

# Treatment response with social interactions: Partial identification via monotone comparative statics

NATALIA LAZZATI  
University of Michigan, Ann Arbor

This paper studies (nonparametric) partial identification of treatment response with social interactions. It imposes conditions motivated by economic theory on the primitives of the model, that is, the structural equations, and shows that they imply shape restrictions on the distribution of potential outcomes via monotone comparative statics. The econometric framework is tractable and allows for counterfactual predictions in models with multiple equilibria. Under three sets of assumptions, we identify sharp distributional bounds on the potential outcomes given observable data. We illustrate our results by studying the effect of police per capita on crime rates in New York state.

**KEYWORDS.** Treatment effects, social interactions, nonparametric bounds, supermodular games, monotone comparative statics, first order stochastic dominance.

**JEL CLASSIFICATION.** C31, D71.

## 1. INTRODUCTION

This paper studies partial identification of treatment response in environments with endogenous social interdependencies. During the last three decades, social interactions have become an essential component of economic analysis. Activities we suspect are subject to strong social pressure include crimes, schooling, and fertility decisions.<sup>1</sup> Models of network goods and two-sided markets display this feature as well. Despite the attention received by this kind of interdependence in many areas of economics, just a few studies of treatment response incorporate the social dimension.<sup>2</sup> That is, standard models assume that each person's outcome varies only with her own treatment. To accommodate the applications above, we allow individual outcomes to also depend on

---

Natalia Lazzati: [nlazzati@umich.edu](mailto:nlazzati@umich.edu)

The author is deeply indebted to Keisuke Hirano and Rabah Amir, and the other members of her dissertation committee: Stanley Reynolds, Mark Walker, and John Wooders. She especially thanks Charles Manski, Lawrence Blume, Steven Durlauf, Dan Ackerberg, Amilcar Menichini, Francesca Molinari, Joerg Stoye, Dan Silverman, Christopher Flinn, Gautam Gowrisankaran, Ivana Komunjer, and the other participants at the Conference on Econometric Analysis of Social Interactions at Northwestern University, ESAM (Singapore), NYU Stern, University of Connecticut, University of Warwick, UCSD, Caltech, Cornell University, University of Arizona, University of Kansas, and University of Michigan. Last, but not least, she is very thankful to three anonymous referees and the editor, Orazio Attanasio, for their helpful suggestions.

<sup>1</sup>See Blume, Brock, Durlauf, and Ionnides (2010) for many other applications.

<sup>2</sup>See, for example, Shaikh and Vytlacil (2011).

the outcomes of other people in the population. Under this setup, we use monotone comparative statics to provide distributional bounds on treatment effects.

Models of endogenous interactions often start with the outline of a system of structural equations that define the outcome of each individual as a function of observable characteristics and the outcomes of the other group members. The solution to the system of equations (when it exists) is the predicted outcome or behavior of the group members. When there are multiple equilibria, observed behavior also depends on the mechanism by which people select among them. We assume the analyst observes a vector of realized outcomes and treatments for a sample of groups. The objective of the researcher is to learn about the potential outcome distribution that would occur in the study population if the groups were to receive a specific treatment. Our approach for identification consists of two steps. We first impose monotone restrictions on the primitives of the model (i.e., the structural equations) and derive their implications on the predicted outcomes of each group. We then translate the monotone comparative statistics results into sharp distributional bounds for the potential outcomes. Our approach is robust to the possibility of multiple equilibria.

Most of the research in econometrics is concerned with the identification of the structural equations. This is indeed the case of recent results in the econometric analysis of game theoretic models; see [Bajari, Hahn, Hong, and Ridder \(2011\)](#) for an updated overview.<sup>3</sup> Following [Manski \(2013\)](#), our objective is (partial) identification of the potential outcome distributions under alternative treatment rules, not the structural functions per se. To achieve this goal, [Manski \(2013\)](#) proceeds by imposing the identification restrictions directly on the equilibrium behavior of the agents. Our approach differs from his in that we impose all shape conditions on the primitives of the structural model and derive their implications on the solution sets. Once this connection is established, our final propositions follow from [Manski \(2013\)](#) with the sole difference that we need to adapt his proofs to multivariate distributions. We thereby provide the microfoundations for some of his novel results and make explicit the strength of the identifying assumptions.

The first two conditions we impose are as follows: the outcome of each individual increases with the outcomes of the others and it varies monotonically with the treatment to be received by the group. These conditions imply clear restrictions on the predicted outcomes: The system of structural equations leads to an increasing function that maps possible outcomes into itself, so that the set of solutions of the model coincides with the set of fixed points of this artificial function. By Tarski's fixed point theorem, the first assumption guarantees the system has a minimal and a maximal solution, that is, the model is coherent (see [Chesher and Rosen \(2012\)](#) for identification of models that may have no solution for some covariate values). The second restriction shifts the function up or down, inducing the extremal solutions to vary monotonically with the potential treatments. These two implications are akin to the main results in the literature of supermodular games (see, e.g., [Milgrom and Roberts \(1990\)](#), [Topkis \(1979\)](#), and [Vives \(1990\)](#)). We motivate coordination on extremal equilibria via simple arguments and propose an

---

<sup>3</sup>We elaborate on this connection at the end of Section 3.3.

alternative restriction that leads to the same comparative statics. Our results extend the routes for inference in Manski (1990, 1997) to models with endogenous interactions by using an approach that relates to game theory. In doing this, we bridge the econometric theory on nonparametric identification with the recent literature on supermodular games.

So far, the monotone assumptions discussed above restrict the response function of each member of the group but are silent with respect to the process of treatment selection. Many studies have established identification results by assuming the outcome functions are statistically independent of realized treatments. Manski and Pepper (2000) weaken that restriction by assuming the outcome functions are stochastically increasing in the realized treatments.<sup>4</sup> We extend this result to interactions-based models by developing an approach that allows for comparisons of equilibrium outcomes for different sets of groups. Our results relate to the methodology proposed by Amir (2008) to contrast Nash equilibria of different games. One important difference between the latter and our result is that we need to compare the distribution of equilibrium points for two sets of games, not only two games. This distributional comparative statics has no direct precedent in the theoretical literature.

We provide identification results for two models of social interactions. In the first setup we study small groups, where each member has a distinctive role, for example, men and women in married couples. The second setup is appropriate to study large neighborhoods with anonymous social interactions, for example, crimes and infectious diseases. Finally, we use the latter framework to study crime rates in New York state for different levels of enforcement that we measure by police per capita, showing the discussed monotone conditions provide quite valuable information.

The rest of the paper is organized as follows. Section 2 presents an initial example that motivates our main results. Sections 3 and 4 study identification of treatment effects for small and large groups, respectively. Section 5 uses our results to study crime rates in New York state. Section 6 concludes and we collect all the proofs in Appendixes A and B. Replication files are available in a supplementary file on the journal website, [http://qeconomics.org/supp/308/code\\_and\\_data.zip](http://qeconomics.org/supp/308/code_and_data.zip).

## 2. AN INITIAL EXAMPLE

This section uses a model similar to the one in Becker (1991) to highlight the objective of our analysis and to differentiate our identification strategy from existing results. Each person  $j \in G$  decides whether to go to a popular restaurant. Here the treatment is the price she would pay for the service and the outcome ( $y_j$ ) is a yes/no indicator taking the value 1 or 0. Consumer  $j$  maximizes her utility, taking as given the behavior of the others. Her demand for the good is given by

$$y_j = f_j \left[ t, (1/|G|) \sum_{i \in G} y_i \right], \quad j \in G, \tag{1}$$

---

<sup>4</sup>Brock and Durlauf (2007) use a similar idea to address partial identification of a model with social interactions in a semiparametric framework.

where  $t$  is the price for the service and  $(1/|G|) \sum_{i \in G} y_i$  is the average of decisions in neighborhood  $G$ . This is the source of endogeneity in our model. An equilibrium in this market is a solution to the system of structural equations (1). We indicate such a solution by

$$\mathbf{y}(t) \equiv [y_j(t), j \in G]. \quad (2)$$

The empirical evidence consists of vectors of prices and individual demands for a set of neighborhoods in the population. The analyst wants to learn about the distribution of individual demands that would occur if groups were to receive a price  $t$ , that is,  $P[y(t)]$ .

The literature on games refers to (1) as the best-reply function of person  $j$  to the profile of actions of other people in her group. Recent results in the econometrics of games provide conditions for identification of the best-reply functions. As in [Manski \(2013\)](#), our purpose is identification of  $P[y(t)]$ , not the best replies per se. Our approach differs from [Manski \(2013\)](#) in that while we impose all our assumptions on the primitives of the model (1), his restrictions are placed directly on the solution set (2). Thus, for instance, while [Manski \(2013\)](#) directly assumes that  $\mathbf{y}(t)$ , as defined in (2), increases in  $t$ , we provide conditions on the primitives (i.e., (1)) that validate this assumption.

The main restrictions we impose, which are subsequently elaborated, are as follows:

A1 (Positive Interactions). Individual  $j$ 's demand increases with the decisions of the other group members.

A2 (Monotone Treatment Response). Individual  $j$ 's demand decreases with the market price.

A3 (Monotone Treatment Selection). The owner of the restaurant chain is more likely to set higher prices to those neighborhoods with stronger demands.

We then use monotone comparative statics to show that the conditions we place on (1) imply clear restrictions on the solution set (2). Within the analysis, these results are captured by a set of lemmas. We finally exploit the shape restrictions we derive on the equilibrium sets to provide nonparametric bounds for the distribution of potential outcomes, or demands,  $P[y(t)]$ . We next formalize and extend all these ideas to frameworks where groups are small and large, respectively. The restaurant example is closer to the second setup.

### 3. IDENTIFICATION OF TREATMENT RESPONSE FOR SMALL GROUPS

This section studies identification of treatment response in situations where groups are small and each group member has a distinctive role. Models of decisions of married couples, small teams of co-workers, and pairs of patients and doctors fit in here.

### 3.1 The model and the analyst's problem

The population  $J$  is partitioned into a finite set of classes  $L = 1, 2, \dots, |L|$ , that is,  $J = (J_1, J_2, \dots, J_{|L|})$ . Each group  $G$  is composed of one individual from each class. We indicate by  $t \in T$  a potential treatment and allow  $t$  to specify policies that may vary across classes. The behavior or achievement of agent  $jl \in G$ ,  $y_{jl} \in Y \subseteq \mathbb{R}$  with  $jl \in J_l$ , depends on the treatment and the behavior of the other group members,  $\mathbf{y}_{-jl} = (y_{jm}, jm \in G, m \neq l)$ . That is,

$$y_{jl} = f_{jl}(t, \mathbf{y}_{-jl}) \quad \text{with } jl \in G. \quad (3)$$

We assume social interactions occur within groups and group membership is known to the econometrician. If the underlying model is a complete information game, then we can think of (3) as the best-reply function of player  $jl$  to the profile of actions of the other players in  $G$ . A simultaneous solution to the system (3) is a vector of potential outcomes

$$\mathbf{y}(t) = (y_{jl}(t), jl \in G) \quad \text{with } \mathbf{y}(t) \in Y^{|L|}. \quad (4)$$

We typify groups through the outcome functions of the group members and relate the distribution of types of groups to the underlying process of group formation. The next example illustrates the probabilistic approach we describe afterward.

**EXAMPLE 1.** We model a scenario of a tobacco prevention program. The population has six people, three boys and three girls. Thus,  $J = (J_1, J_2)$ , where  $J_l = (1l, 2l, 3l)$  with  $l = 1, 2$  is the set of boys and girls, respectively. Each group is a dating couple and its members decide whether to smoke. The treatment is 1 if the couple receives information about the risks of smoking and is 0 otherwise. An individual's decision about smoking depends on the treatment received and the smoking decision of his or her partner. Thus,  $T = \{0, 1\}$ ,  $Y = \{0, 1\}$ , and  $f_{jl}: T \times Y \rightarrow Y$  with  $j = 1, 2, 3$  and  $l = 1, 2$ .

Couples differ with respect to their outcome functions, that is, in terms of the way they react to the treatments and partners' decisions. In this example, we have  $2^4$  (or 16) possible types of people—for each of the four possible configurations of treatments and partner's decisions, there are two possible actions—and then  $16^2$  (or 256) types of couples. Let us assume the population has outcome functions as those in Figure 1. In this figure, the 1 in square brackets means the second boy will smoke if he does not receive information and his girlfriend smokes.

The population described in Figure 1 involves only two types of girls and boys (out of 16) and then four types of couples (out of 256). The distribution of types of couples depends on both the smoking and the dating preferences of the individuals. Suppose these individuals prefer to date rather than remaining alone and to engage with individuals with similar smoking tastes. Thus, two out of the three couples will have members who smoke if and only if they are uninformed and their partners smoke, and the other couple will have members who smoke if and only if their partners smoke irrespective of the treatment. The other two types of groups have no chance of being formed given the assumed dating preferences.

Outcomes	Treatments and Partners' Decisions			
	(0, 0)	(0, 1)	(1, 0)	(1, 1)
Boys				
$y_{11}$	0	1	0	0
$y_{21}$	0	[1]	0	0
$y_{31}$	0	1	0	1
Girls				
$y_{12}$	0	1	0	0
$y_{22}$	0	1	0	0
$y_{32}$	0	1	0	1

FIGURE 1. Structural functions of the population.

In our model, individuals may differ with respect to the way they react to the treatments and the decisions of the other group members. We let  $T$  be countable and  $Y$  be finite with  $\underline{Y} = \min(Y)$  and  $\overline{Y} = \max(Y)$ .<sup>5</sup> Thus, there are countably many different systems of structural equations (3) that can describe a group. We say a group is of type  $k$  if its members have structural equations  $\mathbf{f}_k(t, \mathbf{y}) \equiv [f_{kl}(t, \mathbf{y}_{-l}), l \in L]$  with  $f_{kl}: T \times Y^{|L|-1} \rightarrow Y$ . We let  $K$  indicate the set of possible types of groups. Since random group formation is often hard to motivate, we do not impose such a restriction. Instead, we think that the relative proportion of different agents within each class and the underlying matching process define the distribution of types of groups in the universe of groups  $\mathcal{U}$ . Let  $\pi_k$  denote the fraction of type- $k$  groups, so that  $\boldsymbol{\pi} \equiv (\pi_k, k \in K)$  is the distribution of types in  $\mathcal{U}$ . In Example 1,  $K = \{1, 2, \dots, 256\}$  and  $\boldsymbol{\pi} = (2/3, 1/3, 0, \dots, 0)$ . We assume group formation is independent of  $t$ , in the sense that potential treatments do not affect the distribution of types of groups,  $\boldsymbol{\pi}$ .<sup>6</sup> Nevertheless, realized treatments may provide rich information about the type of group that could have generated the data (see condition A3 in Section 2).

In summary,  $[(\mathbf{f}_k, \pi_k), k \in K]$  describes the universe of groups,  $\mathcal{U}$ , in the population  $J$ . Groups have observable realized treatments  $\tau^m$  and outcomes  $\mathbf{y}^m \equiv (y_1^m, y_2^m, \dots, y_{|L|}^m)$ , so that the available data are  $[(\tau^m, \mathbf{y}^m), m \in M]$ . The researcher wants to learn about the joint outcome distribution that would occur in the population  $J$  if the groups were to receive a treatment  $t$ , that is,  $P[\mathbf{y}(t)]$ , where  $\mathbf{y}(t) \equiv [y_1(t), y_2(t), \dots, y_{|L|}(t)]$  is a random vector.

<sup>5</sup>These two restrictions are introduced for simplicity. We conjecture that it may be possible to extend the results to uncountable sets by taking appropriate care of measurability conditions. Manski (2007) uses a similar approach to develop partial identification of counterfactual choice probabilities.

<sup>6</sup>Manski (2013) describes a similar condition by saying that reference groups are treatment-invariant, and thus nonmanipulable (i.e., the social planner cannot use the treatments to change a person's influence group).

**REMARK.** Though the primitives in our model,  $\mathbf{f}_k(t, \mathbf{y})$ , do not incorporate covariate information, we can assume that this information has been taken into account. We just need to interpret the primitives as conditional on some covariate values.

The next example provides an application of the model we just presented.

**EXAMPLE 2.** This model can be used to study retirement decisions of husbands and wives. Let  $J_1$  be the set of men and  $J_2$  be the set of women, with  $\mathcal{U}$  defined as the set of all married couples in  $J$ . Let the outcome of interest be the retirement age and let the treatment be the income tax. Many studies argue that endogenous interactions are important within couples as spouses will obtain more pleasure from retirement if they retire together.

### 3.2 Monotone assumptions and their implications

We next impose three sets of conditions on the structural equations and derive their implications in terms of equilibrium behavior. We then use these results to provide bounds for  $P[\mathbf{y}(t)]$ .

*Coherence of the model* This section shows equilibrium existence and relates observed and equilibrium outcomes with the primitives of the model.

Without further restrictions, the system of equations (3) might have no solution. When this happens our equilibrium concept has no predictive power. We next provide a restriction that precludes this possibility. Throughout, we indicate by  $\geq$  the standard coordinatewise order.

**A1 (Positive Interactions).** For each  $k \in K$ ,  $l \in L$ , and  $t \in T$ ,  $f_{kl}(t, \mathbf{y}) \geq f_{kl}(t, \mathbf{y}')$  for all  $\mathbf{y}, \mathbf{y}' \in Y^{|L|-1}$  with  $\mathbf{y} \geq \mathbf{y}'$ .

Condition A1 requires the outcome of each individual to increase with the outcomes of the other group members. In Example 2, this condition requires that each member of the married couple be more willing to retire whenever his or her partner retires.

Among the models that satisfy A1, we have the supermodular games.<sup>7</sup> Let  $U_{kl}(y_l, t, \mathbf{y})$  be the payoff function of a type- $l$  agent in a type- $k$  group who gets treatment  $t$  and let  $f_{kl}(t, \mathbf{y}) = \arg \max_{y_l \in Y} U_{kl}(y_l, t, \mathbf{y})$ . Then A1 holds if, for all  $y_l \geq y'_l$  and  $\mathbf{y} \geq \mathbf{y}'$ ,

$$U_{kl}(y_l, t, \mathbf{y}) - U_{kl}(y'_l, t, \mathbf{y}) \geq U_{kl}(y_l, t, \mathbf{y}') - U_{kl}(y'_l, t, \mathbf{y}'). \quad (5)$$

This condition states that the extra payoff of selecting a high over a low action increases with the action profile of the others.<sup>8</sup> This is the distinctive requirement of any supermodular game.<sup>9</sup>

Condition A1 is also satisfied by models with positive externalities, for example, peer effects in the classroom.

---

<sup>7</sup>See, for example, Milgrom and Roberts (1990), Vives (1990), and Topkis (1979).

<sup>8</sup>This analysis implicitly assumes the maximizer of  $U_{kl}(y_l, t, \mathbf{y})$  exists and is unique.

<sup>9</sup>Condition (5) could be relaxed by using the results in Milgrom and Shannon (1994).

Let  $\phi(t, k)$  denote the solution set of the system of structural equations for a given  $t, k$ . The next result states the existence of extremal equilibria.

**LEMMA 1.** *If A1 holds, then  $\phi(t, k)$  has a least and a greatest solution for all  $k \in K$  and  $t \in T$ .*

We next offer a sketch of proof. Let  $M_{t,k} : Y^{|L|} \rightarrow Y^{|L|}$  be defined as

$$\begin{aligned} M_{t,k}(y_1, y_2, \dots, y_{|L|}) &= (y'_{k1}, y'_{k2}, \dots, y'_{k|L|}) \\ \text{with } y'_{kl} &= f_{kl}(t, y_{-l}) \text{ for all } l \in L. \end{aligned} \tag{6}$$

By construction, the set of fixed points of  $M_{t,k}$  coincides with  $\phi(t, k)$ . Then this proof reduces to showing that the set of fixed points of (6) has a least and a greatest element. Under positive interactions,  $M_{t,k}$  is increasing. Thus, the result follows by Tarski's fixed point theorem.<sup>10</sup>

While condition A1 guarantees equilibrium existence, it does not eliminate the possibility of multiple solutions. Thus, to complete the model, we need to introduce an equilibrium selection mechanism. In this study, we restrict attention to deterministic equilibrium selection rules that depend on  $k$  and  $t$ . These types of selection mechanisms are widely used in economic theory. One of the reasons is that they can often be motivated by either Pareto dominance or simple learning processes in which the individuals adjust their choices based on the observed behavior of the others.<sup>11</sup> The experimental literature on coordination in games offers substantial support to our approach; see, for example, Cooper, DeJong, Forsythe, and Ross (1992). In contrast to our restriction, stochastic selection rules have been used for identification and testability in previous studies; see, for example, Bajari et al. (2011) and Echenique and Komunjer (2009).

Let  $y_k(t)$  be the element of  $\phi(t, k)$  that is selected by a type- $k$  group that gets treatment  $t$ . Thus, the probability that the vector of potential outcomes falls in  $B \subseteq \mathbb{R}^{|L|}$  is given by

$$P[y(t) \in B] = \sum_{k \in K} 1[y_k(t) \in B] \pi_k. \tag{7}$$

Let  $\pi_{k|\tau}$  indicate the fraction of groups in  $\mathcal{U}$  that are of type  $k$  conditional on  $\tau$ , and let  $P(\tau)$  denote the distribution of realized treatments across groups. We institute the convention that  $\pi_{k|\tau} \equiv 0$  if the conditioning event does not hold. The probability that the joint vector of realized outcomes falls in a set  $B \subseteq \mathbb{R}^{|L|}$  is given by

$$P(y \in B) = \sum_{s \in T} P(y \in B | \tau = s) P(\tau = s) \tag{8}$$

with  $P(y \in B | \tau = s) = \sum_{k \in K} 1[y_k(s) \in B] \pi_{k|\tau=s}$ . Then  $P(y)$  is a mixture of realized outcome distributions conditional on observed treatments, with  $P(\tau)$  as the mixing probability function. We next evaluate the implications of two other monotone restrictions.

---

<sup>10</sup>A similar approach has been used by Ackerberg and Gowrisankaran (2006), Jia (2008), and De Paula (2009).

<sup>11</sup>See, for example, Amir and Lazzati (2011), Monderer and Shapley (1996), and Oyama, Sandholm, and Tercieux (2012).

*Monotone treatment response* We now assume the structural equations are weakly increasing in potential treatments. As we explain below, we get opposite results if we reverse the signs of these effects.

A2 (Monotone Treatment Response). For each  $k \in K$ ,  $l \in L$ , and  $\mathbf{y} \in Y^{|L|-1}$ ,  $f_{kl}(t, \mathbf{y}) \geq f_{kl}(t', \mathbf{y})$  for all  $t, t' \in T$  with  $t \geq t'$ .

When the outcome functions correspond to the best replies of a supermodular game, A2 is satisfied if, for all  $y_l \geq y'_l$  and  $t \geq t'$ ,

$$U_{kl}(y_l, t, \mathbf{y}) - U_{kl}(y'_l, t, \mathbf{y}) \geq U_{kl}(y_l, t', \mathbf{y}) - U_{kl}(y'_l, t', \mathbf{y}). \quad (9)$$

The interpretation of this condition is similar to that of condition (5). In Example 2, condition A2 would require that people be more willing to retire when the income tax is larger.

Without further restrictions, monotone treatment response allows counterfactual predictions at the extremal elements of the solution set. That is, if A1 and A2 hold, then mapping  $M_{t,k}$  in (6) shifts up as  $t$  increases and the extremal solutions in  $\phi(t, k)$  move in the same direction. It follows that if the groups always select either the smallest or the largest equilibrium, then we can predict the directional effect of  $t$  on the distribution of potential outcomes. The same result holds under an alternative condition on the structural equations. We formalize these restrictions next and elaborate on them afterward.

S1 (Equilibrium Selection). One of the following conditions holds: (i) each group selects either the smallest or the largest element of  $\phi(t, k)$  and the selection rule (weakly) increases in  $t$  or (ii)  $\mathbf{f}_k(t, \inf \phi(t', k)) \geq \sup \phi(t', k) \forall t, t' \in T$  such that  $t \geq t'$  and  $\forall k \in K$ .

Condition S1(i) can be motivated in at least two ways. The first justification is based on Pareto dominance. The models we discuss here often allow us to Pareto rank the equilibrium outcomes (or, as in Brock and Durlauf (2001), to rank them in terms of expected welfare). If the structural functions correspond to the best replies of a supermodular game, then the largest equilibrium Pareto dominates all the others if, for all  $\mathbf{y} \geq \mathbf{y}'$ ,

$$U_{kl}(y_l, t, \mathbf{y}) \geq U_{kl}(y_l, t, \mathbf{y}'),$$

that is, when the payoff functions display positive spillovers. The opposite is true whenever the spillovers are negative. Cooper et al. (1992) provide evidence that this selection is more likely to hold under preplay communication among the agents. Though strong, this selection rule has already been used in the empirical literature (see, e.g., Gowrisankaran and Stavins (2004)). As we explain at the end of this section, this assumption could be relaxed with a richer data set that contains exclusion restrictions at the individual level.

The second justification relies on a simple dynamic process. Let us define an adaptive dynamics  $\mathcal{A}(\mathbf{y}, t, k)$  as a sequence  $\{\mathbf{y}^i\}_{i=0}^\infty$  such that

$$\mathbf{y}^0 = \mathbf{y}, \quad \mathbf{y}^i = \mathbf{f}_k(t, \mathbf{y}^{i-1}), \quad i \geq 1.$$

This dynamic process begins with an initial vector  $\mathbf{y}^0$ , then assumes people behave according to their respective outcome functions inducing a new vector  $\mathbf{y}^1$ , and the process repeats indefinitely. The literature on games refers to this process as either best-response dynamics or myopic best reply. We say the adaptive dynamics are pessimistic if the elements of  $\mathcal{A}(\mathbf{y}, t, k)$  with  $\mathbf{y} = \underline{Y}^{[L]}$  describe the evolution of outcomes as we apply treatment  $t$ . We say they are optimistic if this evolution is alternatively described by  $\mathcal{A}(\mathbf{y}, t, k)$  with  $\mathbf{y} = \overline{Y}^{[L]}$ . It can be shown that if the adaptive dynamics are pessimistic, then this process converges to the smallest element of  $\phi(t, k)$ , and it converges to the largest element whenever they are optimistic. Thus, S1(i) can be motivated by assuming that the cross-sectional distribution of outcomes arises as the limit of either pessimistic or optimistic adaptive dynamics.<sup>12</sup>

Condition S1(ii) is quite different from the last one. It ensures that  $\phi(t, k)$  is strongly increasing in  $t$ , in the sense that when  $t$  increases, the smallest element of the new solution set is higher than the largest element of the old one.<sup>13</sup> Thus, any selection from the equilibrium set is necessarily increasing in the potential treatment. The drawback of this condition is that it is nonprimitive, that is, to check it requires knowledge of the extremal elements of the solution sets. It is trivially satisfied when the equilibrium is unique, a common assumption in the empirical literature (see, e.g., Glaeser, Sacerdote, and Scheinkman (1996)) that would hold here if, for instance, the system of structural equations were a contraction. Condition S1(ii) requires a very strong treatment effect.

We next state that under all previous restrictions the equilibrium outcomes increase in  $t$ .

**LEMMA 2.** *If A1, A2, and S1 hold, then  $\mathbf{y}_k(t) \geq \mathbf{y}_k(t')$  for all  $t, t' \in T$  with  $t \geq t'$  and  $k \in K$ .*

**REMARK.** Lemma 2 exploits two types of monotone shape restrictions on the outcome function of each group member: First, monotonicity regarding other people choices, that is, A1; second, monotonicity regarding the treatment, that is, A2. These two types of monotone restrictions play different roles and deserve different attention. Condition A1 guarantees the existence of extremal equilibria via Tarski's fixed point theorem. Given the approach we follow, the potential outcomes need to be increasing in other people choices. In other words, our method only applies to models with positive strategic interactions. On the other hand, condition A2 allows us to obtain the monotone comparative statics results that we exploit for counterfactual predictions. In this case, the potential outcomes can be either increasing or decreasing in the treatment vector. When they are increasing and S1 holds,  $\mathbf{y}_k(t) \geq \mathbf{y}_k(t')$  for all  $t \geq t'$ . When they are decreasing and the dual of S1 holds,  $\mathbf{y}_k(t) \leq \mathbf{y}_k(t')$  for all  $t \geq t'$ .<sup>14</sup> The proof for the decreasing case is anal-

---

<sup>12</sup>Suppose we want to know what would happen if we increased the treatment level of a certain group from  $t'$  to  $t$ . If the evolution of outcomes is described by  $\mathcal{A}(\mathbf{y}, t, k)$  with  $\mathbf{y} = \mathbf{y}(t')$ , then the monotone comparative statics in Lemma 2 holds as well. To observe this possibility in our model, we should simply allow the selection rule to depend on the previous treatment.

<sup>13</sup>See Echenique and Sabarwal (2003).

<sup>14</sup>By the dual of S1 we mean that one of the following conditions holds: (i) each group selects either the smallest or the largest element of  $\phi(t, k)$  and the selection rule (weakly) decreases in  $t$  or (ii)  $\mathbf{f}_k(t, \sup \phi(t', k)) \leq \inf \phi(t', k) \forall t, t' \in T$  such that  $t \geq t'$  and  $\forall k \in K$ .

ogous to the proof for the increasing one: we simply need to reverse the order of the treatment (i.e., to use  $-t$  instead of  $t$ ) and apply identical steps.

The previous assumptions restrict the response function of each group member, but are silent with respect to the process of treatment selection. The next restriction fills this gap.

*Monotone treatment selection* Many studies have established identification by assuming the outcome functions are statistically independent of realized treatments, that is,  $P[\mathbf{y}(t) | \tau = t] = P[\mathbf{y}(t)]$ . In our case, this condition would be plausible if an explicit randomization mechanism had been used to assign treatments to the groups, so that  $\pi_{k|\tau} = \pi_k$ . Recognizing that this assumption is often hard to justify, [Manski and Pepper \(2000\)](#) alternatively assume the outcome distributions are stochastically increasing in realized treatments, referring to this condition as monotone treatment selection. We next extend their finding to interactions-based models.

To facilitate the introduction of the required conditions in this section, we need to define a natural partial order on  $(\mathbf{f}_k, k \in K)$ .

**DEFINITION 3.** We say  $\mathbf{f}_k \geq \mathbf{f}_{k'}$  if  $f_{kl}(t, \mathbf{y}) \geq f_{k'l}(t, \mathbf{y}) \forall l \in L$  and  $\forall (t, \mathbf{y}) \in T \times Y^{|L|-1}$ .

That is, we say  $\mathbf{f}_k$  is larger than  $\mathbf{f}_{k'}$  if it is pointwise greater. Appendix A (Proofs) shows that (under A1) if  $\mathbf{f}_k \geq \mathbf{f}_{k'}$ , then the least and the greatest solutions in  $\phi(t, k)$  are larger than the corresponding solutions in  $\phi(t, k')$ . Let  $\mathbf{f}$  denote a random vector of functions with support on  $(\mathbf{f}_k, k \in K)$ , and let us define  $P(\mathbf{f} = \mathbf{f}_k | \tau) \equiv \pi_{k|\tau}$ . The next assumption allows us to compare equilibrium distributions for two sets of groups that received different treatments.

**A3 (Monotone Treatment Response).** We have  $P(\mathbf{f} | \tau = s) \geq_{st} P(\mathbf{f} | \tau = s')$  for all  $s \geq s'$ .

This condition simply states that the structural functions are statistically increasing in realized treatments. Depending on the context, an alternative interpretation would be that groups that self-select into higher treatments have stochastically larger structural functions than those that self-select into lower ones. Here again the identification power of monotone treatment selection relies on an additional restriction.

**S2 (Equilibrium Selection).** One of the following conditions holds: (i) groups select either the smallest or the largest element of the solution set and (for each  $t \in T$ ) the selection rule is the same  $\forall k \in K$  or (ii)  $\mathbf{f}_k(t, \inf \phi(t, k')) \geq \sup \phi(t, k') \forall k, k' \in K$  such that  $\mathbf{f}_k \geq \mathbf{f}_{k'}$ .

Condition S2(i) requires the selection rule to be the same across groups, a natural restriction given that we want to compare potential outcomes across them. Whenever the model allows us to claim that either the smallest or the largest equilibrium Pareto dominates the others—as in the case of supermodular games with negative and positive spillovers, respectively—this condition can be justified by assuming a selection

mechanism based on Pareto dominance. Condition S2(ii) guarantees that for every pair  $k, k' \in K$  such that  $\mathbf{f}_k \geq \mathbf{f}_{k'}$ , the solution set  $\phi(t, k)$  is strongly greater than  $\phi(t, k')$  in the sense that the smallest element of the first solution set is higher than the largest element of the second set.

If A1 holds, then the added benefit of S2 is that now  $\mathbf{y}_k(t) \geq \mathbf{y}_{k'}(t)$  for all  $k, k' \in K$  such that  $\mathbf{f}_k \geq \mathbf{f}_{k'}$ . The next lemma derives from this observation. This distributional comparative statics result has no direct precedent in the theoretical literature of games.

**LEMMA 4.** *Assume A1, A3, and S2 hold. Let  $s, s' \in T$ . Then, for all  $t \in T$ ,*

$$s \geq s' \implies P[\mathbf{y}(t) | \tau = s] \geq_{\text{st}} P[\mathbf{y}(t) | \tau = s']. \quad (10)$$

**REMARK.** The remark below Lemma 2 applies here as well. That is, we obtain the opposite inequality if we assume negative treatment selection and correspondingly accommodate S2.

The next section uses all previous results to provide bounds for  $P[\mathbf{y}(t)]$ . Before doing so, we connect our approach with the econometric literature on games.

*Connection with games* Let us interpret  $y_{jl} = f_{jl}(t, \mathbf{y}_{-jl})$  with  $jl \in G$  as the best-reply function of player  $jl$  to the profile of actions of the other players in her group. Our approach differs from the current econometric literature on games in that we impose conditions on the best-reply functions, but do not intend to identify them. In fact, given the data we assume is available, we cannot do so. In a companion paper, Lazzati (2014) imposes similar monotone conditions on the primitives of the model and identifies bounds for the best-reply functions by taking advantage of exclusion restrictions.<sup>15</sup> She then uses the extremal selections of these bounds to provide partial identification results for equilibrium behavior. This second paper is based on a different data set, but allows us, on the other hand, to dispense with any restriction regarding the equilibrium selection mechanism.<sup>16</sup>

### 3.3 Identification region for $P[\mathbf{y}(t)]$

The identification problem can be formulated in terms of a simple decomposition that follows from the law of total probability:

$$P[\mathbf{y}(t)] = P[\mathbf{y}(t) | \tau = t]P(\tau = t) + P[\mathbf{y}(t) | \tau \neq t]P(\tau \neq t). \quad (11)$$

Under A1, the empirical evidence reveals  $P[\mathbf{y}(t) | \tau = t] = P(\mathbf{y} | \tau = t)$ ,  $P(\tau = t)$ , and  $P(\tau \neq t)$ .<sup>17</sup> The sampling process alone remains silent about the potential outcome distribution for those groups that have realized treatments that differ from the potential

<sup>15</sup>By exclusion restrictions, we mean variables that, while affecting the payoff of one member of the group, can be excluded from directly affecting the payoffs of the other group members. These exclusion restrictions play a key role in identifying best-reply functions in the empirical literature on games.

<sup>16</sup>See also Molinari and Rosen (2008).

<sup>17</sup>Recall that we assume that each group type behaves according to one element of the solution set for each potential treatment that they might receive. If the solution set is nonempty, then the latter guarantees that equality (11) holds. Condition A1 ensures that the solution set has at least one element.

one, that is,  $P[\mathbf{y}(t) \mid \tau \neq t]$ . However, the set of possible vectors of outcomes is bounded from below and above by  $\underline{Y}^{|L|}$  and  $\bar{Y}^{|L|}$ , respectively. Then, without any other restriction, the degenerate distributions  $P(\underline{Y}^{|L|})$  and  $P(\bar{Y}^{|L|})$  are sharp lower and upper bounds for  $P[\mathbf{y}(t) \mid \tau \neq t]$ .

Let  $\Delta_{Y^{|L|}}$  be the set of multivariate distributions with domain on  $Y^{|L|}$  and let  $H$  be the set of distributions that are consistent with our assumptions given the data, that is, the sharp identification region for  $P[\mathbf{y}(t)]$ . In the next result, “st” means stochastic dominance.<sup>18</sup>

**PROPOSITION 5.** *Assume A1 holds. Then, for all  $t \in T$ ,*

$$\begin{aligned} H\{P[\mathbf{y}(t)]\} &= \{\delta \in \Delta_{Y^{|L|}} : P(\mathbf{y} \mid \tau = t)P(\tau = t) + P(\bar{Y}^{|L|})P(\tau \neq t) \\ &\quad \geq_{\text{st}} \delta \geq_{\text{st}} P(\mathbf{y} \mid \tau = t)P(\tau = t) + P(\underline{Y}^{|L|})P(\tau \neq t)\}. \end{aligned}$$

At the beginning of the last section, we showed equilibrium existence and related observed and equilibrium outcomes with the primitives of the model. Given these results, Proposition 5 builds directly on Manski (1990, 2013). The only difference is that, in our model,  $\mathbf{y}(t)$  indicates a vector of outcomes. Thus,  $\geq_{\text{st}}$  partially orders multivariate distribution functions and the proof of Proposition 5 correspondingly modifies those of Manski (1990, 2013).

The endpoints of Proposition 5 almost always differ, so  $P[\mathbf{y}(t)]$  is typically only partially identified. Moreover, if  $\Pr(\tau \neq t) = 1$ , then these bounds are sharp but completely uninformative. We next show that the identification region can be substantially tightened by adding either monotone treatment response or monotone treatment selection.

If A1, A2, and S1 hold, then, by Lemma 2, the predicted outcomes  $\mathbf{y}_k(t)$  increase in  $t$ . Let us consider a type- $k$  group with empirical evidence  $(\tau^m, \mathbf{y}^m)$ . If  $\tau^m \leq t$ , then  $\mathbf{y}^m$  is a sharp lower bound for  $\mathbf{y}_k(t)$ ; otherwise, the empirical evidence is uninformative and  $\underline{Y}^{|L|}$  is the sharp lower bound. Alternatively, if  $\tau^m \geq t$ , then  $\mathbf{y}^m$  is a sharp upper bound for  $\mathbf{y}_k(t)$ ; otherwise, the sharp upper bound is just the greatest possible vector of outcomes,  $\bar{Y}^{|L|}$ . Since  $k$  was arbitrarily chosen, this analysis extends to all groups in  $\mathcal{U}$  and justifies the next result.

**PROPOSITION 6.** *Assume A1, A2, and S1 hold. Then, for all  $t \in T$ ,*

$$\begin{aligned} H\{P[\mathbf{y}(t)]\} &= \{\delta \in \Delta_{Y^{|L|}} : P(\mathbf{y} \mid \tau \geq t)P(\tau \geq t) + P(\bar{Y}^{|L|})P(\tau \not\geq t) \\ &\quad \geq_{\text{st}} \delta \geq_{\text{st}} P(\mathbf{y} \mid \tau \leq t)P(\tau \leq t) + P(\underline{Y}^{|L|})P(\tau \not\leq t)\}. \end{aligned}$$

We now invoke monotone treatment selection. If A1, A3, and S2 hold, then  $P(\mathbf{y} \mid \tau = t)$  is a lower bound for  $P[\mathbf{y}(t) \mid \tau \geq t]$  and an upper bound for  $P[\mathbf{y}(t) \mid \tau \leq t]$ . These observations are direct implications of Lemma 4. The identification result is as follows.

---

<sup>18</sup>See Appendix A for the typical multivariate characterization of first order stochastic dominance.

**PROPOSITION 7.** *Assume A1, A3, and S2 hold. Then, for all  $t \in T$ ,*

$$\begin{aligned} H\{P[\mathbf{y}(t)]\} &= \{\delta \in \Delta_{Y^{|L|}} : P(\mathbf{y} | \tau = t)P(\tau \leq t) + P(\bar{Y}^{|L|})P(\tau \not\leq t) \\ &\geq_{st} \delta \geq_{st} P(\mathbf{y} | \tau = t)P(\tau \geq t) + P(\underline{Y}^{|L|})P(\tau \not\geq t)\}. \end{aligned}$$

The multivariate standard stochastic order is closed with respect to marginalization. Thus, Propositions 5, 6, and 7 also imply sharp bounds for the marginal distributions of potential outcomes for individuals who belong to a subset of classes  $S \subset L$ , that is,  $P[\mathbf{y}_S(t)]$ , where  $\mathbf{y}_S(t)$  is the restriction of  $\mathbf{y}(t)$  to  $S$ .<sup>19</sup> In the next result, we write  $L/S$  for the set of classes in  $L$  different from  $S$ . Corollary 8 (without proof) formalizes this claim for the monotone treatment selection restriction; a similar idea can be applied to our first two propositions.<sup>20</sup>

**COROLLARY 8.** *Assume A1, A3, and S2 hold. Then*

$$\begin{aligned} H\{P[\mathbf{y}_S(t)]\} &= \{\delta \in \Delta_{Y^{|S|}} : P(\mathbf{y}_S | \tau = t)P(\tau \leq t) + P(\bar{Y}^{|S|})P(\tau \not\leq t) \\ &\geq_{st} \delta \geq_{st} P(\mathbf{y}_S | \tau = t)P(\tau \geq t) + P(\underline{Y}^{|S|})P(\tau \not\geq t)\} \end{aligned}$$

for all  $t \in T$ , where  $P(\mathbf{y}_S | \cdot) = E_{\mathbf{y}_{L/S}}[P(\mathbf{y} | \cdot)]$ .

Although the sets of assumptions (A1, A2, and S1) and (A1, A3, and S2) are not individually refutable, the combined restriction can be shown to be false given the data. To see why, note that if these two sets of restrictions hold, then, for all  $s > s'$ ,

$$\begin{aligned} P(\mathbf{y} | \tau = s) &= P(\mathbf{y}(s) | \tau = s) \geq_{st} P(\mathbf{y}(s) | \tau = s') \\ &\geq_{st} P(\mathbf{y}(s') | \tau = s') = P(\mathbf{y} | \tau = s'), \end{aligned}$$

where the right inequality follows by (A1, A2, and S1) and the left inequality follows by (A1, A3, and S2). Thus, the joint restriction implies that  $P(\mathbf{y} | \tau = s) \geq_{st} P(\mathbf{y} | \tau = s')$  for all  $s > s'$ .<sup>21</sup> It is interesting to note that this joint restriction is no longer testable when the monotone treatment response and the monotone treatment selection assumptions work in opposite directions, as in the case of crimes we study in Section 5. To see why, let

<sup>19</sup>Corollary 8 derives from Proposition 7, using Theorem 3.3.10 of Müller and Stoyan (2002, p. 94).

<sup>20</sup>We can also construct sharp bounds for other measures of interest, such as differences of expectations  $E[\mathbf{y}(t)] - E[\mathbf{y}(t')]$  with  $t > t'$ . If we only assume monotone treatment selection or (A1, A3, and S2), then the sharp upper bound for the latter is just the difference between the upper bound of the expectation of the vector of potential outcomes for the high treatment level and the lower bound for the expectation of the vector of potential outcomes for the low treatment level; the sharp lower bound can be obtained in a similar fashion. Under monotone treatment selection, the difference in expectations cannot be signed a priori. The reason is that this assumption does not impose any restriction on the distribution of potential outcomes across different potential treatments. The result changes if we use monotone treatment response or (A1, A2, and S1). In this case, the lower bound for the difference of expectations is always positive.

<sup>21</sup>In a recent article, Lee, Linton, and Whang (2009) propose a test for stochastic monotonicity that could be used to check the validity of the joint requirement.

us assume that the monotone treatment response is negative and the monotone treatment selection is positive. Then, for each  $s > s'$ , all we can say is

$$P(\mathbf{y} | \tau = s) \geq_{st} P(\mathbf{y}(s) | \tau = s') \quad \text{and} \quad P(\mathbf{y} | \tau = s') \geq_{st} P(\mathbf{y}(s) | \tau = s'),$$

where the right inequality follows from negative treatment response and the left inequality follows from monotone treatment selection. This information does not suffice to compare the observable distributions  $P(\mathbf{y} | \tau = s)$  and  $P(\mathbf{y} | \tau = s')$ .

All our identified bounds can be consistently estimated from sample data of a subset of groups under regularity conditions. The analyst just needs to substitute all the empirical distributions with their sample analogs. The techniques developed by [Andrews and Soares \(2010\)](#), [Imbens and Manski \(2004\)](#), [Rosen \(2008\)](#), and [Stoye \(2009\)](#), among others, could be adapted to construct confidence sets for the identified regions. The asymptotic properties of the estimates depend on sampling a large number of groups.

#### 4. IDENTIFICATION OF TREATMENT RESPONSE FOR LARGE GROUPS

In this section, groups are large (but finite) and social interactions are anonymous. Models of crimes, schooling, infectious diseases, and addictions fit well here. An important difference between this section and the previous one is that groups may differ in the number of people. This feature makes the comparative statics harder.

##### 4.1 The model and the analyst's problem

For each treatment  $t \in T$  to be received by group  $G$  in the population  $J$ , the vector of potential outcomes  $\mathbf{y}(t) \equiv [y_j(t), j \in G] \in Y^{|G|}$  solves the system of structural equations

$$y_j = f_j[t, P_G(y)] \quad \text{with } j \in G, \tag{12}$$

where  $P_G(y)$  is the distribution of outcomes induced by vector  $\mathbf{y} \equiv (y_j, j \in G)$ .<sup>22</sup>

We say a group is of type  $k$  if its members have structural equations  $\mathbf{f}_k \equiv [f_{kl}(\cdot), l \in G_k]$ , where  $f_{kl}: T \times \Delta Y \rightarrow Y$ . Let  $\pi_k$  denote the fraction of individuals who belong to type- $k$  groups, so that  $\boldsymbol{\pi} \equiv (\pi_k, k \in K)$  is the discrete distribution of types. Since groups are allowed to differ with respect to the number of members,  $\pi_k$  depends on both the fraction of groups that are of type  $k$  and the relative size of this type of group. Then  $[(\mathbf{f}_k, \pi_k), k \in K]$  characterizes the universe of groups in  $\mathcal{U}$ . Groups have realized treatments and outcomes given by  $[(\tau^m, \mathbf{y}^m), m \in M]$ . The objective of analysis is to learn about the counterfactual distribution of individual outcomes in the population for treatment  $t$ , that is,  $P[y(t)]$ .

**EXAMPLE 3.** This model can be used to study crime rates and social interactions. Let  $J$  be the citizens of a given state and let us define a group as the set of people who live in one of its cities. The outcome of interest is the decision to commit a crime at the

---

<sup>22</sup>Formally, for all sets  $B \subset \mathbb{R}$ ,  $P_G(y \in B) = \sum_{j \in G} 1(y_j \in B)(1/|G|)$ .

individual level and the treatment is police per capita at the city level. Distributional endogenous interactions are important in this model, as a higher crime participation rate by members of a city leads to fewer resources being spent on apprehending each criminal, which lowers his probability of punishment and further increases his incentives to commit a crime (see Sah (1991)).

#### 4.2 Monotone assumptions and their implications

This section imposes three alternative sets of conditions that are naturally satisfied in various social interactions models and derives their main implications in terms of equilibrium behavior.

*Coherence of the model* The first assumption imposes a monotone condition on  $(\mathbf{f}_k, k \in K)$ , which, as A1, guarantees equilibrium existence.

A1' (Positive Interactions). For each  $k \in K$ ,  $l \in G_k$ , and  $t \in T$ ,  $f_{kl}[t, P(y)] \geq f_{kl}[t, P(y')]$  for all  $P(y), P(y') \in \Delta_Y$  with  $P(y) \geq_{st} P(y')$ .

Let  $\varphi(t, k)$  indicate the solution set for  $t, k$ . Since the analyst's objective is to perform inference on the individual outcome distributions in  $J$ , we let the elements of  $\varphi(t, k)$  be the distribution functions induced by the solution vectors. In the next lemma, least and greatest are with respect  $\geq_{st}$ .

LEMMA 9. *If A1' holds, then  $\varphi(t, k)$  has a least and a greatest solution for all  $k \in K$  and  $t \in T$ .*

We write  $P[y_k(t)]$  for the element of  $\varphi(t, k)$  that is selected by a type- $k$  group that receives treatment  $t$ . The probability that the potential outcome falls in a set  $B \subseteq \mathbb{R}$  is given by

$$P[y(t) \in B] = \sum_{k \in K} P[y_k(t) \in B] \pi_k \quad (13)$$

and it is linear in  $\pi_k$ . Let  $\pi_{k|\tau}$  be the proportion of individuals in  $J$  who belong to type- $k$  groups conditional on the realized treatment  $\tau$ . We let  $P(\tau)$  denote the distribution of realized treatments in  $J$ . The probability that the realized outcome falls in a set  $B \subseteq \mathbb{R}$  is given by

$$P(y \in B) = \sum_{s \in T} P[y(s) \in B | \tau = s] P(\tau = s), \quad (14)$$

where  $P[y(s) \in B | \tau = s] = \sum_{k \in K} P[y_k(s) \in B] \pi_{k|\tau=s}$ .

We next elaborate on the power of two extra monotone restrictions. As we explained earlier, though we assume that the monotone treatment response and selection restrictions are positive, we can easily accommodate the opposite cases.

*Monotone treatment response* We now assume that individual outcomes increase in  $t$  on  $T$ .

A 2' (Monotone Treatment Response). For each  $k \in K$ ,  $l \in G_k$ , and  $P(y) \in \Delta_Y$ ,  $f_{kl}[t, P(y)] \geq f_{kl}[t', P(y)]$  for all  $t, t' \in T$  with  $t \geq t'$ .

If A1' and A2' hold, then the smallest and the largest distributions in  $\varphi(t, k)$  increase in  $t$  with respect to  $\geq_{st}$ . Thus, here again, extremal equilibrium selection rules are enough to make counterfactual predictions as we vary the treatment. We introduce an alternative assumption that leads to the same result. To this end, let us define  $F_k$  for all  $B \subseteq \mathbb{R}$  as

$$F_k(y \in B | t, P(y)) = (1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t, P(y)] \in B\}. \quad (15)$$

Thus,  $F_k$  is an aggregate response function that indicates the fraction of people in a type- $k$  group whose outcomes would lie in the set  $B$  for some  $t$  and some initial  $P(y)$ .

S1' (Equilibrium Selection). One of the following conditions holds: (i) each group selects either the smallest or the largest element of  $\varphi(t, k)$  and the selection rule (weakly) increases in  $t$  or (ii)  $F_k(y | t, \inf \phi(t', k)) \geq_{st} \sup \phi(t', k) \forall t, t' \in T$  such that  $t > t'$  and  $\forall k \in K$ .

We can motivate S1'(i) in the same way as we did with S1(i). The following result is the analog to Lemma 2.

LEMMA 10. If A1', A2', and S1' hold, then  $P[y_k(t)] \geq_{st} P[y_k(t')] \forall t, t' \in T$  with  $t > t'$  and  $k \in K$ .

We finally study monotone treatment selection.

*Monotone treatment selection* This section provides sufficient conditions to validate the use of realized treatments as monotone instrumental variables. To this end, we introduce a partial order on  $(F_k, k \in K)$ , as defined in (15).

DEFINITION 11. We say  $F_k \geq F_{k'}$  if  $F_k[y | t, P(y)] \geq_{st} F_{k'}[y | t, P(y)] \forall [t, P(y)] \in T \times \Delta_Y$ .

According to the last definition,  $F_k$  is greater than  $F_{k'}$  if the outcome distribution of the type- $k$  group is stochastically higher than the outcome distribution of the type- $k'$  group for any conditioning event. Let  $F$  denote a random function with support  $(F_k, k \in K)$ , and define  $P(F = F_k | \tau) \equiv \pi_{k|\tau}$  for all  $k \in K$ . The next condition is similar to condition A3.

A3' (Monotone Treatment Response). We have  $P(F | \tau = s) \geq_{st} P(F | \tau = s')$  for all  $s \geq s'$ .

Condition A3' states that the proportion of individuals who belong to groups with weakly higher structural functions increases with realized treatments. The motivation for this constraint is well captured by Example 3: in the study of crime rates, it is reasonable to think that the public authority is more likely to invest more (per capita) on the criminal apprehension system of those cities where people's willingness to commit crimes is believed to be higher. (As we explained earlier, our framework can easily accommodate the case of negative treatment selection.) Here again, before elaborating on the identification power of adding the monotone treatment selection condition, we need to introduce another restriction.

S2' (Equilibrium Selection). One of the following conditions holds: (i) groups select either the smallest or the largest element of the solution set and (for each  $t \in T$ ) the selection rule is the same  $\forall k \in K$  or (ii)  $F_k[y | t, \inf \varphi(t, k')] \geq_{st} \sup \varphi(t, k') \forall k, k' \in K$  such that  $F_k \geq F_{k'}$ .

The next lemma is the analog to Lemma 4.

**LEMMA 12.** *Assume A1', A3', and S2' hold. Let  $s, s' \in T$ . Then, for all  $t \in T$ ,*

$$s \geq s' \implies P[y(t) | \tau = s] \geq_{st} P[y(t) | \tau = s']. \quad (16)$$

The next section uses all of our previous results to provide bounds for  $P[y(t)]$ .

### 4.3 Identification region for $P[y(t)]$

The bounds for  $P[y(t)]$  are quite similar to those of  $P[\mathbf{y}(t)]$ . We only need to substitute  $\mathbf{y}(t)$  by  $y(t)$  and A1, (A1, A2, S1), and (A1, A3, S2) by A1', (A1', A2', S1'), and (A1', A3', S2') in Propositions 5, 6, and 7 to obtain Propositions 5', 6', and 7', respectively. Appendix A formalizes these claims. Thus, instead of repeating these results, we elaborate on two interesting differences between the two models.

Quantiles are often parameters of interest in applied studies. For  $\alpha \in (0, 1)$ , the  $\alpha$ -quantile of  $P[y(t)]$  is defined as  $Q_\alpha[y(t)] \equiv \inf_{y'} \{E[1[y(t) \leq y']] \geq \alpha\}$ . The characterization of the standard stochastic order in terms of the expectations of increasing functions induces a partial order on the quantiles of random variables: if a distribution function stochastically dominates another distribution function, then all the quantiles of the former are larger than the corresponding quantiles of the latter. Hence, the analogs of Propositions 5, 6, and 7 for the case of large groups can be easily reformulated in terms of the quantiles of the pertinent distributions, for example, the medians. Manski (1997) provides functional forms to construct bounds for quantiles in an individualistic model. The lack of objective basis for ordering multivariate observations is a major difficulty in extending the previous definition to random vectors. There are several attempts in the statistical literature toward multidimensional generalizations of univariate quantiles, each of which captures distinct aspects of interest. We remained silent about quantiles in Section 3.3 as it is not immediate that the multivariate standard stochastic order has clear monotone predictions for all the quantiles according to all the existing definitions.

We mentioned in Section 3.3 that all our identified bounds can be consistently estimated from sample data of a subset of groups under regularity conditions. When groups are large, the same result holds if we sample a subset of individuals from each group in the random sample of groups. The reason is that, in this second case, we are not interested in obtaining the joint distribution of outcomes of the group members; see, for example, [Brock and Durlauf \(2000\)](#).

## 5. APPLICATION: CRIME RATES AND SOCIAL INTERACTIONS

This section illustrates our previous results by applying them to the analysis of crime rates in New York state. [Becker \(1968\)](#) studies individual decisions to commit crimes from an economic perspective. He develops a cost–benefit analysis and argues that a key ingredient in an individual’s choice of whether to become a criminal is his perceived probability of punishment. Subsequent work emphasizes the importance of positive social interactions in motivating criminal behavior (see [Glaeser, Sacerdote, and Scheinkman \(1996\)](#) and the literature therein).

We use the model in Section 4 to study crimes across cities. Let each person in a given city decide whether to commit a crime. We define the treatment as police per capita at the city level and define the outcome as a yes/no indicator that takes the value 1 or 0, that is,  $Y = \{0, 1\}$ . [Sah \(1991\)](#) presents a model where one individual’s choice to commit a crime lowers the probability that any other individual ends up arrested. Since the police cannot be in two places at the same time, the higher is the criminal activity in a given city, the lower is the probability of being punished. His argument justifies A1'.<sup>23</sup> In addition, each individual’s decision to commit a crime decreases with the amount of police per capita in his own city. (This is a natural direct effect of the treatment.) Then the dual version of A2' holds here as well. It is also reasonable to think that the public authority is more likely to invest more on the criminal apprehension system of those cities where people’s willingness to commit crimes is believed to be higher.<sup>24</sup> The last statement validates condition A3'. We will also assume that the dual of S1' and S2' hold.

The analyst wants to learn about the fraction of people who would commit a crime in New York state if all its cities were to be assigned a given level of police per capita. The next subsection describes the data we use and the subsequent subsection shows our findings.

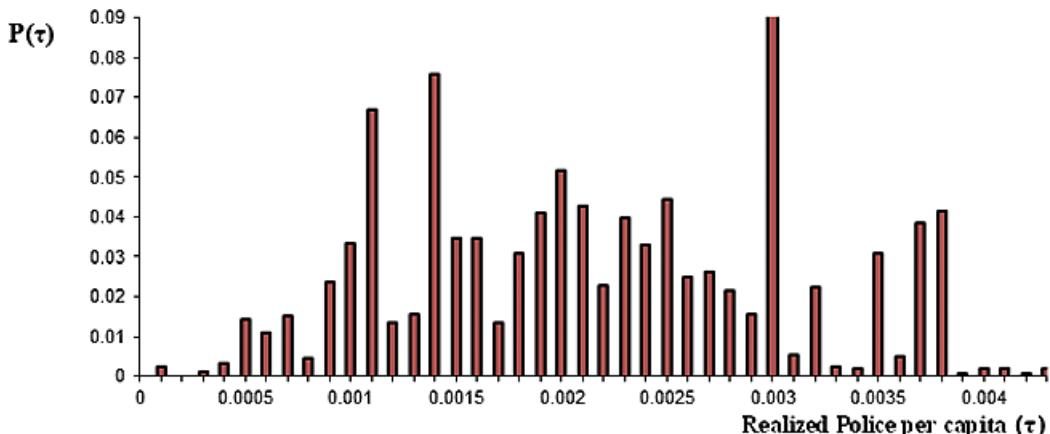
### 5.1 Data set

Our data source is the Uniform Crime Reporting (UCR) program of the Federal Bureau of Investigation (FBI) for the year 2009. The UCR program informs crimes reported and verified. The data set also provides information about the number of police at the city

---

<sup>23</sup>This assumption is justified in the criminology literature by using many different arguments (for example, the theory of differential association by [Sutherland \(1974\)](#)). Our results do not depend on the particular mechanism by which the interactions arise. Thus, any of the existing justifications can be used to validate A1'.

<sup>24</sup>[Lazzati and Menichini \(2014\)](#) provide a theoretical model of optimal police allocation that could be used to justify this restriction.



Source: FBI, UCR (Uniform Crime Reporting) program for year 2009

FIGURE 2. Distribution of realized levels of police per capita.

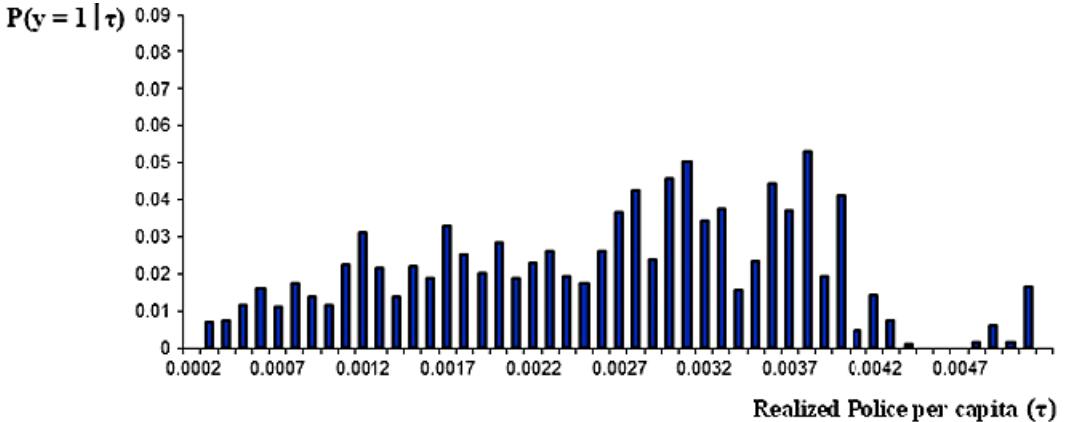
level and the population of each city. We decided to eliminate New York City from the sample, as its features (e.g., number of people) are markedly different from the characteristics of the other cities. Our data cover 47% of the remaining population of New York state (314 cities). To perform the analysis, we discretized the level of police per capita in multiples of 0.0001.

Figure 2 displays the data for 99.4% of the sampled population: for expositional ease, the figure does not include a few observations with extremely high levels of police per capita, but these observations were taken into account in the estimation. For levels of police per capita between 0 and 0.005,  $P(\tau)$  indicates the fraction of individuals in the sample who received treatment  $\tau$  during the year 2009. Since a large part of the sampled population—specifically, 91.02%—received treatments between 0.001 and 0.004, the next section estimates the outcomes of interest for levels of police per capita in that range of values. The asymptotic properties of our estimators should be thought of as the number of sampled cities approaches the whole universe of cities in New York state.

In Figure 3,  $P(y = 1 | \tau)$  indicates the fraction of people who committed a crime in 2009 conditional on living in a city with a level of police per capita  $\tau$ . Throughout this application, we assume each individual has committed at most one crime during 2009.

**REMARK.** The assumption that each individual has committed at most one crime during 2009 is surely strong. This condition could be relaxed by a richer data set that incorporates information on offender reincidence. This issue could also be fixed by interpreting the potential outcome as the decision of committing at least one crime, but results should correspondingly be adapted to this alternative interpretation.<sup>25</sup>

<sup>25</sup>I thank one of the referees for this suggestion.



Source: FBI, UCR (Uniform Crime Reporting) program for year 2009

FIGURE 3. Criminal activity conditional on realized treatments.

## 5.2 Findings

Since the outcome of interest is binary (i.e.,  $Y = \{0, 1\}$ ), the fraction of people who would commit a crime for a given level of police per capita  $t$  (i.e.,  $P[y(t) = 1]$ ) contains all the information we need to describe the distribution of potential outcomes,  $P[y(t)]$ .

We consider four levels of police per capita (i.e., four treatments): 0.001, 0.002, 0.003, and 0.004. For each of them, Figure 4 reports the lower and the upper estimations of the identified bounds for  $P[y(t) = 1]$  under three sets of assumptions.<sup>26</sup> These bounds are reported in the third and the fourth columns respectively, and are expressed in percentage points. They are the sample analogs of Propositions 5', 6', and 7', respectively. In addition, we provide (in the last column) confidence sets for  $P[y(t) = 1]$  at the 95% of confidence level. To compute them, we follow the methodology in [Imbens and Manski \(2004\)](#) and [Stoye \(2009\)](#); see Appendix B for further details. We next highlight our findings.<sup>27</sup>

We can clearly see in Figure 4 that A1' alone is practically uninformative in the four cases. The mere addition of either (dual of A2', S1') or (A3', S2') substantially improves all the predictions. For instance, for  $t = 0.004$ , A1' alone predicts  $100 \times P[y(t) = 1] \in (0, 100)$ , while the introduction of (dual of A2', S1') reduces that interval to (0, 5.04). The reason is that a very small fraction (i.e., 0.25%) of the population has realized treatment  $\tau = 0.004$ .

<sup>26</sup>In an individualistic model, [Manski and Pepper \(2000\)](#) study returns to schooling assuming both the monotone treatment response and the monotone treatment selection restrictions are satisfied. They show that the bounds under the joint restriction are much tighter than the simple intersection of the bounds under the individual assumptions. We computed the sharp bounds under the joint restriction for the crime application. In our case, they were almost identical to the intersection of the bounds under the individual restrictions. We speculate that this result, which contrasts with that in [Manski and Pepper \(2000\)](#), relates to the fact that, in our case, the monotone treatment response and selection restrictions work in opposite directions.

<sup>27</sup>For  $t = 0.004$ , one of the estimators takes a value slightly below 0 and a second estimator takes a value slightly above 100. We wrote 0 and 100 instead, as the true parameter of interest will always take values between 0 and 100.

Police Per Capita	Assumptions	Lower Bound	Upper Bound	95% Confidence Set
$t = 0.001$	A1'	0.06	96.74	(0.03, 98.37)
	A1', dual of A2', S1'	2.90	89.25	(2.72, 92.10)
	A1', A3', S2'	1.79	89.26	(1.74, 92.11)
$t = 0.002$	A1'	0.12	94.97	(0.07, 96.98)
	A1', dual of A2', S1'	2.18	54.94	(1.95, 59.47)
	A1', A3', S2'	1.39	55.04	(1.29, 59.56)
$t = 0.003$	A1'	0.50	89.74	(0.35, 92.48)
	A1', dual of A2', S1'	1.23	18.42	(1.02, 21.74)
	A1', A3', S2'	1.26	20.03	(1.07, 23.29)
$t = 0.004$	A1'	0.00	99.8	(0.00, 100.0)
	A1', dual of A2', S1'	0.02	04.1	(0.00, 05.04)
	A1', A3', S2'	0.01	01.47	(0.00, 02.42)

Source: FBI, UCR (Uniform Crime Reporting) program for year 2009.

FIGURE 4. Lower and upper bounds for  $P[y(t) = 1]$  in percent.

and then the empirical evidence alone is ineffective to identify the potential outcomes. However, by adding monotone treatment response, all the data are used in the estimation: part of the observations improve the upper bound, another segment helps to estimate the lower bound, and a third group of observations is used to construct both. For this treatment level, (A3', S2') provides even more information as the interval shrinks to (0, 2.42). The reason for this result is that the groups that received this large treatment displayed very low criminal activity.

The bounds we just provided can be improved if we select less conservative, though still credible, upper bounds for the largest possible fraction of people who could commit a crime at any treatment level. We refer to the latter probability as Pmax. Figure 5 provides bounds for the fraction of people who would commit a crime at a level of police per capita  $t = 0.002$  for three different values of Pmax.<sup>28</sup> Our results show that the bounds become substantially more informative as we reduce Pmax. Similar results hold for the other treatment levels.

*Welfare analysis* Though the previous results show that the estimated intervals for criminal activity decrease as police per capita increases, increasing police per capita has an opportunity cost in terms of other social policies that could be pursued with the budget assigned to police resources.<sup>29</sup> Thus, to decide whether to increase the police resources, the public authority may want to compare the overall cost of the available alternatives. We next show that our previous bounds can help the public authority to make a decision, even when some of the bounds are quite wide. Our model is based on Becker (1968).

<sup>28</sup>I thank one referee for this suggestion.

<sup>29</sup>I thank a referee for suggesting this welfare analysis.

Values of Pmax	Assumptions	Lower Bound	Upper Bound	95% Confidence Set
Pmax = 1	A1'	0.12	94.97	(0.07, 96.98)
	A1', dual of A2', S1'	2.18	54.94	(1.95, 59.47)
Pmax = 0.50	A1', A3', S2'	1.39	55.04	(1.29, 59.56)
	A1'	0.12	47.55	(0.07, 48.52)
	A1', dual of A2', S1'	2.18	27.96	(1.95, 30.18)
Pmax = 0.25	A1', A3', S2'	1.39	28.06	(1.29, 30.27)
	A1'	0.12	23.83	(0.07, 24.30)
	A1', dual of A2', S1'	2.18	14.47	(1.95, 15.53)
	A1', A3', S2'	1.39	14.57	(1.29, 15.62)

Source: FBI, UCR (Uniform Crime Reporting) program for year 2009.

FIGURE 5. Lower and upper bounds for  $P[y(t = 0.002) = 1]$  in percent.

Let  $O(t)$  be the number of offences in New York state when its cities receive a level of police per capita  $t$ . If the number of people is  $N$ , then  $O(t) \approx NP[y(t) = 1]$ . Let  $H(O(t))$  be the corresponding social harm. We assume  $H$  is linear so that  $H(O(t)) = hO(t)$ , where  $h > 0$  is the social cost per crime committed. In addition, we let  $c > 0$  be the opportunity cost per unit of police officer. The public authority wants to compare the overall cost of two levels of police per capita  $t, t' \in T$  with  $t > t'$ ; that is, it is interested in estimating

$$\Delta(t, t') = (hNP[y(t) = 1] + cNt) - (hNP[y(t') = 1] + cNt').$$

To provide bounds for  $\Delta(t, t')$ , we let  $P_l[y(t) = 1]$  and  $P_u[y(t) = 1]$  be the lower and the upper bounds for  $P[y(t) = 1]$ . The bounds for  $\Delta(t, t')$  can be obtained as

$$\begin{aligned} & hN\{P_u[y(t) = 1] - P_l[y(t') = 1]\} + cN(t - t') \\ & \geq \Delta(t, t') \geq hN\{P_l[y(t) = 1] - P_u[y(t') = 1]\} + cN(t - t'). \end{aligned}$$

Though they may not be sharp, these bounds can be easily implemented. The lower bound can be either positive or negative. When it is positive, our analysis suggests that it is not worth it to increase the level of police even if it reduces the criminal activity.

To illustrate how useful this analysis might be, we next estimate the lower bound of  $\Delta(t, t')$  for two alternative changes in treatment levels. We assume that  $h$  is \$400 and the opportunity cost of a police officer  $c$  is \$105,000.<sup>30</sup> To estimate the lower bounds, we use the results in Figure 4 under the restriction (A1', dual of A2', S1'). We get

$$\Delta(t = 0.003, t' = 0.002) \geq -N128.8 \quad \text{and} \quad \Delta(t = 0.004, t' = 0.003) \geq N18.04.$$

This result shows that increasing police per capita in all cities of New York state from 0.002 to 0.003 might reduce the social cost. It also states that further increments from

<sup>30</sup>Estimates of the average social cost of crimes substantially differ across studies. Most of them agree that the estimates for most of the crimes are very low. However, they also agree that the average is pushed up by a small number of crimes that involve extremely large monetary losses.

0.003 to 0.004 will never do so. As we mentioned earlier, this outcome is interesting as it shows that even when our bounds are wide, they can still be useful to guide policy decisions.

## 6. CONCLUDING REMARKS

This paper provides identification results for treatment response models with endogenous social interactions by means of monotone comparative statics. In doing so, we bridge the theory of identification of treatment effects that exploits monotone restrictions with recent results on games with strategic complementarities.

The approach to partial identification is nonparametric and allows for counterfactual predictions under multiple equilibria. Moreover, it relies neither on random treatment assignment nor on random assignment of individuals to the groups. Our results derive from shape restrictions on the primitives of the model that lead to monotone comparative statics of the equilibrium sets. The identification regions have the form of intervals, that is, we identify two extreme distributions that are functions of the observable data such that the true distribution must lie between them in terms of stochastic dominance. The bounds we provide are sharp and, by applying them to the study of crimes in New York state, we show that they can also be very informative.

An additional contribution of this study is to develop a flexible and tractable probabilistic framework for the model that typifies groups through the relevant features of the group members and can accommodate various processes of group formation. Last, this article also contributes to the literature on identification of nonparametric simultaneous equations models.<sup>31</sup>

## APPENDIX A: PROOFS

To make this paper self-contained, we provide the three theorems we invoke.

**TARSKI'S FIXED POINT THEOREM (TFP).** *If  $X$  is a complete lattice and  $f : X \rightarrow X$  is an increasing function, then  $f$  has a fixed point. Moreover, the set of fixed points of  $f$  has a smallest and a largest element (Tarski (1955)).*

**MILGROM AND ROBERTS' THEOREM (MR).** *If  $X$  is a complete lattice,  $S$  is a partially ordered set, and  $f : X \times S \rightarrow X$  is an increasing function, then the least and greatest fixed points of  $f$  are increasing in  $s$  on  $S$  (Milgrom and Roberts (1990)).*

The concept of first order (or standard) stochastic dominance is based on upper sets. Let us consider  $(\Omega, \geq)$ , where  $\Omega$  is a set and  $\geq$  defines a partial order on it. A subset  $U \subset \Omega$  is an upper set if and only if  $x' \in U$  and  $x \geq x'$  imply  $x \in U$ .

**FIRST ORDER STOCHASTIC DOMINANCE THEOREM (FOSD).** *Let  $X, X' \in \mathbb{R}^n$  be two random vectors. The following two statements are equivalent:*

---

<sup>31</sup>See, for example, Matzkin (2008) and the literature therein.

- (i) We have  $P(X \in U) \geq P(X' \in U)$  for all upper set  $U \subset \mathbb{R}^n$ .
- (ii) We have  $E[f(X)] \geq E[f(X')]$  for all increasing functions  $f(\cdot)$  (such that the expectations exist).

We write  $P(X) \geq_{st} P(X')$  if these conditions hold. This concept extends to arbitrary partially ordered domains up to measurability considerations (see Mosler and Scarsini (1991)).

**PROOF OF LEMMA 1.** Consider the mapping

$$\begin{aligned} M_{t,k} : Y^{|L|} &\rightarrow Y^{|L|}, \\ (y_1, y_2, \dots, y_{|L|}) &\mapsto (y'_{k1}, y'_{k2}, \dots, y'_{k|L|}), \end{aligned} \tag{17}$$

where  $y'_{kl} = f_{kl}(t, \mathbf{y}_{-l})$  for  $l = 1, 2, \dots, |L|$  and  $\mathbf{y}_{-l} \equiv (y_1, \dots, y_{l-1}, y_{l+1}, \dots, y_{|L|})$ . By construction, the solution set to the system of equations  $\mathbf{f}_k, \phi(t, k)$ , coincides with the set of fixed points of  $M_{t,k}$ . Then the proof of Lemma 1 reduces to show that the set of fixed points of  $M_{t,k}$  has a least and a greatest element. Here  $(Y^{|L|}, \geq)$  is a complete lattice for the coordinatewise (partial) order  $\mathbf{y} \geq \mathbf{y}'$  if  $y_l \geq y'_l$  for all  $l \in L$ . By A1,  $M_{t,k}$  is increasing. Hence, Lemma 1 follows by TFP.  $\square$

**PROOF OF LEMMA 2.** Lemma 1 shows that if A1 holds, then  $\phi(t, k)$  has a least and a greatest element, and A2 ensures that the system of equations  $\mathbf{f}_k$  increases in  $t$  for any fixed  $k \in K$ . Then  $M_{t,k}$ , as defined in (17), increases in  $t$  on  $T$ , and the fact that the extremal elements of  $\phi(t, k)$  increase in  $t$  follows by MR. Thus, condition S1(i) is sufficient to our claim. The fact that S1(ii) is sufficient to our claim holds by Echenique and Sabarwal (2003).  $\square$

The proof of Lemma 4 requires an additional result.

**LEMMA 13.** Assume A1 and S2 hold. Let  $k, k' \in K$ . If  $\mathbf{f}_k \geq \mathbf{f}_{k'}$  according to Definition 3, then  $\mathbf{y}_k(t) \geq \mathbf{y}_{k'}(t)$  for all  $t \in T$ .

**PROOF.** Lemma 1 shows that if A1 holds, then  $\phi(t, k)$  and  $\phi(t, k')$  have a least and a greatest element. Let  $S = \{0, 1\}$ ,  $M_{t,k}(0) \equiv M_{t,k'}$ , and  $M_{t,k}(1) \equiv M_{t,k}$  with  $M_{t,k'}$  and  $M_{t,k}$  defined as in (17). Notice that  $\mathbf{f}_k \geq \mathbf{f}_{k'}$  implies  $M_{t,k}(s)$  is increasing in  $s$  on  $S$ . Then, by MR—since  $\phi(t, k)$  and  $\phi(t, k')$  are the sets of fixed points of  $M_{t,k}$  and  $M_{t,k'}$ , respectively—the least (greatest) vector in  $\phi(t, k)$  is larger than the least (greatest) vector in  $\phi(t, k')$   $\forall t \in T$ . The fact that  $\mathbf{y}_k(t) \geq \mathbf{y}_{k'}(t) \forall t \in T$  holds by S2(i). The fact that S2(ii) is sufficient to our claim holds by Echenique and Sabarwal (2003).  $\square$

**PROOF OF LEMMA 4.** Let  $s, s' \in T$  with  $s \geq s'$ , and let us assume that A1, A3, and S2 hold. Fix an upper set  $U \subset \mathbb{R}^{|L|}$ . By definition,

$$P[\mathbf{y}(t) \in U \mid \tau = s] = \sum_{k \in K} 1[\mathbf{y}_k(t) \in U] \pi_{k|\tau=s}. \tag{18}$$

Recall that  $P(\mathbf{f} = \mathbf{f}_k \mid \tau) \equiv \pi_{k|\tau}$ . By Lemma 13,  $1[\mathbf{y}_k(t) \in U] \geq 1[\mathbf{y}_{k'}(t) \in U]$  if  $\mathbf{f}_k \geq \mathbf{f}_{k'}$ . Then  $P[\mathbf{y}(t) \in U \mid \tau = s]$  is the expectation of an increasing function of  $\mathbf{f}$  conditional on  $\tau = s$ . By A3,  $P(\mathbf{f} \mid \tau = s) \geq_{st} P(\mathbf{f} \mid \tau = s')$ . Then, by FOSD(ii),  $P[\mathbf{y}(t) \in U \mid \tau = s] \geq P[\mathbf{y}(t) \in U \mid \tau = s']$ . The result follows by FOSD(i) as  $U$  was arbitrarily selected.  $\square$

**PROOF OF PROPOSITION 5.** We first show

$$P[\mathbf{y}(t)] \geq_{st} P(\mathbf{y} \mid \tau = t)P(\tau = t) + P(\underline{Y}^{|L|})P(\tau \neq t). \quad (19)$$

Let  $U \subset \mathbb{R}^{|L|}$  be an upper set and let us consider the next two steps:

$$\begin{aligned} P[\mathbf{y}(t) \in U] &= P[\mathbf{y}(t) \in U \mid \tau = t]P(\tau = t) + P[\mathbf{y}(t) \in U \mid \tau \neq t]P(\tau \neq t) \\ &\geq P(\mathbf{y} \in U \mid \tau = t)P(\tau = t) + P(\underline{Y}^{|L|} \in U)P(\tau \neq t). \end{aligned} \quad (20)$$

The empirical evidence reveals  $P[\mathbf{y}(t) \in U \mid \tau = t] = P(\mathbf{y} \in U \mid \tau = t)$ . The inequality is true as  $\underline{Y}^{|L|}$  is a lower bound for any vector of potential outcomes. Since  $U$  was arbitrarily selected, our first claim holds by FOSD(i). The proof for the upper bound is similar, so we omit it.

To show that the identified area is sharp, first notice that we restrict  $P[\mathbf{y}(t)]$  to the set of all possible distributions that are consistent with the nature of the outcomes, that is,  $P[\mathbf{y}(t)] \in \Delta_{Y^{|L|}}$ . In addition, given the data, our initial assumptions are consistent with both  $P[\mathbf{y}(t) \mid \tau \neq t] = P(\underline{Y}^{|L|})$  and  $P[\mathbf{y}(t) \mid \tau \neq t] = P(\bar{Y}^{|L|})$ . Then  $P[\mathbf{y}(t)]$  can coincide with any element of  $\Delta_{Y^{|L|}}$  that lies between the lower and the upper bound, and our claim follows.  $\square$

The proof of Proposition 6 requires an intermediate result that relates to Lemma 2.

**COROLLARY 14.** *Assume A1, A2, and S1 hold. Then we have  $P[\mathbf{y}(t) \mid \tau \leq t] \geq_{st} P(\mathbf{y} \mid \tau \leq t)$  and  $P(\mathbf{y} \mid \tau \geq t) \geq_{st} P[\mathbf{y}(t) \mid \tau \geq t]$ .*

**PROOF.** Let  $U \subset \mathbb{R}^{|L|}$  be an upper set and consider the next steps:

$$\begin{aligned} P[\mathbf{y}(t) \in U \mid \tau \leq t] &= \sum_{s \in T} \left\{ \sum_{k \in K} 1[\mathbf{y}_k(t) \in U] \pi_{k|\tau=s} \right\} 1(s \leq t)P(\tau = s \mid \tau \leq t) \\ &\geq \sum_{s \in T} \left\{ \sum_{k \in K} 1[\mathbf{y}_k(s) \in U] \pi_{k|\tau=s} \right\} 1(s \leq t)P(\tau = s \mid \tau \leq t) \\ &= P(\mathbf{y} \in U \mid \tau \leq t). \end{aligned} \quad (21)$$

Under A1, A2, and S1, the inequality follows by Lemma 2, as it implies  $\mathbf{y}_k(t) \geq \mathbf{y}_k(s)$  for all  $t \geq s$ . Since  $U$  was arbitrarily selected, the first claim follows by FOSD(i). The proof for the second claim is similar, so we omit it.  $\square$

**PROOF OF PROPOSITION 6.** We next show

$$P[\mathbf{y}(t)] \geq_{st} P(\mathbf{y} \mid \tau \leq t)P(\tau \leq t) + P(\underline{Y}^{|L|})P(\tau \not\leq t). \quad (22)$$

Let  $U \subset \mathbb{R}^{|L|}$  be an upper set and let us consider the next three steps:

$$\begin{aligned} P[\mathbf{y}(t) \in U] &= P[\mathbf{y}(t) \in U \mid \tau \leq t]P(\tau \leq t) + P[\mathbf{y}(t) \in U \mid \tau \not\leq t]P(\tau \not\leq t) \\ &\geq P(\mathbf{y} \in U \mid \tau \leq t)P(\tau \leq t) + P[\mathbf{y}(t) \in U \mid \tau \not\leq t]P(\tau \not\leq t) \\ &\geq P(\mathbf{y} \in U \mid \tau \leq t)P(\tau \leq t) + P(\underline{Y}^{|L|} \in U)P(\tau \not\leq t). \end{aligned} \quad (23)$$

Under **A1**, **A2**, and **S1**, the first inequality follows by Corollary 14 and FOSD(i). The second inequality is true as  $\underline{Y}^{|L|}$  is a lower bound for any vector of outcomes. Since  $U$  was arbitrarily selected, the claim holds by FOSD(i). The proof for the upper bound is similar, so we omit it.

Sharpness follows by the same argument as that invoked in Proposition 5.  $\square$

The proof of Proposition 7 requires an intermediate result that relates to Lemma 4.

**COROLLARY 15.** *Assume A1, A3, and S2 hold. Then we have  $P[\mathbf{y}(t) \mid \tau \geq t] \geq_{st} P(\mathbf{y} \mid \tau = t)$  and  $P(\mathbf{y} \mid \tau = t) \geq_{st} P[\mathbf{y}(t) \mid \tau \leq t]$  for all  $t \in T$ .*

**PROOF.** Let  $U \subset \mathbb{R}^{|L|}$  be an upper set, and consider the next steps:

$$\begin{aligned} P[\mathbf{y}(t) \in U \mid \tau \geq t] &= \sum_{s \in T} P[\mathbf{y}(t) \in U \mid \tau = s]1(s \geq t)P(\tau = s \mid \tau \geq t) \\ &\geq \sum_{s \in T} P[\mathbf{y}(t) \in U \mid \tau = t]1(s \geq t)P(\tau = s \mid \tau \geq t) \\ &= P[\mathbf{y}(t) \in U \mid \tau = t] \sum_{s \in T} 1(s \geq t)P(\tau = s \mid \tau \geq t) \\ &= P(\mathbf{y} \in U \mid \tau = t). \end{aligned} \quad (24)$$

Under **A1**, **A3**, and **S2**, the inequality holds by Lemma 4. The third line holds as  $P[\mathbf{y}(t) \in U \mid \tau = t]$  is independent of  $s$ , and the last line is true as  $\sum_{s \in T} 1(s \geq t) \times P(\tau = s \mid \tau \geq t) = 1$ . Since  $U$  was arbitrarily selected, the first claim holds by FOSD(i).

The proof of the second claim is similar, so we omit it.  $\square$

**PROOF OF PROPOSITION 7.** We next show

$$P[\mathbf{y}(t)] \geq_{st} P(\mathbf{y} \mid \tau = t)P(\tau \geq t) + P(\underline{Y}^{|L|})P(\tau \not\geq t). \quad (25)$$

Let  $U \subset \mathbb{R}^{|L|}$  be an upper set and let us consider the next three steps:

$$\begin{aligned} P[\mathbf{y}(t) \in U] &= P[\mathbf{y}(t) \in U \mid \tau \geq t]P(\tau \geq t) + P[\mathbf{y}(t) \in U \mid \tau \not\geq t]P(\tau \not\geq t) \\ &\geq P(\mathbf{y} \in U \mid \tau = t)P(\tau \geq t) + P[\mathbf{y}(t) \in U \mid \tau \not\geq t]P(\tau \not\geq t) \\ &\geq P(\mathbf{y} \in U \mid \tau = t)P(\tau \geq t) + P(\underline{Y}^{|L|} \in U)P(\tau \not\geq t). \end{aligned} \quad (26)$$

Under **A1**, **A3**, and **S2**, the first inequality follows by Corollary 15, and the last inequality holds as  $\underline{Y}^{|L|}$  is a lower bound for any vector of potential outcomes. Since  $U$  was arbitrarily selected, our first claim holds by FOSD(i). The proof for the upper bound is similar, so we omit it.

Sharpness follows by the same argument as that invoked in Proposition 5. □

**PROOF OF LEMMA 9.** Consider the mapping

$$\begin{aligned} N_{t,k} : \Delta_Y &\rightarrow \Delta_Y, \\ P(y) &\rightarrow P(y'), \end{aligned} \tag{27}$$

where  $P(y' \in B) = (1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t, P(y)] \in B\}$  for all  $B \subseteq \mathbb{R}$ . It is immediate to notice that the range of  $N_{t,k}$  is as given. By construction, the extremal elements of  $\varphi(t, k)$  coincide with the extremal fixed points of  $N_{t,k}$ . Then the proof of Lemma 9 reduces to show that the set of fixed points of  $N_{t,k}$  has a least and a greatest element.

By Echenique (2003, Lemma 1),  $(\Delta_Y, \geq_{st})$  is a complete lattice. By assumption A1',

$$(1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t, P(y)] \in U\} \geq (1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t, Q(y)] \in U\}$$

if  $P(y) \geq_{st} Q(y)$  for all upper set  $U \subset \mathbb{R}$ . Then, by FOSD(i),  $P(y') \geq_{st} Q(y')$ . Thus,  $N_{t,k}$  is monotone increasing and the claim holds by TFP. □

**PROOF OF LEMMA 10.** Lemma 9 shows that if A1' holds, then  $\varphi(t, k)$  has a least and a greatest element, and A2' ensures the system of equations  $f_k[t, P(y)]$  increases in  $t$  for any fixed  $k \in K$ . Then, for all  $t > t'$ ,

$$(1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t, P(y)] \in U\} \geq (1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t', P(y)] \in U\}$$

for any upper level set  $U$ . It follows, by FOSD(i), that  $N_{t,k}$ , as defined in (27), increases with respect to FOSD in  $t$  on  $T$ . The fact that the extremal elements of  $\varphi(t, k)$  increase in  $t$  follows by MR. Condition S1'(i) is then sufficient to our claim. The fact that S1'(ii) is sufficient to our claim holds by Echenique and Sabarwal (2003). □

The proof of Lemma 12 requires an additional result.

**LEMMA 16.** *Assume A1' and S2' hold. Let  $k, k' \in K$ . If  $F_k \geq F_{k'}$  according to Definition 11, then  $P[y_k(t)] \geq_{st} P[y_{k'}(t)]$  for all  $t \in T$ .*

**PROOF.** Lemma 9 shows if A1' holds, then  $\varphi(t, k)$  and  $\varphi(t, k')$  have a least and a greatest element. Let  $S = \{0, 1\}$ ,  $N_{t,k}(0) \equiv N_{t,k'}$ , and  $N_{t,k}(1) \equiv N_{t,k}$ , with  $N_{t,k'}$  and  $N_{t,k}$  defined as in (27). Notice that  $F_k \geq F_{k'}$  implies  $N_{t,k}(s)$  is increasing in  $s$  on  $S$  with respect to FOSD. Then, by MR—since the extremal elements of  $\varphi(t, k)$  and  $\varphi(t, k')$  coincide with the extremal fixed points of  $N_{t,k}$  and  $N_{t,k'}$ , respectively—the least (greatest) distribution in  $\varphi(t, k)$  is larger than the least (greatest) distribution in  $\varphi(t, k')$   $\forall t \in T$ . The fact that  $P[y_k(t)] \geq_{st} P[y_{k'}(t)] \forall t \in T$  holds by S2'(i). The fact that S2'(ii) is sufficient to our claim holds by Echenique and Sabarwal (2003). □

**PROOF OF LEMMA 12.** Let  $s, s' \in T$  with  $s \geq s'$ , and let us assume that **A1'**, **A3'**, and **S2'** hold. Fix an upper set  $U \subset \mathbb{R}$ . By definition,

$$P[y(t) \in U | \tau = s] = \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=s}. \quad (28)$$

Recall  $P(F = F_k | \tau) \equiv \pi_{k|\tau}$ . By Lemma 16,  $P[y_k(t) \in U] \geq P[y_{k'}(t) \in U]$  if  $F_k \geq F_{k'}$ . Then  $P[y(t) \in U | \tau = s]$  is the expectation of an increasing function of  $F$  conditional on  $\tau = s$ . By **A3'**, we know that  $P(F | \tau = s) \geq_{st} P(F | \tau = s')$ . Then, by **FOSD(ii)**, we get  $P[y(t) \in U | \tau = s] \geq P[y(t) \in U | \tau = s']$ . The result follows by **FOSD(i)** as  $U$  was arbitrarily selected.  $\square$

Having shown Lemma 9, the proof of Proposition 5' is similar to that of Proposition 5. Thus we omit it.

The proof of Proposition 6' requires an intermediate result that relates to Lemma 10.

**COROLLARY 17.** *Assume **A1'**, **A2'**, and **S1'** hold. Then we have  $P[y(t) | \tau \leq t] \geq_{st} P(y | \tau \leq t)$  and  $P(y | \tau \geq t) \geq_{st} P[y(t) | \tau \geq t]$ .*

**PROOF.** Let  $U \subset \mathbb{R}$  be an upper set and consider the next steps:

$$\begin{aligned} P[y(t) \in U | \tau \leq t] &= \sum_{s \in T} \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=s} \right\} 1(s \leq t) P(\tau = s | \tau \leq t) \\ &\geq \sum_{s \in T} \left\{ \sum_{k \in K} P[y_k(s) \in U] \pi_{k|\tau=s} \right\} 1(s \leq t) P(\tau = s | \tau \leq t) \\ &= P(y \in U | \tau \leq t). \end{aligned} \quad (29)$$

Under **A1'**, **A2'**, and **S1'**, the inequality follows by Lemma 10, as they imply  $P[y_k(t)] \geq_{st} P[y_k(s)]$  for all  $t \geq s$ . Since  $U$  was arbitrarily selected, the first claim follows by **FOSD(i)**. The proof for the second claim is similar, so we omit it.  $\square$

Having shown Corollary 17, the proof of Proposition 6' is similar to that of Proposition 6. Thus we omit it.

The proof of Proposition 7' requires an intermediate result that relates to Lemma 12.

**COROLLARY 18.** *Assume **A1'**, **A3'**, and **S2'** hold. Then we have  $P[y(t) | \tau \geq t] \geq_{st} P(y | \tau = t)$  and  $P(y | \tau = t) \geq_{st} P[y(t) | \tau \leq t]$  for all  $t \in T$ .*

**PROOF.** Let  $U \subset \mathbb{R}$  be an upper set and consider the next steps:

$$\begin{aligned} P[y(t) \in U | \tau \geq t] &= \sum_{s \in T} \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=s} \right\} 1(s \geq t) P(\tau = s | \tau \geq t) \\ &\geq \sum_{s \in T} \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=t} \right\} 1(s \geq t) P(\tau = s | \tau \geq t) \end{aligned} \quad (30)$$

$$\begin{aligned}
&= \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=t} \right\} \sum_{s \in T} 1(s \geq t) P(\tau = s | \tau \geq t) \\
&= P(y \in U | \tau = t).
\end{aligned}$$

Under **A1'**, **A3'**, and **S2'**, the inequality follows by Lemma 12. The third line is true as  $P[y(t) \in U | \tau = t]$  is independent of  $s$ , and the last line is true as  $\sum_{s \in T} 1(s \geq t) P(\tau = s | \tau \geq t) = 1$ . Since  $U$  was arbitrarily selected, the first claim holds by **FOSD(i)**. The proof of the second claim is similar, so we omit it.  $\square$

Having shown Corollary 18, the proof of Proposition 7' is similar to that of Proposition 7. Thus we omit it.

## APPENDIX B

### B.1 Estimators for the extremal bounds of $P[y(t) = 1]$

The parameter of interest is the fraction of people who would commit a crime in New York state at a level of police per capita  $t$ , that is,  $P[y(t) = 1]$ . We observe criminal activity and realized levels of police per capita for a random sample of cities in New York state of size  $N$ .

Let us denote by  $P_l[y(t) = 1]$  and  $P_u[y(t) = 1]$  the identified lower and upper bounds for  $P[y(t) = 1]$  under the sustained assumptions. We indicate by  $\Delta(t)$  the distance between these bounds, that is,  $\Delta(t) \equiv P_u[y(t) = 1] - P_l[y(t) = 1]$ . We distinguish estimators by using carets.

In what follows,  $\tau_m$  is the realized level of police per capita in city  $m$ ;  $d_m$  is the number of crimes in city  $m$  divided by the number of people in city  $m$ ;  $n_m$  is the number of people in city  $m$  divided by the number of people in the sample; and  $N$  is the number of cities in the sample. We next provide expressions for the estimators of interest under three sets of restrictions.

*I. Positive interactions* The propositions below specify Propositions 5', 6', and 7' to the analysis of crimes.

**PROPOSITION 19.** *Assume **A1'** holds. Then, for all  $t \in T$ ,*

$$\begin{aligned}
H\{P[y(t) = 1]\} &= \{\delta \in [0, 1] : P(y = 1 | \tau = t)P(\tau = t) + P(\tau \neq t) \\
&\geq \delta \geq P(y = 1 | \tau = t)P(\tau = t)\}.
\end{aligned} \tag{31}$$

The estimators for the bounds are just the sample analogs of (31); that is,  $\forall t \in T$ ,

$$\widehat{P}_u[y(t) = 1] = \sum_{m \in M} [d_m 1(\tau_m = t) + 1(\tau_m \neq t)] n_m,$$

$$\widehat{P}_l[y(t) = 1] = \sum_{m \in M} d_m 1(\tau_m = t) n_m.$$

## *II. Positive interactions and (negative) monotone treatment response*

**PROPOSITION 20.** *Assume A1' and the duals of (A2' and S1') hold. Then, for all  $t \in T$ ,*

$$\begin{aligned} H\{P[y(t) = 1]\} &= \{\delta \in [0, 1] : P(y = 1 | \tau \leq t)P(\tau \leq t) + P(\tau \not\leq t) \\ &\geq \delta \geq P(y = 1 | \tau \geq t)P(\tau \geq t)\}. \end{aligned} \tag{32}$$

The estimators for the bounds are just the sample analogs of (32); that is,  $\forall t \in T$ ,

$$\begin{aligned} \hat{P}_u[y(t) = 1] &= \sum_{m \in M} [d_m 1(\tau_m \leq t) + 1(\tau_m > t)] n_m, \\ \hat{P}_l[y(t) = 1] &= \sum_{m \in M} d_m 1(\tau_m \geq t) n_m. \end{aligned}$$

## *III. Positive interactions and monotone treatment selection*

**PROPOSITION 21.** *Assume A1', A3', and S2' hold. Then, for all  $t \in T$ ,*

$$\begin{aligned} H\{P[y(t) = 1]\} &= \{\delta \in [0, 1] : P(y = 1 | \tau = t)P(\tau \leq t) + P(\tau \not\leq t) \\ &\geq \delta \geq P(y = 1 | \tau = t)P(\tau \geq t)\}. \end{aligned} \tag{33}$$

The estimators for the bounds are just the sample analogs of (33); that is,  $\forall t \in T$ ,

$$\begin{aligned} \hat{P}_u[y(t) = 1] &= \sum_{m \in M} \left[ \left( \frac{\sum_{m \in M} d_m 1(\tau_m = t) n_m}{\sum_{m \in M} 1(\tau_m = t) n_m} \right) 1(\tau_m \leq t) + 1(\tau_m > t) \right] n_m, \\ \hat{P}_l[y(t) = 1] &= \sum_{m \in M} \left[ \left( \frac{\sum_{m \in M} d_m 1(\tau_m = t) n_m}{\sum_{m \in M} 1(\tau_m = t) n_m} \right) 1(\tau_m \geq t) \right] n_m. \end{aligned}$$

**REMARK.** Since all the previous estimators take the form of weighted averages, then the corresponding vectors of variances and correlation,  $(\hat{\sigma}_l^2, \hat{\sigma}_u^2, \hat{\rho})$ , can be easily estimated.

### B.2 Confidence intervals for $P[y(t) = 1]$

We estimate confidence regions by following the approaches of [Imbens and Manski \(2004\)](#) and [Stoye \(2009\)](#). [Imbens and Manski \(2004\)](#) show uniform validity of their confidence region under the following assumption (Assumption 1 adapts their condition to our problem).

ASSUMPTION 1.

(i) *There exist estimators  $\widehat{P}_l[y(t) = 1]$  and  $\widehat{P}_u[y(t) = 1]$  that satisfy*

$$\sqrt{N} \begin{bmatrix} \widehat{P}_l[y(t) = 1] - P_l[y(t) = 1] \\ \widehat{P}_u[y(t) = 1] - P_u[y(t) = 1] \end{bmatrix} \xrightarrow{d} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_l^2 & \rho\sigma_l\sigma_u \\ \rho\sigma_l\sigma_u & \sigma_u^2 \end{bmatrix} \right)$$

*uniformly in  $P \in \Psi$ , and there are estimators  $(\widehat{\sigma}_l^2, \widehat{\sigma}_u^2, \widehat{\rho})$  that converge to their population values uniformly in  $P \in \Psi$ .*

(ii) *For all  $P \in \Psi$ ,  $\underline{\sigma}^2 \leq \sigma_l^2 \leq \bar{\sigma}^2$  and  $\underline{\sigma}^2 \leq \sigma_u^2 \leq \bar{\sigma}^2$  for some positive and finite  $\underline{\sigma}^2$  and  $\bar{\sigma}^2$ , and  $P_u[y(t) = 1] - P_l[y(t) = 1] \leq \bar{\Delta} < \infty$ .*

(iii) *For all  $\varepsilon > 0$ , there are  $\nu > 0$ ,  $K$ , and  $N_0$  such that  $N \geq N_0$  implies  $\Pr(\sqrt{N}|\bar{\Delta} - \Delta| > K\Delta^\nu) < \varepsilon$  uniformly in  $P \in \Psi$ .*

Stoye (2009) explains that Assumption 1(iii) requires  $\widehat{\Delta}(t)$  to be superefficient at 0, a quite strong condition. Thus, he imposes conditions (i) and (ii), and shows that the results of Imbens and Manski (2004) are still valid if we substitute condition (iii) by the next requirement.

ASSUMPTION 1(iii'). *There exists a sequence  $\{a_N\}$  such that  $a_N \rightarrow 0$ ,  $a_N\sqrt{N} \rightarrow \infty$ , and  $\sqrt{N}|\bar{\Delta} - \Delta_N| \xrightarrow{P} 0$  for all sequences of distributions  $P_N \subseteq \Psi$  with  $\Delta_N \leq a_N$ .*

Though condition (iii') is often hard to validate, Stoye (2009) provides a sufficient condition that clearly holds in our setup.

LEMMA 22. *Let Assumption 1(i) and (ii) hold, and assume*

$$\Pr(\widehat{P}_u[y(t) = 1] \geq \widehat{P}_l[y(t) = 1]) = 1.$$

*Then Assumption 1(iii') is implied.*

We next introduce the confidence intervals we use in the analysis of crimes.

Imbens and Manski (2004) propose the confidence region

$$CI_\alpha \equiv \left[ \widehat{P}_l[y(t) = 1] - \frac{c_\alpha \widehat{\sigma}_l}{\sqrt{N}}, \widehat{P}_u[y(t) = 1] + \frac{c_\alpha \widehat{\sigma}_u}{\sqrt{N}} \right],$$

where  $c_\alpha$  solves

$$\Phi\left(c_\alpha + \frac{\sqrt{N}\widehat{\Delta}}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right) - \Phi(-c_\alpha) = 1 - \alpha.$$

Stoye (2009) shows uniform validity of  $CI_\alpha$  under Assumption 1(i), (ii), and (iii').<sup>32</sup>

---

<sup>32</sup>Stoye (2009) proposes an alternative confidence region that takes into account the bivariate nature of the estimators. In our applications, both approaches give almost identical results.

## REFERENCES

- Ackerberg, D. and G. Gowrisankaran (2006), “Quantifying equilibrium network externalities in the ACH banking industry.” *RAND Journal of Economics*, 37, 738–761. [56]
- Amir, R. (2008), “Comparative statics of Nash equilibria.” Report. [51]
- Amir, R. and N. Lazzati (2011), “Network effects, market structure and industry performance.” *Journal of Economic Theory*, 146, 2389–2419. [56]
- Andrews, D. and G. Soares (2010), “Inference for parameters defined by moment inequalities using generalized moment selection.” *Econometrica*, 78, 119–157. [63]
- Bajari, P., H. Hahn, H. Hong, and G. Ridder (2011), “A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models.” *International Economic Review*, 52, 807–824. [50, 56]
- Becker, G. (1968), “Crime and punishment: An economic approach.” *Journal of Political Economy*, 76, 169–217. [67, 70]
- Becker, G. (1991), “A note on restaurant pricing and other examples of social influences on price.” *Journal of Political Economy*, 99, 1109–1116. [51]
- Blume, L., W. Brock, S. Durlauf, and Y. Ionides (2010), “Identification of social interactions.” In *Handbook of Social Economics*, Vol. 1B (J. Benhabib, A. Bisin, and M. Jackson, eds.), 853–964, North-Holland, Amsterdam. [49]
- Brock, W. and S. Durlauf (2000), “Interactions-based models.” In *Handbook of Econometrics*, Vol. 5 (J. Heckman and E. Leamer, eds.), 3297–3380, North-Holland, Amsterdam. [67]
- Brock, W. and S. Durlauf (2001), “Discrete choice with social interactions.” *Review of Economic Studies*, 68, 235–260. [57]
- Brock, W. and S. Durlauf (2007), “Identification of binary choice models with social interactions.” *Journal of Econometrics*, 140, 52–75. [51]
- Chesher, A. and A. Rosen (2012), “Simultaneous equations models for discrete outcomes, coherence, completeness, and identification.” Working Paper CWP21/12, CeMMAP. [50]
- Cooper, R., D. DeJong, R. Forsythe, and T. Ross (1992), “Communication in coordination games.” *Quarterly Journal of Economics*, 107, 739–771. [56, 57]
- De Paula, Á. (2009), “Inference in a synchronization game with social interactions.” *Journal of Econometrics*, 148, 56–71. [56]
- Echenique, F. (2003), “Mixed equilibria in games of strategic complementarities.” *Economic Theory*, 22, 33–44. [76]
- Echenique, F. and I. Komunjer (2009), “Testing models with multiple equilibria by quantile methods.” *Econometrica*, 77, 1281–1297. [56]

- Echenique, F. and T. Sabarwal (2003), “Strong comparative statics of equilibria.” *Games and Economic Behavior*, 42, 307–314. [58, 73, 76]
- Glaeser, E., B. Sacerdote, and J. Scheinkman (1996), “Crime and social interactions.” *Quarterly Journal of Economics*, 111, 507–548. [58, 67]
- Gowrisankaran, G. and J. Stavins (2004), “Network externalities and technology adoption: Lessons from electronic payments.” *RAND Journal of Economics*, 35, 260–276. [57]
- Imbens, G. and C. Manski (2004), “Confidence intervals for partially identified parameters.” *Econometrica*, 72, 1845–1857. [63, 69, 79, 80]
- Jia, P. (2008), “What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry.” *Econometrica*, 76, 1263–1316. [56]
- Lazzati, N. (2014), “Simultaneous choice models: The sandwich approach to nonparametric analysis.” Available at <http://ssrn.com/abstract=2450793>. [60]
- Lazzati, N. and A. Menichini (2014), “Hot spot policing: A theoretical study of place-based strategies to crime.” Available at <http://ssrn.com/abstract=2277876>. [67]
- Lee, S., O. Linton, and Y.-J. Whang (2009), “Testing for stochastic monotonicity.” *Econometrica*, 77, 585–602. [62]
- Manski, C. (1990), “Nonparametric bounds on treatment effects.” *American Economic Review*, 80, 319–323. [51, 61]
- Manski, C. (1997), “Monotone treatment response.” *Econometrica*, 65, 1311–1334. [51, 66]
- Manski, C. (2007), “Partial identification of counterfactual choice probabilities.” *International Economic Review*, 48, 1393–1410. [54]
- Manski, C. (2013), “Identification of treatment response with social interactions,” *Econometrics Journal*, 16, S1–S23. [50, 52, 54, 61]
- Manski, C. and J. Pepper (2000), “Monotone instrumental variables: With an application to the returns to schooling.” *Econometrica*, 68, 997–1010. [51, 59, 69]
- Matzkin, R. (2008), “Identification in nonparametric simultaneous equations models.” *Econometrica*, 76, 945–978. [72]
- Milgrom, P. and J. Roberts (1990), “Rationalizability, learning and equilibrium in games with strategic complementarities.” *Econometrica*, 58, 1255–1277. [50, 55, 72]
- Milgrom, P. and C. Shannon (1994), “Monotone comparative statics.” *Econometrica*, 62, 157–180. [55]
- Molinari, F. and A. Rosen (2008), “The identification power of equilibrium in games: The supermodular case.” *Journal of Business & Economic Statistics*, 26, 297–302. [60]
- Monderer, D. and L. Shapley (1996), “Potential games.” *Games and Economic Behavior*, 14, 124–143. [56]

- Mosler, K. and M. Scarsini (1991), "Some theory of stochastic dominance." In *Stochastic Orders and Decisions Under Risk*, IMS Lecture Notes Monograph Series, Vol. 19, Inst. Math. Statist., Hayward, CA. [73]
- Müller, A. and D. Stoyan (2002), *Comparison Methods for Stochastic Models and Risks*. Wiley, Chichester. [62]
- Oyama, D., W. Sandholm, and O. Tercieux (2012), "Sampling best response dynamics and deterministic equilibrium selection." Available at <http://ssrn.com/abstract=1985856>. [56]
- Rosen, A. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146, 107–117. [63]
- Sah, R. (1991), "Social osmosis and patterns of crime." *Journal of Political Economy*, 99, 1972–1995. [64, 67]
- Shaikh, A. and E. Vytlacil (2011), "Partial identification in triangular systems of equations with binary dependent variables." *Econometrica*, 79, 949–955. [49]
- Stoye, J. (2009), "More on confidence intervals for partially identified parameters." *Econometrica*, 77, 1299–1315. [63, 69, 79, 80]
- Sutherland, E. (1974), *Criminology*, ninth edition. Lippincott, Philadelphia, PA. [67]
- Tarski, A. (1955), "A lattice theoretical fixpoint theorem and its applications." *Pacific Journal of Mathematics*, 5, 285–309. [72]
- Topkis, D. (1979), "Equilibrium points in nonzero-sum  $n$ -person submodular games." *SIAM Journal of Control and Optimization*, 17, 773–787. [50, 55]
- Vives, X. (1990), "Nash equilibrium with strategic complementarities." *Journal of Mathematical Economics*, 19, 305–321. [50, 55]

---

Submitted September, 2012. Final version accepted January, 2014.