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

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

September 25, 2024

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 distributions is explained as Nash equilibrium play under the payoff functions specified by (a.1). Then, it must hold that $\beta_2 = 0$,

* Email addresses: nlazzati@ucsc.edu; ecsqkhj@nus.edu.sg; kshirai1985@kwansei.ac.jp

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

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

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

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 Statistical tests of nonparametric and linear models

We have 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. Therefore, we also implement the Kline-Tamer test to the same set of samples to check if each of them is supported by the linear 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. Thus both the Kline-Tamer procedure and our $\mathcal{SC}$-rationalizability test are working as they should: the former rejecting the linear model and the latter not rejecting the nonparametric model.

Power of the nonparametric test. Note that the preceding result suggests that Kline and Tamer ' s approach has very good power to detect model misspecification, since their results find empty sets for the parameters in all the samples generated by $\mathcal{P}$ (as it should, since $\mathcal{P}$ is inconsistent

$(\mathrm{MP}_1, \mathrm{MP}_2, \mathrm{MS}) = (0,0,0)$			
$\mathrm{P}(N,N)$	$\mathrm{P}(N,E)$	$\mathrm{P}(E,N)$	$\mathrm{P}(E,E)$
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 $\mathcal{P}'$

with the linear model). It would be interesting to see if 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. $\mathcal{P}'$ is not too different from $\mathcal{P}$ in Table A.1 and, in fact, it is equal to the empirical distribution of one of the samples from $\mathcal{P}$, with a p-value 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 has strong testing power.

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 set (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 us write for each $i \in \mathcal{N}$, $Y_i = \times_{k=1}^{K(i)} Y_{ik}$ where 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 combi-

[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 $F(a' \wedge a'') - F(a') > (\geqslant)\, 0$.

nation of quasisupermodularity and condition (4) in the main paper 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 $(y''_{i(-J)}, \mathbf{y}''_{-i}, x''_i) > (y'_{i(-J)}, \mathbf{y}'_{-i}, x'_i)$,

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

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 indeed the case:

The next results states that Theorem 1 can be also extended to the case of 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 $y^t_{ik} = \mathrm{argmax}_{y_{ik} \in Y_{ik}} \Pi_{ik}(y_{ik}, \mathbf{z}^t_i)$ 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 $\mathcal{P} = \{\mathrm{P}(\cdot \mid \mathbf{x})\}_{\mathbf{x}\in\widehat{\mathbf{X}}}$ 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

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

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 $(\mathbf{y}, \mathbf{x}) \in \mathbf{Y} \times \widehat{\mathbf{X}}$ 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* $\mathcal{P} = \{\mathrm{P}(\cdot|\mathbf{x})\}_{\mathbf{x}\in\widehat{\mathbf{X}}}$ *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*

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

- *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*

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

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 $\mathbf{x} \in \widehat{\mathbf{X}}$ 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 $\mathbf{x} \in \widehat{\mathbf{X}}$ 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 $\mathbf{x} \in \widehat{\mathbf{X}}$ 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 $\mathbf{x} \in \widehat{\mathbf{X}}$ 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 $\mathbf{x} \in \widehat{\mathbf{X}}$ 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 $\mathbf{x} \in \widehat{\mathbf{X}}$ 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 $\bar{\mathbf{y}}$ and covariate $\bar{\mathbf{x}}$, what is the greatest possible fraction of groups which have $\bar{\mathbf{y}}$ as a pure strategy Nash equilibrium at $\bar{\mathbf{x}}$, 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 $\bar{\mathbf{x}}$? In formal terms, this amounts to identifying the *set* of conditional distributions $\mathrm{P}(\cdot|\bar{\mathbf{x}})$ such that the augmented set of distributions $\mathcal{P} \cup \{\mathrm{P}(\cdot|\bar{\mathbf{x}})\}$ is still $\mathcal{SC}$-rationalizable.

Let $\mathrm{B} : \{\bar{\mathbf{x}}\} \cup \widehat{\mathbf{X}} \to \mathbf{Y}$ be a group type defined on the enlarged domain $\{\bar{\mathbf{x}}\} \cup \widehat{\mathbf{X}}$. Let $\widetilde{\mathcal{B}}$ 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 $\mathcal{P} \cup \{\mathrm{P}(\cdot|\bar{\mathbf{x}})\}$ is $\mathcal{SC}$-rationalizable if and only if we can find a probability distribution $\tilde{\tau} = (\tilde{\tau}^{\mathrm{B}})_{\mathrm{B}\in\widetilde{\mathcal{B}}}$ over $\widetilde{\mathcal{B}}$ such that

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

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

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 $\mathrm{P}(\cdot|\bar{\mathbf{x}})$ obtained from (a.14) is a predicted distribution at $\bar{\mathbf{x}}$. In other words, if we let $\mathbb{P}(\bar{\mathbf{x}})$ be the *set* of predicted distributions at $\bar{\mathbf{x}}$, then $\mathrm{P}(\cdot|\bar{\mathbf{x}})$ is in $\mathbb{P}(\bar{\mathbf{x}})$ if and only if there is $\tilde{\tau}$ that solves (a.13) and (a.14). Since the conditions are linear, $\mathbb{P}(\bar{\mathbf{x}})$ 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 $\bar{\lambda}(\cdot|\mathbf{\Pi}, \mathbf{x})$ (for $\mathbf{x} \in \{\overline{\mathbf{x}}\} \cup \mathbf{X}$) that rationalizes $\mathcal{P}$ and satisfies $\mathrm{P}(\mathbf{y}|\overline{\mathbf{x}}) = \int \bar{\lambda}(\mathbf{y}|\mathbf{\Pi}, \overline{\mathbf{x}})\, d\mathrm{P}_{\mathbf{\Pi}}$ for all $\mathbf{y} \in \mathbf{Y}$. Let $\hat{\lambda}$ be a new equilibrium selection rule where $\hat{\lambda}(\cdot|\mathbf{\Pi}, \mathbf{x}^t) = \bar{\lambda}(\cdot|\mathbf{\Pi}, \mathbf{x})$ 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 $\mathrm{P}(\mathbf{y}|\widehat{\mathbf{x}}) = \int \hat{\lambda}(\mathbf{y}|\mathbf{\Pi}, \widehat{\mathbf{x}})\, d\mathrm{P}_{\mathbf{\Pi}}$ 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**

[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.

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.

Significance of strategic interaction. 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

$(\bar{\mathbf{y}}, \bar{\mathbf{x}})$ in terms of the RM axiom on the extended domain.

A5. Omitted details from statistical tests

A5.1. Key lemmata

The bootstrap procedures with tightening described in Kitamura and Stoye (2018) (and its variations in Smeulders et al. (2021), Deb et al. (2023), and ours) largely depend on representation result in Kitamura and Stoye (2018) (Lemma 4.1 in their paper). It is worth restating this result as well as a modified version by Smeulders et al. (2021) here, given its relevance to what we are doing.

Let $\mathbf{B}$ be an $m \times n$ matrix and denote its set of column vectors by $\mathcal{B}$. The convex cone generated by $\mathbf{B}$ is the set

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

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

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

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 $\mathbf{d} \in \mathbf{D}^{\leqslant}$ when $\mathbf{d}$ is a row vector of $\mathbf{D}^{\leqslant}$, and the same goes for $\mathbf{D}^{=}$.

Kitamura and Stoye (2018) (in their Lemma 4.1) show 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.

This 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 $\mathcal{A}$.

Since we adopt the computation procedure based on column generation, we need a version of Lemma A.1 where strictly positive weights are required only for a specific subset of $\mathcal{B}$. The following result is found 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*

$$\mathcal{A}'_\kappa = \left\{ \mathbf{B}\tau : \tau^{\mathbf{b}} \geqslant \frac{\kappa}{|\mathcal{B}'|} \text{ for all } \mathbf{b} \in \mathcal{B}' \text{ and } \tau^{\mathbf{b}} \geqslant 0 \text{ for all } \mathbf{b} \in \mathcal{B} \backslash \mathcal{B}' \right\} \tag{a.19}$$

can be represented as

$$\mathcal{A}'_\kappa = \left\{ \mathbf{p} \in \mathbb{R}^m : \mathbf{D}^{\leqslant}\mathbf{p} \leqslant -\kappa \mathbf{u}' \text{ and } \mathbf{D}^{=}\mathbf{p} = \mathbf{0} \right\}, \tag{a.20}$$

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

As in the case of (a.18), the $\mathcal{H}$-representation (a.20) is obtained by using exactly the same matrix $\mathbf{D} =^t [\mathbf{D}^{\leqslant}, \mathbf{D}^{=}]$ as (a.16). In the rest of this section, we rely on this lemma instead of Lemma A.1. It matters how we obtain the set $\mathcal{B}'$ obeying the requirement in the lemma, which could depend on the targeted application. A procedure for constructing a suitable $\mathcal{B}'$ in our game setting is provided in the next subsection.

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. This matrix has $|\mathcal{B}|$ columns of length $|\mathbf{Y} \times \widehat{\mathbf{X}}|$ (recall that the set of group types obeying RM axiom is denoted by $\mathcal{B}$). As we pointed out in Section 4.1, our null hypothesis that $\mathbf{p}$ is in the set of $\mathcal{SC}$-rationalizable distributions defined as

$$\mathbb{P}^{\mathrm{SC}} = \left\{ \mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}} \right\} \tag{a.21}$$

is equivalent to $\mathbf{p}$ being in the cone $\mathcal{A} = \{\mathbf{B}\tau : \tau \geqslant \mathbf{0}\}$. To show the validity of the critical value, Kitamura and Stoye (2018) introduced the following assumptions on data generating process (which we first introduced 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 impose another condition to guarantee the stable behavior of the test statistic. We now explain this condition in the context of our model. Suppose an action profile $\mathbf{y}$ is picked at random for each realization of $\mathbf{x}$ from data. This random vector $\mathbf{b} \in \{0,1\}^{|\mathbf{Y} \times \hat{\mathbf{X}}|}$ can be interpreted as a single-valued group type; clearly, $\mathrm{E}[\mathbf{b}]$ is equal to the empirical choice frequency. By the equivalence between (a.15) and (a.16), for each realization of $\mathbf{b}$, we have $\mathbf{b} \in \mathcal{B} \iff \mathbf{D}\mathbf{b} \leqslant \mathbf{0}$. 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 $\mathbf{D}\mathbf{b} \leqslant \mathbf{0}$, and let $\mathbf{g} =^t (g_1, g_2, ..., g_l) = \mathbf{D}\mathbf{b}$. Kitamura and Stoye (2018) impose the following condition on $\mathbf{g}$.

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 for some positive constants c_1 and c_2.

With these assumptions in place, the validity of our bootstrap procedure is ensured. Indeed, suppose that $\sqrt{N}(\mathbf{q} - \mathbf{p}) \xrightarrow{d} N(0, S)$, and let $\hat{S}$ consistently estimate S. With η^* defined in (14), let $\tilde{\eta}^*$ be given by

$$\tilde{\eta}^* = \eta^* + \frac{1}{\sqrt{N}} N(0, \hat{S})$$

and let

$$\tilde{J}_N := \min_{\eta \in \mathcal{A}'_{\kappa_N}} N(\tilde{\eta}^* - \eta) \cdot (\tilde{\eta}^* - \eta) \tag{a.22}$$

$$= \min_{\substack{\mathbf{D}^{\leqslant}\eta \leqslant -\kappa_N \bar{\mathbf{u}}' \\ \mathbf{D}^{=}\eta = 0}} N(\tilde{\eta}^* - \eta) \cdot (\tilde{\eta}^* - \eta). \tag{a.23}$$

Note that $\mathbf{D}^{\leqslant}$, $\mathbf{D}^{=}$ and $\mathbf{u}'$ are the matrices and vector providing the representation in (a.20) for $\kappa = \kappa_N$, where κ_N is a tuning parameter which obeying $\kappa_N \downarrow 0$ and $\sqrt{N}\kappa_N \uparrow \infty$ as $N \to \infty$. Then we have the following result, which says (in essence) that the distribution of $\tilde{J}_N$ approximates the distribution of the test statistic J_N and ensures the validity of the critical value in Section 4.1 of the main paper.

THEOREM A.2. *Choose $\kappa_N > 0$ such that $\kappa_N \downarrow 0$ and $\sqrt{N}\kappa_N \uparrow \infty$ as $N \to \infty$ and suppose that $\mathcal{B}' \subset \mathcal{B}$ satisfies the requirement in Lemma A.2. Then, under Assumptions 1 and 2, it holds that*

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

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

The result stated here is found in Smeulders et al. (2021) and justifies the bootstrap procedure we use, where strictly positive weights are required only for the elements of $\mathcal{B}'$. This result is a modification of Theorem 4.2 in Kitamura and Stoye (2018); that theorem justifies a bootstrap procedure which requires positive weights on all elements of $\mathcal{B}$.

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]

Improvements on the testing procedure. We explain two simple modifications to the column generation method that could improve the computation time.

Column generation involves progressively adding single-crossing group types to improve on $J_{N,0}$. This involves solving (18) (in the main paper), which can be hard if $|\widehat{\mathbf{X}}|$ is large. In fact, to improve on $J_{N,0}$, it suffices to find $\mathbf{b} \in \mathcal{B}$ such that $(\mathbf{q} - \eta_0) \cdot (\mathbf{b} - \eta_0) > 0$. In our program for the empirical application in Section 5, we impose a time limit for solving (18) and use the best feasible solution found within that time.[4] It is only when this solution satisfies $(\mathbf{q} - \eta_0) \cdot (\mathbf{b} - \eta_0) \leqslant 0$ that we continue to solve the maximization problem exactly.

The final step in our statistical test involves calculating the bootstrap test statistic $J_N^{(r)}$ and the p-value. As noted by Smeulders et al. (2021), computation time can be further reduced by not

[3] It is straightforward to check that the dimension of the space spanned by the set of all logically possible group types is precisely $|\mathbf{Y} \times \widehat{\mathbf{X}}| - |\mathbf{Y}| + 1$ 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.

[4] Using Rglpk package on R, our program gives 0.2 seconds for solving (18) at each repetition. This constraint can be binding when $\widehat{\mathbf{X}}$ is large, leading to a large integer programming problem.

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, when calculating $J_N^{(r)}$ via column generation, we can terminate the procedure once a value of $J_{N,0}^{(r)}$ becomes lower than J_N.

A5.3. Supplementary notes for Section 4.2

Validity of critical value. 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

$$\mathbb{P}^{SC}(\beta;\mathcal{B}^*) = \left\{ \mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}} \text{ and } \sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta \right\}. \tag{a.25}$$

Assuming that we have calculated the test statistic $J_N(\beta)$, the next step is to construct the bootstrap sample. To do this, we first 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. We set

$$\Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*) = \left\{ \tau \in \Delta^{\mathcal{B}} : \tau^{\mathbf{b}} \geqslant \frac{\beta\kappa_N}{|\mathcal{B}' \cap \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \cap \mathcal{B}^* \text{ and } \tau^{\mathbf{b}} \geqslant \frac{(1-\beta)\kappa_N}{|\mathcal{B}' \backslash \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \backslash \mathcal{B}^* \right\}, \tag{a.26}$$

and

$$\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*) = \left\{ \mathbf{B}\tau : \tau \in \Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*) \text{ and } \sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta \right\}, \tag{a.27}$$

where $\mathcal{B}'$ must be chosen to satisfy the requirement in Lemma A.2 and also that $\mathcal{B}' \cap \mathcal{B}^*$ and $\mathcal{B}' \backslash \mathcal{B}^*$ are nonempty sets. We can then calculate

$$\eta^*(\beta) = \underset{\eta \in \mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*)}{\operatorname{argmin}} N(\mathbf{q}-\eta)\cdot(\mathbf{q}-\eta) = \underset{\substack{\tau \in \Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*) \\ \sum_{\mathbf{b}\in\mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta}}{\operatorname{argmin}} N(\mathbf{q}-\mathbf{B}\tau)\cdot(\mathbf{q}-\mathbf{B}\tau). \tag{a.28}$$

(which is the counterpart to η^* in the statistical test presented in Section 4.1). For each $r = 1, 2, ..., R$, we can generate a bootstrap sample $\mathbf{q}^{(r)}$ using the standard nonparametric bootstrap re-sampling from $\eta^*(\beta)$ and re-center this to $\widehat{\mathbf{q}}^{(r)} = (\mathbf{q}^{(r)} - \mathbf{q}) + \eta^*(\beta)$. We define the bootstrap test statistic $J_N^{(r)}(\beta)$ by

$$J_N^{(r)}(\beta) := \min_{\eta \in \mathbb{P}^{\rm SC}_{\kappa_N}(\beta;\mathcal{B}^*)} N(\widehat{\mathbf{q}}^{(r)} - \eta) \cdot (\widehat{\mathbf{q}}^{(r)} - \eta) = \min_{\substack{\tau \in \Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*) \\ \sum_{\mathbf{b} \in \mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta}} N(\widehat{\mathbf{q}}^{(r)} - \mathbf{B}\tau) \cdot (\widehat{\mathbf{q}}^{(r)} - \mathbf{B}\tau). \tag{a.29}$$

With the empirical distribution of $J_N^{(r)}(\beta)$ we can calculate the p-value and determine if the null hypothesis is rejected.

To justify this procedure, we rely on the argument in Deb et al. (2023), with a modification to account for the fact that our tightening procedure is different because we apply the tightening only to elements of $\mathcal{B}'$ rather than all the elements of $\mathcal{B}$ (see (a.26)). (Recall that we allow for the possibility that $\mathcal{B}$ is not computed in its entirety.)

Suppose that $\sqrt{N}(\mathbf{q} - \mathbf{p}) \xrightarrow{d} N(0, S)$, with $\hat{S}$ being a consistent estimator for S. Define $\tilde{\eta}^*(\beta)$ by

$$\tilde{\eta}^*(\beta) = \eta^*(\beta) + \frac{1}{\sqrt{N}} N(0, \hat{S})$$

and let

$$\tilde{J}_N(\beta) = \min_{\eta \in \mathbb{P}^{\rm SC}_{\kappa_N}(\beta;\mathcal{B}^*)} N(\tilde{\eta}^*(\beta) - \eta) \cdot (\tilde{\eta}^*(\beta) - \eta) \tag{a.30}$$

$$= \min_{\substack{\tau \in \Delta^{\mathcal{B}}_{\kappa_N}(\beta;\mathcal{B}^*) \\ \sum_{\mathbf{b} \in \mathcal{B}^*} \tau^{\mathbf{b}} \geqslant \beta}} N(\tilde{\eta}^*(\beta) - \mathbf{B}\tau) \cdot (\tilde{\eta}^*(\beta) - \mathbf{B}\tau). \tag{a.31}$$

The following result says (in essence) that the distribution of $\tilde{J}_N(\beta)$ approximates the distribution of the test statistic $J_N(\beta)$ and justifies our calculation of the critical value.

Theorem A.3. *Choose $\kappa_N > 0$ such that $\kappa_N \downarrow 0$ and $\sqrt{N}\kappa_N \uparrow \infty$ as $N \to \infty$ and suppose that $\mathcal{B}' \subset \mathcal{B}$ satisfies the requirement in Lemma A.2 and $\mathcal{B}' \cap \mathcal{B}^*$ and $\mathcal{B}' \backslash \mathcal{B}^*$ are nonempty. Then, under Assumptions 1 and 2, it holds that*

$$\liminf_{N \to \infty} \inf_{(\beta, \mathbf{p}) \in \mathcal{F}} Pr(J_N(\beta) \leqslant \hat{c}_{1-\alpha}) = 1 - \alpha, \tag{a.32}$$

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 $\tilde{J}_N(\beta)$ with $0 \leqslant \alpha \leqslant \frac{1}{2}$ ($\overline{\mathbb{P}}$ is defined as in Theorem A.2 and $\mathbb{P}^{SC}$ is defined in (a.21)).

Proof. The proof proceeds in the same way as the proof of Theorem 4 in Deb et al. (2023), with some adjustments that we explain next. A key step in the proof of Deb et al. (2023) is to obtain the $\mathcal{H}$-representation of *their* versions of $\mathbb{P}^{\mathrm{SC}}(\beta;\mathcal{B}^*)$ and $\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*)$. For us to use their proof, we first need to obtain $\mathcal{H}$-representations of these two sets, as we have defined it (by (a.25) and (a.27) respectively). Indeed, we claim that

$$\mathbb{P}^{\mathrm{SC}}(\beta;\mathcal{B}^*) = \left\{\mathbf{p} : \mathbf{D}\mathbf{p} \leqslant 0, \widetilde{\mathbf{D}}\mathbf{p} \leqslant \mathbf{v}(\beta), \mathbf{E}_{\mathbf{x}}\mathbf{p} = \mathbb{1}\right\}, \text{ and} \tag{a.33}$$

$$\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*) = \left\{\mathbf{p} : \mathbf{D}^{\leqslant}\mathbf{p} \leqslant -\kappa_N\mathbf{u}(\beta), \mathbf{D}^{=}\mathbf{p} = 0,\ \widetilde{\mathbf{D}}\mathbf{p} \leqslant \mathbf{v}(\beta), \mathbf{E}_{\mathbf{x}}\mathbf{p} = \mathbb{1}\right\} \tag{a.34}$$

where

(i) the set $\{\mathbf{p} : \mathbf{D}\mathbf{p} \leqslant 0\}$ is the $\mathcal{H}$-representation of the cone $\{\mathbf{B}\tau : \tau \geqslant 0\}$, with $\mathbf{D}^{\leqslant}$ and $\mathbf{D}^{=}$ respectively corresponding to the inequality and equality constraints;

(ii) the set $\{\mathbf{p} : \mathbf{D}^{\leqslant}\mathbf{p} \leqslant -\kappa_N\mathbf{u}(\beta)$ and $\mathbf{D}^{=}\mathbf{p} = 0\}$ is the $\mathcal{H}$-representation of

$$\left\{\mathbf{B}\tau : \tau^{\mathbf{b}} \geqslant \frac{\beta\kappa_N}{|\mathcal{B}' \cap \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \cap \mathcal{B}^*,\ \tau^{\mathbf{b}} \geqslant \frac{(1-\beta)\kappa_N}{|\mathcal{B}'\backslash\mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}'\backslash\mathcal{B}^*,\ \text{and } \tau \geqslant 0\right\};$$

(iii) the set $\{\mathbf{p} : \widetilde{\mathbf{D}}\mathbf{p} \leqslant \mathbf{v}(\beta)\}$ coincides with $\{\mathbf{B}\tau : \sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} \geqslant \beta,\ \tau \in \mathbb{R}^{|\mathcal{B}|}_{+}\}$; and

(iv) $\mathbf{E}_{\mathbf{x}}\mathbf{p} = \mathbb{1}$ corresponds to the adding-up condition of choice frequencies at each covariate $\mathbf{x}$. Note that $\mathbb{1}$ stands for the vector of 1's.

The existence of the representation in (i) is immediate from Minkowski-Weyl duality and the condition (iv) is also clear. Condition (ii) follows from Lemma A.2.[5] There exists $\widetilde{\mathbf{D}}$ and $\mathbf{v}(\beta)$ such that condition (iii) holds since $\{\mathbf{B}\tau : \sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} \geqslant \beta,\ \tau \in \mathbb{R}^{|\mathcal{B}|}_{+}\}$ is obviously a polyhedron (in the sense of Ziegler (1995)).

With these representations of $\mathbb{P}^{\mathrm{SC}}(\beta;\mathcal{B}^*)$ and $\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*)$, we obtain the result by (essentially) mimicking the proof in Deb et al. (2023). Note that the counterpart to (iii) in their proof is defined by the equality $\widetilde{\mathbf{D}}\mathbf{p} = \mathbf{v}(\beta)$, since they define $\mathbb{P}^{\mathrm{SC}}(\beta;\mathcal{B}^*)$ and $\mathbb{P}^{\mathrm{SC}}_{\kappa_N}(\beta;\mathcal{B}^*)$ by using $\sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} = \beta$, rather than $\sum_{\mathbf{b}\in\mathcal{B}^*}\tau^{\mathbf{b}} \geqslant \beta$. As a result of this, some equalities to need to be converted to inequalities when mimicking their proof, but this can be done without introducing difficulties. **QED**

[5] In applying Lemma A.2, it does not matter even if the lower bounds on $\tau^{\mathbf{b}}$ vary across $\mathbf{b} \in \mathcal{B}'$.

Summary. For a fixed $\mathcal{B}^* \subset \mathcal{B}$ and $\beta \in (0,1)$, we provide a step-by-step procedure for obtaining the p-value for the null hypothesis described in (19). The upper bound of β is obtained through binary search.

I Obtain the test statistic $J_N(\beta)$ defined by (20) (in the main paper) as follows:

(i) Based on $\mathcal{B}_0$, solve the minimization problem (21) to get $J_{N,0}(\beta)$ and $\eta_0(\beta)$.

(ii) Check the value of (23). IF it is strictly positive, then update $\mathcal{B}_0$ by adding a solution of (23) and go to (i).[6] ELSE, adopt the resulting $J_{N,0}(\beta)$ as $J_N(\beta)$ and STOP.

II Obtain the tightened estimator $\eta^*(\beta)$ in (a.28) as follows:

(i) Calculate $\mathcal{B}' \subset \mathcal{B}$ obeying the requirements in Theorem A.3.[7]

(ii) Set $\mathcal{B}'$ as $\mathcal{B}_0$, and run the procedure in Step I, replacing $\Delta^{\mathcal{B}_0}$ in problem (21) with

$$\Delta^{\mathcal{B}_0}_{\kappa_N}(\beta;\mathcal{B}^*) = \Big\{\tau \in \Delta^{\mathcal{B}_0} : \tau^{\mathbf{b}} \geqslant \frac{\beta\kappa_N}{|\mathcal{B}' \cap \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \cap \mathcal{B}^* \\ \text{and } \tau^{\mathbf{b}} \geqslant \frac{(1-\beta)\kappa_N}{|\mathcal{B}' \backslash \mathcal{B}^*|} \text{ for } \mathbf{b} \in \mathcal{B}' \backslash \mathcal{B}^* \Big\}.$$

When it stops, the resulting $\eta_0(\beta)$ is $\eta^*(\beta)$.

III Obtain the bootstrap test statistics $J^{(r)}_N(\beta)$ defined in (a.29) for $r = 1, 2, ..., R$:

(i) Obtain the re-centered bootstrap sample $\widehat{\mathbf{q}}^{(r)} = (\mathbf{q}^{(r)} - \mathbf{q}) + \eta^*(\beta)$.

(ii) Set $\mathcal{B}'$ as $\mathcal{B}_0$ and run the procedure in Step I replacing $\mathbf{q}$ and $\Delta^{\mathcal{B}}(\beta;\mathcal{B}^*)$ in problem (21) with $\widehat{\mathbf{q}}^{(r)}$ and $\Delta^{\mathcal{B}_0}_{\kappa_N}(\beta;\mathcal{B}^*)$, respectively. When it stops, the resulting $J_{N,0}(\beta)$ is $J^{(r)}_N(\beta)$.

IV Lastly, calculate the p-value $p = \#\{J^{(r)}_N(\beta) > J_N(\beta)\}/R$.

[6] Note that, in this case, a solution of (23) is a *pair* of group types $\{\mathbf{b}^*, \mathbf{b}\}$ described in Proposition 3.

[7] Theorem A.3 requires $\mathcal{B}'$ to satisfy the requirements in Lemma A.2, $\mathcal{B}' \cap \mathcal{B}^* \neq \varnothing$, and $\mathcal{B}' \backslash \mathcal{B}^* \neq \varnothing$. If $\mathcal{B}'$, constructed as in the preceding subsection, does not satisfy the latter two requirements, then we need to (manually) add some group types to $\mathcal{B}'$ so that they are satisfied. (If a set satisfies the requirements of Lemma A.2, adding more types to that set will not upset the requirements.) In practice, we find that this problem does not arise because we would have accumulated (through performing the statistical test of the model using column generation) a very large subset of $\mathcal{B}$ satisfying the requirements of Lemma A.2. We could then choose *this* set to be $\mathcal{B}'$; it is likely that this set would satisfy $\mathcal{B}' \cap \mathcal{B}^* \neq \varnothing$ and $\mathcal{B}' \backslash \mathcal{B}^* \neq \varnothing$.

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 $\mathcal{SC}$-rationalizability 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.[8] 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.[9] 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 their parametric estimation for some of the discretization patterns displayed in Table A.3. (See

[8] 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.

[9] 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 $|\mathbf{X}| = 144$.

$MP_{LCC} \times MP_{OA} \times MS$	$\lvert\widehat{\mathbf{X}}\rvert$	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

$MP_{LCC} \times MP_{OA} \times MS$	$\lvert\widehat{\mathbf{X}}\rvert$	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

Section A1 for more details on Kline and Tamer's estimation procedure.) While their approach is mainly developed for inference of partially identified parameters, they point out that it could also be used for "specification testing." In their procedure the empirical choice frequencies are repeatedly sampled and, for each sample, one could derive the set of model parameters. Kline and Tamer point out that the frequency with which the identified set (of parameters) is nonempty could be thought of as a form of specification testing. The results obtained via this procedure are broadly consistent with ours. Following Kline and Tamer, we created 251 random samples from the empirical choice frequencies (which we call $\mathcal{Q}$); the frequency of obtaining a nonempty set of estimated coefficients in these samples is (i) 1 for $2 \times 2 \times 2$, (ii) 0.928 for $2 \times 2 \times 3$, and (iii) 0 for $3 \times 3 \times 3$ (at a tolerance level of 0.075). In particular, this approach would conclude that the parametric model is misspecified in case (iii), which is consistent with the result from our test since it rejects the nonparametric model in this case.

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

$$\mathcal{B}^\dagger = \{\mathrm{B} \in \mathcal{B} : \overline{\mathbf{y}} \in \mathrm{NE}(\overline{\mathbf{x}}, \mathbf{\Pi}) \text{ for any } \mathbf{\Pi} \in \mathcal{SC} \text{ that rationalizes B}\}\,. \tag{a.35}$$

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* $\mathrm{B}(\mathbf{x}'') = (N, E)$.[10]

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*
(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

[10]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.

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 $\Big((|\mathbf{Y} \times \widehat{\mathbf{X}}| + 1) + K(\overline{\mathbf{y}}, \overline{\mathbf{x}})\Big) \times |\mathbf{Y} \times \widehat{\mathbf{X}}|$, where

$$K(\overline{\mathbf{y}}, \overline{\mathbf{x}}) = \#\{\mathbf{x}' \in \widehat{\mathbf{X}} : x'_1 \geqslant \overline{x}_i\} \times \#\{\mathbf{x}'' \in \widehat{\mathbf{X}} : x''_2 \leqslant \overline{x}_2\}, \text{ if } \mathbf{y} = (N, E), \tag{a.36}$$

$$= \#\{\mathbf{x}' \in \widehat{\mathbf{X}} : x'_1 \leqslant \overline{x}_1\} \times \#\{\mathbf{x}'' \in \widehat{\mathbf{X}} : x''_2 \geqslant \overline{x}_2\}, \text{ if } \mathbf{y} = (E, N). \tag{a.37}$$

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

$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$	$(0,0,0)$		$(0,1,0)$		$(1,0,0)$		$(1,1,0)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
$\max \Pr[\bar{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \bar{\mathbf{x}})]$	0.699	0.544	0.815	0.503	0.503	0.644	0.558	0.555
$\min \Pr[\bar{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \bar{\mathbf{x}})]$	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
$(\text{MP}_{LCC}, \text{MP}_{OA}, \text{MS})$	$(0,0,1)$		$(0,1,1)$		$(1,0,1)$		$(1,1,1)$	
Action profile	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)	(N,E)	(E,N)
$\max \Pr[\bar{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \bar{\mathbf{x}})]$	0.841	0.616	0.913	0.496	0.485	0.661	0.523	0.497
$\min \Pr[\bar{\mathbf{y}} \in \mathbf{NE}(\mathbf{\Pi}, \bar{\mathbf{x}})]$	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

of (N, E).[11] 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, $\bar{\mathbf{x}} = (0, 0, 1)$.

References

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2023): Revealed price preference: theory and empirical analysis. *Review of Economic Studies*, 90, 707–743.

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, 437–455.

Ziegler, G.M. (1995): *Lectures on polytopes.* Graduate texts in mathematics, Springer.

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