sets for identified sets. *Econometrica*, 86, 1965-2018.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): Average and quantile effects in nonseparable panel models. *Econometrica*, 81, 535-580.

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 (2022): Revealed price preference: theory and empirical analysis. *arXiv:1801.02702*.

Evans, W., and I. Kessides (1994): Living by the 'Golden Rule': Multimarket Contact in the U.S. Airline Industry,' *Quarterly Journal of Economics*, 109, 341–366.

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

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

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

Honoré, B. E., and E. Tamer (2006): Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74, 611-629.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Online Appendix

A1. Connection to the linear model

A1.1. An $\mathcal{SC}$-rationalizable distribution inconsistent with the linear model

As we pointed out in Section 2 in the main paper, there is a set of population distributions $\mathcal{P} = \{\mathrm{P}(\mathbf{y}|\mathbf{x})\}_{\mathbf{x}\in\hat{\mathbf{X}}}$ that is $\mathcal{SC}$-rationalizable, but not compatible with the linear specification often adopted in the literature. We now consider a closely related example of such a set of distributions, but with exactly the same framework as the entry game in Section 5 of the main paper. That is, $\mathcal{N} = \{1,2\}$, $y_i \in \{N, E\}$, and $x_i = (\mathrm{MP}_i, \mathrm{MS}) \in \{0,1\} \times \{0,1\}$ for $i = 1, 2$, with the value of MS being shared by both players.

The set of distributions $\mathcal{P}$ summarized in Table A.1 is $\mathcal{SC}$-rationalizable (one can confirm this using a part of our program). However, it is inconsistent with pure strategy Nash equilibrium play under the following specification of payoff functions. For $i = 1$, 2, we assume that the payoff of not entering (N) is always zero and the payoff of entering (E) is given by

$$\pi_i(E, y_{-i}, x_i, \varepsilon) = \alpha_i + \beta_i \,\mathrm{MP}_i + \gamma_i \,\mathrm{MS} + \delta_i \mathbf{1}(y_{-i} = E) + \varepsilon_i, \tag{a.1}$$

with $(\beta_i, \gamma_i) > 0$ and $\delta_i < 0$. In addition, suppose that the joint distribution of $(\varepsilon_1, \varepsilon_2)$ is absolutely continuous and fully supported, which is satisfied by many distributions employed in the literature (such as the joint normal distribution).

To see the inconsistency, it suffices to look at the subtables of Table A.1 with $(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0, 0, 1)$, and $(0, 1, 1)$. Suppose by way of contradiction that this set of distribution is explained as Nash equilibrium play under the payoff functions specified by (a.1). Then, it must hold that $\beta_2 = 0$, since $\mathrm{P}(N, N|0, 1, 1) - \mathrm{P}(N, N|0, 0, 1) = 0$. Indeed, letting μ be the induced probability measure of $(\varepsilon_1, \varepsilon_2)$, this difference is equal to the difference between

$$\mu\left(\{(\varepsilon_1, \varepsilon_2) : \alpha_1 + \gamma_1 + \varepsilon_1 < 0 \text{ and } \alpha_2 + \beta_2 + \gamma_2 + \varepsilon_2 < 0\}\right)$$

and

$$\mu\left(\{(\varepsilon_1, \varepsilon_2) : \alpha_1 + \gamma_1 + \varepsilon_1 < 0 \text{ and } \alpha_2 + \gamma_2 + \varepsilon_2 < 0\}\right).$$

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,0,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.250	0.250	0.333	0.167

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,1,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.167	0.333	0.167	0.333

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,0,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.167	0.250	0.416	0.167

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,1,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.083	0.333	0.250	0.333

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,0,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.083	0.416	0.250	0.250

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,1,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.083	0.333	0.167	0.416

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,0,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.083	0.333	0.250	0.333

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,1,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.083	0.250	0.167	0.500

Table A.1: An $\mathcal{SC}$-rationalizable $\mathcal{P}$ inconsistent with the linear model

Then, our hypothesis on the distribution ensures that $\beta_2 = 0$. On the other hand, we also have $\mathrm{P}(E,E|0,1,1) - \mathrm{P}(E,E|0,0,1) > 0$, which implies that the difference between

$$\mu\left(\{(\varepsilon_1, \varepsilon_2) : \alpha_1 + \gamma_1 + \delta_1 + \varepsilon_1 \geqslant 0 \text{ and } \alpha_2 + \beta_2 + \gamma_2 + \varepsilon_2 \geqslant 0\}\right)$$

and

$$\mu\left(\{(\varepsilon_1, \varepsilon_2) : \alpha_1 + \gamma_1 + \delta_1 + \varepsilon_1 \geqslant 0 \text{ and } \alpha_2 + \gamma_2 + \varepsilon_2 \geqslant 0\}\right)$$

is positive. This, in turn, implies that $\beta_2 > 0$, contradicting the preceding argument. Note that, here, we only use the frequencies of (N,N) and (E,E), which always arise as unique equilibria, and hence a selection scheme from multiple equilibria does not matter.

A1.2 Comparison based on simulation data

We heve shown that the set of distributions depicted in Table A.1 is $\mathcal{SC}$-rationalizable, but incompatible with the parametric specification in (a.1). In what follows, using simulation data, we examine whether this difference is in fact empirically relevant. Specifically, we generate random samples using the set of distributions in Table A.1 and check how often $\mathcal{SC}$-rationalizability is rejected and how often the linear specification is rejected. In Kline and Tamer (2016), they provide a procedure to estimate the identified set of coefficients in (a.1); i.e. $\theta := (\alpha_1, \beta_1, \gamma_1, \delta_1; \alpha_2, \beta_2, \gamma_2, \delta_2)$. As they pointed out in their paper, their procedure can also work as a test for model (mis-)specification by checking whether the estimated identified set is nonempty. We also implement our test to the same set of samples to check if each of them is actually supported by our model.

Before proceeding to the result, we briefly refer to how Kline and Tamer's test works. They consider a nonegative function that summarizes the relationship between $\mathcal{P}$ and θ, say, $M(\theta, \mathcal{P})$, under suitable sign restrictions (in the current case, $(\beta_i, \gamma_i) > 0$ and $\delta_i < 0$ for $i = 1, 2$). This function is designed so that θ can generate $\mathcal{P}$ if and only if $M(\theta, \mathcal{P}) = 0$. The set $\Theta_I = \{\theta : M(\theta, \mathcal{P}) = 0\}$ is then interpreted as the identified set for θ. To deal with empirical distributions $\mathcal{Q}$, they introduce an exogenous tolerance parameter $\rho > 0$ so that the set estimation of Θ_I is such that $\widehat{\Theta}_I = \{\theta : M(\theta, \mathcal{Q}) \leqslant \rho\}$, and one can conclude that the specification is valid if $\widehat{\Theta}_I \neq \varnothing$. Since there is no guide in the paper to select this parameter, we report results for the tolerance parameter they use, which is $\rho = 0.075$.

Recall that the structure of the model behind $\mathcal{P}$ in Table A.1 is the same as the empirical data set in Section 5 of the main paper. We generate 100 samples from $\mathcal{P}$ such that each sample has the same size $N = 7882$ and the same fraction of realization of each $\mathbf{x} \in \{0, 1\}^3$ as the data set in Section 5. We find that 92 samples (out of the 100) pass our test with 5% significance level, while Kline and Tamer's procedure cannot find nonempty identified sets for any of these samples (i.e. all samples fail their test). Thus, the linear specification itself is rejected and having a nonparametric test that allows for nonlinearity is important.

Remark. Note that the preceding result suggests that Kline and Tamer's test has very good testing power, since their test actually rejects samples generated by $\mathcal{P}$, which is inconsistent with the linear model. It may be interesting to see whether our test can also reject samples generated by distributions inconsistent with *our* model. To check this, we generated 100 samples from another set of distributions, which we shall refer to as $\mathcal{P}'$, which is not $\mathcal{SC}$-rationalizable. In particular, to see whether our test can detect subtle inconsistency, we use $\mathcal{P}'$ summarized in Table A.2, which is not too different from $\mathcal{P}$ in Table A.1. In fact, this $\mathcal{P}'$ is equal to the empirical distribution of one of the samples from $\mathcal{P}$, with a p-value is equal to 0.01. Out of 100 samples drawn from $\mathcal{P}'$, we find that 97 samples *fail* our test, which implies that our test also has reasonably strong testing power.

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,0,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.255	0.265	0.337	0.143

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,1,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.151	0.336	0.193	0.330

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,0,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.167	0.240	0.441	0.152

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,1,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.078	0.332	0.254	0.337

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,0,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.080	0.445	0.233	0.242

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,1,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.093	0.319	0.155	0.434

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,0,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.088	0.302	0.245	0.366

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (1,1,1)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
0.087	0.263	0.164	0.487

Table A.2: Non $\mathcal{SC}$-rationalizable distribution $\mathcal{P}'$

A2. Multi-dimensional action spaces

In the main paper, we assume that the set of actions of each player $i \in \mathcal{N} = \{1, 2, ..., n\}$, Y_i, is a finite and totally ordered (in other words, it is a finite chain). In this section, we show that all our results are valid, as long as every Y_i is a product of finite chains. In what follows, let for each $i \in \mathcal{N}$, $Y_i = \times_{k=1}^{K(i)} Y_{ik}$ with every Y_{ik} is a finite chain.

As shown by Milgrom and Shannon (1994), the counterpart of the Basic Theorem for multi-dimensional action spaces requires *quasisupermodularity* in addition to the single crossing differences.[1] To be precise, $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i) = \operatorname{argmax}_{y_i \in Y_i} \Pi_i(y_i, \mathbf{y}_{-i}, x_i)$ is monotone in $(\mathbf{y}_{-i}, x_i)$ if the payoff function Π_i is quasisupermodular in y_i and obeys single crossing differences in $(y_i; \mathbf{y}_{-i}, x_i)$. Given that Y_i is assumed to be a product of chains, it is straightforward to show that the combination of quasisupermodularity and condition (4) is equivalent to the following stronger version of single crossing differences: for every nonempty set $J \subset \{1, 2, \ldots, K(i)\}$, $y''_{iJ} > y'_{iJ}$ and $(y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i) > (y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i)$,

$$\Pi_i \left(y''_{iJ}, y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i\right) > \Pi_i \left(y'_{iJ}, y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i\right) \tag{a.2}$$
$$\Longrightarrow \Pi_i \left(y''_{iJ}, y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i\right) > \Pi_i \left(y'_{iJ}, y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i\right).$$

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

Note that, here, y_{iJ} and $y_{i(-J)}$ denote the subvectors on J and its complement respectively that together constitute y_i. In other words, if over some subset of dimensions J, the agent prefers a higher action y''_{iJ} to a lower one y'_{iJ}, keeping fixed the actions on the other dimensions and the covariates, then that preference is maintained if actions on the other dimensions and/or the covariates are raised. Let $\mathcal{SC}$ be the set of profiles of payoff functions $\mathbf{\Pi} = (\Pi_1, \Pi_2, ..., \Pi_n)$ in which every Π_i obeys the single crossing differences in the sense of (a.2). The following result is the multi-dimensional analog to the Basic Theorem in the main paper.

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

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

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

The notions of (generalized) group types, single crossing group types and the RM axiom can all be straightforwardly extended to the case where each player has a multidimensional action space.

With this theorem, It is clear that all our notions can be trivially adjusted to the current setting using exactly the same notation. It is also easy to see that all our results are valid, if the RM axiom still characterizes a group type consistent with the model even in multi-dimensional setting. In the rest of this section, we show that this is affirmative:

The next results states that Theorem 1 can be also extended to the case with multi-dimensional action spaces.

Theorem A.1. *A generalized group type* $\mathrm{B} : \widehat{\mathbf{X}} \rightrightarrows \mathbf{Y}$ *is a single-crossing group type if and only if it satisfies the RM axiom.*

Proof. The Basic Theorem guarantees that if B is a single-crossing group type, then it obeys the RM axiom. It remains for us to show the converse. Our strategy is to explicitly construct payoff functions that rationalize B and satisfies single crossing differences in the sense of (a.2). Our strategy is to construct a payoff function $\Pi : Y_i \times \mathbf{Y}_{-i} \times X_i \to \mathbb{R}$ of the form

$$\Pi_i(y_i, \mathbf{y}_{-i}, x_i) = \sum_{k=1}^{K(i)} \Pi_{ik}(y_{ik}, \mathbf{y}_{-i}, x_i), \tag{a.3}$$

with each $\Pi_{ik}(y_{ik}, \mathbf{y}^t_{-i}, x^t_i)$ having increasing differences: for every $y''_{ik} > y'_{ik}$, and $(\mathbf{y}''_{-i}, x''_i) > (\mathbf{y}'_{-i}, x'_i)$,

$$\Pi_{ik}(y''_{ik}, \mathbf{y}''_{-i}, x''_i) - \Pi_{ik}(y'_{ik}, \mathbf{y}''_{-i}, x''_i) \geqslant \Pi_{ik}(y''_{ik}, \mathbf{y}'_{-i}, x'_i) - \Pi_{ik}(y'_{ik}, \mathbf{y}'_{-i}, x'_i). \tag{a.4}$$

It is easy to see that, then, Π_i also obeys the increasing differences, which in turn implies single crossing differences in the sense of (a.2). We also ensure that for each $(\mathbf{y}_{-i}, x_i)$, $\mathrm{BR}_i(\mathbf{y}_{-i}, x_i)$ is a singleton for every $x_i \in X_i$.

Similar to the proof of Theorem 1 in the main paper, we introduce the following notation. For each $i \in \mathcal{N}$, $\mathbf{Z}_i = \mathbf{Y}_{-i} \times X_i$. Since $\widehat{\mathbf{X}}$ is finite, gathering together with the finiteness of every Y_i, the graph of $\mathrm{B}(\mathbf{x})$, which we represent as $\mathcal{G}(\mathrm{B}) := \left\{(\mathbf{y}, \mathbf{x}) : \mathbf{y} \in \mathrm{B}(\mathbf{x}) \text{ for some } \mathbf{x} \in \widehat{\mathbf{X}}\right\}$ is also finite. Hence, with a suitable finite set of indices $\mathcal{T} = \{1, 2, ..., T\}$, it can be written as $\mathcal{G}(\mathrm{B}) = \{(\mathbf{y}^t, \mathbf{x}^t) : \mathbf{y}^t \in \mathrm{B}(\mathbf{x}^t) \text{ for some } t \in \mathcal{T}\}$. Notice that, letting $\mathbf{z}^t_i := (\mathbf{y}^t_{-i}, x^t_i)$ for $t \in \mathcal{T}$, each $(\mathbf{y}^t, \mathbf{x}^t) \in \mathcal{G}(\mathrm{B})$ can be written as $(y^t_i, \mathbf{z}^t_i)$ for every $i \in \mathcal{N}$.

Repeating the procedure in the proof of Theorem 1 in the main paper, for every $i \in \mathcal{N}$ and $k = 1, 2, ..., K(i)$, we obtain $\Pi_{ik} : Y_{ik} \times \mathbf{Z}_i \to \mathbb{R}$ such that $\Pi_i(y''_{ik}, \mathbf{z}'') - \Pi_i(y'_{ik}, \mathbf{z}'') \geqslant \Pi_i(y''_{ik}, \mathbf{z}') - \Pi_i(y'_{ik}, \mathbf{z}')$ for all $y''_{ik} > y'_{ik}$, and that $y^t_{ik} = \operatorname{argmax}_{y_{ik} \in Y_{ik}} \Pi_{ik}(y_{ik}, \mathbf{z}^t_i)$ for every $t \in \mathcal{T}$. (Just replace Y_i there with Y_{ik}.) Using these Π_{ik}'s, it is clear that for every $t \in \mathcal{T}$, a multi-dimensional action $y^t_i = (y^t_{i1}, ..., y^t_{iK(i)})$ is the unique maximizer of $\Pi_i(y_i, \mathbf{z}_i) = \sum_{k=1}^{K(i)} \Pi_{ik}(y_{ik}, \mathbf{z}_i)$ at $\mathbf{z}^t_i$ for every $t \in \mathcal{T}$. Lastly, with Y_i taking only finitely many values, we can always guarantee that $\Pi_i(\cdot, \mathbf{z})$ has *strict* preference over Y_i at every value of $\mathbf{Z}_i$ by perturbing f_{ik} if necessary. **QED**

A3. More results on inference and predictions

This section contains results omitted from Section 3.4 of the main paper. In the first subsection, we explain how we can obtain a tight bound on the probability that an agent has a given ranking between a pair of actions. The second subsection expands on the discussion of Nash equilibrium predictions in Section 3.4 (Application 2) and also establishes that *the set of Nash equilibrium predictions increases with the covariate*, in a sense related to first order stochastic dominance.

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

$$\mathrm{P}(\mathbf{y}|\mathbf{x}) = \sum_{\{\mathrm{B}\in\mathcal{B}:\ \mathrm{B}(\mathbf{x})=\mathbf{y}\}} \tau^{\mathrm{B}} \text{ for all } \mathbf{y} \in \mathbf{Y} \text{ and } \mathbf{x} \in \hat{\mathbf{X}}. \tag{a.5}$$

A3.1. Predicting player preferences

We are interested in estimating the proportion of groups in the population where agent i prefers some action y_i'' over another action y_i', when the covariate takes a specific value x_i^* and other players are playing a given profile of strategies $\mathbf{y}_{-i}^*$. In formal terms, letting $\mathbf{z}^* = (\mathbf{y}_{-i}^*, x_i^*)$, we would like to identify the maximal and minimal possible probabilities of

$$S = \{\mathbf{\Pi} \in \mathcal{SC} : \Pi_i \text{ satisfies } \Pi_i(y_i'', \mathbf{z}^*) > \Pi_i(y_i', \mathbf{z}^*)\}. \tag{a.6}$$

Letting $(\mathbf{y}, \mathbf{x}) \in \mathbf{Y} \times \hat{\mathbf{X}}$ so that $\mathbf{y} = (y_i, \mathbf{y}_{-i}^*)$ and $\mathbf{x} = (x_i^*, \mathbf{x}_{-i})$, it is clear that $\Pr(S) \geqslant \mathrm{P}(\mathbf{y}|\mathbf{x})$. But, in fact, we can obtain a sharper lower bound for $\Pr(S)$ by exploiting the assumption that players have single crossing payoff functions.

Proposition A.1. *Suppose that $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x})\}_{\mathbf{x}\in\hat{\mathbf{X}}}$ is $\mathcal{SC}$-rationalizable by some distribution $\mathrm{P}_{\mathbf{\Pi}}$. Then for S defined by (a.6),*

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

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

- *if $y_i'' > y_i'$ then $\mathbf{m}(y_i'', y_i') = \min \sum_{\mathrm{B}\in\underline{\mathcal{B}}} \tau^{\mathrm{B}}$ subject to τ solving (a.5), with*

$$\underline{\mathcal{B}} = \{\mathrm{B} \in \mathcal{B} :\ \mathrm{B}(\mathbf{x}) = (y_i'', \mathbf{y}_{-i}) \textit{ and } (\mathbf{y}_{-i}, x_i) \leqslant \mathbf{z}^* \textit{ for some } \mathbf{x} \in \hat{\mathbf{X}}\ \}; \tag{a.7}$$

- *if $y_i'' < y_i'$ then $\mathbf{m}(y_i'', y_i') = \min \sum_{\mathrm{B}\in\overline{\mathcal{B}}} \tau^{B}$ subject to τ solving (a.5), with*

$$\overline{\mathcal{B}} = \{\mathrm{B} \in \mathcal{B} :\ \mathrm{B}(\mathbf{x}) = (y_i'', \mathbf{y}_{-i}) \textit{ and } (\mathbf{y}_{-i}, x_i) \geqslant \mathbf{z}^* \textit{ for some } \mathbf{x} \in \hat{\mathbf{X}}\ \}. \tag{a.8}$$

Proof. We consider the case where $y_i'' > y_i'$, since the case where $y_i'' < y_i'$ proceeds in analogous fashion. Let

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

where $\underline{\mathcal{B}}$ is defined by (a.7). Then, for each $\mathbf{\Pi} \in S'$ and $\mathrm{B} \in \underline{\mathcal{B}}$ rationalized by it, there exists some $\mathbf{x} \in \hat{\mathbf{X}}$ for which $\mathrm{B}(\mathbf{x}) = (y_i'', \mathbf{y}_{-i})$ and $(\mathbf{y}_{-i}, x_i) \leqslant \mathbf{z}^*$, and $\Pi_i(y_i'', \mathbf{y}_{-i}, x_i) > \Pi_i(y_i', \mathbf{y}_{-i}, x_i)$. Since $(\mathbf{y}_{-i}, x_i) \leqslant \mathbf{z}^*$ and Π_i obeys single-crossing differences, the above implies that $\Pi_i(y_i'', \mathbf{z}^*) > \Pi_i(y_i', \mathbf{z}^*)$. Thus, the weight on S' should be weakly smaller than that on S, and we conclude that

$$\sum_{\mathrm{B} \in \underline{\mathcal{B}}} \tau^{\mathrm{B}} = \sum_{\mathrm{B} \in \underline{\mathcal{B}}} \int \mathrm{P}(\mathrm{B}|\mathbf{\Pi}) d\mathrm{P}_{\mathbf{\Pi}} = \int \sum_{\mathrm{B} \in \underline{\mathcal{B}}} \mathrm{P}(\mathrm{B}|\mathbf{\Pi}) d\mathrm{P}_{\mathbf{\Pi}} \leqslant \int_{S'} d\mathrm{P}_{\mathbf{\Pi}} \leqslant \int_{S} d\mathrm{P}_{\mathbf{\Pi}},$$

where $\mathrm{P}(\mathrm{B}|\mathbf{\Pi})$ stands for the (unobserved) probability that B realizes conditional on $\mathbf{\Pi}$. Note that the first equality follows, since for each B, $\tau^B = \int \mathrm{P}(\mathrm{B}|\mathbf{\Pi}) d\mathrm{P}_{\mathbf{\Pi}}$ holds whenever τ solves (a.5), that the penultimate inequality holds, since $\sum_{\mathrm{B} \in \underline{\mathcal{B}}} \mathrm{P}(\mathrm{B}|\mathbf{\Pi})$ does not exceed 1 and equals 0 if $\mathbf{\Pi} \notin S'$, and that the final inequality flows from $S' \subset S$. Given that τ must obey (a.5), a lower bound on $\sum_{\mathrm{B} \in \underline{\mathcal{B}}} \tau^{\mathrm{B}}$ is $\mathbf{m}(y_i'', y_i')$, which proves our claim. **QED**

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

$$\mathbf{m}(y_i'', y_i') \leqslant \int_S d\mathrm{P}_{\mathbf{\Pi}} \leqslant 1 - \mathbf{m}(y_i', y_i''). \tag{a.9}$$

We can calculate $\mathbf{m}(y_i'', y_i')$ and $\mathbf{m}(y_i', y_i'')$ from the conditional choice distributions by solving the relevant linear program. The next result strengthens Proposition A.1 by showing that the bounds in (a.9) are tight.

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

$$\mathbf{m}(y_i'', y_i') = \int_S dP_{\mathbf{\Pi}}; \tag{a.10}$$

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

$$\int_S dP_{\mathbf{\Pi}} = 1 - \mathbf{m}(y_i', y_i''). \tag{a.11}$$

Proof. Notice that (a.11) is equivalent to there being a distribution $\mathrm{P}_{\mathbf{\Pi}}$ with support on $\mathcal{SC}$ such that $\int_{\hat{S}} d\mathrm{P}_{\mathbf{\Pi}} = \mathbf{m}(y_i', y_i'')$ where

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

Therefore, to prove (a.9), it suffices to establish (a.10).

We first consider the case where $y_i'' > y_i'$. Suppose that $\tau = \underline{\tau}$ solves $\min \sum_{B \in \underline{\mathcal{B}}} \tau^{\mathrm{B}}$ subject to τ satisfying (7) in the main paper, with $\underline{\mathcal{B}}$ given by (a.7), so that $\mathbf{m}(y_i'', y_i') = \sum_{B \in \underline{\mathcal{B}}} \underline{\tau}^{\mathrm{B}}$. We know from our proof of Theorem 2 (see the discussion immediately preceding the statement of the theorem in Section 3.2) that $\mathcal{P}$ can be rationalized by a distribution $\mathrm{P}^*_{\mathbf{\Pi}}$ that gives weight of τ^{B} to a profile $\mathbf{\Pi}^{\mathrm{B}} \in \mathcal{SC}$ that rationalizes B; by taking strictly increasing transformations if necessary, we can guarantee that $\mathbf{\Pi}^{\mathrm{B}} \neq \mathbf{\Pi}^{\mathrm{B}'}$ for any $\mathrm{B} \neq \mathrm{B}'$. If $\mathrm{B} \in \underline{\mathcal{B}}$, then any $\mathbf{\Pi}^{\mathrm{B}}$ that rationalizes B will satisfy $\Pi_i^{\mathrm{B}}(y_i'', \mathbf{z}^*) > \Pi_i^{\mathrm{B}}(y_i', \mathbf{z}^*)$, so $\int_S d\mathrm{P}^*_{\mathbf{\Pi}} \geqslant \mathbf{m}(y_i'', y_i')$. We claim that (a.10) in fact holds for the distribution $\mathrm{P}^*_{\mathbf{\Pi}}$. To show this, it suffices to prove that if $\mathrm{B} \notin \underline{\mathcal{B}}$ then there is $\mathbf{\Pi}^{\mathrm{B}} \in \mathcal{SC}$ rationalizing B such that Π_i^{B} satisfies

$$\Pi_i^{\mathrm{B}}(y_i'', \mathbf{z}^*) < \Pi_i^{\mathrm{B}}(y_i', \mathbf{z}^*), \tag{a.12}$$

so that $\mathbf{\Pi}^{\mathrm{B}} \notin S$. In what follow, we fix some $\mathrm{B} \in \mathcal{B} \backslash \underline{\mathcal{B}}$ and explicitly construct $\mathbf{\Pi}^{\mathrm{B}}$ that rationalizes B and Π_i satisifes (a.12).

Since B is chosen from $\mathcal{B}$, the existence of $\mathbf{\Pi} \in \mathcal{SC}$ that rationalizes it is ensured. Hence, the only issue is whether we can find $\mathbf{\Pi}$ so that Π_i obeys (a.12). To construct such Π_i, we start from specifying the ordinal contents of it. Let us define $\mathbf{Z} := \mathbf{Y}_{-i} \times \mathrm{proj}_i \mathbf{X}$, and denote a typical element $(\mathbf{y}_{-i}, x_i)$ by $\mathbf{z}$. For $\mathbf{z} = (\mathbf{y}_{-i}, x_i)$, if there is some $\mathbf{x} \in \widehat{\mathbf{X}}$ such that x_i is the i-th component of it and $\mathbf{y}_{-i}$ is specified by $\mathrm{B}(\mathbf{x})$, then we denote it by $\mathbf{z}(\mathbf{x})$. Similarly, when $y_i \in Y_i$ is specified by $\mathrm{B}(\mathbf{x})$ at some $\mathbf{x} \in \widehat{\mathbf{X}}$, then we denote it by $y_i(\mathbf{x})$. Now, define the binary relation $\succ$ on $Y_i \times \mathbf{Z}$ as follows: for any pair $(\bar{y}_i, \mathbf{z})$ and $(\hat{y}_i, \mathbf{z})$ with $\bar{y}_i > \hat{y}_i$,

(i) $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$, if there is $\mathbf{x} \in \hat{\mathbf{X}}$ for which $\mathbf{z}(\mathbf{x}) \leqslant \mathbf{z}$ and $y_i(\mathbf{x}) = \bar{y}_i$.

(ii) $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$, if there is $\mathbf{x} \in \hat{\mathbf{X}}$ for which $\mathbf{z}(\mathbf{x}) \geqslant \mathbf{z}$ and $y_i(\mathbf{x}) = \hat{y}_i$.

(iii) $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$, for all other cases.

We claim that the above defined $\succ$ has the following properties: (**P1**) $\succ$ rationalizes the group type B; (**P2**) $(y_i', \mathbf{z}^*) \succ (y_i'', \mathbf{z}^*)$; (**P3**) any two distinct $(\bar{y}_i, \mathbf{z})$ and $(\hat{y}_i, \mathbf{z})$ are strictly comparable; (**P4**) $\succ$ is transitive on $Y_i \times \{\mathbf{z}\}$ for any $\mathbf{z} \in \mathbf{Z}$; (**P5**) $\succ$ has the single-crossing property in the sense that if $(y_i^{\star\star}, \mathbf{z}) \succ (y_i^{\star}, \mathbf{z})$ for some $y_i^{\star\star} > y_i^{\star}$ then $(y_i^{\star\star}, \tilde{\mathbf{z}}) \succ (y_i^{\star}, \tilde{\mathbf{z}})$ for any $\tilde{\mathbf{z}} > \mathbf{z}$. Assuming that these properties hold, it is clear that any function Π_i that represents $\succ$ (in the sense that $\Pi_i(y_i^{\star\star}, \mathbf{z}) > \Pi_i(y_i^{\star}, \mathbf{z})$ whenever $(y_i^{\star\star}, \mathbf{z}) \succ (y_i^{\star}, \mathbf{z})$) will be a payoff function that obeys single-crossing differences, rationalizes i's actions, and (because of (**P2**)) satisfies (a.12). Note that the existence of a representation for $\succ$ is clear since $\succ$ satisfies (**P3**) and (**P4**) and Y_i is a finite set.

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

To recap, we have shown that if $y_i'' > y_i'$ then the distribution $\mathrm{P}^*_{\mathbf{\Pi}}$ rationalizes the data and satisfies (a.10). It remains for us to prove the same result for $y_i'' < y_i'$. Using an analogous proof strategy, we need to show that for any $\mathrm{B} \notin \overline{\mathcal{B}}$, we can find $\mathbf{\Pi}^{\mathrm{B}} \in \mathcal{SC}$ rationalizing B such that Π_i^{B} satisfies (a.12) and so $\mathbf{\Pi} \notin S$. The proof proceeds by defining $\succ$ in the following way: for any pair $(\bar{y}_i, \mathbf{z})$ and $(\hat{y}_i, \mathbf{z})$ with $\hat{y}_i < \bar{y}_i$, (i) if there is $\mathbf{x} \in \hat{\mathbf{X}}$ such that $\mathbf{z}(\mathbf{x}) \leqslant \mathbf{z}$ and $y_i(\mathbf{x}) = \bar{y}_i$, then $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$; (ii) if there is $\mathbf{x} \in \hat{\mathbf{X}}$ such that $\mathbf{z}(\mathbf{x}) \geqslant \mathbf{z}$ and $y_i(\mathbf{x}) = \hat{y}_i$, then $(\hat{y}_i, \mathbf{z}) \succ (\bar{y}_i, \mathbf{z})$; (iii) if neither (i) nor (ii) holds then $(\bar{y}_i, \mathbf{z}) \succ (\hat{y}_i, \mathbf{z})$. In other words, the definition is the same as the one for the other case, except that (iii) has been modified. One could check that (**P1**) to (**P5**) hold and, in particular, (the new version of) (iii) guarantees (**P2**) since we now assume $y_i'' < y_i'$. With these properties on $\succ$, there is a function Π_i that represents $\succ$ and it will be a payoff function that obeys single-crossing differences, rationalizes i's actions, and satisfies (a.12). **QED**

A3.2. Nash Equilibrium predictions

In Section 3.4 of the main paper we posed the following question: given a strategy profile $\overline{\mathbf{y}}$ and covariate $\overline{\mathbf{x}}$, what is the greatest possible fraction of groups which have $\overline{\mathbf{y}}$ as a pure strategy Nash equilibrium at $\overline{\mathbf{x}}$, among all the possible $\mathcal{SC}$-rationalizations of $\mathcal{P}$? In this section, we pose a more general question: as a result of Nash equilibrium play with monotone best responses, what are the possible distributions of joint actions at the covariate value $\overline{\mathbf{x}}$? In formal terms, this amounts to identifying the *set* of conditional distributions $\mathrm{P}(\cdot|\overline{\mathbf{x}})$ such that the augmented set of distributions $\mathcal{P} \cup \{\mathrm{P}(\cdot|\overline{\mathbf{x}})\}$ is still $\mathcal{SC}$-rationalizable.

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

$$\mathrm{P}(\mathbf{y}|\mathbf{x}) = \sum\nolimits_{\{\mathrm{B}\in\tilde{\mathcal{B}}\,:\,\mathrm{B}(\mathbf{x})=\mathbf{y}\}} \tilde{\tau}^{\mathrm{B}} \text{ for each } \mathbf{y} \in \mathbf{Y} \text{ and } \mathbf{x} \in \hat{\mathbf{X}}, \text{ and} \tag{a.13}$$

$$\mathrm{P}(\mathbf{y}|\overline{\mathbf{x}}) = \sum\nolimits_{\{\mathrm{BT}\in\tilde{\mathcal{B}}\,:\,\mathrm{B}(\overline{\mathbf{x}})=\mathbf{y}\}} \tilde{\tau}^{\mathrm{B}} \text{ for each } \mathbf{y} \in \mathbf{Y}. \tag{a.14}$$

Note that the left hand side of the equations in (a.13) are distributions in $\mathcal{P}$, so those equations constitute conditions that $\tilde{\tau}$ has to satisfy. For any $\tilde{\tau}$ that satisfies those conditions, the resulting $\mathrm{P}(\cdot|\overline{\mathbf{x}})$ obtained from (a.14) is a predicted distribution at $\overline{\mathbf{x}}$. In other words, if we let $\mathbb{P}(\overline{\mathbf{x}})$ be the *set* of predicted distributions at $\overline{\mathbf{x}}$, then $\mathrm{P}(\cdot|\overline{\mathbf{x}})$ is in $\mathbb{P}(\overline{\mathbf{x}})$ if and only if there is $\tilde{\tau}$ that solves (a.13) and (a.14). Since the conditions are linear, $\mathbb{P}(\overline{\mathbf{x}})$ is a convex set and its properties can be found by further investigating the linear program.

The following result states that $\mathbb{P}(\overline{\mathbf{x}})$ is nonempty so long as $\mathcal{P}$ is $\mathcal{SC}$-rationalizable; in other words, that there *is* a solution to (a.13) and (a.14). This requires a short proof using the Basic Theorem. The result also tells us that $\mathbb{P}(\overline{\mathbf{x}})$ is, in a sense, increasing with respect to first order stochastic dominance.[2]

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

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

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

A4. Matrix representation of $\mathcal{B}^*$

We outline how we obtain the matrix characterization of $\mathcal{B}^*$ referred to in Section 3.4, which is in turn needed to implement the estimation we describe in Section 4.2.

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

Significance of strategic interactions. We need to construct the set of group types, $\mathcal{B}^*$, for which a subset of agents, $\mathcal{N}' \subset \mathcal{N}$, have payoff functions that do not depend on the strategies of any other agent. As explained in Section 3.4, these group types can be characterized by a stronger version of the RM axiom: for each $i \in \mathcal{N}'$, $\mathbf{y}'' \in \mathrm{B}(\mathbf{x}'')$, $\mathbf{y}' \in \mathrm{B}(\mathbf{x}')$, and $x_i'' > x_i' \Longrightarrow y_i'' \geqslant y_i'$. The standard RM axiom is required for the other agents. In this case, the matrix C^* and the column vector θ^* that characterize $\mathcal{B}^*$ can be constructed similarly to C and θ in the case of $\mathcal{B}$ (as provided in Proposition 2). We only need to incorporate in the definition of $\mathcal{R}(\mathbf{y}, \mathbf{x})$ the variation of the RM axiom.

Probability bounds for equilibrium actions. For a given $\overline{\mathbf{y}} \in \mathbf{Y}$ and $\overline{\mathbf{x}} \in \mathbf{X}$, let $\mathcal{B}^*$ be the set of group types that *can* support $\overline{\mathbf{y}}$ as a Nash equilibrium action profile at $\mathbf{x} = \overline{\mathbf{x}}$. As referred to in Section 3.4, a group type B is contained in $\mathcal{B}^*$, if and only if the (possibly) multi-valued group type $\overline{\mathrm{B}} : \widehat{\mathbf{X}} \cup \{\overline{\mathbf{x}}\} \rightrightarrows \mathbf{Y}$ defined as follows obeys the RM-axiom: let $\overline{B}$ be so that $\overline{B}(\overline{\mathbf{x}}) = \mathrm{B}(\overline{\mathbf{x}}) \cup \{\overline{\mathbf{y}}\}$ and $\overline{\mathrm{B}}(\mathbf{x}) = \mathrm{B}(\mathbf{x})$ for every $\mathbf{x} \in \widehat{\mathbf{X}} \backslash \{\overline{\mathbf{x}}\}$. (Note that $\overline{\mathrm{B}}(\overline{\mathbf{x}}) = \{\overline{\mathbf{y}}\}$, if $\overline{\mathbf{x}} \notin \widehat{\mathbf{X}}$.) Using the vector notation of B we can subsequently define the set $\mathcal{R}(\overline{\mathbf{y}}, \overline{\mathbf{x}})$ as follows

$$\mathcal{R}(\overline{\mathbf{y}}, \overline{\mathbf{x}}) = \left\{ (\mathbf{y}, \mathbf{x}) \in \mathbf{Y} \times \widehat{\mathbf{X}} : b_{(\overline{\mathbf{y}}, \overline{\mathbf{x}})} = 1 \Longrightarrow b_{(\mathbf{y}, \mathbf{x})} = 0 \text{ for all } \mathbf{b} \in \mathcal{B}^* \right\}.$$

Recalling the definition of the RM-axiom, $(\mathbf{y}, \mathbf{x}) \in \mathcal{R}(\overline{\mathbf{y}}, \overline{\mathbf{x}})$, if and only if there exists an agent i such that $(\mathbf{y}_{-i}, x_i) > (<)(\overline{\mathbf{y}}_{-i}, \overline{x}_i)$ and $y_i < (>)\overline{y}_i$.

Using this, in turn, define a vector $\zeta \in \{0, 1\}^{|\mathbf{Y} \times \widehat{\mathbf{X}}|}$ such that for each $(\mathbf{y}, \mathbf{x}) \in \mathbf{Y} \times \widehat{\mathbf{X}}$, $\zeta_{(\mathbf{y}, \mathbf{x})} = \mathbf{1}\left((\mathbf{y}, \mathbf{x}) \in \mathcal{R}(\overline{\mathbf{y}}, \overline{\mathbf{x}})\right)$. Then, let C^* be a $(|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1) \times |\mathbf{Y} \times \widehat{\mathbf{X}}|$-matrix such that the first $|\mathbf{Y} \times \widehat{\mathbf{X}}| \times |\mathbf{Y} \times \widehat{\mathbf{X}}|$-matrix equals the matrix C constructed in the proof of Proposition 2 (in the main paper), and the additional $(|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1)$-th row is equal to ζ (defined above). Finally, let the $(|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1)$-dimensional column vector θ^* be such that $\theta^* = (\theta, 0)$, where θ is as defined in the proof of Proposition 2 in the main paper. It follows that, a given group type $\mathbf{b} \in \overline{\mathcal{B}}$ is in the set $\mathcal{B}^*$ if and only if $C^*\mathbf{b} \leqslant \theta^*$. This inequality ensures that $C\mathbf{b} \leqslant \theta$, which is equivalent to $\mathbf{b}$ obeying RM axiom, and $\zeta \cdot \mathbf{b} \leqslant 0$, which is in turn equivalent to $\mathbf{b}$ not containing a behavior contradicting $(\overline{\mathbf{y}}, \overline{\mathbf{x}})$ in terms of the RM axiom on the extended domain.

A5. Omitted details from statistical tests

A5.1. Key lemma by Kitamura and Stoye (2018)

The bootstrap procedures with tightening *a lá* Kitamura and Stoye (2018) (including Smeulders et al. (2021), Deb et al. (2022), and ours) largely depend on Lemma 4.1 in Kitamura and Stoye (2018). It is worth restating the lemma here, given its relevance to what we are doing.

Letting $\mathbf{B}$ be an $m \times n$ matrix, a convex cone generated by $\mathbf{B}$ is represented as

$$\mathcal{A} = \{\mathbf{B}\tau : \tau \geqslant \mathbf{0}\}, \tag{a.15}$$

which is referred to as the $\mathcal{V}$-representation (meaning $\mathcal{V}$ertex) of a convex cone. By Minkowski-Weyl duality, $\mathcal{A}$ has an alternative representation, which is called the $\mathcal{H}$-representation (meaning $\mathcal{H}$yperplane) such that

$$\mathcal{A} = \{\mathbf{p} \in \mathbb{R}^m : \mathbf{D}\mathbf{p} \leqslant \mathbf{0}\} \tag{a.16}$$

for some $l \times m$ matrix $\mathbf{D}$. In the constraints in (a.16), some are inequality conditions while others are in fact equality conditions. To distinguish them, we let

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}^{\leqslant} \\ \mathbf{D}^{=} \end{bmatrix},$$

where $\mathbf{D}^{\leqslant}$ and $\mathbf{D}^{=}$ correspond respectively to the inequality and equality constraints. Abusing notation, we sometimes write $\mathbf{d} \in \mathbf{D}^{\leqslant}$ to represent that $\mathbf{d}$ is a row vector of $\mathbf{D}^{\leqslant}$, and the same goes for $\mathbf{D}^{=}$. Note that $\mathbf{d} \in \mathbf{D}^{\leqslant}$ means the existence of some $\mathbf{p} \in \mathcal{A}$ such that $\mathbf{d} \cdot \mathbf{p} < 0$, while, for each $\mathbf{d} \in \mathbf{D}^{=}$, it holds that $\mathbf{d} \cdot \mathbf{p} = 0$ for all $\mathbf{p} \in \mathcal{A}$.

Kitamura and Stoye (2018) shows that the tightening of a convex cone in the $\mathcal{V}$-representation is inherited by the $\mathcal{H}$-representation in the following sense.

LEMMA A.1. *Let $\mathcal{A}$ be a convex cone represented as (a.15) and (a.16) using the matrices $\mathbf{B}$ and $\mathbf{D}$.*

For $\kappa > 0$, define $\mathcal{A}_\kappa$ such that

$$\mathcal{A}_\kappa = \left\{ \mathbf{B}\tau : \tau - \left(\frac{\kappa}{n}\right) \mathbf{I}_n \geqslant \mathbf{0} \right\}, \tag{a.17}$$

where $\mathbf{I}_n$ is the n-dimensional vector of 1's. Then, $\mathcal{A}_\kappa$ can be represented as

$$\mathcal{A}_\kappa = \left\{ \mathbf{p} \in \mathbb{R}^m : \mathbf{D}^{\leqslant}\mathbf{p} \leqslant -\kappa\mathbf{u} \text{ and } \mathbf{D}^{=}\mathbf{p} = \mathbf{0} \right\}, \tag{a.18}$$

with $\mathbf{u}$ being a strictly positive vector (the length of $|\mathbf{D}^{\leqslant}|$).

It is worth noting that the lemma ensures the dual representation (a.18) using the same $\mathbf{D}^{\leqslant}$ and $\mathbf{D}^{=}$ as the $\mathcal{H}$-representation (a.16) of the original convex cone $\mathcal{A}$. We rely on Theorem 4.2 in Kitamura and Stoye (2018) and Theorem 4 in Deb et al. (2022) to ensure the validity of the critical values in Section 4, but a common crucial step in these theorems is in fact establishing the representation along the lines of (a.18).

A5.2. Supplementary notes for Section 4.1

Validity of critical value. Let $\mathbf{B}$ be the matrix of which column vectors correspond to (single-valued) group types obeying RM axiom. Then, $m = |\mathbf{Y} \times \widehat{\mathbf{X}}|$ and $n = |\mathcal{B}|$ (recall that the set of group types obeying RM axiom is denoted by $\mathcal{B}$). Ignoring the summing up condition for now, the essential part of the set $\mathbb{P}^{\mathrm{SC}}$ is $\mathcal{A}$, and similarly, for $\kappa > 0$, the essential part of $\mathbb{P}^{\mathrm{SC}}_\kappa$ is $\mathcal{A}_\kappa$ (see also Theorem 3.1 in Kitamura and Stoye (2018)). To show the validity of the critical value, Kitamura and Stoye (2018) introduced the following assumptions on data generating process. (They are also referred to in Section 4 of the main paper).

Assumption 1. *Let $N_{\mathbf{x}}$ be a number of observations with covariates $\mathbf{x}$, and $N = \sum_{\mathbf{x} \in \widehat{\mathbf{X}}} N_{\mathbf{x}}$. Then, for each $\mathbf{x} \in \widehat{\mathbf{X}}$, $\frac{N_{\mathbf{x}}}{N} \to \rho_{\mathbf{x}}$ as $N \to \infty$, where $\rho_{\mathbf{x}} > 0$.*

Assumption 2. *The empirical distribution is obtained from N repeated cross sections of random samples for each realization of covariates $\mathbf{x}$.*

Kitamura and Stoye (2018) further specify a class of distribution, on which they impose some conditions to guarantee stable behavior of the test statistic. Our argument here follows the one in their paper precisely. By the preceding assumptions, we collect a random sample of action profiles $\mathbf{y}$ for each realization of $\mathbf{x}$. Combining a choice made by a group facing each $\mathbf{x}$ in a sample, we can obtain a single-valued group type $\mathbf{b} \in \{0,1\}^{|\mathbf{Y}\times\hat{\mathbf{X}}|}$ as a random vector (it holds by definition that $\mathrm{E}[\mathbf{b}]$ is equal to the empirical choice frequency). Using the $\mathcal{H}$-representation, for each realization of $\mathbf{b}$, it is obvious that $\mathbf{b} \in \mathcal{B} \iff \mathbf{Db} \leqslant \mathbf{0}$. Rows of $\mathbf{D}$ correspond to the restrictions from RM axiom; some of them are satisfied by definition for any $\mathbf{b}$ representing a single-valued group type (such as the sum-up condition for each $\mathbf{x}$), while others are nontrivial restrictions. Let $\mathcal{K}^R$ be the set of indices of rows corresponding to the latter restrictions in $\mathbf{Db} \leqslant \mathbf{0}$, and $\mathbf{g} =^t (g_1, g_2, ..., g_{|\mathcal{K}^R|}) = \mathbf{Db}$. Kitamura and Stoye (2018) introduced the following condition (see their paper for the detailed argument).

Condition 1. For each $k \in \mathcal{K}^R$, $\mathrm{var}(g_k) > 0$ and $\mathrm{E}\left[|g_k/\sqrt{\mathrm{var}(g_k)}^{2+c_1}|\right] < c_2$ hold, where c_1 and c_2 are positive constants.

With these preparations, Theorem 4.2 in Kitamura and Stoye (2018) can be presented as follows:

THEOREM A.2. *Choose $\kappa_N > 0$ such that $\kappa_N \downarrow 0$ and $\sqrt{N}\kappa_N \uparrow \infty$ as $N \to \infty$. Then, under Assumptions 1 and 2, it holds that*

$$\liminf_{N\to\infty} \inf_{\mathbf{p}\in\overline{\mathbb{P}}\cap\mathcal{A}} Pr(J_N \leqslant \hat{c}_{1-\alpha}) = 1-\alpha, \tag{a.19}$$

where $\overline{\mathbb{P}}$ is the set of all population distributions $\mathbf{p}$ (i.e. $\mathbf{p} \in [0,1]^{|\mathbf{Y}\times\hat{\mathbf{X}}|}$ such that $\sum_{\mathbf{y}\in\mathbf{Y}} p_{\mathbf{y},\mathbf{x}} = 1$ for each $\mathbf{x} \in \hat{\mathbf{X}}$) obeying Condition 1, and $\hat{c}_{1-\alpha}$ is the $1-\alpha$ quantile of J_N with $0 \leqslant \alpha \leqslant \frac{1}{2}$.

Lemma A.1 plays a key role in proving this theorem. However, since we adopt the computation procedure based on column generation, we need to appeal to the version of Theorem A.2 by Smeulders et al. (2021), which obtains essentially the same asymptotic property, but only requires strictly positive weights on some specific subset of column vectors of $\mathbf{B}$. The key for their result is the following amended version of Lemma A.1 (Lemma 1 in Smeulders et al. (2021)).

LEMMA A.2. *Let $\mathcal{A}$ be defined as (a.15) and (a.16), and suppose that $\mathcal{B}' \subset \mathcal{B}$ satisfies the following property: for every $\mathbf{d} \in \mathbf{D}^{\leqslant}$, there exists some $\mathbf{b} \in \mathcal{B}'$ such that $\mathbf{d}\cdot\mathbf{b} < 0$. Then, for each $\kappa > 0$, the* *set*

$$\mathcal{A}'_\kappa = \left\{\mathbf{B}\tau : \tau_{\mathbf{b}} \geqslant \frac{\kappa}{|\mathcal{B}'|} \textit{ for all } \mathbf{b} \in \mathcal{B}' \textit{ and } \tau_{\mathbf{b}} \geqslant 0 \textit{ for all } \mathbf{b} \in \mathcal{B}\backslash\mathcal{B}'\right\} \tag{a.20}$$

can be represented as

$$\mathcal{A}'_\kappa = \left\{\mathbf{p} \in \mathbb{R}^m : \mathbf{D}^{\leqslant}\mathbf{p} \leqslant -\kappa\mathbf{u}' \textit{ and } \mathbf{D}^{=}\mathbf{p} = \mathbf{0}\right\}, \tag{a.21}$$

with $\mathbf{u}'$ *being a strictly positive vector.*

It is pointed out by Smeulders et al. (2021) that this lemma allows us to obtain a similar result to Theorem A.2, with the original proof by Kitamura and Stoye (2018) being valid as it stands (except for using Lemma A.2 instead of Lemma A.1). Then, in turn, as long as Assumptions 1 and 2 are maintained, $\kappa_N > 0$ is chosen such that $\kappa_N \downarrow 0$ and $\sqrt{N}\kappa_N \uparrow \infty$ as $N \to \infty$, and $\mathcal{B}'$ obeys the requirement in Lemma A.2, the fraction $p = \#\{J_N^{(r)} > J_N\}/R$ obtained by our procedure in Section 4.1 works as the valid critical value.

Construction of the set $\mathcal{B}'$. In the main paper, we require that $\mathcal{B}'$ contains a basis of the space spanned by $\mathcal{B}$, since it works as a sufficient condition for the requirement in Lemma A.2.

LEMMA A.3. *Suppose that $\mathcal{B}'$ contains a basis of the space spanned by $\mathcal{B}$, and let $\mathbf{D}^{\leqslant}$ be the same as Lemma A.2. Then, for each $\mathbf{d} \in \mathbf{D}^{\leqslant}$, there exists some $\mathbf{b} \in \mathcal{B}'$ such that $\mathbf{d} \cdot \mathbf{b} < 0$.*

Proof. By way of contradiction, suppose that there exists some $\mathbf{d} \in \mathbf{D}^{\leqslant}$ such that $\mathbf{d} \cdot \mathbf{b} = 0$ for all $\mathbf{b} \in \mathcal{B}'$. Since $\mathcal{B}'$ contains all linear basis of $\mathcal{B}$, this implies that $\mathbf{d} \cdot \mathbf{p} = 0$ for all $\mathbf{p} \in \mathcal{A}'_\kappa$. However, this means that $\mathbf{d} \in \mathbf{D}^{=}$, and since $\mathbf{D}^{\leqslant} \cap \mathbf{D}^{=} = \varnothing$, this is a contradiction. **QED**

A possible procedure for obtaining $\mathcal{B}'$ is as follows. First, recall that by Proposition 2 in the main paper, a group type is in $\mathcal{B}$ if and only if it solves the integer programming problem $C\mathbf{b} \leqslant \theta$. Let $\mathcal{B}''$ be some linearly independent set of group types in $\mathcal{B}$ (for example, taking any singleton set as $\mathcal{B}''$, it is linearly independent). We could check the existence of group types in $\mathcal{B}$ which are linearly independent of the ones in $\mathcal{B}''$ by checking if there is $\mathbf{b} \in \mathcal{B}$ (equivalently, that solve $C\mathbf{b} \leqslant \theta$) and a real-valued vector $\mathbf{w}$ such that $\mathbf{B}'' \cdot (\mathbf{b} - \mathbf{B}''\mathbf{w}) = 0$ and $\mathbf{b} \neq \mathbf{B}''\mathbf{w}$, where $\mathbf{B}''$ refers to the matrix made out of the vectors in $\mathcal{B}''$. If such a group type $\mathbf{b}$ can be found, then we add it to $\mathcal{B}''$

and repeat the procedure. This process will stop when there are no vectors in $\mathcal{B}$ which are linearly independent of the ones in $\mathcal{B}''$, at which point we obtain a basis for $\mathcal{B}$ (and hence, we adopt the resulting $\mathcal{B}''$ as $\mathcal{B}'$). Notice that while it could be practically hard to completely list the elements of $\mathcal{B}$, listing $\mathcal{B}'$ is less demanding, since the dimension of this space grows a lot more slowly than the number of actions and covariate values.[3]

A5.3. Supplementary notes for Section 4.2

The main issue here is the validity of the critical value for each $\beta \in (0,1)$ used in Section 4.2. We can largely appeal to Theorem 4 in Deb et al. (2022) but, as in the preceding subsection, a small twist is needed because of the use of column generation.

Recall that for a given $\beta \in (0,1)$ and $\mathcal{B}^* \subset \mathcal{B}$, the null hypothesis is that the data set $\mathbf{q}$ is a sample from some element of the set

$$\mathbb{P}^{SC}(\beta;\mathcal{B}^*) = \left\{ \mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}} \text{ and } \sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta \right\}, \tag{a.22}$$

and we need to construct a suitable tightening for the domain of τ, which is denoted by $\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*)$ in the main paper. Maintaining Assumptions 1 and 2, Theorem 4 in Deb et al. (2022) still holds by obtaining J_N from the tightening $\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*)$ such that

$$\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*) = \left\{ \tau \in \Delta^{\mathcal{B}} : \tau^{\mathbf{b}} \geqslant \frac{(1-\beta)\kappa_N}{|\mathcal{B}' \cap \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \cap \mathcal{B}^* \text{ and } \tau^{\mathbf{b}} \geqslant \frac{\beta\kappa_N}{|\mathcal{B}' \backslash \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \backslash \mathcal{B}^* \right\}, \tag{a.23}$$

where κ_N and $\mathcal{B}' \subset \mathcal{B}$ respectively obey the properties referred to in the preceding subsection.

THEOREM A.3. *Choose κ_N in the same way as Theorem A.2. Then, under Assumptions 1 and 2, it holds that*

$$\liminf_{N\to\infty} \inf_{(\beta,\mathbf{p})\in\mathcal{F}} Pr(J_N(\beta) \leqslant \hat{c}_{1-\alpha}) = 1-\alpha, \tag{a.24}$$

[3] It is straightforward to check that the dimension of the space spanned by the set of all logically possible group types is precisely $|\mathbf{Y} \times \hat{\mathbf{X}}| - |\mathbf{Y}| + 1$ and so obviously the dimension of the space spanned by $\mathcal{B}$ can be no higher. In fact, the span of $\mathcal{B}$ coincides with that of the set of all logically possible group types, even though $\mathcal{B}$ is a proper subset.

where $\mathcal{F} = \left\{(\beta, \mathbf{p}) : \beta \in (0,1), \mathbf{p} \in \overline{\mathbb{P}} \cap \mathbb{P}^{SC}\right\}$ *and* $\hat{c}_{1-\alpha}$ *is the* $1-\alpha$ *quantile of* $J_N(\beta)$ *with* $0 \leqslant \alpha \leqslant \frac{1}{2}$ *(*$\overline{\mathbb{P}}$ *is defined as Theorem A.2).*

Given this theorem, other than using $\Delta^{\mathcal{B}}_{\kappa_N}(\beta; \mathcal{B}^*)$ instead of $\Delta^{\mathcal{B}}_{\kappa_N}$, the critical value for each $\beta \in (0,1)$ is obtained by exactly the same bootstrap procedure as the one outlined in Section 4.1. The only departure from the tightening in Deb et al. (2022) is that we require strictly positive weights for elements in $\mathcal{B}' \subset \mathcal{B}$, rather than all of $\mathcal{B}$. This difference is material in the part of the proof of Deb et al. (2022) that depends on Lemma A.1, since we do not have the entire set $\mathcal{B}$ due to the fact that we use of column generation. Nevertheless, proceeding as in the previous subsection, by choosing $\mathcal{B}'$ obeying the requirement in Lemma A.2, we can appeal to that lemma instead of Lemma A.1 (see also Lemma A.3). Apart from this variation, the proof proceeds in the same way as the analogous one in Deb et al. (2022).

Remark 1. As seen from (a.23), we need to choose $\mathcal{B}'$ so that both $\mathcal{B}' \cap \mathcal{B}^* \neq \varnothing$ and $\mathcal{B}' \backslash \mathcal{B}^* \neq \varnothing$. If $\mathcal{B}'$ (constructed as in the preceding subsection) does not satisfy these conditions, we need to (manually) add any group type contained in these intersections. Nevertheless, this aspect would not matter in practice, especially if the statistical test has been done in advance. In that case, one can expect to have a reasonably 'rich' subset of $\mathcal{B}$ as a result of the column generation procedure constructed for the test, which can be used as a starter set $\mathcal{B}'$ for the estimation of β.

Remark 2. Note that Deb et al. (2022) define $\mathbb{P}^{SC}(\beta; \mathcal{B}^*)$ by using the equality constraint $\sum_{\mathbf{b} \in \mathcal{B}^*} \tau^{\mathbf{b}} = \beta$, rather than $\sum_{\mathbf{b} \in \mathcal{B}} \tau^{\mathbf{b}} \geqslant \beta$. We adopt the inequality condition since it facilitates the implementation of the column generation procedure. In particular, our procedure depends on Proposition 3 in the main paper, which in turn depends on a set of inequalities. (At this point, we do not have the counterpart of this proposition for the equality constraint.) Using inequalities should not cause any substantial change in the proof of Theorem A.3 beside the fact that some equalities in the proof need to be converted to inequalities in a rather obvious way (see Online Appendix of Deb et al. (2022)).

A6. Additional empirical analysis

A6.1. Finer discretization of covariates

In the main paper, we initially discretize each covariate into two values, following Kline and Tamer (2016). Specifically, the market presence variables (MP_{LCC} and MP_{OA}) and the market size variable MS take value 1 if their actual values are above median amongst observed data. In order to show that our test can actually deal with a larger model, we also implemented our test with all these covariates being discretized into four values using quartile points. As we pointed out in the main paper, the data set passes the $\mathcal{SC}$-rationalizabiiity test with the binary discretization, but it fails with a quartile discretization.

To see the effect of finer discretization one by one, we consider the cases in which (i) MP_{LCC} and MP_{OA} are split into four values, while MS is kept binary, and (ii) MS is split into four values, while MP_{LCC} and MP_{OA} are kept binary. We obtain that the former is rejected with p-value 0, while the latter is supported with p-value 0.412. We also implement our test with other types of discretizations. The results are summarized in Table A.3.[4] These results imply that our behavioral hypothesis is vulnerable to finer discretizations of the market presence variables, while it is (to some extent) robust to finer discretizations of the market size variable.

When covariates are more finely discretized, the number of observations in each bin becomes small, which makes the estimation noisier. Given this, we use choice distributions for $\mathbf{x}$'s with more than or equal to 50 observations.[5] In other words, for each pattern of discretization, we let $\widehat{\mathbf{X}} = \{\mathbf{x} \in \mathbf{X} : N_{\mathbf{x}} \geqslant 50\}$, where $N_{\mathbf{x}}$ is the number of observations with covariates $\mathbf{x}$. The size of $\widehat{\mathbf{X}}$ in each case is also reported in Table A.3. The results here also imply that our test still have good testing power even with relatively small size of $\widehat{\mathbf{X}}$. For example, for the case of $6 \times 6 \times 6$, we only use observations for 46 types of realization of covariates out of 216 possible realization of covariates, but the hypothesis of $\mathcal{SC}$-rationalizablity is refuted with p-value being equal to 0.

Lastly, to compare the results of our test with that of Kline and Tamer (2016), we also implement

[4] As seen from the table, not all results are listed and the results for other combinations are available from the authors. Also, one may implement them using our R program.

[5] The conclusions of the tests (pass/fail with 5% significance level) in Table A.3 remains the same even if we let $\widehat{\mathbf{X}} = \{\mathbf{x} \in \mathbf{X} : N_{\mathbf{x}} \geqslant 100\}$, except for the cases with $6 \times 6 \times 4$ and $6 \times 6 \times 6$. In those cases, the test fails to reject the data, possibly because $\widehat{\mathbf{X}}$ becomes too small relative to the original size of $\mathbf{X}$, which weaken the testing power. For example, in $6 \times 6 \times 4$, $|\widehat{\mathbf{X}}| = 12$, when $|\mathbf{X}| = 144$.

$\mathrm{MP}_{LCC} \times \mathrm{MP}_{OA} \times \mathrm{MS}$	$\lvert\widehat{\mathbf{X}}\rvert$	p-value
$2 \times 2 \times 2$	8	0.138
$2 \times 2 \times 3$	12	0.195
$2 \times 2 \times 4$	16	0.412
$2 \times 2 \times 6$	24	0.195
$2 \times 2 \times 8$	32	0.020
$3 \times 3 \times 2$	18	0.000

$\mathrm{MP}_{LCC} \times \mathrm{MP}_{OA} \times \mathrm{MS}$	$\lvert\widehat{\mathbf{X}}\rvert$	p-value
$3 \times 3 \times 3$	27	0.017
$4 \times 4 \times 2$	32	0.000
$4 \times 4 \times 4$	60	0.000
$6 \times 6 \times 2$	65	0.000
$6 \times 6 \times 4$	70	0.000
$6 \times 6 \times 6$	46	0.000

Table A.3: Tests under various discretization

their parametric estimation for some patterns of discretization in Table A.3 (see Section A1 for more details on Kline and Tamer's estimation). The results are broadly consistent. We find that with tolerance level 0.075, the probability of obtaining nonempty set of estimated coefficients is (i) 1 for $2 \times 2 \times 2$, (ii) 0.928 for $2 \times 2 \times 3$, and (iii) 0 for $3 \times 3 \times 3$. In particular, note that the model is rejected in case (iii).

A6.2. Lower probability bounds for equilibrium actions

In Section 5 of the main paper, as an application of counterfactual analysis discussed in Section 3.4, we estimated the maximum probability of a given action profile $\overline{\mathbf{y}}$ being an equilibrium action at the covariate $\overline{\mathbf{x}} \in \widehat{\mathbf{X}}$. We motivated this exercise by considering a policy maker who could influence the equilibrium selection mechanism but not player payoffs and asked the extent to which she could shift the action profile towards the outcome $\overline{\mathbf{y}}$. In formal terms, we considered the weight on the set

$$\mathcal{B}^* = \{\mathrm{B} \in \mathcal{B} : \text{there is } \mathbf{\Pi} \in \mathcal{SC} \text{ that rationalizes B such that } \overline{\mathbf{y}} \in \mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})\}$$

(see Section 3.4, Application 2). $\mathcal{B}^*$ is obviously a superset of (and possibly a *strict* superset of)

$$\mathcal{B}^0 = \{\mathrm{B} \in \mathcal{B} : \mathrm{B}(\overline{\mathbf{x}}) = \overline{\mathbf{y}}\}$$

(the set of types that actually play $\overline{\mathbf{y}}$ at $\overline{\mathbf{x}}$). What we refer to as $\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$ in Section 5 (see also Table A.4 below) is the upper limit of the confidence interval on

$$\sum_{\mathrm{B} \in \mathcal{B}^*} \tau^{\mathrm{B}}$$

subject to $(\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$ solving the model (which we can estimate by the procedure set out in Section 4.2). This represents the most *optimistic* estimate of what the policy maker can do, not only because it assumes the greatest possible weight on $\mathcal{B}^*$ but also because it assumes (given the definition of $\mathcal{B}^*$) that every group type which *can* have payoff functions for which $\overline{\mathbf{y}}$ is an equilibrium, actually does have such payoff functions.

It is also interesting to investigate the most *conservative* assessment of what the policy maker can do, assuming that equilibrium selection rules are freely manipulable. This is given by the weight on the set

$$\mathcal{B}^{\dagger} = \{\mathrm{B} \in \mathcal{B} : \overline{\mathbf{y}} \in \mathrm{NE}(\overline{\mathbf{x}}, \mathbf{\Pi}) \text{ for any } \mathbf{\Pi} \in \mathcal{SC} \text{ that rationalizes B}\}\,. \tag{a.25}$$

This is the set of types for which $\overline{\mathbf{y}}$ *must* be an equilibrium at the covariate $\overline{\mathbf{x}}$. Obviously,

$$\mathcal{B}^0 \subseteq \mathcal{B}^{\dagger} \subseteq \mathcal{B}^*.$$

Below, we explain how we may characterize $\mathcal{B}^{\dagger}$ in the case of the application in Section 5. This allows us to calculate the lower limit of the confidence interval on

$$\sum_{\mathrm{B}\in\mathcal{B}^{\dagger}} \tau^{\mathrm{B}},$$

subject to $(\tau^{\mathrm{B}})_{\mathrm{B}\in\mathcal{B}}$ solving the model, which we denote by $\min \Pr[\overline{\mathbf{y}} \in \mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$ (see also Table A.4 below); it is the most conservative estimate of far the policy maker can shift the equilibrium towards $\overline{\mathbf{y}}$ without manipulating payoffs.

CLAIM A.1. *Let* $\overline{\mathbf{x}} \in \widehat{\mathbf{X}}$. *When* $\overline{\mathbf{y}} = (N, E)$, *the group type* B *is in* $\mathcal{B}^{\dagger}$ *if and only if* $\mathrm{B} \in \mathcal{B}$ *and either*
(i) $\mathrm{B}(\overline{\mathbf{x}}) = (N, E)$ *or*
(ii) $\mathrm{B}(\overline{\mathbf{x}}) = (E, N)$ *and there exists some* $\mathbf{x}' = (x_1', x_2')$ *and* $\mathbf{x}'' = (x_1'', x_2'')$ *with* $x_1' \geqslant \overline{x}_1$ *and* $x_2'' \leqslant \overline{x}_2$ *for which* $\mathrm{B}(\mathbf{x}') = (N, E)$ *and* $\mathrm{B}(\mathbf{x}'') = (N, E)$.[6]

Similarly, when $\overline{\mathbf{y}} = (E, N)$, *the group type* B *is in* $\mathcal{B}^{\dagger}$ *if and only if* $\mathrm{B} \in \mathcal{B}$ *and either*

[6] In fact, the crucial part of condition (ii) is that player 1 chooses N when $x_1' \geqslant \bar{x}_1$ while player 2 chooses E when $x_2'' \leqslant \bar{x}_2$. However, $\mathrm{B}(\mathbf{x}') = (N, N)$ and $\mathrm{B}(\mathbf{x}'') = (E, E)$ are excluded by the RM axiom, and hence it suffices to consider the situation described in the statement.

(i') $\mathrm{B}(\overline{\mathbf{x}}) = (E, N)$ *or*

(ii') $\mathrm{B}(\overline{\mathbf{x}}) = (N, E)$ *and there exists some* $\mathbf{x}' = (x_1', x_2')$ *and* $\mathbf{x}'' = (x_1'', x_2'')$ *with* $x_1' \leqslant \overline{x}_1$ *and* $x_2'' \geqslant \overline{x}_2$ *for which* $\mathrm{B}(\mathbf{x}') = (E, N)$ *and* $\mathrm{B}(\mathbf{x}'') = (E, N)$.

Proof. We only prove the case of $\overline{\mathbf{y}} = (N, E)$, since the other case can be shown in a similar vein. It is clear that if B obeys (i), then any $\mathbf{\Pi} \in \mathcal{SC}$ rationalizing it must support (N, E) as an equilibrium action at $\overline{\mathbf{x}}$. Suppose that B obeys (ii). Then, $\mathrm{B}(\mathbf{x}') = (N, E)$ implies that $N \in \mathrm{BR}_1(E, x_1')$, and the monotonicity of best response implies that $N \in \mathrm{BR}_1(E, \overline{x}_1)$ holds. Also, $\mathrm{B}(\mathbf{x}') = (N, E)$ and the monotonicity implies that $E \in \mathrm{BR}_2(N, \overline{x}_2)$, and hence, for any $\mathbf{\Pi} \in \mathcal{SC}$ rationalizing B, (N, E) must be supported as an equilibrium action at $\overline{\mathbf{x}} = (\overline{x}_1, \overline{x}_2)$.

Conversely, if neither (i) nor (ii) holds, $\mathrm{B} \in \mathcal{B}$ can be rationalized by some $\mathbf{\Pi} \in \mathcal{SC}$ for which (N, E) is not an equilibrium action at $\overline{\mathbf{x}}$. It is trivial that $\mathrm{B}(\overline{\mathbf{x}}) = (N, N)$ or (E, E), then (N, E) cannot be supported as an equilibrium action, since (N, N) and (E, E) cannot be a part of multiple equilibria in our setting. When $\mathrm{B}(\overline{\mathbf{x}}) = (E, N)$ and player 1 never plays N for any $x_1' \geqslant \overline{x}_1$, one can find his (single-crossing) payoff function so that $\Pi_1(E, y_2, \overline{x}_1) > \Pi_1(N, y_2, \overline{x}_1)$ for $y_2 \in \{E, N\}$; indeed, our construction of the payoff function in the proof of Theorem 1 would satisfy this property. Similarly, when $\mathrm{B}(\overline{\mathbf{x}}) = (E, N)$ and player 2 never plays E for any $x_2' \leqslant \overline{x}_2$, one can find his (single-crossing) payoff function so that $\Pi_2(N, y_1, \overline{x}_2) > \Pi_1(E, y_1, \overline{x}_2)$ for $y_1 \in \{N, E\}$. By doing so, a profile of payoff function $\mathbf{\Pi} = (\Pi_1, \Pi_2)$ does not support (N, E) as an equilibrium action at $\overline{\mathbf{x}}$. **QED**

For each $\overline{\mathbf{x}} \in \widehat{\mathbf{X}}$ and $\overline{\mathbf{y}} \in \{(N, E), (E, N)\}$, we would like to estimate the minimum possible fraction of group types in $\mathcal{B}^\dagger$. In order to apply the procedure in Section 4.2 of the main paper, we estimate the *maximum* possible fraction on $\mathcal{B}^{\dagger\dagger} := \mathcal{B} \backslash \mathcal{B}^\dagger$, where $\mathcal{B}^{\dagger\dagger}$ corresponds to the set of types for which $\overline{\mathbf{y}}$ *may not* be an equilibrium action at $\overline{\mathbf{x}}$, and subtract it from 1. Similar to the case of maximum probability bounds (dealt with in Section A4), we employ the matrix characterization of group types in $\mathcal{B}^{\dagger\dagger}$. Specifically, we shall construct $C^{\dagger\dagger}$ and $\theta^{\dagger\dagger}$ such that a group type $\mathbf{b} \in \mathcal{B}^{\dagger\dagger}$ if and only if $C^{\dagger\dagger}\mathbf{b} \leqslant \theta^{\dagger\dagger}$.

Construction of $C^{\dagger\dagger}$ and $\theta^{\dagger\dagger}$. For each $\overline{\mathbf{y}} \in \{(N, E), (E, N)\}$ and $\overline{\mathbf{x}} \in \widehat{\mathbf{X}}$, define the matrix $C^{\dagger\dagger}$ as follows. Let us first note that the size of this matrix is equal to $\Big((|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1) + K(\overline{\mathbf{y}}, \overline{\mathbf{x}})\Big) \times |\mathbf{Y} \times \widehat{\mathbf{X}}|$,

$(\mathrm{MP}_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS})$	$(0,0,0)$		$(0,1,0)$		$(1,0,0)$		$(1,1,0)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
$\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$	0.699	0.544	0.815	0.503	0.503	0.644	0.558	0.555
$\min \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$	0.658	0.115	0.784	0.154	0.385	0.234	0.512	0.121
Observed Prob.	0.682	0.006	0.785	0.003	0.367	0.253	0.542	0.050
$(\mathrm{MP}_{LCC}, \mathrm{MP}_{OA}, \mathrm{MS})$	$(0,0,1)$		$(0,1,1)$		$(1,0,1)$		$(1,1,1)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
$\max \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$	0.841	0.616	0.913	0.496	0.485	0.661	0.523	0.497
$\min \Pr[\overline{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$	0.806	0.220	0.881	0.237	0.385	0.278	0.484	0.106
Observed Prob.	0.832	0.001	0.910	0.000	0.326	0.306	0.501	0.021

Table A.4: Probability bounds for equilibrium action profiles

where

$$K(\overline{\mathbf{y}}, \overline{\mathbf{x}}) = \#\{\mathbf{x}' \in \widehat{\mathbf{X}} : x_1' \geqslant \overline{x}_i\} \times \#\{\mathbf{x}'' \in \widehat{\mathbf{X}} : x_2'' \leqslant \overline{x}_2\}, \text{ if } \mathbf{y} = (N, E), \tag{a.26}$$

$$= \#\{\mathbf{x}' \in \widehat{\mathbf{X}} : x_1' \leqslant \overline{x}_1\} \times \#\{\mathbf{x}'' \in \widehat{\mathbf{X}} : x_2'' \geqslant \overline{x}_2\}, \text{ if } \mathbf{y} = (E, N). \tag{a.27}$$

We set the first $|\mathbf{Y} \times \widehat{\mathbf{X}}| \times |\mathbf{Y} \times \widehat{\mathbf{X}}|$-submatrix as the matrix C (for characterizing RM group types) constructed in the proof of Proposition 2. The $(|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1)$-th row is for checking condition (i) in Claim A.1, and the remaining $K(\overline{\mathbf{y}}, \overline{\mathbf{x}})$ rows are for checking condition (ii) in Claim A.1. When $\overline{\mathbf{y}} = (N, E)$, this part of the matrix is constructed as follows (the case of $\overline{\mathbf{y}} = (E, N)$ is similar):

1. In the $(|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1)$-th row, the coordinate corresponding to $((N, E), \overline{\mathbf{x}}))$ takes value 1, and others are set to 0.

2. Each of remaining $K(\overline{\mathbf{y}}, \overline{\mathbf{x}})$ rows should be related to each combination $(\mathbf{x}', \mathbf{x}'')$ for which $x_1' \geqslant \overline{x}_1$ and $x_2'' \leqslant \overline{x}_2$. Regarding each row as such, the coordinate corresponding to $((E, N), \overline{\mathbf{x}})$ is set to 1, the coordinates corresponding to $((N, E), \mathbf{x}')$ and $((N, E), \mathbf{x}'')$ are set to 0.1, and others are 0.

Finally, the vector for the RHS, $\theta^{\dagger\dagger}$, is defined as the $\Big((|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1) + K(\overline{\mathbf{y}}, \overline{\mathbf{x}})\Big)$ dimensional column vector, of which the first $|\mathbf{Y} \times \widehat{\mathbf{X}}|$-dimensional subvector is equal to the vector θ constructed in the proof of Proposition 2, the $(|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1)$-th element is 0, and all other elements are set to 1.1.

Estimates of $\min \mathbf{Pr}[\overline{\mathbf{y}} \in \mathrm{NE}(\mathbf{\Pi}, \overline{\mathbf{x}})]$**.** Given the matrix representation we can estimate the greatest possible weight on $\mathcal{B}^{\dagger\dagger}$ using the procedure in Section 4.2 (and Appendix A5.3), and thus the lowest possible probability for $\overline{\mathbf{y}}$ to be an equilibrium action at $\overline{\mathbf{x}}$; formally, we obtain the lower limit of the confidence interval on $\sum_{\mathrm{B}\in\mathcal{B}^\dagger}\tau^{\mathrm{B}}$. This is denoted by $\min\Pr[\overline{\mathbf{y}}\in\mathbf{NE}(\mathbf{\Pi},\overline{\mathbf{x}})]$ in Table A.4. Notice that there is not much difference between $\min\Pr[(N,E)\in\mathbf{NE}(\mathbf{\Pi},\overline{\mathbf{x}})]$ and the empirical frequency of (N,E).[7] On the other hand, $\min\Pr[(E,N)\in\mathbf{NE}(\mathbf{\Pi},\mathbf{x})]$ is appreciably larger than the empirical frequency of (E,N) at some covariate values, for example, $\overline{\mathbf{x}}=(0,0,1)$.

References

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2022): Revealed price preference: theory and empirical analysis. *arXiv:1801.02702*.

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

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

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

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

[7] Even though $\mathcal{B}^0\subseteq\mathcal{B}^\dagger$ it is possible for $\min\Pr[\overline{\mathbf{y}}\in\mathbf{NE}(\mathbf{\Pi},\overline{\mathbf{x}})]$ to be lower than the empirical frequency of $\overline{\mathbf{y}}$ due to sampling variation, and this is observed in some entries in the table.