A revealed preference theory of monotone choice and strategic complementarity

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

March 16, 2017

Abstract: We develop a revealed preference characterization of those data sets which are consistent with the agent obeying single crossing differences and the interval dominance order. This characterization can be extended to test for strategic complementarity in games and to make out-of-sample predictions of Nash equilibria. In keeping with the literature on axiomatic or revealed preference characterizations, these results require repeated observations of the same agent or group of agents. To facilitate the empirical application of these tests, we formulate stochastic versions that are applicable to cross sectional data. Lastly, to illustrate how our results could be used, we apply them to model spousal and regulatory influence on smoking behavior.

Keywords: monotone comparative statics, single crossing property, interval dominance, supermodular games, lattices

JEL classification numbers: C6, C7, D7

* Email addresses: nlazzati@ucsc.edu; john.quah@jhu.edu; kshirai1985@gmail.com
For helpful discussions and comments, the authors are grateful to G. Bulman, J. Fox, K. Hirano, A. Kajii, C. Manski, J. Stoye, B. Strulovici, S. Takahashi and M. Walker, as well as to seminar audiences at Arizona, Johns Hopkins, Kobe, Kyoto, Louvain (CORE), NYU, the National University of Singapore (NUS), Northwestern University, University of Paris (Dauphine), Queensland, Shanghai University of Finance and Economics, Singapore Management University, UC San Diego, and to conference participants at the Cowles Conference on Heterogenous Agents and Microeconometrics (Yale, 2015), the SAET Conference (Cambridge (UK), 2015), the World Congress of the Econometric Society (Montreal, 2015) and the SWET Conference (UC Riverside, 2016). Koji Shirai gratefully acknowledges financial support from the Japan Society for Promotion of Science (Grant-in-Aid for JSPS Fellows); he would also like to thank St Hugh's College (Oxford), the Oxford Economics Department, and the NUS Economics Department for their hospitality during his extended visits to these institutions when he was a JSPS fellow.

1 Introduction

In both theoretical and empirical work, economists often require that the action chosen by an agent will increase with another variable, so that the two may be regarded as complements. The theory of monotone comparative statics provides conditions on preferences that guarantee this behavior. To put this more formally, suppose an agent i chooses an action from a set S which is a subset of $\mathbb{R}$. The agent's preference over different actions in S depends on some exogenous variable ξ_i, which is drawn from a partially ordered set. It is known that the agent's optimal action increases with ξ_i if the agent's preference obeys *single crossing differences* (Milgrom and Shannon, 1994); loosely speaking, this property requires that, whenever $x_i'' > x_i'$ and the agent prefers x_i'' to x_i' at some value of the exogenous variable, then an increase in the exogenous variable will preserve the agent's preference for x_i'' over x_i'.[1] When S is an interval of $\mathbb{R}$ then a property weaker than single crossing differences called the *interval dominance order* (Quah and Strulovici, 2009) is sufficient to guarantee the monotonicity of the optimal action; this property says that whenever x_i'' is preferred to all actions in the interval $[x_i', x_i'']$, then this preference is preserved when the exogenous variable increases.

Given the central role played by single crossing differences and the interval dominance order in monotone comparative statics, it is important that we develop a nonparametric procedure for testing whether observed data is in fact consistent with one or both of these properties. The first and most basic objective of this paper is to develop such a test.

1.1 Testing for single crossing differences

Suppose an observer has a finite data set, where at observation t, the exogenous variable is ξ_i^t and the agent i chooses x_i^t from the feasible action set A_i^t, which we assume is a compact interval of $\mathbb{R}$. We show that if agent i has a preference that respects the interval dominance order, then the data set $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t=1}^T$ must obey an intuitive and easy-to-check property we call the *axiom of revealed complementarity* (ARC). Conversely, if $\mathcal{O}_i$ obeys this property, then the agent's choices can be rationalized by a preference obeying the interval dominance order; indeed, it can

[1] Milgrom and Shannon (1994) refers to this as the 'single crossing property.' The term 'single crossing differences,' which is more descriptive and also analogous to 'increasing differences,' follows Milgrom (2004).

be rationalized by a preference obeying the stronger property of single crossing differences.[2]

Loosely speaking, ARC says the following: suppose that at some value of the exogenous variable, agent i reveals a preference for x_i'' over x_i' with $x_i'' > x_i'$ (this can be a direct revelation in the sense that x_i'' was chosen when x_i' was feasible at some observation, or it could be revealed indirectly via transitive closure); then the agent cannot reveal a preference for x_i' over x_i'' at a higher exogenous variable. Our result is reminiscent of canonical results such as Afriat's or Richter's Theorem, which say that so long as the preference pairs revealed by the data contains no cycles, then these (typically incomplete) revealed preferences admit a completion. The distinctive issue in our context is that, in a sense, there is not a single preference over actions but many preferences corresponding to different values of the exogenous variable; therefore, our proof involves first characterizing all the preference information conveyed by the data, and then completing these preferences simultaneously in a way that obeys single crossing differences.

In games with strategic complementarity (see Milgrom and Roberts (1990) and Vives (1990)) players' strategies are complements in the sense that an agent's best response increases with the action of other players in the game. These games are known to be very well-behaved; for example, they always have pure strategy Nash equilibria and, in fact, there is always a largest and a smallest pure strategy Nash equilibrium.

In this context, suppose that for each player i ($i = 1, 2, ..., n$), we observe the feasible action set A_i^t, the action chosen by the player, $x_i^t \in A_i^t$, and an exogenous variable y_i^t (drawn from a poset) that affects player i's action. An observation t may be succinctly written as (x^t, y^t, A^t) (where $x^t = (x_i^t)_{i=1}^n$, etc.) such that x^t is the observed action profile in the treatment (y^t, A^t). Then we can ask whether the data set $\mathcal{O} = \{(x^t, y^t, A^t)\}_{t=1}^T$ is consistent with the hypothesis that the observations constitute pure strategy Nash equilibria in games with strategic complementarity; note that this hypothesis is internally consistent since we know that such equilibria must exist in this class of games. The answer to our question is straightforward given the single-agent results: all we need to do is to check that each player's choices obey ARC, in the sense that, for all i,

[2]Readers familiar with Afriat's Theorem may notice a parallel in the following sense: the generalized axiom of revealed preference (GARP) is necessary whenever the consumer is maximizing a locally nonsatiated preference and it is sufficient to guarantee the *stronger* conclusion that there is a continuous, strictly increasing, and concave function rationalizing the data. In our case, ARC is necessary for strict interval dominance and sufficient for strict single crossing differences (which is a stronger property).

$\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t=1}^T$ (with $\xi_i^t = (x_{-i}^t, y_i^t)$) obeys ARC. (From player i's perspective, the variables affecting his preference are the realized value of y_i and the actions of other players.)

When the data set $\mathcal{O}$ obeys ARC (in the sense that every player obeys ARC), it would be natural to exploit this data to make predictions of the outcome in a new game, with different feasible action sets $A^0 = (A_i^0)_{i=1}^n$ and different exogenous variables $y^0 = (y_i^0)_{i=1}^n$, assuming that the players' preferences obey single crossing differences and remain unchanged. We provide a procedure for working out the set of all possible Nash equilibria in this new game. We also show that this set has properties that echo those of a set of Nash equilibria in a game with strategic complementarity: while the set itself may not have a largest or smallest element, its closure does have a largest and a smallest element and these extremal elements increase with y^0.

1.2 Tests on cross sectional data

So far we have considered an observer who records the behavior of an agent or a group of agents across a sequence of different treatments. It is not always possible to obtain data of this type in empirical settings. Suppose instead that, at each treatment, we observe the joint actions taken by a large population of n-player groups, so the data set is $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t=1}^T$, where μ^t is a distribution on A^t. Then the natural generalization of our notion of rationalization is to require that the population be decomposable into segments such that (i) all groups within a segment have the same equilibrium play at any one treatment and (ii) the equilibrium play is consistent with strategic complementarity. This rationalization concept captures the idea that treatments have been randomly assigned across the whole population of groups, so that the distribution of 'group types' is the same across treatments; note, however, that it places *no restrictions* on that distribution, nor on how groups select among pure strategy Nash equilibria.

We show that it is possible to check whether $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t=1}^T$ is consistent with strategic complementarity in this sense by solving a certain system of linear equations. When a data set passes this test, we provide (following Manski (2007)), a procedure to partially identify the distribution of equilibrium responses in the population under a new treatment.

Lastly, to illustrate the use of our techniques, we apply them to model the influence of spouses and workplace smoking regulations on smoking behavior. The US census provides information on tobacco use in married couples and smoking policies at their workplaces (whether it is permitted or not).[3] From this we obtain the (sample) joint distribution of smoking behavior among couples and smoking policies at their workplaces. We then test the hypothesis that someone is more likely to smoke if his/her spouse is smoking and his/her workplace permits smoking, with the distribution of smoking patterns arising as a Nash equilibrium between spouses. We show that the census data are not exactly rationalizable in this way; however, using an econometric procedure developed recently by Kitamura and Stoye (2016), we find that this failure is not statistically significant, so the hypothesis of strategic complementarity cannot be rejected.

1.3 Related literature

Topkis (1998, Theorem 2.8.9) considers a correspondence $\varphi : T \to \mathbb{R}^\ell$ mapping elements of a totally ordered set T (interpreted as the set of parameters) to compact sublattices of $\mathbb{R}^l$. He shows that this correspondence is increasing in the strong set order if and only if there is $f : \mathbb{R}^\ell \times T \to \mathbb{R}$ such that $\varphi(t) = \arg\max_{x\in\mathbb{R}^\ell} f(x,t)$, where f is supermodular in x and has increasing differences in (x,t). The rationalizability concept used by Topkis is more stringent than the one we employ since the optimal choices must coincide with (rather than simply contain) $\varphi(t)$. In the case where φ is a choice *function*, it is not hard to see that such a rationalization is possible even when T is a partially (rather than totally) ordered set; this has been noted by Carvajal (2004) who also applies it to a game setting (where T, being the profile of actions of other players in the game, will generally not be totally ordered).

In this paper, we also permit the set of parameters to be partially rather than totally ordered. However, we confine ourselves to the case where actions are totally ordered (in essence, elements of $\mathbb{R}$ rather than $\mathbb{R}^\ell$), while allowing for observations of the choices made from different *subsets* of the set of all possible actions. Consequently, at a given parameter value, the observer may have partial information on the agent's ranking over different actions rather than simply the globally optimal action. In this respect, the problem is more complicated than the one posed by Topkis, because the rationalizing preference we construct has to agree with this wider range of preference information (in addition to obeying single crossing differences).

The extension of our revealed preference tests to cross-sectional data sets with unobserved

[3] The data were collected in the early 90s, when significant numbers of workplaces still permitted smoking.

heterogeneity follows an approach that has been taken by other authors (see, for example, McFadden and Richter (1991) and Manski (2007)). Manski (2007) also discusses making predictions in unobserved treatments and our approach to this issue is, in its essentials, the same as his.

Echenique and Komunjer (2009) develop a structural model that could be used to test for strategic complementarity in certain special classes of games, including two person games. Their test relies on a stochastic equilibrium selection rule that places strictly positive probability on the extremal elements of the set of Nash equilibria and checks certain observable properties implied by strategic complementarity; the sufficiency of those properties (for rationalizability) is not addressed. Our rationalizability tests do not require assumptions on equilibrium selection. Aradillas-Lopez (2011) provides nonparametric probability bounds for Nash equilibrium actions for a class of games with characteristics that are similar to, but distinct from, games with strategic complementarity. There are also papers where actions are assumed to be strategic complements or substitutes in order to sharpen inference or predictions; for example, Kline and Tamer (2012), Molinari and Rosen (2008), Uetake and Watanabe (2013), and Lazzati (2015). By and large, the emphasis in these papers is not to test for strategic complementarity but to exploit it as an assumption; indeed the model may not include the type of exogenous treatment variation that makes the assumption refutable.[4]

1.4 Organization of the paper

Section 2 gives a quick review of standard results in monotone comparative statics and supermodular games. Sections 3 and 4 discuss the revealed preference theory motivated by the standard results discussed in Section 2, in the context of (respectively) individual and group decisions. The extension to cross-sectional data sets is explained in Section 5, while Section 6 applies the results to study smoking decisions among couples.

[4]There is also an econometric literature on testing for complementarities in firm activity (see, for example, Athey and Stern (1998)). The data environments in those contexts lead to econometric approaches that are quite different from the method used in Sections 5 and 6 of this paper.

2 Basic concepts and theory

Our objective in this section is to give a quick review of some basic concepts and results in monotone comparative statics and of their application to games with strategic complementarities. This will motivate the revealed preference theory developed later in the paper.

2.1 Monotone choice on intervals

Let $X_i \subset \mathbb{R}$ be the set of all conceivable actions of an agent i. A *feasible action set* of agent i is a subset A_i of X_i. We assume that A_i is compact in $\mathbb{R}$ and that it is an *interval of* X_i. We say that a set $A_i \subseteq X_i$ is an *interval of* X_i if, whenever x_i'', $x_i' \in A_i$, with $x_i'' > x_i'$, then, for any element $\tilde{x}_i \in X_i$ such that $x_i'' > \tilde{x}_i > x_i'$, $\tilde{x}_i \in A_i$. Given that A_i is both compact and an interval, we can refer to it as a *compact interval.* It is clear that there must be $\underline{a}_i$ and $\bar{a}_i$ in A_i such that $A_i = \{x_i \in X_i : \underline{a}_i \leqslant x_i \leqslant \bar{a}_i\}$ and it is sometimes convenient to denote A_i by $[\underline{a}_i, \bar{a}_i]$. We denote by $\mathcal{A}_i$ the collection of all compact intervals of X_i. We assume that agent i's choice over different actions in a feasible action set A_i is affected by a parameter ξ_i, where ξ_i is drawn from a partially ordered set (or poset, for short) $(\Xi_i, \geqslant)$; ξ_i may include certain exogenous variables and/or the actions of other agents (when we extend the analysis to a game). For the sake of notational simplicity, we are using the same notation for the orders on X_i and Ξ_i and for any other ordered sets; we do not anticipate any danger of confusion.

We call a binary relation $\succsim_i$ on $X_i \times \Xi_i$ a *preference* of agent i if, for every fixed $\xi_i \in \Xi_i$, $\succsim_i$ is a complete, reflexive and transitive relation on X_i. A preference $\succsim_i$ is *regular* if, for all $A_i \in \mathcal{A}_i$ and ξ_i, the set $\mathrm{BR}_i(\xi_i, A_i)$ defined by

$$\mathrm{BR}_i(\xi_i, A_i) = \{x_i' \in A_i : \ (x_i', \xi_i) \succsim_i (x_i, \xi_i) \text{ for all } x_i \in A_i\}, \tag{1}$$

is nonempty and compact in $\mathbb{R}$. Regularity obviously holds in the important case where every bounded set of X_i is finite (for example, if $X_i \subseteq \mathbb{N}$) and, more generally, it holds if $\succsim_i$ is continuous at every $\xi_i \in \Xi_i$. We refer to $\mathrm{BR}_i(\xi_i, A_i)$ as agent i's *best response* or *optimal choice* at (ξ_i, A_i) and it is *monotone* or *increasing* in ξ_i if, for every $\xi_i'' > \xi_i'$,

$$x_i'' \in \mathrm{BR}_i(\xi_i'', A_i) \text{ and } x_i' \in \mathrm{BR}_i(\xi_i', A_i) \Longrightarrow x_i'' \geqslant x_i'. \tag{2}$$

The preference $\succsim_i$ is said to obey *strict interval dominance* (SID) if, for every $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$,

$$(x_i'', \xi_i') \succsim_i (x_i, \xi_i') \text{ for all } x_i \in [x_i', x_i''] \Longrightarrow (x_i'', \xi_i'') \succ_i (x_i', \xi_i''), \tag{3}$$

where $\succ_i$ is the asymmetric part of $\succsim_i$, i.e., $(x_i, \xi_i) \succ_i (y_i, \xi_i)$ if $(x_i, \xi_i) \succsim_i (y_i, \xi_i)$ and $(y_i, \xi_i) \not\succsim_i (x_i, \xi_i)$. We denote the symmetric part of $\succsim_i$ by $\sim_i$, i.e., $(x_i, \xi_i) \sim_i (y_i, \xi_i)$ if $(x_i, \xi_i) \succsim_i (y_i, \xi_i)$ and $(y_i, \xi_i) \succsim_i (x_i, \xi_i)$. The following result is a straightforward adaptation of Theorem 1 in Quah and Strulovici (2009). We shall re-prove it here because of its central role in this paper.

THEOREM A. *Suppose $\succsim_i$ is a regular preference on $X_i \times \Xi_i$. Then agent i has a monotone best response correspondence if and only if $\succsim_i$ obeys strict interval dominance.*

Proof. To show that $\succsim_i$ obeys SID, suppose that, for some $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$, the left side of (3) holds. Letting $A_i = [x_i', x_i'']$, we obtain $x_i'' \in \mathrm{BR}_i(\xi_i', A_i)$. Hence, by (2), it also holds that $x_i'' \in \mathrm{BR}_i(\xi_i'', A_i)$. If $(x_i'', \xi_i'') \sim_i (x_i', \xi_i'')$ were to hold, then $x_i' \in \mathrm{BR}_i(\xi_i'', A_i)$. However, then we have that $x_i'' \in \mathrm{BR}_i(\xi_i', A_i)$, $x_i' \in \mathrm{BR}_i(\xi_i'', A_i)$, and $x_i' < x_i''$, which contradicts (2). Therefore, $(x_i'', \xi_i'') \succ_i (x_i', \xi_i'')$. Conversely, suppose $\xi_i'' > \xi_i'$, $x_i'' \in \mathrm{BR}_i(\xi_i'', A_i)$ and $x_i' \in \mathrm{BR}_i(\xi_i', A_i)$. If $x_i'' < x_i'$, then $(x_i', \xi_i') \succsim_i (x_i, \xi_i')$ for every $x_i \in [x_i'', x_i'] \subseteq A_i$. SID guarantees that $(x_i', \xi_i'') \succ_i (x_i'', \xi_i'')$, which contradicts the assumption that $x_i'' \in \mathrm{BR}_i(\xi_i'', A_i)$. $\square$

Readers familiar with the theory of monotone comparative statics will notice that our definition of monotonicity in (2) is stronger than the standard notion, which merely requires that $\mathrm{BR}_i(\xi_i'', A_i)$ dominates $\mathrm{BR}_i(\xi_i', A_i)$ in the *strong set order*, which means that, for any $x_i'' \in \mathrm{BR}_i(\xi_i'', A_i)$ and $x_i' \in \mathrm{BR}_i(\xi_i', A_i)$, $\max\{x_i'', x_i'\} \in \mathrm{BR}_i(\xi_i'', A_i)$ and $\min\{x_i'', x_i'\} \in \mathrm{BR}_i(\xi_i', A_i)$. This weaker notion of monotonicity can be characterized by preferences obeying *interval dominance* (rather than strict interval dominance), which is defined as follows: for every $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$,

$$(x_i'', \xi_i') \succsim_i (\succ_i)\, (x_i, \xi_i') \text{ for every } x_i \in [x_i', x_i''] \Longrightarrow (x_i'', \xi_i'') \succsim_i (\succ_i)\, (x_i', \xi_i''). \tag{4}$$

(The reader can verify this claim by a straightforward modification of the proof of Theorem A or by consulting Theorem 1 in Quah and Strulovici (2009).) Throughout this paper we have chosen to work with a stronger notion of monotonicity; the weaker notion does not permit meaningful revealed preference analysis because it does not exclude the possibility that an agent is simply indifferent to all actions at every ξ_i. In this sense, our stronger assumption here is analogous to the assumption of local non-satiation made in Afriat's Theorem.[5]

Note that interval dominance coincides with strict interval dominance if, instead of regularity, we require $\succsim_i$ to have the stronger property that $\mathrm{BR}_i(\xi_i, A_i)$ is nonempty *and unique* for every ξ_i and compact interval A_i. Throughout this paper, we have chosen neither to require nor to guarantee the uniqueness of the best response since it is unnecessarily restrictive: on continuous domains it effectively implies that the payoff function over actions is quasiconcave.

The interval dominance order is Quah and Strulovici's (2009) generalization of single crossing differences, due to Milgrom and Shannon (1994). Just as there is strict interval dominance, so there is a strict version of single crossing differences. We say that a preference relation $\succsim_i$ has *strict single crossing differences (SSCD)* if, for every $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$,

$$(x_i'', \xi_i') \succsim_i (x_i', \xi_i') \Longrightarrow (x_i'', \xi_i'') \succ_i (x_i', \xi_i''). \tag{5}$$

It is clear that every preference that obeys SSCD will also satisfy SID. Hence, it is obvious from Theorem A that if $\succsim_i$ is a regular preference on $X_i \times \Xi_i$ that obeys SSCD, then agent i has a monotone best response correspondence $\mathrm{BR}_i(\xi_i, A_i)$ for every interval $A_i \in \mathcal{A}_i$.[6]

2.2 Strategic complementarity

Let $N = \{1, 2, ..., n\}$ be the set of agents in a game, and let $X_i \subset \mathbb{R}$ be the set of all conceivable actions of agent i. We assume that i has a feasible action set A_i that is a compact interval of X_i; we denote the family of compact intervals of X_i by $\mathcal{A}_i$. Agent i's choice over different feasible actions is affected by the actions of other players and also by an exogenous variable y_i, which is drawn from a poset $(Y_i, \geqslant)$. Let $\Xi_i = X_{-i} \times Y_i$, where $X_{-i} = \times_{j \neq i} X_j$. A typical element of Ξ_i is denoted by $\xi_i = (x_{-i}, y_i)$ and Ξ_i is a poset if we endow it with the product order. We assume that agent i has a preference $\succsim_i$ on $X_i \times \Xi_i$, in the sense defined in Section 2.1.

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

[6] In fact, SSCD of a preference ensures more than that: it is necessary and sufficient for the monotonicity of a best response correspondence on *arbitrary* feasible action sets and not only interval feasible action sets. On the relationship between single crossing differences and the interval dominance order, see Quah and Strulovici (2009).

Given a profile of regular preferences $(\succsim_i)_{i\in N}$, a *joint feasible action set* $A \in \mathcal{A} = \times_{i\in N}\mathcal{A}_i$, and a profile of exogenous variables $y \in Y = \times_{i\in N} Y_i$, we can define a game

$$\mathcal{G}(y, A) = \left[(y_i)_{i\in N}, (A_i)_{i\in N}, (\succsim_i)_{i\in N}\right].$$

We say that the family of games $\mathbb{G} = \{\mathcal{G}(y, A)\}_{(y,A)\in Y\times\mathcal{A}}$ exhibits *strategic complementarity* if, for every $A \in \mathcal{A}$, the best response of each agent i (as given by (1)) is monotone in $\xi_i = (x_{-i}, y_i)$. It is clear from Theorem A that *the family of games* $\mathbb{G} = \{\mathcal{G}(y, A)\}_{(y,A)\in Y\times\mathcal{A}}$ *exhibits strategic complementarity if and only if* $\succsim_i$ *is an SID preference for every agent* i.

Example 1. Consider a Bertrand oligopoly with n firms, with each firm producing a single differentiated product. Firm i has constant marginal cost $c_i > 0$, faces the demand function $D_i(p_i, p_{-i}) : \mathbb{R}_{++} \times \mathbb{R}^{n-1}_{++} \to \mathbb{R}_+$, and chooses its price $p_i > 0$ to maximize profit $\Pi_i(p_i, p_{-i}, c_i) = (p_i - c_i)D_i(p_i, p_{-i})$. Suppose that its own-price elasticity of demand,

$$-\frac{p_i}{D_i(p_i, p_{-i})}\frac{\partial D_i}{\partial p_i}(p_i, p_{-i})$$

is strictly falling with respect to p_{-i} (the prices charged by other firms); this captures the idea that the other firms' products are substitutes for firm i's product. Then Π_i obeys SSCD in $(p_i; p_{-i}, c_i)$ and, hence, on any compact interval of prices, firm i's profit-maximizing prices are monotone in (p_{-i}, c_i).[7] If this property holds for every firm in the industry, then the collection of Bertrand games generated by different feasible price sets to each firm and different exogenous variables, $c = (c_i)_{i\in N}$, constitutes a collection of games exhibiting strategic complementarity.

The following result summarizes some of the properties of Nash equilibria in a game with strategic complementarity. For our purposes, the most important feature of these games is that they have pure strategy Nash equilibria, so it is not a priori unreasonable to hypothesize that players are playing a pure strategy Nash equilibrium in such a game.

Theorem B. *Suppose* $\mathbb{G} = \{\mathcal{G}(y, A)\}_{(y,A)\in Y\times\mathcal{A}}$ *exhibits strategic complementarity. Then, for every game* $\mathcal{G}(y, A) \in \mathbb{G}$*, the set of pure strategy Nash equilibria* $E(y, A)$ *is nonempty and there*

[7]Specifically, they guarantee that for any $p_i'' > p_i'$, $\ln \Pi(p_i'', p_{-i}, c_i) - \ln \Pi(p_i', p_{-i}, c_i)$ is strictly increasing in (p_{-i}, c_i), which implies SSCD (see, Milgrom and Shannon (1994)).

is a largest and a smallest Nash equilibrium, both of which are increasing in y.

The set of Nash equilibria of $\mathcal{G}(y, A)$ coincides with the fixed points of the joint best response correspondence $\mathrm{BR}(\cdot, y, A) : A \rightrightarrows A$, where, denoting (x_{-i}, y_i) by ξ_i,

$$\mathrm{BR}(x, y, A) = (\mathrm{BR}_1(\xi_1, A_1), \mathrm{BR}_2(\xi_2, A_2), ..., \mathrm{BR}_n(\xi_n, A_n)) .$$

Both the non-emptiness and structure of $E(y, A)$ flow from the fact that this is a very well-behaved correspondence. Indeed, under strategic complementarity, $\mathrm{BR}_i(\xi_i, A_i)$ is increasing in ξ_i (in the sense of (2), for all i) and so $\mathrm{BR}(x, y, A)$ is increasing in (x, y).[8]

3 Revealed monotone choice

Consider an observer who collects a finite data set from agent i, where each observation consists of the action chosen by the agent, the set of feasible actions, and the value of the parameter. Formally, the data set is $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$, where $\mathcal{T} = \{1, 2, ..., T\}$. This means that, at observation t, the agent is subjected to the *treatment* $(\xi_i^t, A_i^t) \in \Xi_i^t \times \mathcal{A}_i$ and chooses the action $x_i^t \in A_i^t$.[9] We say that $\mathcal{O}_i$ (or simply, agent i) is *consistent with monotonicity* or *monotone-rationalizable* if there is a regular and SID preference $\succsim_i$ on $X_i \times \Xi_i$ such that for every $t \in \mathcal{T}$, $(x_i^t, \xi_i^t) \succsim_i (x_i, \xi_i^t)$ for every $x_i \in A_i^t$. The motivation for this definition is clear given Theorem A: if $\mathcal{O}_i$ is monotone-rationalizable then we have found a preference that (i) accounts for the observed behavior of the agent and (ii) guarantees that the agent's optimal choice based on this preference is increasing in ξ_i, on any feasible action set that is an interval. Our principal objective in this section is to characterize those data sets that are monotone-rationalizable.

3.1 The axiom of revealed complementarity

We first define the revealed preference relations induced by $\mathcal{O}_i$. The *direct revealed preference* relation $\succsim_i^R$ is defined as follows: $(x_i'', \xi_i) \succsim_i^R (x_i', \xi_i)$ if $(x_i'', \xi_i) = (x_i^t, \xi_i^t)$ and $x_i' \in A_i^t$ for some $t \in \mathcal{T}$. The *indirect revealed preference* relation $\succsim_i^{RT}$ is the transitive closure of $\succsim_i^R$, i.e., $(x_i'', \xi_i) \succsim_i^{RT}$

[8] For a proof of Theorem B see Milgrom and Roberts (1990) or Topkis (1998).

[9] In Manski (2007) different treatments correspond to different feasible sets. It is clear, given our focus on monotonicity, that we should consider treatments that also involve changes in the parameter affecting preference.

(x_i', ξ_i) if there exists a finite sequence $z_i^1, z_i^2, ..., z_i^k$ in X_i such that

$$(x_i'', \xi_i) \succsim_i^R (z_i^1, \xi_i) \succsim_i^R (z_i^2, \xi_i) \succsim_i^R ... \succsim_i^R (z_i^k, \xi_i) \succsim_i^R (x_i', \xi_i). \tag{6}$$

The motivation for this terminology is clear. If agent i has preference $\succsim_i$ and, at some treatment (ξ_i, A_i), the agent chooses x_i'' when $x_i' \in A_i$, then it must be the case that $(x_i'', \xi_i) \succsim_i (x_i', \xi_i)$. Furthermore, given that $\succsim_i$ is transitive, if $(x_i'', \xi_i) \succsim_i^{RT} (x_i', \xi_i)$ then $(x_i'', \xi_i) \succsim_i (x_i', \xi_i)$.[10] We are now ready to introduce the axiom that characterizes monotone-rationalizability.

Definition 1. *The data set $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t=1}^T$ obeys the* Axiom of Revealed Complementarity (ARC) *if, for every $s, t \in \mathcal{T}$,*

$$\xi_i^t > \xi_i^s,\ x_i^t < x_i^s,\ \text{ and } (x_i^s, \xi_i^s) \succsim_i^{RT} (x_i^t, \xi_i^s) \Longrightarrow (x_i^t, \xi_i^t) \not\succsim_i^{RT} (x_i^s, \xi_i^t). \tag{7}$$

It is clear that ARC is a non-vacuous restriction on data. So long as the number of observations $\mathcal{O}_i$ is finite, checking whether (x_i^s, ξ_i^s) and (x_i^t, ξ_i^s) are related by $\succsim_i^{RT}$ is a finite procedure and, consequently, so is checking for ARC. It is also clear that there are no computational difficulties, whether theoretical or practical, associated with the implementation of this test.

We first show that ARC is necessary for monotone-rationalizability. This in turn requires a result showing that $\succsim_i^{RT}$ has what we call the *interval property.* A relation $\mathcal{R}$ on $X_i \times \Xi_i$ has this property if, whenever $(x_i, \xi_i)\, \mathcal{R}\, (\tilde{x}_i, \xi_i)$, for x_i, $\tilde{x}_i$ in X_i, then $(x_i, \xi_i)\, \mathcal{R}\, (z_i, \xi_i)$ for any z_i between x_i and $\tilde{x}_i$, i.e., $x_i < z_i < \tilde{x}_i$ or $\tilde{x}_i < z_i < x_i$. (In fact, this property also plays an important role in proving the sufficiency of ARC.)

Lemma 1. *The relation $\succsim_i^{RT}$ in $X_i \times \Xi_i$ induced by $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t=1}^T$ has the interval property.*

Proof. If $(x_i'', \xi_i) \succsim_i^R (x_i', \xi_i)$, then there is A_i^t such that $x_i'' = x_i^t$ and $x_i' \in A_i^t$. Since A_i^t is an interval, it is clear that $(x_i'', \xi_i) \succsim_i^R (x_i, \xi_i)$ for any x_i between x_i'' and x_i'. Now suppose $(x_i'', \xi_i) \succsim_i^{RT} (x_i', \xi_i)$, but $(x_i'', \xi_i) \not\succsim_i^R (x_i', \xi_i)$. Then, we have a sequence like (6). Suppose also that $x_i'' > x_i'$ and consider x_i such that $x_i'' > x_i > x_i'$. (The case where $x_i'' < x_i'$ can be handled in a similar way.) Letting $z_i^0 = x_i''$ and $z_i^{k+1} = x_i'$, we know that there exists at least one $0 \leqslant m \leqslant k$

[10] Note, however, that $\succsim_i^R$ and $\succsim_i^{RT}$ are not generally complete on X_i for every fixed ξ_i; as such, these relations are not preferences as we have defined them.

such that $z_i^m \geqslant x_i \geqslant z^{m+1}$. Since $(z_i^m, \xi_i) \succsim_i^R (z_i^{m+1}, \xi_i)$, it must hold that $(z_i^m, \xi_i) \succsim_i^R (x_i, \xi_i)$. This in turn implies that $(x_i'', \xi_i) = (z_i^0, \xi_i) \succsim_i^{RT} (x_i, \xi_i)$, since $(z_i^0, \xi_i) \succsim_i^{RT} (z_i^m, \xi_i)$. $\square$

Proof of the necessity of ARC. Indeed, suppose there are observations s and t such that $\xi_i^t > \xi_i^s$, $x_i^t < x_i^s$, and $(x_i^s, \xi_i^s) \succsim_i^{RT} (x_i^t, \xi_i^s)$. By Lemma 1, $\succsim_i^{RT}$ has the interval property, and so $(x_i^s, \xi_i^s) \succsim_i^{RT} (x_i, \xi_i^s)$ for all $x_i \in [x_i^t, x_i^s]$. Since $\mathcal{O}_i$ is SID-rationalizable, there is an SID preference $\succsim_i$ on $X_i \times \Xi_i$ such that $(x_i^s, \xi_i^s) \succsim_i (x_i, \xi_i^s)$ for all $x_i \in [x_i^t, x_i^s]$. The SID property on $\succsim_i$ guarantees that $(x_i^s, \xi_i^t) \succ_i (x_i^t, \xi_i^t)$, which means $(x_i^t, \xi_i^t) \not\succsim_i^{RT} (x_i^s, \xi_i^t)$. $\square$

Our more substantial claim is that ARC is also *sufficient* for monotone-rationalizability. In fact, an even stronger property is true: whenever a data set obeys ARC, it is rationalizable by an SSCD (and not just SID) preference.[11] The next result summarizes our main findings.

Theorem 1. *The following statements on the data set $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ are equivalent:*

(a) *$\mathcal{O}_i$ is monotone-rationalizable.*

(b) *$\mathcal{O}_i$ obeys ARC.*

(c) *$\mathcal{O}_i$ is rationalizable by a regular and SSCD preference relation on $X_i \times \Xi_i$.*

Remark: It is known that SSCD is sufficient (and, in fact, also necessary) for an agent's optimal action to be increasing with the parameter ξ_i on any *arbitrary* constraint set drawn from X_i (see Edlin and Shannon (1998)). It follows that when $\mathcal{O}_i$ is monotone-rationalizable, we can find a preference that both explains the data and guarantees that the optimal choices based on this preference will be increasing with the parameter, on any (not necessarily interval) constraint set.

Since every SSCD preference is also an SID preference, (c) implies (a), and we have just shown that (a) implies (b). It remains for us to show that (b) implies (c). Our proof involves first working out the (incomplete) revealed preference relations on $X_i \times \Xi_i$ that *must* be satisfied by any SID preference that rationalizes the data and then explicitly constructing a rationalizing preference on $X_i \times \Xi_i$ that completes that incomplete relation and obeys SSCD.

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

Given $\mathcal{O}_i$, the *single crossing extension* of the indirect revealed preference relation $\succsim_i^{RT}$ is another binary relation $\succ_i^{RTS}$ defined in the following way: (i) for $x_i'' > x_i'$, $(x_i'', \xi_i) \succ_i^{RTS} (x_i', \xi_i)$ if there is $\xi_i' < \xi_i$ such that $(x_i'', \xi_i') \succsim_i^{RT} (x_i', \xi_i')$ and (ii) for $x_i'' < x_i'$, $(x_i'', \xi_i) \succ_i^{RTS} (x_i', \xi_i)$, if there is $\xi_i'' > \xi_i$ such that $(x_i'', \xi_i'') \succsim_i^{RT} (x_i', \xi_i'')$.

Let $\succsim_i^{RTS}$ be the relation given by $\succsim_i^{RTS} = \succsim_i^{RT} \cup \succ_i^{RTS}$.[12] It follows immediately from its definition that $\succsim_i^{RTS}$ *has strict single crossing differences*, in the following sense: if $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$ *or* $x_i'' < x_i'$ and $\xi_i'' < \xi_i'$, then

$$(x_i'', \xi_i') \succsim_i^{RTS} (x_i', \xi_i') \Longrightarrow (x_i'', \xi_i'') \succ_i^{RTS} (x_i', \xi_i''). \tag{8}$$

In addition, let $\succsim_i^{RTST}$ be the transitive closure of $\succsim_i^{RTS}$, i.e., $(x_i'', \xi_i) \succsim_i^{RTST} (x_i', \xi_i)$ if there exists a sequence $z_i^1, z_i^2, ..., z_i^k$ such that

$$(x_i'', \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \succsim_i^{RTS} ... \succsim_i^{RTS} (z_i^k, \xi_i) \succsim_i^{RTS} (x_i', \xi_i). \tag{9}$$

If we can find at least one strict relation $\succ_i^{RTS}$ in the sequence (9), then, we let $(x_i'', \xi_i) \succ_i^{RTST} (x_i', \xi_i)$.[13] The relevance of the binary relations $\succsim_i^{RTST}$ and $\succ_i^{RTST}$ flows from the following result, which says that any rationalizing preference for agent i must respect the ranking implied by them.

Proposition 1. *Suppose that the preference $\succsim_i$ obeys SID and rationalizes $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$. Then $\succsim_i$* extends *$\succsim_i^{RTST}$* and *$\succ_i^{RTST}$ in the following sense:*

$$(x_i'', \xi_i) \succsim_i^{RTST} (\succ_i^{RTST}) (x_i', \xi_i) \Longrightarrow (x_i'', \xi_i) \succsim_i (\succ_i) (x_i', \xi_i) \tag{10}$$

Proof. Without loss of generality, we may let $x_i'' > x_i'$. Since $\succsim_i$ is transitive, it is clear that we need only show that $(x_i'', \xi_i) \succsim_i (\succ_i) (x_i', \xi_i)$ whenever $(x_i'', \xi_i) \succsim_i^{RTS} (\succ_i^{RTS}) (x_i', \xi_i)$. If $(x_i'', \xi_i) \succsim_i^{RTS} (\succ_i^{RTS}) (x_i', \xi_i)$ then there exists some $\xi_i' \leqslant (<) \, \xi_i$ such that $(x_i'', \xi_i') \succsim_i^{RT} (x_i', \xi_i')$. By the interval property of $\succsim_i^{RT}$, we obtain $(x_i'', \xi_i') \succsim_i^{RT} (x_i, \xi_i')$ for all $x_i \in [x_i', x_i'']$. Since $\succsim_i$ rationalizes $\mathcal{O}_i$, we also have $(x_i'', \xi_i') \succsim_i (x_i, \xi_i')$ for all $x_i \in [x_i', x_i'']$. By SID of $\succsim_i$, we obtain $(x_i'', \xi_i) \succsim_i (\succ_i) (x_i', \xi_i)$ for $\xi_i' \leqslant (<) \, \xi_i$. $\square$

[12]Note that $\succ_i^{RTS}$ is *not* the asymmetric part of $\succsim_i^{RTS}$.

[13]Once again, note that $\succ_i^{RTST}$ is *not* the asymmetric part of $\succsim_i^{RTST}$.

At this point, it is reasonable to ask if we could go beyond the revealed preference relations we have constructed and consider the single crossing extension of $\succsim_i^{RTST}$, the transitive closure of that extension, and so on. The answer to that is 'no' because, as we shall show in Lemma 2, $\succsim_i^{RTST}$ obeys SSCD when $\mathcal{O}_i$ obeys ARC, so it does not admit a nontrivial single crossing extension. By Proposition 1, it is clear that, in order for $\mathcal{O}_i$ to be monotone rationalizable, the binary relation $\succsim_i^{RTST}$ must have the following property: for any (x_i', ξ_i) and (x_i'', ξ_i) in $X_i \times \Xi_i$,

$$(x_i', \xi_i) \succsim_i^{RTST} (x_i'', \xi_i) \Longrightarrow (x_i'', \xi_i) \not\succ_i^{RTST} (x_i', \xi_i). \tag{11}$$

If not, we obtain simultaneously, $(x_i', \xi_i) \succsim_i (x_i'', \xi_i)$ and $(x_i'', \xi_i) \succ_i (x_i', \xi_i)$, which is impossible. The following lemma says that these two properties of $\succsim_i^{RTST}$ hold whenever $\mathcal{O}_i$ obeys ARC.

Lemma 2. *Suppose that $\mathcal{O}_i$ obeys ARC. Then $\succsim_i^{RTST}$ obeys SSCD and property (11).*

Since $\succsim_i^R \subseteq \succsim_i^{RTST}$, it is clear that Proposition 1 has a converse: if there is a regular and SID preference $\succsim_i$ on $X_i \times \Xi_i$ that obeys (10), then this preference rationalizes $\mathcal{O}_i$. This observation, together with Lemma 2, suggest that a reasonable way of constructing a rationalizing preference is to *begin with* $\succsim_i^{RTST}$ and $\succ_i^{RTST}$ and then complete these incomplete relations in a way that gives a preference with the required properties, which is precisely the approach we take. Define the binary relation $\succsim_i^*$ on $X_i \times \Xi_i$ in the following manner:

$$\begin{aligned} (x_i'', \xi_i) \succsim_i^* (x_i', \xi_i) \text{ if } & (x_i'', \xi_i) \succsim_i^{RTST} (x_i', \xi_i) \\ & \text{or } (x_i'', \xi_i) \parallel_i^{RTST} (x_i', \xi_i) \text{ and } x_i' \geqslant x_i'', \end{aligned} \tag{12}$$

where $(x_i'', \xi_i) \parallel_i^{RTST} (x_i', \xi_i)$ means neither $(x_i'', \xi_i) \succsim_i^{RTST} (x_i', \xi_i)$ nor $(x_i', \xi_i) \succsim_i^{RTST} (x_i'', \xi_i)$. The following result (which we prove in the Appendix with the help of Lemma 2) completes our argument that (b) implies (c) in Theorem 1.

Lemma 3. *Suppose that $\mathcal{O}_i$ obeys ARC. The binary relation $\succsim_i^*$ is an SSCD preference that rationalizes $\mathcal{O}_i$. On every set $K \subset X_i$ that is compact in $\mathbb{R}$ and for every $\xi_i \in \Xi_i$,* $\mathrm{BR}_i(\xi_i, K, \succsim_i^*)$ *is nonempty and finite; in particular, $\succsim_i^*$ is a regular preference.*

In the case where X_i is a finite set it is obvious that $\succsim_i^*$ has a utility representation, in the sense that there is a real-valued function $u_i(\cdot, \xi_i)$ (defined on X_i and parameterized by ξ_i) such that $u_i(x_i'', \xi_i) \geqslant u_i(x_i', \xi_i)$ if and only if $(x_i'', \xi_i) \succsim_i^* (x_i', \xi_i)$. Though somewhat less obvious, it turns out that this is also true in the case where X_i is a closed interval of $\mathbb{R}$.

Proposition 2. *Suppose that $\mathcal{O}_i$ obeys ARC and X_i is a closed interval of $\mathbb{R}$. Then the preference $\succsim_i^*$ admits a utility representation.*

3.2 ARC, interval constraint sets, and SSCD

ARC is an easy-to-understand property and it is reminiscent of other conditions in revealed preference theory such as the generalized axiom of revealed preference (GARP), which features in Afriat's Theorem, or the congruence axiom in Richter's Theorem. The proof of Theorem 1 we outlined also seems broadly familiar, in the sense that the basic revealed preference relations are extended in a very natural way and then it is completed in a way that satisfies SSCD. So familiar indeed, that someone with intuition developed from the knowledge of those classic results could be forgiven for thinking that Theorem 1 is obvious. But there is more than what meets the eye in Theorem 1 and intuition can be misleading; indeed, any *correct* intuition will have to distinguish between arbitrary constraint sets and interval constraint sets because the result is not true in the former case, as the following example shows.

Example 2. Let $X_i = \{u_i, v_i, w_i\}$ with $u_i < v_i < w_i$, and let $A_i^1 = \{u_i, w_i\}$, $A_i^2 = \{u_i, v_i\}$, and $A_i^3 = \{v_i, w_i\}$. Note that A_i^1 is not an interval of X_i. Suppose that $\xi_i^1 < \xi_i^2 < \xi_i^3$, and that $x_i^1 = w_i$, $x_i^2 = u_i$, and $x_i^3 = v_i$. Then $(w_i, \xi_i^1) \succsim_i^R (u_i, \xi_i^1)$, $(u_i, \xi_i^2) \succsim_i^R (v_i, \xi_i^2)$, and $(v_i, \xi_i^3) \succsim_i^R (w_i, \xi_i^3)$. The indirect revealed preference relation $\succsim_i^{RT}$ is equal to the direct revealed preference relation $\succsim_i^R$ in this example and, clearly, this set of three observations obeys ARC. However, it cannot be rationalized by an SSCD preference. Suppose, instead, that an SSCD preference $\succsim_i$ rationalizes the data. Then, it must hold that $(w_i, \xi_i^1) \succsim_i (u_i, \xi_i^1)$ and, by SSCD, $(w_i, \xi_i^2) \succ_i (u_i, \xi_i^2)$. In addition, we have $(u_i, \xi_i^2) \succsim_i (v_i, \xi_i^2)$ and so $(w_i, \xi_i^2) \succ_i (v_i, \xi_i^2)$. Since $\succsim_i$ obeys SSCD, we obtain $(w_i, \xi_i^3) \succ_i (v_i, \xi_i^3)$, which contradicts the direct revealed preference $(v_i, \xi_i^3) \succsim_i (w_i, \xi_i^3)$. Notice that even though ARC holds in this data set, (11) is violated; indeed, we have $(w_i, \xi_i^1) \succ_i^{RTS} (u_i, \xi_i^1)$, $(u_i, \xi_i^2) \succsim_i^{RTS} (v_i, \xi_i^2)$, and $(v_i, \xi_i^3) \succ_i^{RTS} (w_i, \xi_i^3)$.

The proof of Theorem 1 relies on Lemma 2 and the proof of that lemma (which says that

$\succsim_i^{RTS}$ admits no cycles, which is (11), and that $\succsim_i^{RTST}$ obeys SSCD) relies on the assumption that the constraint sets are intervals. In Example 2, even though ARC holds, the property (11) is violated; indeed, we have $(w_i, \xi_i^1) \succ_i^{RTS} (u_i, \xi_i^1)$, $(u_i, \xi_i^2) \succsim_i^{RTS} (v_i, \xi_i^2)$, and $(v_i, \xi_i^3) \succ_i^{RTS} (w_i, \xi_i^3)$.

At this point, a well-developed intuition may give a further suggestion: that a data set could be rationalized by an SSCD preference so long as none of the extended revealed preferences admit cycles. In other words, if we successively construct the relations $\succ_i^{RTSTST}$ and $\succsim_i^{RTSTST}$, and so forth, the process must terminate, i.e., there is a finite n such that $\succ_i^{RT(ST)^n S} = \succ_i^{RT(ST)^n}$ and $\succsim_i^{RT(ST)^n S} = \succsim_i^{RT(ST)^n}$. Since a necessary condition for $\mathcal{O}_i = \{(x_i^t, \xi_i^t, B_i^t)\}_{t \in \mathcal{T}}$ (where B_i^t are not necessarily intervals) to be rationalizable by an SSCD preference is that (11) holds for $\succsim_i^{RT(ST)^n}$ and $\succ_i^{RT(ST)^n}$ (taking the place of $\succsim_i^{RTST}$ and $\succ_i^{RTST}$), one may expect that this property is also sufficient for rationalizability. However, a counterexample of Kukushkin, Quah and Shirai (2016) shows that this conjecture is also false. In other words, it is *not* inevitable that the absence of cycles guarantees the existence of a completion obeying SSCD; the fact that that holds in Theorem 1 relies on the constraint sets being intervals.

3.3 Out-of-sample predictions of best responses

Suppose an observer collects a data set $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ that is monotone rationalizable, and then, maintaining that hypothesis, asks the following question: what do the observations in $\mathcal{O}_i$ say about the *set of possible choices* of agent i in some treatment $(\xi_i^0, A_i^0) \in \Xi_i \times \mathcal{A}_i$?[14] If $\mathcal{O}_i$ obeys ARC, then we know that the set of all SID preferences that rationalize $\mathcal{O}_i$, call it $\mathcal{P}_i^\star$, is nonempty. For each $\succsim_i \in \mathcal{P}_i^\star$, the set of best responses at (ξ_i^0, A_i^0) is $\mathrm{BR}_i(\xi_i^0, A_i^0, \succsim_i)$, and hence the set of *possible best responses* at (ξ_i^0, A_i^0) is given by

$$\mathrm{PR}_i(\xi_i^0, A_i^0) = \bigcup_{\succsim_i \in \mathcal{P}_i^\star} \mathrm{BR}_i\left(\xi_i^0, A_i^0, \succsim_i\right). \tag{13}$$

It follows from Theorem 1 that,

$$\mathrm{PR}_i(\xi_i^0, A_i^0) = \left\{\tilde{x}_i \in A_i^0 : \overline{\mathcal{O}_i} = \mathcal{O}_i \cup \left\{\left(\tilde{x}_i, \xi_i^0, A_i^0\right)\right\} \text{ obeys ARC}\right\}, \tag{14}$$

[14]The environment (ξ_i^0, A_i^0) may – or may not – be distinct from the ones already observed in the data set; the latter can still be an interesting question since optimal choices are not unique.

where $\overline{\mathcal{O}_i}$ is $\mathcal{O}_i$ augmented by the (fictitious) observation $\{(\tilde{x}_i, \xi_i^0, A_i^0)\}$. The following proposition shows that $\mathrm{PR}_i(\xi_i^0, A_i^0)$ coincides with the undominated elements with respect to $\succsim_i^{RTST}$.

Proposition 3. *Suppose that $\mathcal{O}_i$ obeys ARC. For any $\xi^0 \in \Xi_i$, it holds that*

$$\mathrm{PR}_i(\xi_i^0, A_i^0) = \{x_i \in A_i^0 : \nexists\, \hat{x}_i \in A_i^0 \;\; \textit{such that} \;\; (\hat{x}_i, \xi_i^0) \succ_i^{RTST} (x_i, \xi_i^0)\}. \tag{15}$$

Proof. It follows from (14) that (15) holds provided we can show the following: $\overline{\mathcal{O}}_i = \mathcal{O}_i \cup \{(\tilde{x}_i, \xi_i^0, A_i^0)\}$ violates ARC if and only if there is $\hat{x}_i \in A_i^0$ such that $(\hat{x}_i, \xi_i^0) \succ_i^{RTST} (\tilde{x}_i, \xi_i^0)$. Let $\succsim_i^{\overline{R}}$, $\succsim_i^{\overline{RT}}$, $\succsim_i^{\overline{RTS}}$, and $\succsim_i^{\overline{RTST}}$ be the revealed preference relations derived from $\overline{\mathcal{O}}_i = \mathcal{O}_i \cup \{(\tilde{x}_i, \xi_i^0, A_i^0)\}$; by definition, these must contain the analogous revealed preference relations of $\mathcal{O}_i$, i.e., $\succsim_i^{R}$, $\succsim_i^{RT}$, $\succsim_i^{RTS}$, and $\succsim_i^{RTST}$. Suppose there is $\hat{x}_i \in A_i^0$ such that $(\hat{x}_i, \xi_i^0) \succ_i^{RTST} (\tilde{x}_i, \xi_i^0)$ and so $(\hat{x}_i, \xi_i^0) \succ_i^{\overline{RTST}} (\tilde{x}_i, \xi_i^0)$. On the other hand, since $\hat{x}_i \in A_i^0$, we have $(\tilde{x}_i, \xi_i^0) \succsim_i^{\overline{R}} (\hat{x}_i, \xi_i^0)$. This is a violation of the property (11) and, by Lemma 2, $\overline{\mathcal{O}}_i$ violates ARC. Conversely, suppose that $\overline{\mathcal{O}}_i = \mathcal{O}_i \cup \{(\tilde{x}_i, \xi_i^0, A_i^0)\}$ violates ARC. Since $\mathcal{O}_i$ obeys ARC, this violation can only occur in two ways: there is $\hat{x}_i \in X_i$ such that $(\tilde{x}_i, \xi_i^0) \succsim_i^{\overline{RT}} (\hat{x}_i, \xi_i^0)$ and $(\hat{x}_i, \bar{\xi}_i) \succsim_i^{RT} (\tilde{x}_i, \bar{\xi}_i)$ with either (1) $\hat{x}_i < \tilde{x}_i$ and $\bar{\xi}_i > \xi_i^0$ or (2) $\hat{x}_i > \tilde{x}_i$ and $\bar{\xi}_i < \xi_i^0$. We need to show that $\tilde{x}_i$ is dominated (with respect to $\succ_i^{RTST}$) by some element in A_i^0. In either cases (1) or (2), since $(\hat{x}_i, \bar{\xi}_i) \succsim_i^{RT} (\tilde{x}_i, \bar{\xi}_i)$, we obtain $(\hat{x}_i, \xi_i^0) \succ_i^{RTS} (\tilde{x}_i, \xi_i^0)$. If $\hat{x}_i \in A_i^0$, we are done. If $\hat{x}_i \notin A_i^0$ then, given that $(\tilde{x}_i, \xi_i^0) \succsim_i^{\overline{RT}} (\hat{x}_i, \xi_i^0)$, there exists $\bar{x}_i \in A_i^0$ such that $(\bar{x}_i, \xi_i^0) \succsim_i^{RT} (\hat{x}_i, \xi_i^0)$. Thus $(\bar{x}_i, \xi_i^0) \succ_i^{RTST} (\tilde{x}_i, \xi_i^0)$. □

It is very convenient to have Proposition 3 because computing $\succsim_i^{RTST}$ is straightforward and thus it is also straightforward to obtain the set of possible responses at a given treatment.

Example 3. Consider two observations as depicted in Figure 1, where A_i^1 and A_i^2 are the feasible sets of agent i at observations 1 and 2 respectively, while ξ_i^1 and ξ_i^2 are the parameter values at each observation. Let A_i^0 be the blue segment in the figure. It is easy to check that observations 1 and 2 obey ARC, and that the set of possible best responses, $\mathrm{PR}_i(\xi_i^0, A_i^0)$, is the set indicated in the figure. Notice that this set is not closed since $x_i^* \notin \mathrm{PR}_i(\xi_i^0, A_i^0)$. Indeed, $(x_i^2, \xi_i^1) \succ_i^{RTS} (x_i^*, \xi_i^1)$ since $(x_i^2, \xi_i^2) \succsim_i^{R} (x_i^*, \xi_i^2)$. Furthermore, $(x_i^1, \xi_i^1) \succsim_i^{R} (x_i^2, \xi_i^1)$ and so we obtain $(x_i^1, \xi_i^1) \succ_i^{RTST} (x_i^*, \xi_i^1)$.

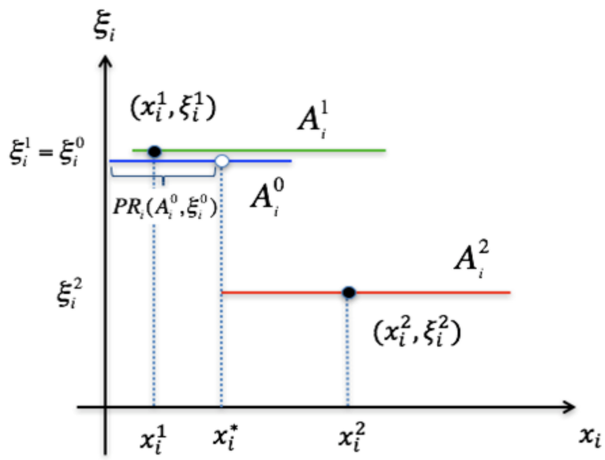

Figure 1: $\mathcal{E}(A^0)$ in Example 3

4 Revealed strategic complementarity

Let $\mathbb{G} = \{\mathcal{G}(y, A)\}_{(y,A)\in Y\times\mathcal{A}}$ be a collection of games, as defined in the Section 2.2. We consider an observer who has a set of observations drawn from this collection. Each observation consists of a triple (x^t, y^t, A^t), where x^t is the action profile observed at the treatment $(y^t, A^t) \in Y \times \mathcal{A}$. The set of observations is finite and is denoted by $\mathcal{O} = \{(x^t, y^t, A^t)\}_{t\in\mathcal{T}}$, where $\mathcal{T} = \{1, 2, ..., T\}$.

Definition 2. *A data set $\mathcal{O} = \{(x^t, y^t, A^t)\}_{t\in\mathcal{T}}$ is* consistent with strategic complementarity *(or* SC-rationalizable*) if there exists a profile of regular and SID preferences $(\succsim_i)_{i\in N}$ such that each observation constitutes a Nash equilibrium, i.e., for every $t \in \mathcal{T}$ and $i \in N$, we have $(x_i^t, x_{-i}^t, y_i^t) \succsim_i (x_i, x_{-i}^t, y^t)$ for all $x_i \in A_i^t$.*

The motivation for this definition is clear. If $\mathcal{O}$ is SC-rationalizable then we have found a profile of preferences $(\succsim_i)_{i\in N}$ such that (i) x^t is a Nash equilibrium of $\mathcal{G}(y^t, A^t)$ and (ii) the family of games $\mathbb{G} = \{\mathcal{G}(y, A)\}_{(y,A)\in Y\times\mathcal{A}}$, where $\mathcal{G}(y, A) = [(y_i)_{i\in N}, (A_i)_{i\in N}, (\succsim_i)_{i\in N}]$ exhibits strategic complementarity (in the sense defined in Section 2.2).[15]

For each agent i, we can define the *agent data set $\mathcal{O}_i = \{(x_i^t, \xi_i^t, A_i^t)\}_{t=1}^T$ induced by $\mathcal{O}$*, where $\xi_i^t = (x_{-i}^t, y_i^t)$. We say that $\mathcal{O} = \{(x^t, A^t, y^t)\}_{t\in\mathcal{T}}$ obeys ARC if $\mathcal{O}_i$ obeys ARC, for every agent

[15] As we pointed out in Section 2.2, all games with strategic complementarity have pure strategy Nash equilibria, so the hypothesis that we observe these equilibria is internally consistent.

i. It is clear that $\mathcal{O}$ is SC-rationalizable if and only if $\mathcal{O}_i$ is monotone-rationalizable for every agent i. This leads to the following result, which is an immediate consequence of Theorem 1 and provides us with an easy-to-implement test of SC-rationalizability.

THEOREM 2. *The data set $\mathcal{O} = \{(x^t, y^t, A^t)\}_{t\in\mathcal{T}}$ is SC-rationalizable if and only if it obeys ARC.*

We turn now to the issue of out-of-sample equilibrium predictions. Given an SC-rationalizable data set $\mathcal{O} = \{(x^t, A^t, y^t)\}_{t\in\mathcal{T}}$, the agent data set $\mathcal{O}_i$ obeys ARC and so the set of regular and SID preferences that rationalize $\mathcal{O}_i$, i.e., $\mathcal{P}_i^\star$, is nonempty. Each observed strategy profile x^t in $\mathcal{O}$ is supported as a Nash equilibrium by any preference profile $(\succsim_i)_{i\in N}$ in $\mathcal{P}^\star = \times_{i\in N}\mathcal{P}_i^\star$. For each $(\succsim_i)_{i\in N} \in \mathcal{P}^\star$, we know from Theorem B that the set of pure strategy Nash equilibria at another game $\mathcal{G}(y^0, A^0)$, which we shall denote by $E(y^0, A^0, (\succsim_i)_{i\in N})$, is nonempty and hence

$$\mathcal{E}(y^0, A^0) = \bigcup_{(\succsim_i)_{i\in N}\in\mathcal{P}^\star} E\left(y^0, A^0, (\succsim_i)_{i\in N}\right)$$

is also nonempty. $\mathcal{E}(y^0, A^0)$ is the set of *possible Nash equilibria of the game* $\mathcal{G}(y^0, A^0)$. This gives rise to two related questions that we shall answer in this section: [1] how can we compute $\mathcal{E}(y^0, A^0)$ from the data? and [2] what can we say about the structure of $\mathcal{E}(y^0, A^0)$?

4.1 Computable characterization of $\mathcal{E}(y^0, A^0)$

$\mathrm{PR}_i(\xi_i, A_i^0)$ denotes the possible best responses of player i in A_i^0 to $\xi_i = (x_{-i}, y_i^0)$ (see (13)); given this, we define the *joint possible response correspondence* $\mathrm{PR}(\cdot, y^0, A^0) : A^0 \rightrightarrows A^0$ by

$$\mathrm{PR}(x, y^0, A^0) = \mathrm{PR}_1(x_{-1}, y_1^0, A_1^0) \times \mathrm{PR}_2(x_{-2}, y_2^0, A_2^0) \ldots \times \mathrm{PR}_n(x_{-n}, y_n^0, A_n^0). \tag{16}$$

The crucial observation to make in computing $\mathcal{E}(y^0, A^0)$ is that just as the set of Nash equilibria in a game coincides with the fixed points of its joint best response correspondence, so the set of possible Nash equilibria, $\mathcal{E}(y^0, A^0)$, coincides with the fixed points of $\mathrm{PR}(\cdot, y^0, A^0)$. Equivalently, one could think of $\mathcal{E}(y^0, A^0)$ as the intersection of the graphs of each player's possible response correspondence, i.e., $\mathcal{E}(y^0, A^0) = \bigcap_{i\in N} \Gamma_i(y^0, A^0)$, where

$$\Gamma_i(y^0, A^0) = \{(x_i, x_{-i}) \in A^0 : x_i \in \mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)\}. \tag{17}$$

Therefore, the computation of $\mathcal{E}(A^0, y^0)$ hinges on the computation of $\mathrm{PR}_i(\cdot, y_i^0, A_i^0) : A_{-i}^0 \rightrightarrows A_i^0$. Two features of this correspondence together make it possible for us to compute it explicitly.

First, we know from Proposition 3 that, for any x_{-i}, the set $\mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ coincides exactly with those elements in A_i^0 that are not dominated (with respect to $\succ_i^{RTST}$) by another element in A_i^0. Since the data set is finite, $\mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ can be constructed after a finite number of steps and, in fact, one could also show that it consists of a finite number of intervals.

Second, the domain of the correspondence $\mathrm{PR}_i(\cdot, y_i^0, A_i^0)$, which is $\times_{j \neq i} A_j^0$, can be partitioned into a finite number of regions such that the correspondence is constant within each region. Specifically, for $j \neq i$, let $A_j^{\mathcal{T}} = \{x_j \in X_j : (x_j, x_{-j}) = x^t \text{ for some } x_{-j} \text{ and } t \in \mathcal{T}\}$. We denote by $\mathcal{I}_j$ the collection consisting of subsets of A_j^0 of the following two types: the singleton sets $\{\tilde{x}\}$, where $\tilde{x}$ is in the set $\underline{A}_j^0 = \left(A_j^{\mathcal{T}} \cap A_j^0\right) \bigcup \max A_j^0 \bigcup \min A_j^0$ and the interval sets $\{x_j \in A_j^0 : \tilde{a} < x_j < \tilde{b}\}$, where $\tilde{a} \in \underline{A}_j^0$ and $\tilde{b}$ is the element in $\underline{A}_j^0$ immediately above $\tilde{a}$; thus $\mathcal{I}_j$ constitutes a finite partition of A_j^0. This in turn means that $\mathcal{H}_i$ is a finite partition of $\times_{j \neq i} A_j^0$, where $\mathcal{H}_i$ consists of hyper-rectangles

$$I_1 \times I_2 \times ... \times I_{i-1} \times I_{i+1} \times ... \times I_N$$

where $I_j \in \mathcal{I}_j$, for $j \neq i$. One could show that for any $H_i \in \mathcal{H}_i$, the following property holds:

$$x'_{-i},\ x''_{-i} \in H_i \Longrightarrow \mathrm{PR}_i(x'_{-i}, y_i^0, A_i^0) = \mathrm{PR}_i(x''_{-i}, y_i^0, A_i^0). \tag{18}$$

Therefore, to compute the correspondence $\mathrm{PR}_i(\cdot, y_i^0, A_i^0)$ we need only find its value via (15) for a typical element within each hyper-rectangle H_i in the *finite* collection $\mathcal{H}_i$.

It follows from these two observations that the graph of player i's possible response correspondence (as defined by (17)) is also given by

$$\Gamma_i(y^0, A^0) = \{(x_i, x_{-i}) \in A^0 : \nexists\, \hat{x}_i \in A_i^0 \text{ such that } (\hat{x}_i, x_{-i}, y_i^0) \succ_i^{RTST} (x_i, x_{-i}, y_i^0)\} \tag{19}$$

and can be explicitly constructed. Furthermore, because $\mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ consists of a finite union of intervals of A_i^0, $\Gamma_i(y^0, A^0)$ is a finite union of hyper-rectangles in A^0. The following theorem, which we prove in the Appendix, summarizes these observations.

Theorem 3. *Suppose a data set $\mathcal{O} = \{(x^t, y^t, A^t)\}_{t=1}^T$ obeys ARC and let $(y^0, A^0) \in Y \times \mathcal{A}$.*

(i) $\text{PR}_i(\cdot, y_i^0, A_i^0)$ *obeys (15) and (18) and, for any $x_{-i} \in \times_{j \neq i} A_j^0$, $\text{PR}_i(x_{-i}, y_i^0, A_i^0)$ consists of a finite union of intervals of A_i^0.*

(ii) *The graph of $\text{PR}_i(\cdot, y_i^0, A_i^0)$, $\Gamma_i(y^0, A^0)$, is a finite union of hyper-rectangles in A^0. Consequently, the set of possible Nash equilibria, $\mathcal{E}(y^0, A^0) = \bigcap_{i \in N} \Gamma_i(y^0, A^0)$, is also a finite union of hyper-rectangles in A^0.*

Example 4. Figure 2(a) depicts two observations, $\{(x^1, A^1)$ and $(x^2, A^2)\}$, drawn from two games involving the same two players. This data set obeys ARC and we would like to compute $\mathcal{E}(A^0)$, where $A_i^0 = A_i^1 \cup A_i^2$ (for $i = 1, 2$). First, we claim that the unshaded area in Figure 2(b) cannot be contained in $\Gamma_1(A^0)$. Indeed, consider the point $x' = (x_1', x_2')$ in the unshaded area, at which $x_1' < x_1^1$, $x_2' > x_2^1$, and $x_1' \in A_1^1$. Therefore, $(x_1^1, x_2^1) \gtrsim_1^R (x_1', x_2^1)$ and so $(x_1^1, x_2^1) \gtrsim_1^{RT} (x_1', x_2^1)$. Since $x_2' > x_2^1$, $(x_1^1, x_2^1) \succ_1^{RTS} (x_1', x_2^1)$, which means that $(x_1', x_2') \notin \Gamma_1(A^0)$. Using (19), it is easy to check that $\Gamma_1(A^0)$ corresponds precisely to the shaded area in Figure 2(b). Similarly, $\Gamma_2(A^0)$ consists of the shaded area in Figure 2(c). The common shaded area, as depicted with the darker shade in Figure 2(d), represents $\mathcal{E}(A^0) = \Gamma_1(A^0) \cap \Gamma_2(A^0)$. Note that the dashed lines are *excluded* from $\mathcal{E}(A^0)$, so this set is not closed.

4.2 The structure of $\mathcal{E}(y^0, A^0)$

As we have pointed out in Section 2.2, the set of pure strategy Nash equilibria in a game with strategic complementarity admits a largest and smallest Nash equilibrium, both of which exhibit monotone comparative statics with respect to exogenous parameters. In this subsection, we show that these properties are largely inherited by the set of predicted pure strategy Nash equilibria $\mathcal{E}(y^0, A^0)$. The next result (which we prove in the Appendix) lists the main structural properties of $\mathcal{E}(y^0, A^0)$; we have consciously presented them in a way that is analogous to Theorem B.

Theorem 4. *Suppose a data set $\mathcal{O} = \{(x^t, y^t, A^t)\}_{t \in \mathcal{T}}$ obeys ARC and let $(y^0, A^0) \in Y \times \mathcal{A}$. Then $\mathcal{E}(y^0, A^0)$, the set of possible pure strategy Nash equilibria of the game $\mathcal{G}(y^0, A^0)$, is nonempty. Its closure admits a largest and a smallest element, both of which are increasing in $y^0 \in Y$.*

Since A^0 is a subcomplete sublattice of $(\mathbb{R}^n, \geqslant)$, *any* set in A^0 will have a supremum and an infimum in A^0. Therefore, the principal claim in Theorem 4 is that the supremum and infimum

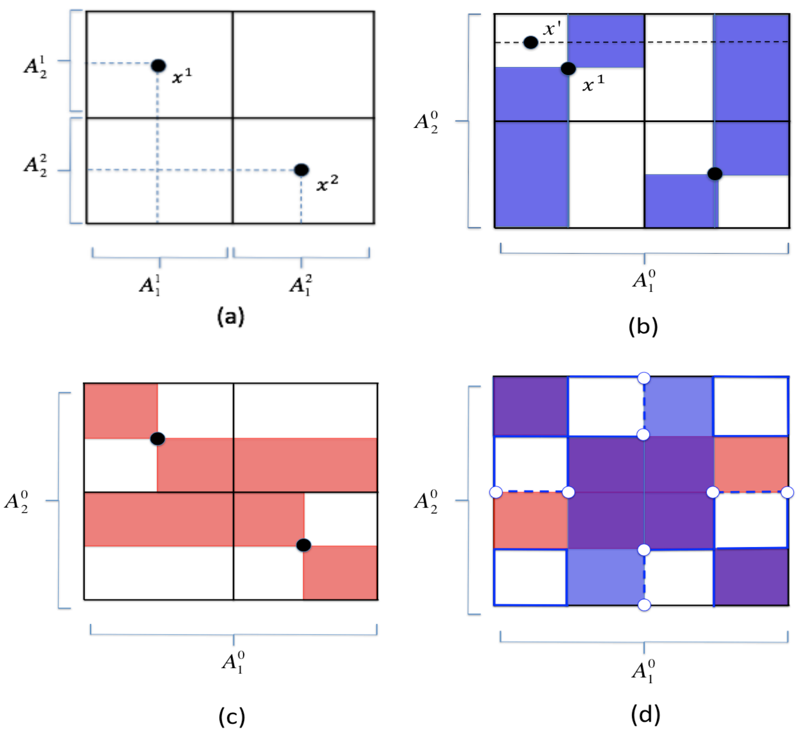

Figure 2: $\mathcal{E}(A^0)$ in Example 4

of the closure of $\mathcal{E}(y^0, A^0)$ are *contained in that set* (and thus arbitrarily close to elements of $\mathcal{E}(y^0, A^0)$): to all intents and purposes, we could speak of a largest and a smallest possible Nash equilibrium. Note that the analogous statement in Theorem B is stronger since it says that the set of pure strategy Nash equilibria *contains* a largest and a smallest element.

Theorem 4 applies also to single agent choice data as a special case. In that context, it says that the supremum and infimum of the set of possible responses both increase with the parameter. Note also that Example 3 in Section 3.3 gives a case where the possible response set *does not* contain its supremum, so the conclusion in Theorem 4 cannot be made as strong as the conclusion in Theorem B. Standard proofs of Theorem B rely on the monotone and closed-valued properties of the best response correspondence. The proof of Theorem 4 makes use of the monotone property of the possible response correspondence, but the proof is more involved than that of Theorem B because this correspondence is not closed-valued.

In the special but important case where A^0 is finite, every subset of A^0 is closed and so it follows immediately from Theorem 4 that $\mathcal{E}(y^0, A^0)$ is a closed set with a largest and smallest element. The conclusion of Theorem 4 may also be strengthened in the case where the feasible action set of every agent is unchanged throughout the observations, i.e., $A^t = A^0 \in \mathcal{A}$ for all $t \in \mathcal{T}$. By (14), a necessary and sufficient condition for $\tilde{x}_i \in A_i^0$ to be contained in $\mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ is that $\overline{\mathcal{O}_i} = \mathcal{O}_i \cup \{(\tilde{x}_i, (x_{-i}, y_i^0), A_i^0)\}$ obeys ARC. If $A^0 = A^t$ for all $t \in \mathcal{T}$, then it is straightforward to check that this is equivalent to $\tilde{x}_i$ having the following property:

$$\text{for all } t \in \mathcal{T},\ \tilde{x}_i \geqslant x_i^t \text{ if } (x_{-i}, y_i^0) > \xi_i^t \text{ and } \tilde{x}_i \leqslant x_i^t \text{ if } (x_{-i}, y_i^0) < \xi_i^t. \tag{20}$$

It follows that $\mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ must be a closed interval in A_i^0 and (by Theorem 3) its graph $\Gamma_i(y^0, A^0)$ is a finite union of *closed* hyper-rectangles. Therefore, $\mathcal{E}(y^0, A^0) = \bigcap_{i \in N} \Gamma_i(y^0, A^0)$ is also closed and, by Theorem 4, it must contain its largest and smallest element.

5 Testing for complementarity with cross sectional data

So far in this paper we have assumed that the observer has access to panel data that gives the actions of the same agent (or, in the case of a game, the same group of agents) across different treatments. Oftentimes, data of this type is not available; instead, we only observe the actions of different agents, with presumably heterogeneous preferences, subject to different treatments. It is possible to extend our revealed preference analysis to this setting, provided we assume that the distribution of preferences is the same in populations subject to different treatments or, put another way, the assignment of agents or groups to treatments is random.

5.1 Stochastic monotone rationalizability

Suppose we observe a population of agents, whom we shall call population i, choosing actions from a subset of a chain X_i. Throughout this section (and unlike previous sections), we require X_i to be a *finite* chain. Agents choose from feasible sets that are intervals of X_i, according to preferences that are affected by a set of parameters Ξ_i. At each observation t, all agents in population i are subject to the same treatment $(\xi_i^t, A_i^t) \in \Xi_i \times \mathcal{A}_i$, though they may choose different actions because they have different preferences. We assume that the true distribution of actions is observable and given by μ_i^t, where $\mu_i^t(x_i)$ denotes the fraction of agents who choose action x_i; we require $\mu_i^t(x_i) = 0$ for all $x_i \notin A_i^t$. The (cross sectional) data set for population

i is a collection of triples $(\mu_i^t, \xi_i^t, A_i^t)$, i.e., $\mathcal{O}_i = \{(\mu_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$, where $\mathcal{T} = \{1, 2, ..., T\}$. Given $\mathcal{O}_i$, we denote the set of observed treatments by M_i, i.e., $M_i = \{(\xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$. We allow for the *same* treatment to occur at different observations; it is possible that $\mu^t \neq \mu^s$ even though the treatments at observations t and s are identical since we do not require agents to have unique optimal actions.[16] We adopt the convention of allowing the same treatment to be repeated in the set M_i if it occurs at more than one observation.

We call $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^T) \in \times_{t \in \mathcal{T}} A_i^t$ a *monotone rationalizable path* on M_i if the induced 'panel' data set $\{(x_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ is monotone-rationalizable (in the sense defined in Section 3) and denote the set of monotone rationalizable paths by $\mathbf{A}_i$. Since we allow for non-unique optimal choices, two distinct monotone rationalizable paths may be rationalized by the same SID preference.

Definition 3. *A data set $\mathcal{O}_i = \{(\mu_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ is* stochastically monotone rationalizable *if there exists a probability distribution Q_i on $\mathbf{A}_i$, the set of monotone rationalizable paths on M_i, such that $\mu_i^t(x_i) = \sum_{\mathbf{x}_i \in \mathbf{A}_i} Q_i(\mathbf{x}_i) 1(x_i^t = x_i)$ for all $t \in \mathcal{T}$ and $x_i \in X_i$.*

When there is no danger of confusion, we shall simply refer to a data set as *monotone rationalizable* when it is stochastically monotone rationalizable. The definition says that the population i can be decomposed into types corresponding to different monotone rationalizable paths, so that the observed behavior of each type (across treatments) is consistent with maximizing an SID preference; it captures the idea that treatments have been randomly assigned across the entire population by requiring that the distribution of types be the same across treatments.[17] This assumption is similar to the exogeneity restriction in the literature on treatment effects and it can be relaxed using similar approaches; for instance, we could assume instead that assignment to the treatment is random only after we conditioning on some pre-treatment covariates.

[16] If it helps, one could think of the index t itself to be part of the treatment, which may influence an agent's selection rule amongst optimal choices, though it has no impact on the agent's preference or the feasible alternatives, which depend only on the 'real' treatment.

[17] While our definition of monotone rationalizable paths excludes the possibility that some group in the population may decide among non-unique optimal actions stochastically, the large population assumption means that this is without loss of generality. If, say, 10% of the population is indifferent between two optimal actions x' and x'' at some observation t, and decides between them by flipping a fair coin, then it simply means that 5% will belong to a type that chooses x' at t and another 5% to a type that choose x'' at t. A data set drawn from a large population of agents with heterogenous SID preferences who use stochastic selection rules (when there are multiple optimal actions) will still be stochastically monotone rationalizable in the sense defined here.

Theorem 1 tell us that a path $\mathbf{x}_i$ on M_i is monotone rationalizable if and only if it is *ARC-consistent* in the sense that $\{(x_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ obeys ARC. This leads immediately to the following result.

THEOREM 5. *A data set $\mathcal{O}_i = \{(\mu_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ is monotone rationalizable if and only if there exists a probability distribution Q_i on $\mathbf{A}_i^*$, the set of ARC-consistent paths on M_i, such that*

$$\mu_i^t(x_i) = \sum_{\mathbf{x}_i \in \mathbf{A}_i^*} Q_i(\mathbf{x}_i) 1(x_i^t = x_i) \text{ for all } t \in \mathcal{T} \text{ and } x_i \in X_i. \tag{21}$$

This theorem sets out a procedure that could, in principle, allow us to determine the monotone-rationalizability of a stochastic data set: first, we need to list all the ARC-consistent paths, and then we solve the linear equations given by (21).[18] Of course, the implementability of this procedure in practice will depend crucially on the number of observations, treatments, and possible actions, which determines the size of the set of ARC-consistent paths. When (21) has a solution, we also recover a distribution on monotone rationalizable paths that is consistent with the data; the set of such distributions is convex and typically non-unique (so we have *partial* identification).

Consider now the special case where the feasible action sets are fixed across all treatments. Then a path $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^T)$ is ARC-consistent if and only if $x_i^t \geqslant x_i^s$ whenever $\xi_i^t > \xi_i^s$. If $\mathcal{O}_i = \{(\mu_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ is monotone rationalizable, the population can be decomposed into ARC-consistent subpopulations; clearly, this implies that whenever $\xi_i^t > \xi_i^s$, the distribution μ_i^t must first order stochastically dominate μ_i^s (which we shall denote by $\mu_i^t \geqslant_{FOSD} \mu_i^s$). Less obviously, the converse is also true, so that monotonicity with respect to first order stochastic dominance *characterizes* monotone rationalizability when the feasible action set is fixed.

THEOREM 6. *Suppose that $A_i^t = X_i$ for all $t \in \mathcal{T}$. Then $\mathcal{O}_i = \{(\mu_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ is monotone rationalizable if and only if $\mu_i^t \geqslant_{FOSD} \mu_i^s$ whenever $\xi_i^t > \xi_i^s$, for t, $s \in \mathcal{T}$.*

5.2 Stochastic strategic complementarity

The results on stochastic monotone rationalizability have an analog in a game-theoretic framework. In this case, we assume that the population consists of groups of n players, with each

[18] In Manski's (2007) terminology, our model is an example of a *linear* behavioral model.

group choosing an action profile from their joint feasible set $A = \times_{i \in N} A_i$, where A_i is an interval of a finite chain X_i. The player in role i takes an action in A_i; the player's preference over his/her actions is affected by the actions of other players in that group and by some exogenous variable drawn from Y_i. We assume that the observer can distinguish amongst players in different roles in the game and can observe their actions separately; for example, in a population of heterosexual couples, the observer can distinguish between the 'husband' player and the 'wife' player and can observe their actions separately.

At observation t, each group in the population chooses an action profile from the joint feasible action set $A^t \in \mathcal{A}$, with the exogenous parameter being $y^t \in Y = \times_{i \in N} Y_i$; thus all groups in the population are subject to the same treatment $(y^t, A^t) \in Y \times \mathcal{A}$, with observed differences in action profiles stemming from heterogenous preferences amongst players within each group and possibly different equilibrium selection rules. We observe a probability distribution μ^t, with support on A^t, where $\mu^t(x)$ denotes the fraction of groups in which the action profile $x \in X$ is played. Therefore, the data set can be written as $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t \in \mathcal{T}}$. We denote the set of observed treatments by M, i.e., $M = \{(y^t, A^t)\}_{t \in \mathcal{T}}$. The possibility of multiple equilibria means that it is both meaningful and interesting to allow for the same treatment to appear at more than one observation. We have explained this at length in Section 5.1 and we shall not repeat it here. We allow identical treatments to appear more than once in M if they correspond to different observations. We refer to $\mathbf{x} = (x^1, x^2, ..., x^T) \in \times_{t \in \mathcal{T}} A^t$ as an *SC-rationalizable path on M* if the induced 'panel' data set $\{x^t, y^t, A^t\}_{t \in \mathcal{T}}$ is SC-rationalizable (in the sense defined in Section 4). The set of SC-rationalizable paths on M is denoted by $\mathbf{A}$.

DEFINITION 4. *A data set $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t \in \mathcal{T}}$ is* stochastically SC-rationalizable *if there is a probability distribution on $\mathbf{A}$ such that $\mu^t(x) = \sum_{\mathbf{x} \in \mathbf{A}} Q(\mathbf{x}) 1(x^t = x)$ for all $t \in \mathcal{T}$ and $x \in X$.*

Unless there is danger of confusion, we shall simply refer to a data set as *SC-rationalizable* when it is stochastically SC-rationalizable. This definition says that the population can be decomposed into 'group types' corresponding to different SC-rationalizable paths, so that we could interpret the action profile for each group as a Nash equilibrium, with players having SID preferences that are the same across observations; it captures the idea that treatments are randomly assigned across the large population of groups, so that the distribution of types is identical across treatments. As in the single agent case, this exogeneity restriction can be relaxed by controlling for covariates. Note also that our definition allows for groups belonging to different SC-rationalizable paths to have members with the same preferences, because of the possibility of multiple equilibria. Lastly, it is worth emphasizing that the definition *imposes no restrictions on what groups can be formed*; for example, if a data set consists of a population of heterosexual couples, then the set of SC-rationalizable paths $\mathbf{A}$ allows for all possible matchings between different types of male and female players.

By Theorem 2, a path on M is SC-rationalizable if and only if it is *ARC-consistent* in the sense that $\{(x^t, y^t, A^t)\}_{t\in\mathcal{T}}$ obeys ARC. This leads immediately to the following result.

THEOREM 7. *A data set $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t\in\mathcal{T}}$ is SC-rationalizable if and only if there exists a probability distribution Q on $\mathbf{A}^*$, the set of ARC-consistent paths on M, such that*

$$\mu^t(x) = \sum_{\mathbf{x}\in\mathbf{A}^*} Q(\mathbf{x})1(x^t = x) \text{ for all } t \in \mathcal{T} \text{ and } x \in X. \tag{22}$$

5.3 Possible equilibrium distributions

Given an SC-rationalizable data set $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t\in\mathcal{T}}$, we may wish to predict behavior at a given treatment $(y^0, A^0) \in Y \times \mathcal{A}$. The prediction consists of all distributions on X that are compatible with the data set $\mathcal{O}$; this can be obtained by identifying those distributions μ^0 (which must have their support on A^0) such that the augmented stochastic data set $\mathcal{O} \cup \{(\mu^0, y^0, A^0)\}$ is SC-rationalizable. We refer to μ^0 as a *possible (Nash) equilibrium distribution* and denote the set of these distributions by $\text{PED}(y^0, A^0)$.[19] It follows immediately from Theorem 7 that μ^0 is a possible equilibrium distribution if and only if there exists a probability distribution $\widetilde{Q}$ on $\mathbf{A}^{**}$, the set of ARC-consistent paths on the set of environments $M \cup \{(y^0, A^0)\}$, such that for every

[19]We allow for $(y^0, A^0) = (y^{t'}, A^{t'})$ for some observation t' and, indeed, it is instructive to consider this case. Then $\mu^{t'}$ is clearly a possible equilibrium distribution but since multiple equilibria are possible, the set of all such distributions can be strictly larger. In other words, in determining whether or not a distribution is a possible equilibrium distribution, we allow for the possibility that groups in the population with multiple Nash equilibria at the treatment $(y^{t'}, A^{t'})$ could switch to an equilibrium different from the one taken at t'.

$t \in \mathcal{T} \cup \{0\}$ and $x \in X$,

$$\mu^t(x) = \sum_{\mathbf{x} \in \mathbf{A}^{**}} \widetilde{Q}(\mathbf{x}) 1(x^t = x) \text{ for all } t \in \mathcal{T} \cup \{0\} \text{ and } x \in X. \tag{23}$$

All the elements of $\text{PED}(y^0, A^0)$ can be obtained by solving the equations (23). The unknown variables in this system are $\widetilde{Q}(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{A}^{**}$ and $\mu^0(x)$ for all $x \in A^0$, and the equations are linear in these variables, which implies that $\text{PED}(y^0, A^0)$ *is a convex set in the simplex* $\Delta_+^{|X|}$. It follows that the possible fraction of the population playing a particular strategy profile $\tilde{x}$ at (y^0, A^0) will take values in an interval, and its limits can obtained by solving the appropriate linear program.

We can also form set estimates of the fraction of players in a particular role who choose a given action. Formally, a distribution μ^0 on X induces a distribution ν_i^0 on the equilibrium actions of player i; for each $\tilde{x}_i \in X_i$,

$$\nu_i^0(\tilde{x}_i) = \sum_{\{x \in A^0 : x_i = \tilde{x}_i\}} \mu^0(x). \tag{24}$$

The set of possible distributions on player i's equilibrium actions, which we shall denote by $\text{PED}_i(y^0, A^0)$ is a convex set in $\Delta_+^{|X_i|}$; this is an immediate consequence of the convexity of $\text{PED}(y^0, A^0)$. Since $\text{PED}_i(\xi^0, A^0)$ is a convex set, the predicted fraction of players in role i who choose a particular action $\tilde{x}_i$ from A_i^0 is given precisely by the closed interval

$$\left[\min\{\nu_i^0(\tilde{x}_i) : \nu_i^0 \in \text{PED}_i(y^0, A^0)\}\ ,\ \max\{\nu_i^0(\tilde{x}_i) : \nu_i^0 \in \text{PED}_i(y^0, A^0)\}\right].$$

By (24), the value of $\max\{\nu_i^0(\tilde{x}_i) : \nu_i^0 \in \text{PED}_i(y^0, A^0)\}$ can be easily obtained by solving the following linear program:

$$\text{maximize } \textstyle\sum_{\{x \in A^0 : x_i = \tilde{x}_i\}} \mu^0(x) \text{ subject to } \{\widetilde{Q}(\mathbf{x})\}_{\mathbf{x} \in \mathbf{A}^{**}} \text{ and } \{\mu^0(x)\}_{x \in X} \text{ satisfying (23).}$$

In a similar vein, we can calculate $\min\{\nu_i^0(\tilde{x}_i) : \nu_i^0 \in \text{PED}_i(y^0, A^0)\}$.

Lastly, we turn our attention to the comparative statics of equilibrium predictions. We know from Theorem 4 that, for panel data sets, an increase in the exogenous variable y^0 leads to higher predictions on the set of possible best responses. The next result extends that theorem to stochastic data sets: it says that the set $\mathrm{PED}(y^0, A^0)$ will vary monotonically with y^0 with respect to first order stochastic dominance.

THEOREM 8. *Suppose* $\mathcal{O} = \{(\mu^t, y^t, A^t)\}_{t\in\mathcal{T}}$ *is an SC-rationalizable data set and let* $(\bar{y}^0, A^0)$ *and* $(\underline{y}^0, A^0)$ *be two treatments with* $\bar{y}^0 > \underline{y}^0$*. Then for every* $\underline{\mu} \in \mathrm{PED}(\underline{y}^0, A^0)$*, there exists* $\bar{\mu} \in \mathrm{PED}(\bar{y}^0, A^0)$ *such that* $\bar{\mu} \geqslant_{FOSD} \underline{\mu}$*, and, for every* $\bar{\mu} \in \mathrm{PED}(\bar{y}^0, A^0)$*, there exists* $\underline{\mu} \in \mathrm{PED}(\underline{y}^0, A^0)$ *such that* $\bar{\mu} \geqslant_{FOSD} \underline{\mu}$.[20]

6 Application: Smoking behavior of married couples

To illustrate how the techniques developed in this paper could be used, we now apply them to analyse the impact of spousal smoking and workplace smoking policies on smoking behavior. Beginning from the mid-80's, restrictions on workplace smoking, principally motivated by the dangers of secondhand tobacco smoke, became increasingly common in the United States. Along with these changes, a literature developed investigating whether these policies also have the effect of reducing smoking rates. Among the influential papers is that of Evans, Farrelly and Montgomery (1999), which also contains a discussion of the relevant literature to that date.

The single equation results in Evans et al., in which smoking bans are assumed to be exogenous, suggest that workplace smoking restrictions lead to a 5.7% decline in smoking prevalence. However, it is obvious that if a worker's unobserved propensity to smoke is correlated with the presence of workplace smoking restrictions, then the single-estimation equations would be subject to an omitted variables bias. Thus, Evans et al. also show that the effects of the ban survive various estimation strategies that address the possibility of smokers selecting into smoke friendly workplaces; amongst other things, they control for several covariates that might signal the health of the individual and also use establishment size as an instrument for workplace smoking policy. They conclude that the omitted variables bias does not substantially alter the estimates obtained from the single equation model.

[20] For two distributions ν and θ on a Euclidean space, we say that ν first order stochastically dominates θ if $\int f(y)d\nu(y) \geqslant \int f(y)d\theta(y)$ for all increasing real-valued functions f. It is known that this holds if and only if, for any set C with the upper comprehensive property (i.e., if $y \in C$ then $z \in C$ for any $z \geqslant y$), its probability under ν is greater than its probability under θ.

Another paper that is related to ours is Cutler and Glaeser (2010). This paper investigates whether people are more likely to smoke if their spouse smokes, with workplace smoking bans serving as an instrument for spousal smoking. Their estimation strategy relies on the assumption that smoking bans are exogenous (which they justify with the findings of Evans et al.). Using a parametric instrumental variable model of smoking behavior, they find a statistically significant spousal effect on smoking choices.

In this section, we test a specific model of spousal interaction and workplace smoking policy effects. Our hypothesis is the following:

(i) *a person's preference between smoking and not-smoking obeys SID (equivalently, SSCD), with the parameter being the spouse's smoking behavior* (smoke or not-smoke) *and the person's workplace smoking policy* (allowed or restricted), *and*

(ii) *a couple's joint smoking behavior emerges as a pure strategy Nash equilibrium.*

Unlike Cutler and Glaeser, our approach is non-parametric and it models explicitly the simultaneity of partner choices. We use a data set that provides us with the smoking decision and the workplace smoking policy for each member of a large population of married couples. Differing workplace smoking policies provide the treatment variation needed for testing the presence of strategic complementarity, using the results in Section 5.2. As in Cutler and Glaeser, the validity of our test requires the distribution of smoking preferences among couples to be the same across workplace smoking policies; we justify this assumption in Section 6.2 and Appendix II using a test of balance, familiar in the literature on treatment effects.

6.1 Data

We employ the Tobacco Use Supplement of the Current Population Survey (TUS-CPS) to get information on both smoking decisions and workplace smoking policies. This is an NCI-sponsored survey of tobacco use that has been administered as part of the US Census Bureau's Current Population Survey every 2 to 3 years since 1992.[21] We focus on the period 1992-1993 because, in contrast to more recent years, a significant proportion of workplaces then did not have smoking

[21] This survey's data were also used in Evans, Farrelly and Montgomery (1999) and Cutler and Glaeser (2010).

Smoking rates		Smoking rates	
Pr (Husband Smokes | Wife Smokes)	59.2%	Pr (Husband Smokes | Smoking Allowed)	31.2%
Pr (Husband Smokes | Wife Doesn't Smoke)	15.6%	Pr (Husband Smokes | Smoking Restricted)	21.9%
Pr (Wife Smokes | Husband Smokes)	46.6%	Pr (Wife Smokes | Smoking Allowed)	24.9%
Pr (Wife Smokes | Husband Doesn't Smoke)	10.0%	Pr (Wife Smokes | Smoking Restricted)	17.6%

Figure 3: Conditional smoking rates

restrictions, which guarantees that we have enough treatment variation. While information on smoking behavior is obtained from everyone in our population of interest, the question on workplace smoking policy is posed only to indoor workers. Thus, we confine our attention to married couples where both members work indoors. After eliminating from our sample all couples where at least one member did not reply to all the questions of interest, we have 5,363 married couples across the US.

Within this sample, the smoking rate is 23.8% among men and 18.7% among women. Smoking is permitted in 19.7% of husbands' workplaces and 15% of wives' workplaces. Figure 3 displays the conditional probabilities of smoking given partner's smoking behavior (left panel) and smoking policy at work (right panel). As we can see, irrespective of gender, the probability of smoking is larger when either the partner smokes or when smoking is permitted in the workplace. Overall, the fraction of spouses that make the same smoking choice —either both smoke or do not smoke— is around 80% of the whole sample. These figures are at least suggestive of the influence of spousal behavior and workplace policy on smoking decisions. To examine this issue more closely, we now apply the test developed in Section 5.2.

6.2 Findings

This section examines the presence of strategic complementarity in two steps. First, we apply the test to the whole sample. We use this first result to clarify the practical implementation of our idea as well as to elaborate on the informativeness of our modelling restrictions. We also use it to show how to deal with small sample issues. Second, we explain why we think it may be important to control for education levels in our application, and show that the initial results remain valid after doing so.

Figure 4 displays the distribution of joint choices regarding smoking decisions for four different

Workplace smoking policy = (0,1)		
Husband/Wife	Non−smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 8.6%	μ (S,S) = 14.9%
Non−smoking (N)	μ (N,N) = 64.6%	μ (S,N) = 11.9%

Workplace smoking policy = (1,1)		
Husband/Wife	Non−smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 11.3%	μ (S,S) = 16.3%
Non−smoking (N)	μ (N,N) = 59.0%	μ (S,N) = 13.4%

Workplace smoking policy = (0,0)		
Husband/Wife	Non−smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 6.9%	μ (S,S) = 9.4%
Non−smoking (N)	μ (N,N) = 71.8%	μ (S,N) = 11.9%

Workplace smoking policy = (1,0)		
Husband/Wife	Non−smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 9.1%	μ (S,S) = 14.9%
Non−smoking (N)	μ (N,N) = 59.1%	μ (S,N) = 16.9%

Figure 4: Joint distribution of smoking choices across smoking policies

workplace smoking policies, which serve as treatments in our analysis. As before, we use μ to indicate the probability of each action profile for each workplace smoking policy. The first argument of μ takes the value of S if the husband smokes and N otherwise; the second argument indicates the smoking decision of his wife. Similarly, the first argument in Workplace Smoking Policy takes the value of 1 if smoking is permitted in the husbands's workplace and 0 if it is restricted; the second argument indicates the smoking policy at the wife's workplace. In this application, the choice set of each person is $\{N, S\}$ and it remains the same across observations. A first look at Figure 4 suggests strong differences of joint spousal smoking choices across the four different joint ban restrictions. (In fact, a simple chi-square test rejects with a p-value of 0 the possibility of equal smoking choices across the four treatments.) We next explore whether these differences are consistent with our model of strategic complementarities.

Notice that in this smoking model there are $4^4 = 256$ possible group paths or types, since for each of the four possible treatment values, there are four joint choices that a married couple can make. It is quite clear that if we allow for all possible types, then even if we require the distribution of types to be independent of treatment, we can still explain *any* observed distribution of outcomes.[22] Our objective is more specific: we wish to test if the data set displayed in Figure 4 is consistent with couples playing two-by-two games of strategic complementarity under different treatments; in other words, we would like to use Theorem 7 to test if this data

[22] Assign to the path $(N, N|0, 0)$, $(S, N|1, 0)$, $(N, S|0, 1)$ and $(S, S|1, 1)$ the probability

$$\mu(N, N|0, 0) \times \mu(S, N|1, 0) \times \mu(N, S|0, 1) \times \mu(S, S|1, 1)$$

and proceed in the same way with the other 255 paths. These probabilities will generate the data.

set is SC-rationalizable.[23]

One could check that of the 256 possible paths, precisely 64 are SC-rationalizable. (Appendix III (online) gives a fuller description of the test procedure.) Ignoring issues of sample size for the moment and treating the observations in Figure 4 as the true distribution of joint actions across the four treatments, we can test for SC-rationalizability by checking if there is a positive solution to the linear system (22), where the solution vector, if it exists, gives the proportion of the population belonging to each of the 64 types. For this test to be valid, we require the distribution of types across the four treatments to be the same, but *no other restrictions are placed on the distribution.* In particular, the test is agnostic about how couples are formed; whether matches are formed randomly or assortatively (in the sense that couples with similar smoking tendencies are more likely to match up[24]) does not affect the validity of the test.[25] Performing this test, we find that there is in fact no solution to the linear system, so the data set is not SC-rationalizable.

This may come as a surprise, since the number of unknowns (64) far exceeds the number of linear constraints and it is tempting to think that the conditions are very permissive. In fact, there is at least one easy-to-understand reason why the data set displayed in Figure 4 is not SC-rationalizable. Notice from Figure 4 that $\mu(N, S|1, 0) = 9.1\% > 8.6\% = \mu(N, S|0, 1)$. This is impossible because, to be consistent with strategic complementarity, any couple type that selects (N, S) under the smoking policy $(1, 0)$ must select (N, S) again under the smoking policy $(0, 1)$.[26]

[23]In studies of the household, one modelling approach is to assume that couples achieve a Pareto optimal outcome, rather than play a Nash equilibrium (see, for example, Chiappori (1988)). However, this solution concept does not impose any discipline in our application. To see why, let us assume that all husbands and wives, respectively, have the following preferences irrespective of the treatment: $(S, N) \succ_h (S, S) \succ_h (N, N) \succ_h (N, S)$ and $(N, S) \succ_w (N, N) \succ_w (S, S) \succ_w (S, N)$. One could check that these two preference relations obey strict single crossing differences between the agent's own action and the action of the partner and also (trivially) the smoking policy at the working place. Moreover, every strategy profile is Pareto optimal with these preferences. As a consequence, these preferences can generate any data set we may observe.

[24]In our model, an individual's type is defined by his/her preference between smoking and non-smoking, conditional on the spouse's smoking behavior and the workplace smoking policy. For example, a possible type for the man is the type who *always* strictly prefers smoking to non-smoking. Our model allows for the possibility that such a man is more likely to match up with a woman of the same type.

[25]In a recent paper, Chiappori, Oreffice, and Quintana-Domeque (2016) show that smoking status can affect matching in the marriage market. Their finding highlights the relevance of our fully agnostic approach to matching or group formation.

[26]Let $\succsim_h$ be the husband's preference and $\succsim_w$ the wife's preference. Then $(N, S|1, 0) \succsim_h (S, S|1, 0)$ implies that $(N, S|0, 1) \succ_h (S, S|0, 1)$, so $(S, S|0, 1)$ is ruled out as an equilibrium. Furthermore, $(N, S|1, 0) \succsim_h (S, S|1, 0)$ implies $(N, N|0, 1) \succ_h (S, N|0, 1)$, so $(S, N|0, 1)$ is impossible as well. Turning now to the wife, since $(N, S|1, 0) \succsim_w (N, N|1, 0)$, we obtain $(N, S|0, 1) \succ_w (N, N|0, 1)$, so $(N, N|0, 1)$ cannot be an equilibrium.

Workplace smoking policy = (0,1)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 8.8%	μ (S,S) = 14.8%
Non-smoking (N)	μ (N,N) = 64.5%	μ (S,N) = 11.8%

Workplace smoking policy = (1,1)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 11.3%	μ (S,S) = 16.3%
Non-smoking (N)	μ (N,N) = 59.0%	μ (S,N) = 13.4%

Workplace smoking policy = (0,0)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 6.9%	μ (S,S) = 9.4%
Non-smoking (N)	μ (N,N) = 71.8%	μ (S,N) = 11.9%

Workplace smoking policy = (1,0)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 8.8%	μ (S,S) = 15.0%
Non-smoking (N)	μ (N,N) = 59.2%	μ (S,N) = 16.9%

Figure 5: Closest SC-rationalizable distribution of smoking choices

Interestingly, if we solve for the data set that is SC-rationalizable and closest (as measured by the sum of square deviations) to the one actually observed, the solution, as displayed in Figure 5, sets $\mu(N, S|1, 0) = \mu(N, S|0, 1) = 8.8\%$.

If we compare the entries in Figures 4 and 5, we see immediately that they are quite close, which makes us wonder whether the observed violation of SC-rationalizability is in fact significant. To address this issue, we adopt the approach proposed by Kitamura and Stoye (2016); they develop a method of evaluating the statistical significance of a data set violating a set of linear constraints that directly applies to our framework.[27] Roughly speaking, the test assumes that the closest compatible distribution displayed in Figure 5 is the true population distribution, and uses a bootstrap procedure to calculate the likelihood of getting a sample like the one we observe. By applying their test, we find that the probability of getting our sample (or a more extreme one), assuming that our modelling restrictions are true, is 0.1605. The latter corresponds to the p-value for the null hypothesis that our modelling assumptions are true. This means that we cannot reject SC-rationalizability at a significance level of 5% or 10%. (See Appendix III (online) for a fuller description of the Kitamura-Stoye procedure and our implementation.[28])

As we mentioned earlier, the validity of our test hinges on the assumption that the distribution of smoking preferences among couples — or types — is the same across workplace smoking

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

[28] The Kitamura-Stoye test relies on a tuning parameter that solves a discontinuity issue arising from the possibility of boundary solutions. In this section, we present p-values corresponding to the tuning parameter suggested by the authors. The p-value estimates are sensitive to the tuning parameter and we present results corresponding to an alternative parameter value in Appendix III (online).

Workplace smoking policy = (0,1)		
Husband/Wife	Non–smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 6.7%	μ (S,S) = 7.7%
Non–smoking (N)	μ (N,N) = 77.4%	μ (S,N) = 8.2%

Workplace smoking policy = (1,1)		
Husband/Wife	Non–smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 6.7%	μ (S,S) = 9.0%
Non–smoking (N)	μ (N,N) = 70.8%	μ (S,N) = 13.5%

Workplace smoking policy = (0,0)		
Husband/Wife	Non–smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 4.9%	μ (S,S) = 4.6%
Non–smoking (N)	μ (N,N) = 81.8%	μ (S,N) = 8.7%

Workplace smoking policy = (1,0)		
Husband/Wife	Non–smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 4.6%	μ (S,S) = 10.2%
Non–smoking (N)	μ (N,N) = 69.6%	μ (S,N) = 15.5%

Figure 6: Joint distribution of smoking choices across smoking policies (HE)

policies. A possible concern is that smokers may choose jobs that are particularly smoke-friendly, which would cause a negative correlation between workplace bans and smoking preferences. To address this concern, we first repeat our initial test controlling on education levels. Specifically, we identify from the entire sample two sub-samples, according to the educational attainment of the couples: one sample of 2643 couples, where both spouses have high education levels (HE) measured as having at least some college education and another sample of 1422 couples, where both spouses have low education levels (LE). Controlling for education levels is potentially important because we notice that the distribution of treatments for HE couples is markedly different from that for LE couples. If, in addition, the propensity to smoke differs significantly across couples with different educational levels, then we can no longer guarantee that the distribution of types in the entire population is the same across treatments.

Figures 6 and 7 display the smoking behavior of HE and LE couples conditional on workplace smoking restrictions. It turns out that the smoking patterns of the HE sample are *exactly* SC-rationalizable. The LE sample is not exactly SC-rationalizable; however, implementing again the Kitamura-Stoye test, we find a p-value of 0.359, so the model cannot be rejected at the 5% or 10% significance level.

Lastly, we need to check that our assumption that the distribution of types is the same across workplace smoking restrictions is plausible among HE couples and, separately, among LE couples. To this end, we show that, within each of the two sub-populations, five characteristics of people that are thought to affect smoking choices are balanced across groups with different workplace

Workplace smoking policy = (0,1)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 12.3%	μ (S,S) = 20.7%
Non-smoking (N)	μ (N,N) =52.0%	μ (S,N) = 15.1%

Workplace smoking policy = (1,1)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 14.0%	μ (S,S) = 21.1%
Non-smoking (N)	μ (N,N) = 53.5%	μ (S,N) = 11.4%

Workplace smoking policy = (0,0)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 9.3%	μ (S,S) = 19.5%
Non-smoking (N)	μ (N,N) = 53.0%	μ (S,N) = 18.2%

Workplace smoking policy = (1,0)		
Husband/Wife	Non-smoking (N)	Smoking (S)
Smoking (S)	μ (N,S) = 12.8%	μ (S,S) = 22.6%
Non-smoking (N)	μ (N,N) = 47.6%	μ (S,N) = 17.0%

Figure 7: Joint distribution of smoking choices across smoking policies (LE)

smoking policies. The motivation for such a test of balance is quite simple (see Imbens (2004) for a more detailed explanation). For instance, Cutler and Glaeser (2010) show that older people have a higher propensity to smoke. If, in addition, people with stronger preferences for smoking were indeed selecting jobs that are particularly smoke-friendly, then observing the workplace smoking policy should help us predict the age of the person. We find no evidence of such effects in our dataset. Specifically, we use Age, Number of children at home, White, Hispanic origin, and Northeast as the explained variables, and workplace smoking bans as the binary explanatory variable.[29] (To give more power to our test, we selected, as the explained variables, all the covariates in our dataset which, according to the tests done by Cutler and Glaeser, have a significant effect on the propensity to smoke.) The two tables in Appendix II show our results for husbands and wives in the two sub-populations. Out of the *20 estimated effects*, only one is significantly different from 0 at the 5% and another at the 10% level. Moreover, an F-test across all characteristics is shown to be non-significant for both the HE couples (p-value = 0.5952) and the LE couples (p-value = 0.2702). Altogether, beyond statistical noise, there do not appear to be meaningful differences in the characteristics of couples subjected to different smoking bans at workplaces. This gives credibility to our exogeneity restriction within the two groups and is in line with the conclusion reached by Evans, Farrelly, and Montgomery (1999).

[29]Note that the last three explained variables (indicating whether or not a person is white, Hispanic, and living in the Northeast of the United States) are binary.

Appendix I

We have shown in Lemma 1 that $\succsim_i^{RT}$ has the interval property. The following extension of that result is needed for the proofs of Lemmas 2 and 3.

LEMMA A1: *The relations $\succsim_i^{RTS}$, $\succ_i^{RTS}$, and $\succsim_i^{RTST}$ on $X_i \times \Xi_i$ have the interval property.*

Proof. Let $x_i'' > x_i > x_i'$. (The case where $x_i'' < x_i < x_i'$ can be proved in a similar way.) If $(x_i'', \xi_i) \succsim_i^{RTS} (\succ_i^{RTS}) (x_i', \xi_i)$ holds, there exists some $\xi_i' \leqslant (<) \, \xi_i$ such that $(x_i'', \xi_i') \succsim_i^{RT} (x_i', \xi_i')$. By the interval property of $\succsim_i^{RT}$, we obtain $(x_i'', \xi_i') \succsim_i^{RT} (x_i, \xi_i')$. Since $x_i'' > x_i$ and $\xi_i' \leqslant (<) \, \xi_i$, we have that $(x_i'', \xi_i) \succsim_i^{RTS} (\succ_i^{RTS}) (x_i, \xi_i)$. So we have shown that $\succsim_i^{RTS}$ and $\succ_i^{RTS}$ have the interval property. Lastly, if $(x_i'', \xi_i) \succsim_i^{RTST} (x_i', \xi_i)$, there exists a sequence $z_i^1, z_i^2, ..., z_i^k$ such that

$$(x_i'', \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \succsim_i^{RTS} ... \succsim_i^{RTS} (z_i^k, \xi_i) \succsim_i^{RTS} (x_i', \xi_i).$$

Letting $z_i^0 = x_i''$ and $z_i^{k+1} = x_i'$, since $x_i'' > x_i > x_i'$, we can find some $0 \leqslant m \leqslant k$ such that $z_i^m \geqslant x_i \geqslant z_i^{m+1}$. By the interval property of $\succsim_i^{RTS}$, we obtain $(z_i^m, \xi_i) \succsim_i^{RTS} (x_i, \xi_i)$. Thus $(x_i'', \xi_i) \succsim_i^{RTST} (x_i, \xi_i)$ since $(x_i'', \xi_i) \succsim_i^{RTST} (z_i^m, \xi_i) \succsim_i^{RTS} (x_i, \xi_i)$. □

LEMMA A2: *Suppose $(x_i'', \xi_i) \succsim_i^{RTST} (x_i', \xi_i)$; then there is z_i^j (for $j = 1, 2, ..., k$) such that*

$$(x_i'', \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \succsim_i^{RTS} ... \succsim_i^{RTS} (z_i^k, \xi_i) \succsim_i^{RTS} (x_i', \xi_i). \tag{25}$$

with

$$x_i'' > z_i^1 > z_i^2 > ... > z_i^k > x_i' \tag{26}$$

if $x_i'' > x_i'$ and the inequality (26) reversed if $x_i'' < x_i'$.[30]

Proof. By the definition of $\succsim_i^{RTST}$, we know there is z_i^j such that (25) holds, so what we need to do is to show that z_i^j obeys (26) if $x_i'' > x_i'$. (The case where $x_i'' < x_i'$ has an analogous proof which we shall skip.) To do this, we choose a chain linking (x_i'', ξ_i) and (x_i', ξ_i) with the property that (writing $z_i^0 = x_i''$ and $z_i^{k+1} = x_i'$) $(z_i^m, \xi_i) \not\succsim_i^{RTS} (z_i^{m'}, \xi_i)$ for $m' > m + 1$; in other words, no link in the chain can be dropped. We claim that (26) must hold in this case. First we note that $z_i^j > x_i'$ for all $j < k + 1$. If not, there is ℓ such that $z_i^\ell \leqslant x_i' < z_i^{\ell-1}$, with $(z_i^{\ell-1}, \xi_i) \succsim_i^{RTS} (z_i^\ell, \xi_i)$;

[30]We consider (26) to hold vacuously if $x_i'' \succsim_i^{RTS} x_i'$.

since $\succsim_i^{RTS}$ has the interval property, we obtain $(z_i^{\ell-1}, \xi_i) \succsim_i^{RTS} (x_i', \xi_i)$ and the chain has been shortened. To show that z_i^j is decreasing, suppose instead that there is m such that $z_i^{m+1} > z_i^m$. Let z_i^{m+n} be the first time after z_i^{m+1} such that $z_i^{m+n} \leqslant z_i^m$. (This must occur since $z_j^m > x_i'$.) Then we have $z_i^{m+n} \leqslant z_i^m < z_i^{m+n-1}$. Since $(z_i^{m+n-1}, \xi_i) \succsim_i^{RTS} (z_i^{m+n}, \xi_i)$, the interval property of $\succsim_i^{RTS}$ guarantees that $(z_i^{m+n-1}, \xi_i) \succsim_i^{RTS} (z_i^m, \xi_i)$. Thus we obtain a cycle

$$(z_i^m, \xi_i) \succsim_i^{RTS} (z_i^{m+1}, \xi_i) \succsim_i^{RTS} \ldots \succsim_i^{RTS} (z_i^{m+n-1}, \xi_i) \succsim_i^{RTS} (z_i^m, \xi_i).$$

Since $\succsim_i^{RTS}$ is cyclically consistent, this chain cannot be related by $\succ_i^{RTS}$ and must be related by $\succsim_i^{RT}$. In particular, $(z_i^{m+n-1}, \xi_i) \not\succ_i^{RTS} (z_i^m, \xi_i)$ and thus $(z_i^{m+n-1}, \xi_i) \not\succ_i^{RTS} (z_i^{m+n}, \xi_i)$ (by the interval property of $\succ_i^{RTS}$). We conclude that $(z_i^m, \xi_i) \succsim_i^{RT} (z_i^{m+n}, \xi_i)$ and thus we can shorten (25) to

$$(x_i'', \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i) \succsim_i^{RTS} \ldots \succsim_i^{RTS} (z_i^m, \xi_i) \succsim_i^{RTS} (z_i^{m+n}, \xi_i) \succsim_i^{RTS} \ldots \succsim_i^{RTS} (z_i^k, \xi_i) \succsim_i^{RTS} (x_i', \xi_i)$$

which contradicts our assumption that no link in the chain can be dropped. $\square$

Proof of Lemma 2: We first prove that (11) holds. (11) is equivalent to $\succsim_i^{RTS}$ being *cyclically consistent*, i.e.,

$$(z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \succsim_i^{RTS} \ldots \succsim_i^{RTS} (z_i^k, \xi_i) \Longrightarrow (z_i^k, \xi_i) \not\succ_i^{RTS} (z_i^1, \xi_i). \tag{27}$$

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

$$(z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \succsim_i^{RTS} \ldots \succsim_i^{RTS} (z_i^k, \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i) \tag{28}$$

$$\Longrightarrow (z_i^1, \xi_i) \not\succ_i^{RTS} (z_i^2, \xi_i) \not\succ_i^{RTS} \ldots \not\succ_i^{RTS} (z_i^k, \xi_i) \not\succ_i^{RTS} (z_i^1, \xi_i)$$

Thus, whenever there is a cycle like (28), it *must* be the case that

$$(z_i^1, \xi_i) \succsim_i^{RT} (z_i^2, \xi_i) \succsim_i^{RT} \ldots \succsim_i^{RT} (z_i^k, \xi_i) \succsim_i^{RT} (z_i^1, \xi_i)$$

We prove (11) by induction on the length of the chain, k, on the left side of (27). Whenever (27) holds for chains of length k or less (equivalently, whenever the cycles in (28) have length k or less), we say that $\succsim_i^{RTS}$ is k-consistent. For 2-consistency, we need to show that

$$(z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \Longrightarrow (z_i^2, \xi_i) \not\succ_i^{RTS} (z_i^1, \xi_i).$$

Suppose that $z_i^1 > z_i^2$; the case of $z_i^1 < z_i^2$ can be dealt with in a similar way. By definition, if $(z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i)$ then there is $\xi_i' \leqslant \xi_i$ such that $(z_i^1, \xi_i') \succsim_i^{RT} (z_i^2, \xi_i')$. On the other hand, if $(z_i^2, \xi_i) \succ_i^{RTS} (z_i^1, \xi_i)$, then there is $\xi_i'' > \xi_i$ such that $(z_i^2, \xi_i'') \succsim_i^{RT} (z_i^1, \xi_i'')$ and so we obtain a violation of ARC. Suppose that $\succsim_i^{RTS}$ is k-consistent for all $k < \bar{k}$. To show that $\bar{k}$-consistency holds, suppose the left side of (27) holds for $k = \bar{k}$ and $z_i^1 < z_i^{\bar{k}}$. Clearly, there must be $m < \bar{k}$ such that $z_i^m < z_i^{\bar{k}}$ and $z_i^{m+1} \geqslant z_i^{\bar{k}}$. We consider two cases: (A) $z_i^m \geqslant z_i^1$ and (B) $z_i^m < z_i^1$. In case (A), by the interval property of $\succsim_i^{RTS}$, we obtain $(z_i^m, \xi_i) \succsim_i^{RTS} (z_i^{\bar{k}}, \xi_i)$. By way of contradiction, suppose also that $(z_i^{\bar{k}}, \xi_i) \succ_i^{RTS} (z_i^1, \xi_i)$. Then the interval property of $\succ_i^{RTS}$ guarantees that $(z_i^{\bar{k}}, \xi_i) \succ_i^{RTS} (z_i^m, \xi_i)$ and so we obtain a violation of 2-consistency. For (B), since $(z_i^m, \xi_i) \succsim_i^{RTS} (z_i^{m+1}, \xi_i)$, the interval property guarantees that $(z_i^m, \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i)$. So we obtain the cycle

$$(z_i^1, \xi_i) \succsim_i^{RTS} (z_i^2, \xi_i) \succsim_i^{RTS} ... \succsim_i^{RTS} (z_i^m, \xi_i) \succsim_i^{RTS} (z_i^1, \xi_i) \tag{29}$$

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

$$(z_i^1, \xi_i) \not\succ_i^{RTS} (z_i^2, \xi_i) \not\succ_i^{RTS} ... \not\succ_i^{RTS} (z_i^m, \xi_i) \not\succ_i^{RTS} (z_i^1, \xi_i)$$

and so we can replace each $\succsim_i^{RTS}$ in (29) by $\succsim_i^{RT}$. Furthermore, $(z_i^m, \xi_i) \not\succ_i^{RTS} (z_i^1, \xi_i)$ guarantees that $(z_i^m, \xi_i) \not\succ_i^{RTS} (z_i^{m+1}, \xi_i)$, by the interval property of $\succ_i^{RTS}$. Therefore, $(z_i^1, \xi_i) \succsim_i^{RT} (z_i^{m+1}, \xi_i)$ and, by the interval property of $\succsim_i^{RT}$, we obtain $(z_i^1, \xi_i) \succsim_i^{RT} (z_i^{\bar{k}}, \xi_i)$. 2-consistency then ensures that $(z_i^{\bar{k}}, \xi_i) \not\succ_i^{RTS} (z_i^1, \xi_i)$. This completes the proof that (11) holds.

By definition, $\succsim_i^{RTST}$ obeys SSCD if whenever $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$ *or* $x_i'' < x_i'$ and $\xi_i'' < \xi_i'$,

$$(x_i'', \xi_i') \succsim_i^{RTST} (x_i', \xi_i') \Longrightarrow (x_i'', \xi_i'') \succ_i^{RTST} (x_i', \xi_i'').$$

We shall concentrate on the case where $x''_i > x'_i$; the other case has a similar proof. If $(x''_i, \xi'_i) \succsim_i^{RTST}$ (x'_i, ξ'_i), then, by Lemma A2, there is z_i^j (for $j = 1, 2, ..., k$) such that

$$(x''_i, \xi'_i) \succsim_i^{RTS} (z_i^1, \xi'_i) \succsim_i^{RTS} (z_i^2, \xi'_i) \succsim_i^{RTS} ... \succsim_i^{RTS} (z_i^k, \xi'_i) \succsim_i^{RTS} (x'_i, \xi'_i).$$

with $x''_i > z_i^1 > z_i^2 > ... > z_i^k > x'_i$. Since $\succsim_i^{RTS}$ obeys SSCD (see (8)), we obtain

$$(x''_i, \xi''_i) \succ_i^{RTS} (z_i^1, \xi''_i) \succ_i^{RTS} (z_i^2, \xi''_i) \succ_i^{RTS} ... \succ_i^{RTS} (z_i^k, \xi''_i) \succ_i^{RTS} (x'_i, \xi''_i)$$

and so $(x''_i, \xi''_i) \succ_i^{RTST} (x'_i, \xi''_i)$. $\square$

Proof of Lemma 3: We first show that $\succsim_i^*$ is a preference that rationalizes $\mathcal{O}_i$. Clearly, $\succsim_i^*$ is complete and reflexive, so to demonstrate that it is a preference we need only show that it is transitive. Indeed, suppose

$$(a_i, \xi_i) \succsim_i^* (b_i, \xi_i) \succsim_i^* (c_i, \xi_i) \succsim_i^* (a_i, \xi_i). \tag{30}$$

There are essentially four possible cases we need to consider:
Case 1. None of the three elements are related by $\succsim_i^{RTST}$. Given the definition of $\succsim_i^*$, this means that $a_i < b_i < c_i < a_i$, which is impossible.
Case 2. $(a_i, \xi_i) \parallel_i^{RTST} (b_i, \xi_i)$, $(b_i, \xi_i) \parallel_i^{RTST} (c_i, \xi_i)$, and $(c_i, \xi_i) \succsim_i^{RTST} (a_i, \xi_i)$. Then (30) can only occur if $a_i < b_i < c_i$, but if this is the case, the interval property of $\succsim_i^{RTST}$ will imply that $(c_i, \xi_i) \succsim_i^{RTST} (b_i, \xi_i)$. So this case is impossible.
Case 3. $(a_i, \xi_i) \parallel_i^{RTST} (b_i, \xi_i)$, $(b_i, \xi_i) \succsim_i^{RTST} (c_i, \xi_i) \succsim_i^{RTST} (a_i, \xi_i)$. This is impossible because, by the transitivity of $\succsim^{RTST}$, we obtain $(b_i, \xi_i) \succsim_i^{RTST} (a_i, \xi_i)$.
Case 4. $(a_i, \xi_i) \succsim_i^{RTST} (b_i, \xi_i) \succsim_i^{RTST} (c_i, \xi_i) \succsim_i^{RTST} (a_i, \xi_i)$. By (11), this is only possible if

$$(a_i, \xi_i) \succsim_i^{RT} (b_i, \xi_i) \succsim_i^{RT} (c_i, \xi_i) \succsim_i^{RT} (a_i, \xi_i),$$

but then we also obtain, by the transitivity of $\succsim_i^{RT}$, $(a_i, \xi_i) \succsim_i^{RT} (c_i, \xi_i)$ and, hence, $(a_i, \xi_i) \succsim_i^*$ (c_i, ξ_i), which establishes the transitivity of $\succsim_i^*$.
Lastly, since $\succsim_i^{RTST} \subset \succsim_i^*$ by construction, it is clear that $\succsim_i^*$ rationalizes $\mathcal{O}_i$.

To show that $\succsim_i^*$ obeys SSCD, let $x_i'' > x_i'$ and $\xi_i'' > \xi_i'$; then

$$\begin{aligned}(x_i'', \xi_i') \succsim_i^* (x_i', \xi_i') &\Longrightarrow (x_i'', \xi_i') \succsim_i^{RTST} (x_i', \xi_i') \\ &\Longrightarrow (x_i'', \xi_i'') \succ_i^{RTST} (x_i', \xi_i'') \\ &\Longrightarrow (x_i'', \xi_i'') \succ_i^* (x_i', \xi_i''),\end{aligned}$$

in which the first implication follows from the definition of $\succsim_i^*$, the second implication from the SSCD property of $\succsim_i^{RTST}$, and the third from the fact that $\succ_i^*$ contains $\succ_i^{RTST}$ (so $\succsim_i^*$ extends $\succsim_i^{RTST}$ in the sense of (10)). The last claim is true because if $(x_i'', \xi_i) \succ_i^{RTST} (x_i', \xi_i)$, then Lemma 2 says that $(x_i', \xi_i) \not\succsim_i^{RTST} (x_i'', \xi_i)$; thus $(x_i', \xi_i) \not\succsim_i^* (x_i'', \xi_i)$ and we obtain $(x_i'', \xi_i) \succ_i^* (x_i', \xi_i)$.

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

Proof of Proposition 2. Let λ by a measure on X_i with the following properties: (i) $\lambda(X_i) < \infty$; (ii) on any nonempty interval I of X_i, $\lambda(I) > 0$; (iii) $\lambda(\{x_i^t\}) > 0$ for all $t \in \mathcal{T}$. For any $(x_i, \xi_i) \in X_i \times \Xi_i$, we define the set $L(x_i, \xi_i) = \{z_i \in X_i : (x_i, \xi_i) \succsim_i^* (z_i, \xi_i)\}$. This set is measurable since $\mathcal{O}_i$ is finite and $\mathcal{A}_i$ consists of compact intervals. Furthermore, λ is a finite measure (according to (i)), so $\lambda(L(x_i, \xi_i))$ is well-defined. We claim that $u_i(x_i, \xi_i) = \lambda(L(x_i, \xi_i))$ represents $\succsim_i^*$. It follows immediately from the definition that $u_i(x_i'', \xi_i) \geqslant u_i(x_i', \xi_i)$ if $(x_i'', \xi_i) \succsim_i^* (x_i', \xi_i)$. So we need only show that $u_i(x_i'', \xi_i) > u_i(x_i', \xi_i)$ if $(x_i'', \xi_i) \succ_i^* (x_i', \xi_i)$. Suppose there exists an observed action, x_i^s, such that $x_i^s \in L(x_i'', \xi_i) \backslash L(x_i', \xi_i)$; then $u_i(x_i'', \xi_i) > u_i(x_i', \xi_i)$ since $\lambda(\{x_i^s\}) > 0$ (by (iii)). If such an x_i^s does not exist, then, in particular, $x_i'' \notin \{x_i^t\}_{t \in \mathcal{T}}$. For $(x_i'', \xi_i) \succsim^* (x_i', \xi_i)$, it must be the case that $(x_i'', \xi_i) \parallel_i^{RTST} (x_i', \xi_i)$ and $x_i'' < x_i'$. We claim that there is a sufficiently small $\epsilon > 0$ such that $x_i'' + \epsilon < x_i'$ and for any $z_i \in [x_i'', x_i'' + \epsilon]$, $(z_i, \xi_i) \parallel_i^{RTST} (x_i', \xi_i)$ and hence $(z_i, \xi_i) \succ_i^* (x_i', \xi_i)$. If this is true, $[x_i'', x_i'' + \epsilon]$ is contained in $L(x_i'', \xi_i) \backslash L(x_i', \xi_i)$ and has positive measure (by (ii)), so again $u_i(x_i'', \xi_i) > u_i(x_i', \xi_i)$. It remains for us to show that $\epsilon > 0$ exists. If it does not exist, then there must be a sequence $x_i^n > x_i''$ and tending towards x_i'' such that $(x_i', \xi_i) \succsim_i^{RTST} (x_i^n, \xi_i)$ (since, with a finite data set, it is impossible for there to be a sequence x_i^n tending x_i'' such that $(x_i^n, \xi_i) \succsim_i^{RTST} (x_i', \xi_i)$). This leads to $(x_i', \xi_i) \succsim_i^{RTST} (x_i'', \xi_i)$, which is a contradiction.[31] □

Proof of Theorem 3. Part (ii) follows straightforwardly from (i), so we shall focus on proving (i), which consists of three claims. Proposition 3 says that (15) holds. To see that (18) holds, first note that $\tilde{x}_i \notin \mathrm{PR}_i(x'_{-i}, y_i^0, A_i^0)$ if and only if $\overline{\mathcal{O}}_i' = \mathcal{O}_i \cup \{(\tilde{x}_i, (x'_{-i}, y_i^0), A_i^0)\}$ violates ARC. Since H_i is not a singleton, it must be an interval and so there is *no* x_i' such that $(x_i', x'_{-i}) = x^t$ for some $t \in \mathcal{T}$. Therefore, $\overline{\mathcal{O}}_i'$ violates ARC if and only if there is $\hat{x}_i \in A_i^0$ and $\bar{x}_{-i}$ such that $(\hat{x}_i, \bar{x}_{-i}, \bar{y}_i) \succsim_i^{RT} (\tilde{x}_i, \bar{x}_{-i}, \bar{y}_i)$ with either (1) $\hat{x}_i < \tilde{x}_i$ and $(\bar{x}_{-i}, \bar{y}_i) > (x'_{-i}, y_i^0)$ or (2) $\hat{x}_i > \tilde{x}_i$ and $(\bar{x}_{-i}, \bar{y}_i) < (x'_{-i}, y_i^0)$. Note that there is $t \in \mathcal{T}$ such that $(\hat{x}_i, \bar{x}_{-i}) = x^t$; in particular, this means that $\bar{x}_{-i} \in \times_{j \neq i} \underline{A}^{\mathcal{T}}$. It follows from our definition of H_i that for any x''_{-i} in H_i, we have $(\bar{x}_{-i}, \bar{y}_i) > (x''_{-i}, y_i^0)$ if $(\bar{x}_{-i}, \bar{y}_i) > (x'_{-i}, y_i^0)$ and $(\bar{x}_{-i}, \bar{y}_i) < (x''_{-i}, y_i^0)$ if $(\bar{x}_{-i}, \bar{y}_i) < (x'_{-i}, y_i^0)$. Thus $\overline{\mathcal{O}}_i'' = \mathcal{O}_i \cup \{(\tilde{x}_i, (x''_{-i}, y_i^0), A_i^0)\}$ violates ARC. We conclude that $\tilde{x}_i \notin \mathrm{PR}_i(x''_{-i}, y_i^0, A_i^0)$ if $\tilde{x}_i \notin \mathrm{PR}_i(x'_{-i}, y_i^0, A_i^0)$, which establishes (18).

Lastly, we show that $\mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ consists of a finite union of intervals of A_i^0. This is equivalent to showing that $A_i^0 \backslash \mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ is a finite union of intervals; an element $\tilde{x}_i$ is in this set if and only if there is $t \in \mathcal{T}$ such that $x_i^t \in A_i^0$ and $(x_i^t, \xi_i^0) \succ_i^{RTST} (\tilde{x}_i, \xi_i^0)$, where $\xi_i^0 = (x_{-i}, y_i^0)$. This in turn holds if and only if there is $s \in \mathcal{T}$ such that either (1) $(x_i^t, \xi_i^0) \succsim_i^{RTST} (x_i^s, \xi_i^0)$ and $(x_i^s, \xi_i^0) \succ_i^{RTS} (\tilde{x}_i, \xi_i^0)$ or (2) $(x_i^t, \xi_i^0) \succ_i^{RTST} (x_i^s, \xi_i^0)$ and $(x_i^s, \xi_i^0) \succsim_i^{RT} (\tilde{x}_i, \xi_i^0)$. Notice for a fixed $s \in \mathcal{T}$, the sets $\{x_i \in A_i^0 : (x_i^s, \xi_i^0) \succ_i^{RTS} (x_i, \xi_i^0)\}$ and $\{x_i \in A_i^0 : (x_i^s, \xi_i^0) \succsim_i^{RT} (x_i, \xi_i^0)\}$ both consist of intervals, because of the interval property on $\succ_i^{RTS}$ and $\succ_i^{RT}$ respectively. It follows that $A_i^0 \backslash \mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$ is a finite union of intervals. □

The proof of Theorem 4 uses the following lemma.

[31] In general, if a sequence x_i^n tends to $x_i'' \in X_i$, and $(x_i', \xi_i) \succsim_i^{RTST} (x_i^n, \xi_i)$ for all n, then $(x_i', \xi_i) \succsim_i^{RTST} (x_i'', \xi_i)$. Analogous closure properties are true of $\succsim_i^{RT}$ and $\succsim_i^{RTS}$. It is straightforward to check that these properties follow from the finiteness of the data set and the compactness of the sets in $\mathcal{A}_i$.

Lemma A3: *Suppose* $\mathcal{O} = \{x^t, y^t, A^t\}_{t=1}^T$ *obeys ARC and let* $A^0 \in \mathcal{A}$. *Then the map* $p_i^{**} : A_{-i}^0 \times Y_i \to A_i^0$ *given by*

$$p_i^{**}(x_{-i}, y_i) = \sup \mathrm{PR}_i(x_{-i}, y_i, A_i^0)$$

has the following properties: (i) it is increasing in $(x_{-i}, y_i) \in A_{-i}^0 \times Y_i$; *(ii) for* x'_{-i} *and* x''_{-i} *in* H_i, $p_i^{**}(x''_{-i}, y_i) = p_i^{**}(x'_{-i}, y_i)$; *and (iii) if, for some* $(\bar{x}_{-i}, \bar{y}_i)$, $p_i^{**}(\bar{x}_{-i}, \bar{y}_i) \in \mathrm{PR}_i(\bar{x}_{-i}, \bar{y}_i, A_i^0)$ *and for some* $(\hat{x}_{-i}, \hat{y}_i) > (\bar{x}_{-i}, \bar{y}_i)$, $p_i^{**}(\bar{x}_{-i}, \bar{y}_i) = p_i^{**}(\hat{x}_{-i}, \hat{y}_i)$, *then* $p_i^{**}(\hat{x}_{-i}, \hat{y}_i) \in \mathrm{PR}_i(\hat{x}_{-i}, \hat{y}_i, A_i^0)$.

Remark: In a similar way, we define $p_i^* : A_{-i}^0 \times Y_i \to A_i^0$ by $p_i^*(x_{-i}, y_i) = \inf \mathrm{PR}_i(x_{-i}, y_i, A_i^0)$. This function will obey properties (i) and (ii) and, instead of property (iii), it will have the following property (iii)′: *if, for some* $(\bar{x}_{-i}, \bar{y}_i)$, $p_i^*(\bar{x}_{-i}, \bar{y}_i) \in \mathrm{PR}_i(\bar{x}_{-i}, \bar{y}_i, A_i^0)$ *and for some* $(\hat{x}_{-i}, \hat{y}_i) < (\bar{x}_{-i}, \bar{y}_i)$, $p_i^*(\bar{x}_{-i}, \bar{y}_i) = p_i^*(\hat{x}_{-i}, \hat{y}_i)$, *then* $p_i^*(\hat{x}_{-i}, \hat{y}_i) \in \mathrm{PR}_i(\hat{x}_{-i}, \hat{y}_i, A_i^0)$.

Proof. Since $\mathrm{PR}_i(x_{-i}, y_i, A_i^0)$ is the union of a collection of best response correspondences (see (13)), each of which is increasing in (x_{-i}, y_i), p_i^{**} must be increasing. Claim (ii) is an immediate consequence of (18) (which was proved in Theorem 3). Lastly, if $p_i^{**}(\bar{x}_{-i}, \bar{y}_i) \in \mathrm{PR}_i(\bar{x}_{-i}, \bar{y}_i, A_i^0)$ then there is $\succsim_i \in \mathcal{P}_i^*$ such that $p_i^{**}(\bar{x}_{-i}, \bar{y}_i) \in \mathrm{BR}_i(\bar{x}_{-i}, \bar{y}_i, A_i^0, \succsim_i)$. Since the best response correspondence is increasing, there is $x'_i \in \mathrm{BR}_i(\hat{x}_{-i}, \hat{y}_i, A_i^0, \succsim_i)$, and thus in $\mathrm{PR}_i(\hat{x}_{-i}, \hat{y}_i, A_i^0)$, such that $x'_i \geqslant p_i^{**}(\bar{x}_{-i}, \bar{y}_i)$. This establishes (iii). □

Proof of Theorem 4: We have already explained at the beginning of Section 4 why $\mathcal{E}(y^0, A^0)$ is nonempty. We shall confine our attention to showing that $\max \overline{\mathcal{E}(y^0, A^0)}$ exists, where $\overline{\mathcal{E}(y^0, A^0)}$ refers to the closure of $\mathcal{E}(y^0, A^0)$; the proof for the other case is similar.[32] Firstly, note that the properties of p_i^{**} listed in Lemma A3 guarantee that there exists a sequence of functions $\{p_i^k(\cdot, y_i^0, A_i^0)\}_{k \in \mathbb{N}}$ selected from $\mathrm{PR}_i(\cdot, y_i^0, A_i^0)$ with the following properties: (i) for x'_{-i} and x''_{-i} in H_i, $p_i^k(x''_{-i}, y_i^0) = p_i^k(x'_{-i}, y_i^0)$; (ii) $p_i^k(x_{-i}, y_i^0, A_i^0)$ is increasing in x_{-i} and in k; (iii) $p_i^k(x_{-i}, y_i^0, A_i^0) = p_i^{**}(x_{-i}, y_i^0, A_i^0)$ if $p_i^{**}(x_{-i}, y_i^0, A_i^0) \in \mathrm{PR}_i(x_{-i}, y_i^0, A_i^0)$; and (iv) $\lim_{k \to \infty} p_i^k(x_{-i}, y_i^0, A_i^0) = p_i^{**}(x_{-i}, y_i^0, A_i^0)$. In other words, there is a sequence of increasing selections from $\mathrm{PR}_i(\cdot, y_i^0, A_i^0)$ that has $p_i^{**}(x_{-i}, y_i^0, A_i^0)$

[32] It is worth pointing out an obvious first approach that will *not* work. Given p_i^{**}, we can define, for each $x \in A^0$, $p^{**}(x, y^0) = (p_i^{**}(x_{-i}, y_i^0))_{i \in N}$, and since p_i^{**} is increasing in x_{-i}, so $p^{**}(x, y^0)$ is increasing in x. By Tarski's fixed point theorem, $p^{**}(\cdot, y^0)$ will have a fixed point and indeed a largest fixed point $x^\star$; thus the existence of $\max \overline{\mathcal{E}(y^0, A^0)}$ is ensured if it could be identified with $x^\star$. However, they are not generally the same points: it is straightforward to construct an increasing (but not compact-valued) correspondence such that its largest fixed point does not coincide with the largest fixed point of its supremum function. Our proof takes a different route.

as its limit, with the sequence being exactly equal to $p_i^{**}(x_{-i}, y_i^0, A_i^0)$ if the latter is a possible response of player i.

The function $p^k(x, y^0, A^0) = (p_i^k(x_{-i}, y_i^0, A_i^0))_{i \in N}$ is increasing in x, since p_i^k is increasing in x_{-i}. By Tarski's fixed point theorem, p^k has a largest fixed point, which we denote by $z^k(y^0, A^0)$. Since $p_i^k(\cdot, y_i^0, A_i^0)$ is a selection from $\mathrm{PR}_i(\cdot, y_i^0, A_i^0)$, $z^k(y^0, A^0) \in \mathcal{E}(y^0, A^0)$. By the monotone fixed points theorem (see Section 2), the sequence $z^k(y^0, A^0)$ is increasing with k. Since A^0 is compact, this sequence must have a limit. This limit, which we denote by $z^{**}(y^0, A^0)$, lies in $\overline{\mathcal{E}(y^0, A^0)}$.

We claim that $z^{**}(y^0, A^0)) \geqslant \tilde{x}$, for any $\tilde{x} \in \mathcal{E}(y^0, A^0)$. Indeed, since $\tilde{x}_i \in \mathrm{PR}_i(\tilde{x}_{-i}, y_i^0, A_i^0)$ for all $i \in N$, for k sufficiently large, $p_i^k(\tilde{x}_{-i}, y_i^0, A_i^0) \geqslant \tilde{x}_i$. Now consider the map p^k confined to the domain $S = \times_{i \in N}\{x_i \in A_i^0 : x_i \geqslant \tilde{x}_i\}$. Since p^k is increasing, the image of p^k also falls on S; in other words, p^k can be considered as a map from S to itself. It is also an increasing map and, by Tarski's fixed point theorem will have a largest fixed point. The largest fixed point of p^k restricted to S must again be $z^k(y^0, A^0)$ and it follows from our construction that $z^k(y^0, A^0) \geqslant \tilde{x}$. In turn this implies that $z^{**}(y^0, A^0) \geqslant \tilde{x}$. So $z^{**}(y^0, A^0)$ is an upper bound of $\mathcal{E}(y^0, A^0)$ and thus also an upper bound of $\overline{\mathcal{E}(y^0, A^0)}$. Given that $z^{**}(y^0, A^0) \in \overline{\mathcal{E}(y^0, A^0)}$, we conclude that $z^{**}(y^0, A^0) = \max \overline{\mathcal{E}(y^0, A^0)}$.

To see that $z^{**}(y, A^0)$ is increasing with respect to the parameter, consider $y'' > y'$. Given the properties of p_i^{**} listed in Lemma A3, we can choose functions $\{p_i^k(\cdot, y_i, A_i^0)\}_{k \in \mathbb{N}}$ selected from $\mathrm{PR}_i(\cdot, y_i, A_i^0)$ (for $y_i = y_i'$ and y_i'') satisfying properties (i) – (iv) and, in addition, $p_i^k(x_{-i}, y_i'', A_i^0) \geqslant p_i^k(x_{-i}, y_i', A_i^0)$ for all x_{-i}. The map $p^k(\cdot, y'', A^0)$ is increasing and, by Tarski's fixed point theorem, it will have a largest fixed point $x^k(y'')$ which also satisfies $x^k(y'') \geqslant x^k(y')$. Taking limits as $k \to \infty$, we obtain $z^{**}(y'') \geqslant z^{**}(y')$. □

Proof of Theorem 6: We may regard μ_i^t as an element of $\Delta = \left\{z \in \mathbb{R}_+^{|X_i|} : \sum_{k=1}^{|X_i|} z_k = 1\right\}$, and $\{\mu_i^t\}_{t \in \mathcal{T}}$ as an element of Δ^T. We denote the set of ARC-consistent paths by $\mathcal{D}$; each path can also be regarded as an element of Δ^T. Let Δ_{ARC}^T (contained in Δ^T) be the set of all $\{\mu_i^t\}_{t \in \mathcal{T}}$ such that $\{(\mu_i^t, \xi_i^t, A_i^t)\}_{t \in \mathcal{T}}$ is monotone rationalizable. By Theorem 5, this set is the convex hull of $\mathcal{D}$ (the set of ARC-consistent paths). Since $A_i^t = X_i$ for all $t \in \mathcal{T}$, $\mathcal{D}$ consists precisely of those paths where a higher parameter leads to a weakly higher action; it follows immediately from this that Δ_{ARC}^T is contained in Δ_{FOSD}^T, the set of $\{(\mu_i^t, \xi_i^t, X_i)\}_{t \in \mathcal{T}}$ that obey first order stochastic dominance in the sense that $\mu^t \geqslant_{FOSD} \mu^s$ whenever $\xi^t > \xi^s$. Both Δ^T_{ARC} and Δ^T_{FOSD} are convex and compact sets in Δ^T. The Krein-Milman Theorem tells us that Δ^T_{FOSD} is the convex hull of its extreme points; therefore, to show that $\Delta^T_{ARC} = \Delta^T_{FOSD}$ (as the proposition claims), we need only show that any extreme point of Δ^T_{FOSD} is an element of $\mathcal{D}$. Equivalently, we shall show the following: if $\{\mu_i^t\}_{t\in\mathcal{T}} \in \Delta^{\mathcal{T}}_{FOSD}$ is not in $\mathcal{D}$, then it is not an extreme point of Δ^T_{FOSD}.

Suppose $\{\mu_i^t\}_{t\in\mathcal{T}} \in \Delta^{\mathcal{T}}_{FOSD}\backslash\mathcal{D}$ and for each $t \in \mathcal{T}$, let $m_i^t \in X_i$ be the median of μ_i^t, i.e., $m_i^t = \inf\left\{a_i : \sum_{a_i \leqslant x_i} \mu_i^t(x_i) \geqslant 0.5\right\}$. Let α_i^t be a distribution defined in the following manner: $\alpha_i^t(x_i) = 2\mu_i^t(x_i)$ if $x_i < m_i^t$; $\alpha_i^t(x_i) = 1 - 2\sum_{x_i<m_i^t} \mu_i^t(x_i)$ if $x_i = m_i^t$; $\alpha_i^t(x_i) = 0$ if $x_i > m_i^t$. We also define the distribution β_i^t: $\beta_i^t(x_i) = 0$ if $x_i < m_i^t$; $\beta_i^t(x_i) = 1-2\sum_{x_i>m_i^t} \mu_i^t(x_i)$ if $x_i = m_i^t$; and $\beta_i^t(x_i) = 2\mu_i^t(x_i)$ if $x_i > m_i^t$. Clearly, it holds that $\mu_i^t = 0.5\alpha_i^t + 0.5\beta_i^t$ for all t. Since $\{\mu_i^t\}_{t\in\mathcal{T}} \notin \mathcal{D}$, there exists $t \in \mathcal{T}$ for which this convex combination is non-degenerate; therefore, $\{\mu_i^t\}_{t\in\mathcal{T}}$ is not an extreme point of Δ^T_{FOSD} if $\{\alpha_i^t\}_{t\in\mathcal{T}}$ and $\{\beta_i^t\}_{t\in\mathcal{T}}$ are both in Δ^T_{FOSD}. We only show this for $\{\alpha_i^t\}_{t\in\mathcal{T}}$ since the other case is similar. Suppose $\xi_i^t > \xi_i^s$ for some $s, t \in \mathcal{T}$. Since $\{\mu_i^t\}_{t\in\mathcal{T}}$ is in Δ^T_{FOSD} it must hold that $m_i^s \leqslant m_i^t$. If $a_i < m_i^s \leqslant m_i^t$, it follows from $\{\mu_i^t\}_{t\in\mathcal{T}} \in \Delta^T_{FOSD}$ that

$$\sum_{x_i \leqslant a_i} \alpha_i^t(x_i) = 2\sum_{x_i \leqslant a_i} \mu_i^t(x_i) \leqslant 2\sum_{x_i \leqslant a_i} \mu_i^s(x_i) = \sum_{x_i \leqslant a_i} \alpha_i^s(x_i).$$

If $a_i \geqslant m_i^s$, then $\sum_{x_i \leqslant a_i} \alpha_i^t(x_i) \leqslant \sum_{x_i \leqslant a_i} \alpha_i^s(x_i) = 1$. We conclude that $\alpha_i^t \geqslant_{FOSD} \alpha_i^s$. $\square$

Proof of Theorem 8: We only show the first claim, since the proof for the second is similar. Let $\mathbf{A}^{**}(y^0)$ be the set of all monotone rationalizable paths on $\{(y^t, A_i^t)\}_{t\in\{0\}\cup\mathcal{T}}$, for $y^0 = \bar{y}^0$ or $\underline{y}^0$. If $\underline{\mu} \in \text{PED}(\underline{y}^0, A^0)$ then there is a probability distribution Q on $\mathbf{A}^{**}(\underline{y}^0)$ such that for every $t \in \mathcal{T}$ and $x \in X$, $\mu^t(x) = \sum_{\mathbf{x}\in\mathbf{A}^{**}(\underline{y}^0)} Q(\mathbf{x})1(x^t = x)$, and $\underline{\mu}(x) = \sum_{\mathbf{x}\in\mathbf{A}^{**}(\underline{y}^0)} Q(\mathbf{x})1(x^0 = x)$. To each path $\mathbf{x}$ in $\mathbf{A}^{**}(\underline{y}^0)$ we associate another path $B(\mathbf{x}) = (b^0, b^1, b^2, ..., b^T)$, where $b^t = x^t$ for $t \in \mathcal{T}$, and b^0 is a predicted Nash equilibrium of the game $\mathcal{G}(\bar{y}^0, A^0)$ such that $b^0 \geqslant x^0$; Theorem 4 guarantees that b^0 exists[33] and, by construction, $B(\mathbf{x})$ is an SC-rationalizable path on the treatment set $\{(y^t, A^t)\}_{t\in\mathcal{T}} \cup \{(\bar{y}^0, A^0)\}$. Clearly, for every $t \in \mathcal{T}$ and $x \in X$, $\mu^t(x) = \sum_{\mathbf{x}\in\mathbf{A}^{**}(\underline{y}^0)} Q(\mathbf{x})1(B(\mathbf{x})^t = x)$ and therefore $\bar{\mu}$, given by $\bar{\mu}(x) = \sum_{\mathbf{x}\in\mathbf{A}^{**}(\underline{y}^0)} Q(\mathbf{x})1(B(\mathbf{x})^0 = x)$, is in $\text{PED}(\bar{y}^0, A^0)$. It is also clear from our choice of b^0 that $\bar{\mu} \geqslant_{FOSD} \underline{\mu}$. $\square$

[33]Since we assume all players have finite strategy sets, the existence of x^0 is guaranteed by Theorem 4.

Appendix II

The next tables apply the empirical design to Husband/Wife characteristics as the dependent variable to test for balance. Each row represents a separate variable. For each education level, we run a SUR model, allowing the errors to be correlated across partners. An F-test for the joint significance of all covariates at the level of Husbands and Wives is presented in the penultimate row. An F-test for the joint significance of all characteristics is presented at the bottom row. No characteristic is significant at 1%; the symbols * and ** represent significance at 10% and 5%.

	High Education					
	Husband			Wife		
Covariate	**Constant**	**Smoking Ban**	**p-value**	**Constant**	**Smoking Ban**	**p-value**
Age	40.076	-0.372 (0.230)	0.106	38.065	-0.390 (0.245)	0.112
White	0.91	0.010 (0.008)	0.238	0.908	-0.0002 (0.001)	0.986
Hispanic	0.025	-0.011 (0.007)	0.121	0.025	-0.005 (0.008)	0.526
# Children	2.025	0.0002 (0.001)	0.902	2.025	-0.0002 (0.002)	0.917
Northeast	0.226	-0.00007 (0.001)	0.936	0.225	0.0001 (0.001)	0.889
p-value		0.2863			0.7121	
p-value	0.5952					

Figure 8: Test for balance (HE)

	Low Education					
	Husband			Wife		
Covariate	**Constant**	**Smoking Ban**	**p-value**	**Constant**	**Smoking Ban**	**p-value**
Age	41.627	-0.485* (0.266)	0.068	39.010	-0.261 (0.282)	0.353
White	0.896	0.010 (0.008)	0.226	0.900	0.010 (0.009)	0.261
Hispanic	0.055	0.016** (0.008)	0.049	0.063	-0.011 (0.009)	0.219
# Children	2.02	0.0001 (0.003)	0.968	2.019	0.002 (0.004)	0.671
Northeast	0.22	-0.0001 (0.002)	0.949	0.220	-0.0009 (0.002)	0.616
p-value		0.1347			0.5813	
p-value	0.2702					

Figure 9: Test for balance (LE)

References

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

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

Carvajal, A. (2004): Testable restrictions of Nash equilibrium with continuous domains. *Discussion paper series, Royal Holloway College, University of London*, 2004-26.

Chiappori, P.-A. (1988): Rational Household Labor Supply. *Econometrica*, 56(1), 63-90.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2016): Bidimensional matching with heterogeneous preferences: education and smoking in the marriage market. *Journal of the European Economic Association*, forthcoming.

Cutler, D., and E. Glaeser (2010): Social interactions and smoking, in *Research Findings in the Economics of Aging*, ed. by D. Wise, 123-141. Chicago: University of Chicago Press.

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

Edlin, A.S. and C. Shannon (1998): Strict monotonicity in comparative statics. *Journal of Economic Theory*, 81, 201–219.

Evans, W., M. Farrelly, and E. Montgomery (1999): Do workplace smoking bans reduce smoking? *The American Economic Review*, 89, 728-747.

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

Kline, B., and E. Tamer (2012): Bounds for best response functions in binary games. *Journal of Econometric*, 166, 92-105.

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

Lazzati, N. (2015): Treatment response with social interactions: identification via monotone comparative statics. *Quantitative Economics*, 6, 49-83.

Imbens, G. (2004): Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86, 4-29.

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

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 J. Roberts (1994): Comparing equilibria. *American Economic Review*, 84, 441-459.

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.

Molinari, F., and A. Rosen (2008): The identification power of equilibrium in games: the supermodular case. *Journal of Business & Economic Statistics*, 26, 297-302.

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

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

Uetake, K., and Y. Watanabe, (2013): A note on identification and estimation of supermodular games. *Advances in Econometrics*, 31, 233-247.

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.

Zhou, L. (1994): The set of Nash equilibria of a supermodular game is a complete lattice. *Games and Economic Behavior*, 7, 295-300.

Appendix III (online, not part of main paper)

AIII.1 Data and Testing Procedures

We make use of the Tobacco Use Supplement to the Current Population Survey (TUS-CPS) to get information on both smoking decisions and workplace smoking policies. This is a National Cancer Institute (NCI) sponsored survey of tobacco use that has been administered as part of the US Census Bureau's Current Population Survey every 2 to 3 years from 1992-1993. We focus on years 1992-1993 because, unlike more recent periods, there was still a significant number of workplaces that permitted smoking. This guarantees we have enough treatment variation. We start by merging the data of years 1992 and 1993. In this initial step, we first recover the information for September 1992 and add the information for January 1993 and July 1993 regarding spouses that do not appear in the previous period(s) of time. In doing so, each couple appears only once in our pooled sample.

While the question about the smoking choice is asked to everyone in our population of interest, the question about the smoking policy in the workplace is asked only to indoor workers. Thus, we restrict attention to married couples where both members work indoors. Finally, we remove from our sample all couples where at least one member does not reply to the questions of interest. The final sample has 5,363 married couples across the US. Of that total, 2,643 couples have both members with high education levels (we explain education levels below), 1,422 couples have both members with low education levels, and the remaining 1,298 couples have one spouse with high education level and the other with low education level. (Both the raw data from the TUS-CPS and the final sample involve large files that are fully available upon request.)

We tabulate responses using the variable definitions contained in the Data Dictionaries of the Current Population Survey for years 1992-1993, as we detail next.

Married Couples: We consider as married couples all pairs where one member is the reference person and the other one responds either *"3. Husband"* or *"4. Wife"* to question A-RRP (Item 18B - Relationship to reference person).

Smoking Decisions: We assign the value N (does not smoke) to all persons that respond *"3. Not at all"* to question A-S34 (Does... now smoke cigarettes every day, some days, or not at all?). We assign value S (smokes) to all persons that respond either *"1. Every day"* or *"2. Some days"* to question A-S34 (Does... now smoke cigarettes every day, some days, or not at all?).

Smoking Restrictions at Workplace: We assign the value 0 (there are smoking restrictions at workplace) to all persons that respond *"1. Yes"* to question A-S68 (Does your place of work have an official policy that restricts smoking in anyway?). We assign a value of 1 (no smoking restrictions at workplace) to all persons that respond *"2. No"* to question A-S68 (Does your place of work have an official policy that restricts smoking in anyway?.

Education: We consider as couples with high education level (HE) all those married couples (defined above) where both members report that they have high education levels; specifically these are couples where both members respond *"40. Some college but not degree"* or above to question A-HGA (Item 18H - Educational attainment). We consider as couples with low education level (LE) all those married couples where both members respond strictly below *"40. Some college but not degree"* to question A-HGA (Item 18H - Educational attainment).

AIII.2 Test and Closest Compatible Distribution

Testing whether a data set is SC-rationalizable involves checking whether a system of linear equations

$$Ax = B \tag{31}$$

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

Matrix A: This matrix is composed of 0s and 1s. Each column describes the behavior (in terms of choices) of a specific SC-rationalizable group path. Recall that a group path specifies the profile of choices that the group makes for each possible vector of parameter values (or treatments). Each row of A corresponds to one of the 16 possible combinations of (joint) smoking choices and treatment values. In our model, A is a 16×64-matrix. In Sheet "Consistent Paths" of the file "Matrices, Data, and Results.xlsx" (included with this submission as a separate file) we describe all possible group paths for our smoking application; the SC-rationalizable group paths get number 1 in column ARC —ARC 1 and ARC 2 check the ARC axiom for the husband and the wife, respectively. In Sheet "Matrix A" of "Matrices, Data, and Results.xlsx" we show how to construct matrix A in our application.

Vector B: The size of this column vector is 16. It is composed of 4 conditional probability distributions. Each conditional distribution specifies the fraction of groups in the data that, for a given treatment, make each of the four possible joint choices. Sheet "Data" of "Matrices, Data, and Results.xlsx" describes all the information from the available data on smoking that we use to construct vector B, and shows how to construct it.

Vector x: This vector represents a probability distribution over the set of SC-rationalizable group paths —whenever the system has a positive solution. In the smoking application, x has 64 elements.

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

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

to check whether system (31) has a positive solution in x. In this specification, inputs A and B are described as above and lb corresponds to a column vector of 64 zeros. When no solution exists, Matlab reports that the primal solution appears infeasible.

For those data vectors B that do not pass this test, we use built-in function "lsqnonneg" in Matlab to find a positive vector $\widehat{x}$, with its components adding up to 1, that minimizes $(B - Ax)' (B - Ax)$. We refer to $A\widehat{x}$ as the closest compatible distribution of choices. Sheet "Results" of "Matrices, Data, and Results.xlsx" describes the closest compatible vectors in columns "Closest" for the three groups —All couples, HE couples and LE couples. As HE passes the test directly, its closest compatible vector is simply B.

AIII.3 Small Sample Inference Procedure

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

$$\textbf{(H)}: \min_{x \in \mathbb{R}_+^K} (B - Ax)' (B - Ax) = 0$$

where K is the number of SC-rationalizable group types. In the smoking application, $K = 64$. A natural sample counterpart of the objective function in **H** is given by

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

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

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

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

Kitamura and Stoye (2016) propose the following bootstrap algorithm to test **H**:

(i) Obtain a vector x^* that solves

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

and compute $\widehat{C}_{\tau_N} = Ax^*$. Here, 1_K is a vector of 1s of dimension K. Following insights from Kitamura and Stoye (2016), we use $\tau_N = \sqrt{\ln(N)/N}$, where N is the size of the sample. We also report results for $\tau_N = 0$. (The tuning parameter τ_N plays the role of a similar tuning parameter in the moment selection approach.)

(ii) Calculate the boostrap estimators under the restriction

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

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

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

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

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

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

We repeat this procedure 6 times. That is, for each of the three groups — All couples, HE couples and LE couples — we implement the test under two specifications of the τ_N-parameter. We obtain the following p-values.

	All Couples	HE	LE
$\tau_N = \sqrt{\ln(N)/N}$	0.1605	1	0.3590
$\tau_N = 0$	0.3705	1	0.5465

In Sheet "Results" of the file "Matrices, Data, and Results.xlsx" we expand on these findings. In particular, we also provide information regarding the closest compatible distribution for $\tau_N = \sqrt{\ln(N)/N}$ (in column "Closest τ") and for $\tau_N = 0$ (in column "Closest").