Treatment Response with Social Interactions: Partial Identification via Monotone Comparative Statics*

Natalia Lazzati †

May 27, 2012

Abstract

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, i.e., 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 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 NY State.

JEL codes: C31, D71

Key words and phrases: Treatment effects, social interactions, nonparametric bounds, supermodular games, monotone comparative statics, first order stochastic dominance.

---

*The author is deeply indebted to Keisuke Hirano and Rabah Amir, as well as to 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 U, NYU Stern, U of Connecticut, U of Warwick, UCSD, Caltech, Cornell U, U of Arizona, and U of Michigan.

†Natalia Lazzati, University of Michigan, Ann Arbor, MI 48109-1220, *e-mail:* nlazzati@umich.edu.

1 Introduction

This paper studies partial identification of treatment response in economic 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.[1] 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.[2] That is, standard models assume that each person's outcome varies only with his own treatment. To accommodate the applications above, we incorporate the social dimension by allowing individual outcomes to also depend on the outcomes of other people in the population. Under this set-up, we use monotone comparative statics to provide distributional bounds on treatment effects.

Models of endogenous interdependencies often start with the outline of a system of structural equations that define the outcome of each individual as a function of a vector of observable characteristics and the outcomes of the other members of the group. The solution to the system of equations (when it exists) is the predicted outcome or behavior of the group members. In the presence of multiple equilibria, observed behavior also depends on the mechanism by which people select among different solutions. Most of the research in econometrics is concerned with the identification of the structural equations. Manski (2011) develops a formal language to study treatment response with social interactions, and uses it to provide new results on identification of potential outcome distributions. Our objective is also (partial) identification of the potential outcome distributions under alternative treatment rules, not the structural functions per se. The approach in this paper differs from Manski (2011) in that we impose all shape conditions on the primitives of the structural model, not on the reduced forms. In doing this we provide the microfoundations for some of his novel results, and make explicit the strength of the identifying assumptions.

[1] See Blume et. al. (2010) for various applications.

[2] See, e.g., Shaikh and Vytlacil (2011) who provide partial identification of average treatment effects for triangular system of equations.

The basis of our model is then a system of simultaneous equations, whose solution corresponds to the predicted outcomes of the group members. We assume the analyst observes a vector of realized outcomes and treatments for a sample of groups—the outcomes that would have been experienced under other treatments are counterfactual. The objective of the researcher is to learn about the joint outcome distribution that would occur in the study population if the groups were to receive a specific treatment. The underlying approach for identification consists of imposing minimal monotone structure to the primitives of the model (i.e., the structural equations) to derive sharp restrictions on the predicted outcomes via stochastic dominance. An important feature of the method we propose is that it allows for making counterfactual predictions under multiple equilibria.

Previous studies have required specific functional forms on the structural equations. Brock and Durlauf (2001a, b, 2007), among others, develop parametric and semi-parametric models and address point and partial (or set) identification of the structural parameters that capture the strength of the social interactions.[3] This method is often hard to motivate by economic theory, so we require no specification of functional forms. Identification is also often obtained via exclusion restrictions, e.g., by assuming the outcome functions are statistically independent of realized treatments, and/or by relying on random assignment of individuals to the groups. Our identification strategy does not rely on these conditions either.

The first two conditions we impose for identification 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. Most of the social interactions models introduce these two conditions as the distinctive features of the phenomenon of interest, so they are often easy to justify on economic grounds. 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 sys-

[3] Brock and Durlauf (2001a, b, 2007) study identification in an incomplete information model that has multiple equilibria for some parameter values. Moffitt (2001) and Graham (2008) address identification of parametric social interaction models as well. Manski (1993) is a direct predecessor of these papers.

tem has a minimal and a maximal solution, i.e., the model is coherent.[4] 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)]. Since coordination on extremal equilibria is often hard to motivate, we show that the monotone comparative statics holds at the selected equilibrium under a reasonable learning behavior. We provide a tractable econometric framework which translates the monotone comparative statics results into sharp distributional bounds for the potential outcomes.

Manski (1990, 1997) derives nonparametric bounds on treatment effects by exploiting natural bounds on outcomes and monotone shape restrictions in settings where a person's outcome varies only with his own treatment. The result in the previous paragraph extends these routes for inference to models with endogenous interactions, by means of an approach that relies on 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.[5] 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 paper and ours is that we need to compare the equilibria of two sets of games, not only two games.

We provide identification results for two models of social interactions. In the first set-up we

[4] Models of interdependent equations for which a solution exists are often named consistent or coherent by the econometric literature.

[5] Brock and Durlauf (2007) use a similar idea to address partial identification of a model with social interactions in a semi-parametric framework.

study small groups, where each member has a distinctive role, e.g., men and women in married couples. The second set-up is appropriate to study large neighborhoods with anonymous social interactions, e.g., crimes and infectious diseases. 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 the rest of the paper. Sections 3.1 and 3.2 study identification of treatment effects for small and large groups, respectively. Section 4 discusses an estimation strategy for the identified distributions. Section 5 uses our results to study crime rates and social interactions in New York State. Section 6 concludes, and we collect all the proofs in Sections 7 and 8.

2 An Initial Example

This section studies a model similar to the one in Becker (1991) to highlight the distinctive features of the interactions-based models that justify our approach. It also helps us to contrast the classical (econometric) identification objective of previous analyses with ours.

Each person $j \in G$ decides whether to go to a popular restaurant. Here the treatment is the price he would pay for the service, and the outcome (y_j) is a yes/no indicator taking the value one or zero. A consumer's demand for the good depends on the price, and the decisions by the other members of his group of reference. The latter is the source of endogeneity. Consumer j maximizes his utility taking as given the behavior of the others. His demand for the good is

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

where t is a potential price for the service and $(1/|G|) \sum_{i \in G} y_i$ is the average of the decisions selected by all members of the same neighborhood. In this model $|G|$ is assumed to be large enough that changes in any y_j hardly affect the average of the choices.[6] An equilibrium in this

[6] This assumption is sustained in Section 3.2 below. It simplifies the theoretical analysis without changing the nature of the results in any fundamental way [see, e.g., Becker (1991)].

market is a solution to the system of structural equations (1)

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

where $y_j(t)$ is the predicted decision of consumer j in G for a potential price t.

In each neighborhood G that has a branch of the popular restaurant chain, transactions take place at some realized price τ. The empirical evidence consists of vectors of prices and individual demands $[(\tau^m, \mathbf{y}^m), m \in M]$, where m indicates a particular neighborhood and M is the set of all neighborhoods in the sample. The analyst wants to learn about the distribution of individual demands that would occur (in the study population) if groups were to receive a price t, i.e., $P[y(t)]$. (This model can easily accommodate price discrimination across consumers.)

The simultaneous equations of system (1) are the primitives of the model. Without further constraints, system (1) might have multiple solutions or no solution. Tamer (2003) has named cases of multiple solutions as incomplete models; models with no solution have *a priori* no predictive power. Previous analysis have posed specific functional forms on system (1), e.g.,

$$y_j = 1\left[\alpha_1 + (\alpha_2/|G|)\textstyle\sum_{i \in G} y_i + \varepsilon_j \geq t\right], \, j \in G \tag{3}$$

where $1(.)$ is an indicator function, and ε_j is a random term that differentiates individuals. Brock and Durlauf (2001a, b, 2007), among others, address point and partial identification of the structural parameter α_2 that captures the strength of the social interactions.[7] Awareness of the parameters and the distribution of latent variables (i.e., ε_j) imply knowledge of the demand functions, and this information permits counterfactual predictions of individual demands for different prices. If the parameters are only partially identified, or the equilibrium is not unique, then the predictions comprise a set of possible outcome distributions. These procedures work when the maintained parametric forms approximate well the true demand functions. We use a nonparametric approach, that relies on economically meaningful assumptions on (1), and allows counterfactual predictions even when there are multiple equilibria.

In this application there are two natural assumptions on the system of structural equations: individual j's demand increases with the decisions of the other members of his group (positive

[7] This example deviates from Brock and Durlauf (2001a, b, 2007) in many respects. It is, nevertheless, the simplest possible representation to clarify how our approach differs from the previous ones.

endogenous interactions); and individual j's demand is a downward sloping function of the market price (monotone treatment response).[8] The motivation for the first condition may be that a person wants to conform with others due to peer pressure or social influence, or because he thinks that people go to the popular restaurant because it has good quality food. The second condition is quite convincing, given the nature of the service we are studying.[9]

We show the last two assumptions imply clear restrictions on the demands at equilibrium. First, for each neighborhood, the system of structural equations has a smallest and a largest solution. Second, under either extremal equilibrium selection rules, adaptive learning behavior, or extra conditions leading to strong comparative statics, the equilibrium demands decrease in the potential price t.[10] We provide a probabilistic framework, that translates this comparative statics result into sharp bounds for the distribution of potential outcomes, or demands, $P[y(t)]$. That is, we identify two extreme distributions that are functions of the observable data, so that any distribution in between cannot be rejected as the true one.

The two discussed conditions assume nothing about the process of treatment selection, i.e., price determination. However, in this model it would be reasonable to think that the owner of the restaurant chain is more likely to set higher prices to those neighborhoods with stronger demands. We show that this assumption also leads to a sharp identification region for $P[y(t)]$ in terms of stochastic dominance. This inference approach relies on the use of realized treatments as monotone instrumental variables.

The next analysis formalizes and extends all previous ideas to frameworks where groups are small and large, respectively. The restaurant example is closer to the second set-up.

[8] Molinari and Rosen (2008) study the identification power of equilibrium in supermodular games, taking advantage of similar properties.

[9] The analyst observes realized prices and quantities in different neighborhoods, and uses the described assumptions to make inference on the counterfactual demands. Since these assumptions are valid irrespective of the behavior of the firms in the market, his predictions are independent of the shape of the supply functions.

[10] Although the model here is similar to Becker (1991) our implications differ from his results because he also assumes a vertical supply function and studies price determination.

3 Identification of Treatment Response Models

3.1 Model of Social Interactions where Identities Matter

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 coworkers, and pairs of patients and doctors fit in here.

3.1.1 Econometric Model and the Analyst's Problem

The population J is partitioned into a finite set of classes L, labelled $1, 2, ..., |L|$, i.e., $J \equiv \left(J_1, J_2, ..., J_{|L|}\right)$ with j_l as a typical element of J_l. Each group is composed of $|L|$ individuals that belong to different classes, e.g., in a model of marital decisions a group is a married couple and classes may refer to gender. This specification is analogous to the one used in the econometric analysis of games (see, e.g., Tamer (2003)). We indicate by $t \in T$ a potential treatment, and allow treatments to specify different policies for individuals that belong to different classes.

The behavior or achievement of individual j_l in G, y_{j_l}, depends on the treatment received by his group and the behavior of the other group members, $\mathbf{y}_{-j_l} \equiv [y_{j_m}, j_m \in G, m \neq l]$, i.e.,

$$y_{j_l} = f_{j_l}\left(t, \mathbf{y}_{-j_l}\right), \, j_l \in G. \tag{4}$$

We assume social interactions occur within groups, and group membership is known to the econometrician. If the underlying model is a (complete information) game, we can think of (4) as the best-reply function of player j_l to the profile of actions of the other players in G. The simultaneous solution to the system of structural equations (4)

$$\mathbf{y}\left(t\right) \equiv \left[y_{j_l}\left(t\right), \, j_l \in G\right] \in Y^{|L|} \tag{5}$$

is the vector of potential outcomes for treatment t.

We typify groups through the relevant feature of the group members, and relate the distribution of types of groups to the underlying process of group formation. The next example sheds light on the probabilistic approach we describe afterwards.

Example 1 *We model a scenario of a tobacco prevention program and smoking decisions. The population has six people, three boys and three girls (i.e., $J \equiv (J_1, J_2)$ where $J_1 = (1_1, 2_1, 3_1)$ is the set of boys and $J_2 = (1_2, 2_2, 3_2)$ is the set of girls). A group is defined as a dating couple.*

Members of each couple decide whether to smoke. The treatment takes a value of one if the couple receives information about the risks of smoking, and zero otherwise. An individual's decision about smoking depends on the treatment received and the smoking decision of his partner. Thus, $T = \{0, 1\}$, $Y = \{0, 1\}$ and $f_{j_l} : T \times Y \to Y$ with $j = 1, 2, 3$ and $l = 1, 2$.

Couples differ in 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 to make—and thereby 16^2 (or 256) types of couples.

Let us assume the population has outcome functions as the ones in Figure 1. Here y_{j_1} and y_{j_2} indicate the outcome that the jth boy and girl, respectively, would select under each of the four possible configurations of treatments and partners' decisions, e.g., the one in square brackets means the second boy would smoke if he does not receive information about the risks of smoking and his girlfriend smokes.

Figure 1: Structural Functions of the Population

Boys

	Treatments and Partners' Decisions			
Outcomes	(0, 0)	(0, 1)	(1, 0)	(1, 1)
y_{11}	0	1	0	0
y_{21}	0	[1]	0	0
y_{31}	0	1	0	1

Girls

	Treatments and Partners' Decisions			
Outcomes	(0, 0)	(0, 1)	(1, 0)	(1, 1)
y_{12}	0	1	0	0
y_{22}	0	1	0	0
y_{32}	0	1	0	1

The population described in Figure 1 involves only two kinds of girls and boys—out of the possible 16—and thereby four types of couples—out of the possible 256. The distribution of types of couples depends not only on the smoking preferences of the individuals in the population, but also on their dating preferences. Suppose, for instance, that these individuals prefer to date rather than remaining alone, and to share their time with partners with similar smoking *tastes. Thus, two out of the three couples will have members that smoke if and only if they are uninformed and their partners smoke, and the other one will have members that smoke if and only if their partners smoke irrespective of the treatment. The other two types of groups involve members with different outcome functions, and have no chance to be formed given the assumed dating preferences.*

Let us formalize the probabilistic framework we use. In the system of equations (4), the structural function f_{j_l} specifies the outcome that individual j_l would experience when facing any potential treatment and any vector of outcomes by the other members of the same group. Within each class two distinct agents may differ with respect to their structural equations, that is, individuals may react likewise or differently to the same incentives. We assume the cardinality of T and Y is countable.[11] Hence, there are countably many distinct structural equations within each class, and countably many different systems of structural equations (4) 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 $f_{kl} : T \times Y^{|L|-1} \to Y$ for $l \in L$. We let K indicate the set of possible types of groups. In Example 1 there are 256 types of groups, i.e., $K = \{1, 2, ..., 256\}$.

The relative proportion of different agents within each class, and the underlying matching process of individuals across classes define the distribution of types of groups in the universe of groups $\mathcal{U}$. Let π_k denote the fraction of groups which are of type k, so that $\boldsymbol{\pi} \equiv (\pi_k, k \in K)$ is the discrete distribution of types in $\mathcal{U}$. As an example of why the underlying sorting mechanism matters, note that economists and sociologists have long observed that individuals choose mates who have socioeconomic profiles similar to their own. If this hypothesis were true and we were interested in the decisions of married couples, we should expect a high proportion of spouses in the population with very similar structural functions, that is, that respond alike when facing the same incentives. Marked differences between spouses would be more often observed if men and women were, alternatively, randomly paired. In Example 1, the distribution of types of groups could be described by $\boldsymbol{\pi} = (2/3, 1/3, 0, ..., 0)$ where $\boldsymbol{\pi}$ has 256 elements. We assume group formation is independent of t, in the sense that potential treatments do not

[11] Our results extend to uncountable sets up to measurability considerations. Manski (2007) uses a similar approach to develop partial identification of counterfactual choice probabilities.

affect the distribution of types of groups, $\boldsymbol{\pi}$.[12] Nevertheless, realized treatments may provide rich information about the type of groups that could have generated the data (see the restaurant example in Section 2).

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

Example 2 *The model described can be used to study retirement decisions of husbands and wives. Let J_1 be the set of men and J_2 the set of women, with the universe of groups $\mathcal{U}$ 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 greater pleasure from retirement if they retire together.*

The next sub-section studies the identification power of monotone shape restrictions that are naturally satisfied in many social-interactions models.

3.1.2 Identification Region for $P[\mathbf{y}(t)]$

The next sections impose three alternative monotone conditions on the structural equations to derive distributional bounds for $P[\mathbf{y}(t)]$ via monotone comparative statics.

Coherence of the Model This sub-section shows equilibrium existence and relates the distribution of potential and realized outcomes to the model predictions. Using these distributions, we then provide sharp bounds for $P[\mathbf{y}(t)]$ in terms of stochastic dominance. The next sub-sections build on this initial result.

[12] Manski (2011) describes a similar condition by saying that reference groups are treatment-invariant, and thus non-manipulable (i.e., the social planner cannot use the treatments to change a person's influence group).

Without further restrictions, the system of equations (4) might have multiple solutions or no solution. The next assumption imposes a shape restriction on $(\mathbf{f}_k, k \in K)$, which precludes the possibility of incoherence. Throughout, $\geq$ indicates the coordinatewise order.[13]

$(A1)$ Let $Y \subseteq \mathbb{R}$ be compact. Let $\mathbf{y}, \mathbf{y}' \in Y^{|L|-1}$. We assume, $\forall t \in T$ and $\forall l \in L$,

$$\mathbf{y} \geq \mathbf{y}' \Longrightarrow f_{kl}(t, \mathbf{y}) \geq f_{kl}(t, \mathbf{y}'). \tag{6}$$

This condition holds for all $k \in K$.[14]

Condition $A1$ requires the outcome of each individual to increase with the vector of outcomes by the other members of the group (i.e., positive endogenous interactions) and the set of feasible outcomes to be a compact subset of the real line. For further reference, we denote by $\underline{Y}$ and $\overline{Y}$ the minimum and maximum of Y respectively.

Within the broad range of models that satisfy this condition, there are two cases of particular interest: First, supermodular games, which are games where the complementarities in the payoffs translate into best-replies that increase in each rival's action. In this case, the system of structural equations (4) should be interpreted as players' best-response functions. Second, models with positive externalities, in which the interactions occur at the level of payoffs, e.g., peer effects in the classroom. In this second case the system of structural equations (4) should be understood as agents' payoffs or achievements.

Let $\phi(t, k)$ denote the solution set of system $\mathbf{f}_k$ for a potential treatment t. In the next lemma, smallest and largest mean coordinatewise smallest and largest.

Lemma 3 *If $A1$ holds, then $\phi(t, k)$ has a smallest and a largest solution $\forall t \in T$ and $\forall k \in K$.*

The system of structural equations leads to a multi-valued function that maps possible outcomes into itself, so that the set of solutions of the model coincides with the set of fixed

[13] When $\geq$ is applied to a subset of $\mathbb{R}$, then it is just the natural order on the reals.

[14] In Sub-section 3.1.1 we assumed Y was countable. Assumption $A1$ imposes further structure on the outcome set. While the restrictions in assumption $A1$ are required by the theoretical methodology we use, the previous condition is imposed to simplify the exposition.

points of this artificial mapping. Under positive endogenous interactions, this function is increasing, and then the proof of Lemma 3 follows directly from Tarski's Fixed Point Theorem [see Section 7 (Proofs)]. Although the requirements of Tarski's theorem are satisfied in many social-interactions models, most of the empirical analysis imposes continuity on the system of structural equations to show existence via Brower's Fixed Point Theorem.[15]

We restrict attention to deterministic equilibrium selection rules, that are a priori allowed to depend on both t and k. We let $\mathbf{y}_k(t)$ indicate the element of $\phi(t,k)$ that is selected by a type-k group when it receives treatment t. Thus, the probability that the vector of potential outcomes falls in a set $B \subseteq \mathbb{R}^{|L|}$ is given by

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

Let $\pi_{k|\tau}$ indicate the fraction of groups in $\mathcal{U}$ which are of type k conditional on τ, and let $P(\tau)$ denote the distribution of realized treatments across groups. We make 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

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

with $P(\mathbf{y} \in B|\tau = s) = \sum_{k \in K} 1[\mathbf{y}_k(s) \in B]\, \pi_{k|\tau=s}$. Then $P(\mathbf{y})$ is a mixture of realized outcome distributions conditional on observed treatments, with $P(\tau)$ as the mixing probability function.

The research objective is inference about the outcome distribution $P[\mathbf{y}(t)]$ that describes treatment response across groups. The identification problem is captured by

$$P[\mathbf{y}(t)] = P[\mathbf{y}(t)|\tau = t]\, P(\tau = t) + P[\mathbf{y}(t)|\tau \neq t]\, P(\tau \neq t). \tag{9}$$

Given our assumptions, the equality (9) follows by the Law of Total Probability.

The empirical evidence reveals $P[\mathbf{y}(t)|\tau = t] = P(\mathbf{y}\,|\tau = t)$, $P(\tau = t)$, and $P(\tau \neq t)$. The sampling process alone remains silent about the potential outcome distribution for those groups that have realized treatments different from the potential one, i.e., $P[\mathbf{y}(t)|\tau \neq t]$. However, by assumption $A1$, the set of possible vectors of outcomes is bounded from below and above

[15] Some exceptions are Ackerberg and Gowrisankaran (2006), Jia (2008), and De Paula (2009).

by $\underline{Y}^{|L|}$ and $\overline{Y}^{|L|}$ respectively. Then the degenerate distributions $P\left(\underline{Y}^{|L|}\right)$ and $P\left(\overline{Y}^{|L|}\right)$ are sharp lower and upper bounds for the counterfactual $P\left[\mathbf{y}\left(t\right)|\tau \neq t\right]$. Having proved equilibrium existence and described the nature of $P\left[\mathbf{y}\left(t\right)\right]$, the next result builds on Manski (1990).

Let $\triangle_{Y^{|L|}}$ denote the set of multivariate distribution functions that are in line with the nature of $Y^{|L|}$, and let H stand for the set of distributions that are consistent with the initial assumptions given the data, i.e., the sharp identification region for $P\left[\mathbf{y}\left(t\right)\right]$. The next identification area is expressed in terms of stochastic dominance (st).[16]

Proposition 4 *Assume A1 holds. Then, for all $t \in T$,*

$$H\left\{P\left[\mathbf{y}\left(t\right)\right]\right\} = \left\{\delta \in \triangle_{Y^{|L|}} : \begin{array}{c} P\left(\mathbf{y}\,|\tau = t\right)P\left(\tau = t\right) + P\left(\overline{Y}^{|L|}\right)P\left(\tau \neq t\right) \\ \geq_{st} \delta \geq_{st} \\ P\left(\mathbf{y}\,|\tau = t\right)P\left(\tau = t\right) + P\left(\underline{Y}^{|L|}\right)P\left(\tau \neq t\right) \end{array}\right\}. \tag{10}$$

Remark. All bounds we provide in this paper, including the ones in Proposition 4, are sharp. This property should be directly understood from the way we defined the set H.

The endpoints of (10) almost always differ, so $P\left[\mathbf{y}\left(t\right)\right]$ is typically only partially identified. Moreover, if $P\left(\tau \neq t\right) = 1$ then the bounds are sharp but uninformative. The next sub-sections introduce two other restrictions that shed extra light on the counterfactual $P\left[\mathbf{y}\left(t\right)|\tau \neq t\right]$.

Identification Using Monotone Treatment Response We now assume the structural equations are weakly increasing in potential treatments. Opposite results apply if we reverse the signs of these effects.

($A2$) Let $(T, \geq)$ be a partially ordered set. Let $t, t' \in T$. We assume, $\forall \mathbf{y} \in Y^{|L|-1}$ and $\forall l \in L$,

$$t \geq t' \Longrightarrow f_{kl}\left(t, \mathbf{y}\right) \geq f_{kl}\left(t', \mathbf{y}\right) \tag{11}$$

This condition holds for all $k \in K$.

Manski (1997) introduces a monotone shape restriction to study identification in models where a person's outcome varies only with his own treatment. He imposes the monotone

[16] See Section 7 (Proofs) for the typical multivariate characterization of first order stochastic dominance.

assumption directly on the predicted outcomes. In our study, the primitives are the structural equations, so our condition is imposed on system $\mathbf{f}_k$, and the monotone restriction on the potential outcomes follows as an implication of the latter. Due to the possibility of multiple equilibria, our result requires a bit more of structure.

Without further restrictions, monotone treatment response allows counterfactual predictions only at the extremal solutions, i.e., if $A1$ and $A2$ hold then the extremal vectors in $\phi(t,k)$ increase in t [see, e.g., Milgrom and Roberts (1990), Theorem 6]. Then, if groups always select either the largest or the smallest equilibrium we can predict the directional effect of t on the distribution of potential outcomes. Since this restriction is quite strong in some applications, we provide two alternative conditions that lead to the same result. One of them imposes a sensible adaptive dynamics on agents' behavior, which guarantees the monotone comparative statics holds at any selected equilibrium. The other one ensures $\phi(t,k)$ is strongly increasing in t, so that the result holds irrespective of the equilibrium selection rule.

Let us define a simple adaptive dynamics $\mathcal{A}(\mathbf{y},t,k)$ as a sequence $\{\mathbf{y}^i\}_{i=0}^{\infty}$ such that

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

This dynamics begins with an initial equilibrium $\mathbf{y}^0$, then assumes people behave according to their respective outcome functions for treatment t inducing a new vector $\mathbf{y}^1$, and the process repeats indefinitely. We say the selection rule is adaptive if the elements of $\mathcal{A}(\mathbf{y}^0,t,k)$ describe the evolution of outcomes as we apply treatment t to the group and the initial outcome is $\mathbf{y}^0$. We think of $\mathbf{y}^0$ as the selected equilibrium for some initial treatment t_0. The next assumption is key to the results below.

($A3$) One of the next 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; (ii) the selection rule is adaptive and T is a chain; or (iii) $\mathbf{f}_k(t,\inf\phi(t',k)) \geq \sup\phi(t',k)\ \forall t > t'$, and $\forall k \in K$.

Conditions $A3$(i) and $A3$(ii) restrict the equilibrium selection mechanism. Since the models discussed here often allow to Pareto rank the equilibrium outcomes [or, as in Brock and Durlauf (2001a), to rank them in terms of expected welfare] selection rules on extremal equilibria are frequently invoked in applied studies [see, e.g., Gowrisankaran and Stavins (2004)]. The second one imposes a reasonable learning to explain how the endogenous outcomes evolve as we vary the treatment. It guarantees that any selected equilibrium (starting at some initial $\mathbf{y}^0$) will necessarily increase (coordinatewise) in t, and rules out selection of unstable equilibria.[17]

Condition $A3$(iii) markedly differs from the other two. When it is combined with $A1$ and $A2$, 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.[18] Thus, any selection from the equilibrium set $\phi(t,k)$ is necessarily increasing in the potential treatment. The drawback of this condition is that it is non-primitive, i.e., to check it requires knowledge of the extremal elements of the solution sets. It amounts for a very strong treatment effect which is more likely to hold if the treatment levels were far away from each other.

Lemma 5 *Assume $A1$, $A2$, and $A3$ hold. Then $\mathbf{y}_k(t) \geq \mathbf{y}_k(t')$, $\forall t, t' \in T$ such that $t > t'$ and $\forall k \in K$.*

If $A1$, $A2$ and $A3$ hold, then, by Lemma 5, 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, $\overline{Y}^{|L|}$. 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 $A3$ hold. Then, for all $t \in T$,*

$$H\{P[\mathbf{y}(t)]\} = \left\{ \delta \in \triangle_{Y^{|L|}} : \begin{array}{c} P(\mathbf{y} \mid \tau \geq t) P(\tau \geq t) + P\left(\overline{Y}^{|L|}\right) P(\tau \ngeq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y} \mid \tau \leq t) P(\tau \leq t) + P\left(\underline{Y}^{|L|}\right) P(\tau \nleq t) \end{array} \right\}. \quad (12)$$

[17]Vives (1990) uses adaptive dynamics to study stability in supermodular games. Inspired by the Correpondence Principle of Samuelson, Echenique (2002) relates the latter idea to monotone comparative statics.

[18]See Echenique and Sabarwal (2003).

The multivariate standard stochastic order is closed with respect to marginalization. Thus, Proposition 6 also implies sharp bounds for the marginal distributions of potential outcomes for those individuals that fit in any subset of classes $S \subset L$, i.e., $P[\mathbf{y}_S(t)]$ where $\mathbf{y}_S(t)$ is the restriction of $\mathbf{y}(t)$ to S. Corollary 7 (without proof) captures this observation.[19] In the next result we write L/S for the set of classes in L different from those in S.

Corollary 7 *Assume A1, A2, and A3 hold. Then,*

$$H\{P[\mathbf{y}_S(\mathbf{t})]\} = \left\{ \delta \in \triangle_{Y^{|S|}} : \begin{array}{c} P(\mathbf{y}_S|\tau \geq t) P(\tau \geq t) + P\left(\overline{Y}^{|S|}\right) P(\tau \ngeq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y}_S|\tau \leq t) P(\tau \leq t) + P\left(\underline{Y}^{|S|}\right) P(\tau \nleq t) \end{array} \right\} \tag{13}$$

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

The same methodology can be used to derive bounds for any increasing function g from $\mathbf{y}(t)$ to $\mathbb{R}^n$ (with $n \leq |L|$), e.g., the mean of the vector of potential outcomes $E[\mathbf{y}(t)]$.[20]

The next sub-section introduces a restriction that allows making counterfactual predictions for a subset of groups by using the empirical evidence of an alternative subset.

Identification Using Monotone Treatment Selection Many studies have established identification results by assuming the outcome functions are statistically independent of realized treatments, i.e., $P[\mathbf{y}(t)|\tau = t] = P[\mathbf{y}(t)]$. 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. This sub-section extends their finding to interactions-based models.

We next introduce a restriction on the primitives which guarantees that groups that have larger realized treatments have potential outcomes that are statistically larger. Before doing so we define a (natural) partial order on $(\mathbf{f}_k, k \in K)$.

[19] Corollary 7 derives from Proposition 6, using Müller and Stoyan (2002), Theorem 3.3.10, p. 94.

[20] Stoye (2010) could also be used to identify spread parameters (e.g., variance) of marginal distributions.

Definition 8 *We say* $\mathbf{f}_k \geq \mathbf{f}_{k'}$ *if* $f_{kl}(t, \mathbf{y}) \geq f_{k'l}(t, \mathbf{y}) \ \forall (t, \mathbf{y}) \in T \times Y^{|L|-1}$ *and* $\forall l \in L$.

That is, we say $\mathbf{f}_k$ is larger than $\mathbf{f}_{k'}$ if it is pointwise greater. Section 7 (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 ones in $\phi(t, k')$. The identification strategy we pursue requires the comparison of equilibrium distributions for two sets of groups that received different treatments (not just two groups). Thus, the analysis that follows extends this intermediate result.

Let $\mathbf{f}$ denote a random vector of functions with support $(\mathbf{f}_k, k \in K)$, and define (for all $k \in K$) $P(\mathbf{f} = \mathbf{f}_k \mid \tau) \equiv \pi_{k|\tau}$. The key assumption is as follows.

$(A4)$ Let $(T, \geq)$ be a partially ordered set. Let $s, s' \in T$. We assume

$$s \geq s' \Longrightarrow P(\mathbf{f} \mid \tau = s) \geq_{st} P(\mathbf{f} \mid \tau = s'). \tag{14}$$

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 weakly larger structural functions than those that self-select into lower ones. Here again, previous to examine the identification power of adding monotone treatment selection, we need to introduce another restriction.

$(A5)$ One of the next two conditions holds (i) groups select either the smallest or the largest element of the solution set, and 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 $A5$(i) requires the selection rule to be the same across groups, a natural restriction given that we want to compare potential outcomes between sets of groups with different realized treatments. Condition $A5$(ii) (when added to $A1$) 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 one.

If $A1$ holds, then the added benefit of $A5$ 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.

Lemma 9 *Assume A1, A4, and A5 hold. Let $s, s' \in T$. Then, for all $t \in T$,*

$$s \geq s' \Longrightarrow P\left[\mathbf{y}(t) \mid \tau = s\right] \geq_{st} P\left[\mathbf{y}(t) \mid \tau = s'\right]. \quad (15)$$

Thus, groups that select larger treatments have potential outcomes that are statistically larger. The empirical evidence reveals $P\left[\mathbf{y}(t) \mid \tau = t\right] = P(\mathbf{y} \mid \tau = t)$, $P(\tau \leq t)$ and $P(\tau \nleq t)$. If $A1$, $A4$ and $A5$ hold, then $P(\mathbf{y} \mid \tau = t)$ is a lower bound for $P\left[\mathbf{y}(t) \mid \tau \geq t\right]$, and it is an upper bound for $P\left[\mathbf{y}(t) \mid \tau \leq t\right]$. These observations, formalized in Corollary 22, in Section 7 (Proofs), are direct implications of Lemma 9. We next state the identification result.

Proposition 10 *Assume A1, A4, and A5 hold. Then, for all $t \in T$,*

$$H\left\{P\left[\mathbf{y}(t)\right]\right\} = \left\{ \delta \in \triangle_{Y^{|L|}} : \begin{array}{c} P(\mathbf{y} \mid \tau = t) P(\tau \leq t) + P\left(\overline{Y}^{|L|}\right) P(\tau \nleq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y} \mid \tau = t) P(\tau \geq t) + P\left(\underline{Y}^{|L|}\right) P(\tau \ngeq t) \end{array} \right\}. \quad (16)$$

These bounds are neither tighter nor looser than the ones in Proposition 6, they are just different.[21] Here again, Proposition 10 implies sharp bounds for the marginal distributions of potential outcomes of individuals in the subset of classes $S \subset L$, i.e., $P\left[\mathbf{y}_S(t)\right]$ where $\mathbf{y}_S(t)$ is the restriction of $\mathbf{y}(t)$ to S. Corollary 11 (without proof) formalizes this claim.

Corollary 11 *Assume A1, A4, and A5 hold. Then,*

$$H\left\{P\left[\mathbf{y}_S(t)\right]\right\} = \left\{ \delta \in \triangle_{Y^{|S|}} : \begin{array}{c} P(\mathbf{y}_S \mid \tau = t) P(\tau \leq t) + P\left(\overline{Y}^{|S|}\right) P(\tau \nleq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y}_S \mid \tau = t) P(\tau \geq t) + P\left(\underline{Y}^{|S|}\right) P(\tau \ngeq t) \end{array} \right\} \quad (17)$$

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

Although the sets of assumptions ($A1$, $A2$, and $A3$) and ($A1$, $A4$, and $A5$) are not individually refutable, the combined restriction can be shown false given the data. Specifically, together they imply that $P(\mathbf{y} \mid \tau = s) \geq_{st} P(\mathbf{y} \mid \tau = s')$ for all $s, s' \in T$ such that $s > s'$. 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.

[21] A more precise region of identification for $P\left[\mathbf{y}(t)\right]$ is obtained when we are confident about $A1 - A5$, so that we can combine both results [see Manski and Pepper (2000) for an individualistic model].

3.2 Model of Anonymous Social Interactions

In this section groups are composed of a large (but finite) number of individuals, and social interactions are anonymous in the sense that each person assigns an identical weight to the outcome of every other member of its reference group. Models of crimes, schooling, infectious diseases, and addictions fit well here.

An important difference between this section and the previous one is that now groups are allowed to differ in the number of people. This feature makes the comparative statics analysis more challenging and calls for a different set of conditions to validate similar results.

3.2.1 Econometric Model and the Analyst's Problem

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

$$y_j = f_j[t, P(y)], \; j \in G \tag{18}$$

where $P(y)$ is the distribution of outcomes induced by a vector $\mathbf{y} \equiv (y_j, j \in G)$.[22] We assume $|G|$ is large enough, so that changes in any individual outcome hardly affect $P(y)$.

We say a group is of type k if its members have structural equations $\mathbf{f}_k \equiv [f_{kl}(.), l \in G_k]$, where $f_{kl} : T \times \triangle_Y \rightarrow Y$ and G_k denotes the set of members for a type-k group. Let π_k denote the fraction of individuals that 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, here π_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 for treatment t (at the individual level), i.e., $P[y(t)]$.

Example 12 *A model as the one described can be used to study crime rates and social interactions. Let J be the set of persons in a given state, and let us define a group as the set of*

[22] Formally, for all set $B \subset \mathbb{R}$, $P(y \in B) = \sum_{j \in G} 1(y_j \in B)(1/|G|)$.

people that live in one of its cities. The outcome of interest is the decision to commit a crime at the individual level, and the treatment is police per capita at the level of the city.

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)].

3.2.2 Identification Region for $P[y(t)]$

The next sub-section studies the identification power of alternative monotone shape restrictions that are naturally satisfied in various social interactions models.

Coherence of the Model The first assumption imposes a monotone shape restriction on $(\mathbf{f}_k, k \in K)$, which, as in Sub-section 3.1.2, guarantees equilibrium existence.

$(A1')$ Let $Y \subseteq \mathbb{R}$ be compact. Let $P(y), P(y') \in \Delta_Y$. We assume, $\forall l \in G_k$ and $\forall t \in T$,

$$P(y) \geq_{st} P(y') \Longrightarrow f_{kl}[t, P(y)] \geq f_{kl}[t, P(y')]. \tag{19}$$

This condition holds for all $k \in K$.

Let $\varphi(t,k)$ denote the set of solutions to system $\mathbf{f}_k$ for a given $t \in T$. Since the analyst's objective is inference on the outcome distributions at the individual level, we let the elements of $\varphi(t,k)$ be the distribution functions induced by the solution vectors. In the next lemma smallest and largest are with respect to the partial order of standard stochastic dominance.

Lemma 13 *Assume $A1'$ holds, then $\varphi(t,k)$ has a smallest and a largest distribution $\forall t \in T$ and $\forall k \in K$.*

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

$$P[y(t) \in B] = \textstyle\sum_{k \in K} P[y_k(t) \in B]\, \pi_k \tag{20}$$

and it is linear in π_k. Let $\pi_{k|\tau}$ be the proportion of individuals in J that belong to type-k groups conditional on the realized treatment τ. 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

$$P(y \in B) = \sum_{s \in T} P[y(s) \in B | \tau = s] P(\tau = s) \tag{21}$$

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

The next proposition is based on the same route for inference of Proposition 4.

Proposition 14 *Assume $A1'$ holds. Then, for all $t \in T$,*

$$H\{P[y(t)]\} = \left\{ \delta \in \triangle_Y : \begin{array}{c} P(y \mid \tau = t) P(\tau = t) + P(\overline{Y}) P(\tau \neq t) \\ \geq_{st} \delta \geq_{st} \\ P(y \mid \tau = t) P(\tau = t) + P(\underline{Y}) P(\tau \neq t) \end{array} \right\}. \tag{22}$$

Identification Using Monotone Treatment Response The next condition assumes individual outcomes increase in t on T.

($A2'$) Let $(T, \geq)$ be a partially ordered set. Let $t, t' \in T$. We assume, $\forall l \in G_k$ and $\forall P(y) \in \Delta_Y$,

$$t \geq t' \Longrightarrow f_{kl}[t, P(y)] \geq f_{kl}[t', P(y)] \tag{23}$$

This condition holds for all $k \in K$.

If $A1'$ and $A2'$ hold, then the smallest and the largest distributions in $\varphi(t, k)$ increase in t with respect to the standard stochastic order. Thus, here again, extremal equilibrium selection rules are enough to make counterfactual predictions as we vary the treatment. We introduce alternative assumptions that lead to the same result.

Let us define an adaptive dynamics $\mathcal{A}(P(y), t, k)$ as a sequence $\{P^i(y)\}_{i=0}^{\infty}$ such that

$$P^0(y) = P(y), \;\; P^i(y \in B) = (1/|G_k|) \sum_{l \in G_k} 1\left(f_{kl}\left[t, P^{i-1}(y)\right] \in B\right), i \geq 1.$$

We say the selection rule is adaptive if the elements of $\mathcal{A}(P^0(y), t, k)$ describe the evolution of the outcomes as we apply treatment from t to the group. We finally define, for all set $B \subseteq \mathbb{R}$,

$$F_k(y \in B \mid t, P(y)) \equiv (1/|G_k|) \sum_{l \in G_k} 1\{f_{kl}[t, P(y)] \in B\}. \tag{24}$$

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 equilibrium $P(y)$. Here again, we think of $P^0(y)$ as an equilibrium distribution for some initial treatment t_0.

($A3'$) One of the next 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; (ii) the selection rule is adaptive and T is a chain; or (iii) $F_k(y \mid t, \inf \phi(t',k)) \geq_{st} \sup \phi(t',k)$ $\forall t > t'$, and $\forall k \in K$.

The following result is the analogue to Proposition 5.

Lemma 15 *Assume A1′, A2′, and A3′ hold. Then* $P[y_k(t)] \geq_{st} P[y_k(t')]$, $\forall t, t' \in T$ *such that* $t > t'$ *and* $\forall k \in K$.

The next proposition captures the main implication of the lemma above.

Proposition 16 *Assume A1′, A2′, and A3′ hold. Then, for all* $t \in T$,

$$H\{P[y(t)]\} = \left\{ \delta \in \triangle_Y : \begin{array}{c} P(y \mid \tau \geq t) P(\tau \geq t) + P(\overline{Y}) P(\tau \ngeq t) \\ \geq_{st} \delta \geq_{st} \\ P(y \mid \tau \leq t) P(\tau \leq t) + P(\underline{Y}) P(\tau \nleq t) \end{array} \right\}. \quad (25)$$

Quantiles are often parameters of interest in applied studies. For $\alpha \in (0,1)$, the α–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 one, then all the quantiles of the former are larger than the corresponding quantiles of the latter. Hence, Proposition 16 can be easily reformulated in terms of the quantiles of the pertinent distributions, e.g., 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 Sub-section 3.1.2 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.

Identification Using Monotone Treatment Selection This sub-section provides sufficient conditions to validate the use of realized treatments as monotone instrumental variables, in the sense of outcome monotonicity. To this end, we need to introduce a partial order on $(F_k, k \in K)$, as defined in equation (24).

Definition 17 *We say* $F_k \geq F_{k'}$ *if* $F_k [y \mid t, P(y)] \geq_{st} F_{k'} [y \mid t, P(y)] \; \forall [t, P(y)] \in T \times \triangle_Y$.

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 one of the type-k' group for any conditioning event. Let F denote a random function with support $(F_k, k \in K)$, and define $P(F = F_k \mid \tau) \equiv \pi_{k|\tau}$ for all $k \in K$. The next condition is an aggregative version of condition $A4$.

$(A4')$ Let T be a partially ordered set. Let $s, s' \in T$. Then,

$$s \geq s' \Longrightarrow P(F \mid \tau = s) \geq_{st} P(F \mid \tau = s'). \tag{26}$$

Condition (26) states that the proportion of individuals that belong to groups with weakly higher structural functions increases with realized treatments. The motivation for this constraint is well captured by Example 12: 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. Here again, before elaborating on the identification power of adding the monotone treatment selection condition, we need to introduce another restriction.

$(A5')$ One of the next two conditions holds (i) groups select either the smallest or the largest element of the solution set, and the selection rule is the same $\forall k \in K$; or (ii) $F_k [y \mid t, \inf \varphi(t, k')] \geq_{st} \sup \varphi(t, k') \; \forall k, k' \in K$ such that $F_k \geq F_{k'}$

The next lemma is the analogue to Lemma 9.

Lemma 18 *Assume A1′, A4′, and A5′ hold. Let $s, s' \in T$. Then, for all $t \in T$,*

$$s \geq s' \Longrightarrow P\left[y\left(t\right) \mid \tau = s\right] \geq_{st} P\left[y\left(t\right) \mid \tau = s'\right]. \tag{27}$$

The following proposition derives as an implication of the last result.

Proposition 19 *Assume A1′, A4′, and A5′ hold. Then, for all $t \in T$,*

$$H\left\{P\left[y\left(t\right)\right]\right\} = \left\{ \delta \in \triangle_Y : \begin{array}{c} P\left(y \mid \tau = t\right) P\left(\tau \leq t\right) + P\left(\overline{Y}\right) P\left(\tau \nleq t\right) \\ \geq_{st} \delta \geq_{st} \\ P\left(y \mid \tau = t\right) P\left(\tau \geq t\right) + P\left(\underline{Y}\right) P\left(\tau \ngeq t\right) \end{array} \right\}. \tag{28}$$

4 Estimation from Sample Data

The bounds identified in Section 3 can be consistently estimated from sample data (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 many others, for partially identified parameters, could be adapted to construct confidence sets for the identified regions.

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 et. al. (1996), and the literature therein].

We use the model in Section 3.2 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 the outcome as a yes/no indicator taking the value one or zero, i.e., $Y = \{0, 1\}$. Sah (1991) presents a model where one individual's choice to become a criminal lowers the probability that any other individual ends up arrested. Since the police cannot be at two places at the same time, the higher the criminal activity in a given city is, the lower the probability of being punished. His argument justifies $A1'$.[23] 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. As we explained in Sub-section 3.2.2, it could also be 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. The last statement validates condition $A4'$. We will also assume $A3'$ and $A5'$ are satisfied.[24]

The analyst wants to learn about the fraction of people that would commit a crime in NY State, if all its cities were to be assigned a given level of police per capita. The next sub-section describes the data we use, and the last one shows our findings.

5.1 Data Set

Our data source is the Uniform Crime Reporting (UCR) program of the 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 level, and the population of each city. We decided to eliminate the city of NY from the sample, as its features (e.g., number of people) are markedly different from the characteristics of the other ones. Our data covers 47% of the remaining population of NY State (314 cities). To perform the analysis we discretized the level of police per capita in multiples of 0.0001.

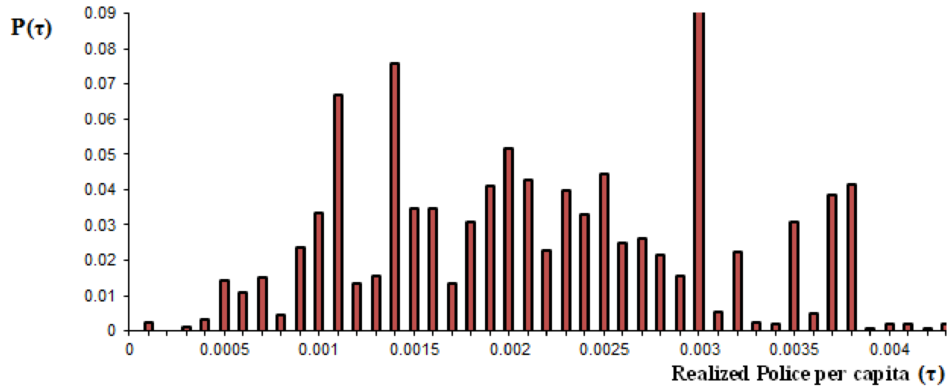
Figure 2 displays the data for the 99.4% of the sampled population—for expositional ease, the figure does not include a few observations with extremely high levels of police per capita, but these observations were taken into account in the estimation. For levels of police per capita

[23] This assumption is justified in the criminology literature by using many different arguments (see, for example, the theory of differential association by E. Sutherland). 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 $A1'$.

[24] Since this application uses the dual version of $A2'$, we should substitute increases by decreases in $A3'$(i), and $\geq_{st}$ by $\leq_{st}$ in $A3'$(iii).

between 0 and 0.005, $P(\tau)$ indicates the fraction of individuals in the sample that received treatment τ 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 sub-section estimates the outcomes of interest for levels of police per capita in that range of values.

Figure 2: Distribution of Realized Levels of Police per capita



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

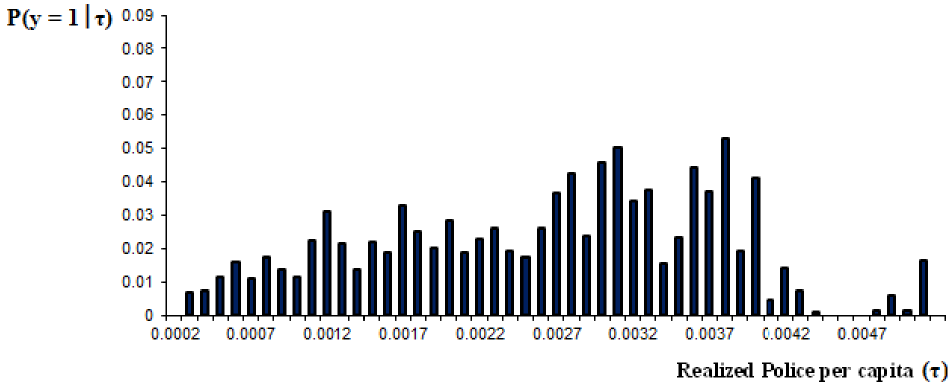
In Figure 3, $P(y = 1 \mid \tau)$ indicates the fraction of people that committed a crime in 2009 conditional on living in a city with a level of police per capita τ. Throughout this application we assume each individual has committed at most one crime during 2009.

5.2 Findings

Since the outcome of interest is binary (i.e., $Y = \{0, 1\}$), the fraction of people that 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.

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. These bounds are reported in the third and the fourth columns respectively, and are expressed in percentage points. They are the sample analogs of the bounds in Propositions 14, 16, and 19, respectively. In addition, we provide (in the last column) confidence sets for $P[y(t) = 1]$ at the 95% of confidence level. To

Figure 3: Criminal Activity Conditional on Realized Treatments



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

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.[25]

Figure 4: Lower and Upper Bounds for $P[y(t)=1]$ in %

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 $A2'$, $A3'$	2.90	89.25	(2.72, 92.10)
	$A1'$, $A4'$, $A5'$	1.79	89.26	(1.74, 92.11)
$t = 0.002$	$A1'$	0.12	94.97	(0.07, 96.98)
	$A1'$, dual of $A2'$, $A3'$	2.18	54.94	(1.95, 59.47)
	$A1'$, $A4'$, $A5'$	1.39	55.04	(1.29, 59.56)
$t = 0.003$	$A1'$	0.50	89.74	(0.35, 92.48)
	$A1'$, dual of $A2'$, $A3'$	1.23	18.42	(1.02, 21.74)
	$A1'$, $A4'$, $A5'$	1.26	20.03	(1.07, 23.29)
$t = 0.004$	$A1'$	0.00	99.80	(0.00, 100.0)
	$A1'$, dual of $A2'$, $A3'$	0.02	04.10	(0.00, 05.04)
	$A1'$, $A4'$, $A5'$	0.01	01.47	(0.00, 02.42)

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

We can clearly see in Figure 4 that $A1'$ alone is practically uninformative in the four cases, as $P[y(t)=1]$ is (by definition) between zero and one for all t. The mere addition of either

[25] For $t = 0.004$, the lower end of one of the confidence regions takes a value slightly below 0, and the upper end of a second one, slightly above 100. We wrote 0 and 100 instead, as the true parameter of interest will always take values between 0 and 100.

(dual of $A2', A3'$) or $(A4', A5')$ 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 $A2', A3'$) reduce that interval to $(0.0, 5.04)$. The reason is that a very small fraction (i.e., 0.25%) of the population has realized treatment $\tau = 0.004$, 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 shed light on the upper bound, other segment helps to estimate the lower bound, and a third group of observations are used to construct both. For this treatment level, $(A4', A5')$ provides even more information as the interval shrinks to $(0.0, 2.42)$. The reason for this being that the groups that received this large treatment displayed a very low criminal activity.[26]

The figure shows a very strong treatment effect, in the sense that (under the last two assumptions) the criminal activity clearly reduces as police per capita increases.[27]

6 Concluding Remarks

This paper provides identification results for treatment response models with endogenous social interactions by means of monotone comparative statics. To this end, we bridge the theory of identification of treatment effects, which exploits monotone restrictions in individualistic models, 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, which 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 one must lie between

[26] We should state here that if the interest of the paper were this particular application, we could add other credible restrictions to the model to further increase the informational content of the bounds.

[27] Since the parameter of interest (i.e., fraction of people that commits a crime) is not point identified, the relevant comparison to measure the treatment effect entails the comparison of two sets. The claim "strong treatment effect" can be supported by using the strong set order to contrast the intervals.

them. The bounds we provide are sharp and, by applying our results to the study of crimes in NY State, we show that they can also be informative.

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

7 Appendix A: Proofs

To make this paper self-contained we provide three theorems that are invoked in the subsequent proofs.

Tarski's Fixed Point Theorem (TFP) *If X is a complete lattice and $f : X \to X$ 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 $f : X \times S \to X$ 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 (FOSD) is based on upper sets. Let us consider $(\Omega, \geq)$, where Ω 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 $x \in U$.

FOSD's Theorem (FOSD) *Let $X, X' \in \mathbb{R}^n$ be two random vectors. The statements*

(i) $P(X \in U) \geq P(X' \in U)$ *for all upper set $U \subset \mathbb{R}^n$; and*

(ii) $E[f(X)] \geq E[f(X')]$ *for all increasing function $f(.)$ (such that the expectations exist)*

are equivalent. 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 3: Consider the mapping

$$
\begin{array}{lll}
M_{t,k}: & Y^{|L|} & \to Y^{|L|} \\
 & \left(y_1, y_2, ..., y_{|L|}\right) & \to \left(y'_{k1}, y'_{k2}, ..., y'_{k|L|}\right)
\end{array}
\tag{29}
$$

where $y'_{kl} = f_{kl}\left(t, \mathbf{y}_{-l}\right)$ for $l = 1, 2, ..., |L|$, and $\mathbf{y}_{-l} \equiv \left(y_1, ..., y_{l-1}, y_{l+1}, ..., y_{|L|}\right)$. It is immediate to notice the range of $M_{t,k}$ is as given. By construction, the solution set to the system of equations $\mathbf{f}_k$, $\phi\left(t, k\right)$, coincides with the set of fixed points of $M_{t,k}$. Then the proof of Lemma 3 reduces to show 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 3 follows by TFP. □

Proof of Proposition 4: We first show

$$
P\left[\mathbf{y}\left(t\right)\right] \geq_{st} P\left(\mathbf{y}\left|\tau = t\right.\right) P\left(\tau = t\right) + P\left(\underline{Y}^{|L|}\right) P\left(\tau \neq t\right). \tag{30}
$$

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

$$
\begin{aligned}
P\left[\mathbf{y}\left(t\right) \in U\right] &= P\left[\mathbf{y}\left(t\right) \in U|\tau = t\right] P\left(\tau = t\right) + P\left[\mathbf{y}\left(t\right) \in U|\tau \neq t\right] P\left(\tau \neq t\right) \\
&\geq P\left(\mathbf{y} \in U|\tau = t\right) P\left(\tau = t\right) + P\left(\underline{Y}^{|L|} \in U\right) P\left(\tau \neq t\right).
\end{aligned} \tag{31}
$$

The empirical evidence reveals $P\left[\mathbf{y}\left(t\right) \in U|\tau = t\right] = P\left(\mathbf{y} \in U|\tau = t\right)$. 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 the identified area is sharp, first notice that we restrict $P\left[\mathbf{y}\left(t\right)\right]$ to the set of all possible distributions that are consistent with the nature of the outcomes, i.e., $P\left[\mathbf{y}\left(t\right)\right] \in \triangle_{Y^{|L|}}$. In addition, given the data, our initial assumptions are consistent with both $P\left[\mathbf{y}\left(t\right)|\tau \neq t\right] = P\left(\underline{Y}^{|L|}\right)$ and $P\left[\mathbf{y}\left(t\right)|\tau \neq t\right] = P\left(\overline{Y}^{|L|}\right)$. Then $P\left[\mathbf{y}\left(t\right)\right]$ can coincide with any element of $\triangle_{Y^{|L|}}$ that lies between the lower and the upper bound, and our claim follows. □

Proof of Lemma 5: Lemma 3 shows that if $A1$ holds then $\phi\left(t, k\right)$ has a least and a greatest element, and $A2$ ensures the system of equations $\mathbf{f}_k$ increases in t for any fixed $k \in K$. Then $M_{t,k}$, as defined in (29), increases in t on T, and the fact that the extremal elements of $\phi\left(t, k\right)$ increase in t follows by MR. Thus, condition $A3$(i) is sufficient to our claim.

Assume condition $A3$(ii) holds, and consider $\mathcal{A}(\mathbf{y}^0, t, k)$. Since T is a chain and $\mathbf{y}^0$ is an equilibrium for some initial t_0, then either $\mathbf{y}^1 = \mathbf{f}_k(t, \mathbf{y}^0) \geq \mathbf{y}^0$ or $\mathbf{y}^1 = \mathbf{f}_k(t, \mathbf{y}^0) \leq \mathbf{y}^0$ holds. Suppose that $\mathbf{y}^1 = \mathbf{f}_k(t, \mathbf{y}^0) \geq \mathbf{y}^0$ (the proof of the other case is similar, thus we omit it). It follows, by $A1$, that the sequence $\{\mathbf{y}^i\}_{i=0}^{\infty}$ is increasing. Since Y is countable, then $f_{kl}(t, \mathbf{y})$ is continuous in $\mathbf{y}$ on $Y^{|L|}$ $\forall l \in L$. Being increasing and bounded, $\{\mathbf{y}^i\}_{i=0}^{\infty}$ has a limit vector $\widehat{\mathbf{y}} = \sup\{\mathbf{y}^i\}$. By the continuity of $\mathbf{f}_k$, $\widehat{\mathbf{y}}$ is a fixed point of $M_{t,k}$. By condition $A3$(ii), $\widehat{\mathbf{y}} = \mathbf{y}_k(t)$. By $A2$, if $t \geq t'$, then

$$\mathbf{f}_k\left(t, \mathbf{y}^0\right) \geq \mathbf{f}_k\left(t', \mathbf{y}^0\right).$$

It follows, by $A1$, that the limit of the sequence $\{\mathbf{y}^i\}_{i=0}^{\infty}$ is higher with t than with t', i.e., $\mathbf{y}_k(t) \geq \mathbf{y}_k(t')$. (Part of this proof is based on Vives (1999), Theorem 2.10(ii).)

The fact that $A3$(iii) is sufficient to our claim holds by Echenique and Sabarwal (2003). □

The proof of Proposition 6 requires an intermediate result that relates to Lemma 5.

Corollary 20 *Assume $A1$, $A2$ and $A3$ hold. Then, we have $P[\mathbf{y}(t) | \tau \leq t] \geq_{st} P(\mathbf{y} | \tau \leq t)$ and $P(\mathbf{y} | \tau \geq t) \geq_{st} P[\mathbf{y}(t) | \tau \geq t]$.*

Proof of Corollary 20: Let $U \subset \mathbb{R}^{|L|}$ be an upper set, and consider the next steps

$$\begin{aligned} P[\mathbf{y}(t) \in U | \tau \leq t] &= \textstyle\sum_{s \in T} \left\{\sum_{k \in K} 1[\mathbf{y}_k(t) \in U]\, \pi_{k|\tau=s}\right\} 1(s \leq t)\, P(\tau = s \mid \tau \leq t) \\ &\geq \textstyle\sum_{s \in T} \left\{\sum_{k \in K} 1[\mathbf{y}_k(s) \in U]\, \pi_{k|\tau=s}\right\} 1(s \leq t)\, P(\tau = s \mid \tau \leq t) \\ &= P(\mathbf{y} \in U | \tau \leq t). \end{aligned} \tag{32}$$

Under $A1$, $A2$ and $A3$, the inequality follows by Lemma 5, 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. □

Proof of Proposition 6: We next show

$$P[\mathbf{y}(t)] \geq_{st} P(\mathbf{y} | \tau \leq t)\, P(\tau \leq t) + P\left(\underline{Y}^{|L|}\right) P(\tau \nleq t). \tag{33}$$

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

$$\begin{aligned} P[\mathbf{y}(t) \in U] &= P[\mathbf{y}(t) \in U|\tau \leq t] P(\tau \leq t) + P[\mathbf{y}(t) \in U|\tau \nleq t] P(\tau \nleq t) \\ &\geq P(\mathbf{y} \in U|\tau \leq t) P(\tau \leq t) + P[\mathbf{y}(t) \in U|\tau \nleq t] P(\tau \nleq t) \\ &\geq P(\mathbf{y} \in U \mid \tau \leq t) P(\tau \leq t) + P\left(\underline{Y}^{|L|} \in U\right) P(\tau \nleq t). \end{aligned} \tag{34}$$

Under $A1$, $A2$ and $A3$, the first inequality follows by Corollary 20 and FOSD(i). The second one 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 the one invoked in Proposition 4. □

The proof of Lemma 9 requires an additional result.

Lemma 21 *Assume $A1$ and $A5$ hold. Let $k, k' \in K$. If $\mathbf{f}_k \geq \mathbf{f}_{k'}$ according to Def. 8, then $\mathbf{y}_k(t) \geq \mathbf{y}_{k'}(t)$ for all $t \in T$.*

Proof of Lemma 21: Lemma 3 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 (29). 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 $A5$(i).

The fact that $A5$(ii) is sufficient to our claim holds by Echenique and Sabarwal (2003). □

Proof of Lemma 9: Let $s, s' \in T$, with $s \geq s'$, and let us assume that $A1$, $A4$, and $A5$ hold. Fix an upper set $U \subset \mathbb{R}^{|L|}$. By definition,

$$P[\mathbf{y}(t) \in U|\tau = s] = \textstyle\sum_{k \in K} 1[\mathbf{y}_k(t) \in U] \pi_{k|\tau=s}. \tag{35}$$

Recall $P(\mathbf{f} = \mathbf{f}_k \mid \tau) \equiv \pi_{k|\tau}$. By Lemma 21, $1[\mathbf{y}_k(t) \in U] \geq 1[\mathbf{y}_{k'}(t) \in U]$ if $\mathbf{f}_k \geq \mathbf{f}_{k'}$. Then $P[\mathbf{y}(t) \in U|\tau = s]$ is the expectation of an increasing function of $\mathbf{f}$ conditional on $\tau = s$. By $A4$, $P(\mathbf{f} \mid \tau = s) \geq_{st} P(\mathbf{f} \mid \tau = s')$. Then, by FOSD(ii), $P[\mathbf{y}(t) \in U|\tau = s] \geq P[\mathbf{y}(t) \in U|\tau = s']$. The result follows by FOSD(i) as U was arbitrarily selected. □

The proof of Proposition 10 requires an intermediate result that relates to Lemma 9.

Corollary 22 *Assume A1, A4 and A5 hold. Then, we have* $P[\mathbf{y}(t)|\tau \geq t] \geq_{st} P(\mathbf{y}|\tau = t)$ *and* $P(\mathbf{y}|\tau = t) \geq_{st} P[\mathbf{y}(t)|\tau \leq t]$ *for all* $t \in T$.

Proof of Corollary 22: Let $U \subset \mathbb{R}^{|L|}$ be an upper set, and consider the next steps

$$\begin{aligned} P[\mathbf{y}(t) \in U|\tau \geq t] &= \textstyle\sum_{s \in T} P[\mathbf{y}(t) \in U|\tau = s] 1(s \geq t) P(\tau = s|\tau \geq t) \\ &\geq \textstyle\sum_{s \in T} P[\mathbf{y}(t) \in U|\tau = t] 1(s \geq t) P(\tau = s|\tau \geq t) \\ &= P[\mathbf{y}(t) \in U|\tau = t] \textstyle\sum_{s \in T} 1(s \geq t) P(\tau = s|\tau \geq t) \\ &= P(\mathbf{y} \in U|\tau = t). \end{aligned} \tag{36}$$

Under $A1$, $A4$ and $A5$, the inequality holds by Lemma 9. The third line holds as $P[\mathbf{y}(t) \in U|\tau = t]$ is independent of s, and the last one 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. □

Proof of Proposition 10: We next show

$$P[\mathbf{y}(t)] \geq_{st} P(\mathbf{y}|\tau = t) P(\tau \geq t) + P\left(\underline{Y}^{|L|}\right) P(\tau \ngeq t). \tag{37}$$

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

$$\begin{aligned} P[\mathbf{y}(t) \in U] &= P[\mathbf{y}(t) \in U|\tau \geq t] P(\tau \geq t) + P[\mathbf{y}(t) \in U|\tau \ngeq t] P(\tau \ngeq t) \\ &\geq P(\mathbf{y} \in U|\tau = t) P(\tau \geq t) + P[\mathbf{y}(t) \in U|\tau \ngeq t] P(\tau \ngeq t) \\ &\geq P(\mathbf{y} \in U|\tau = t) P(\tau \geq t) + P\left(\underline{Y}^{|L|} \in U\right) P(\tau \ngeq t). \end{aligned} \tag{38}$$

Under $A1$, $A4$ and $A5$, the first inequality follows by Corollary 22, and the last one 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 the one invoked in Proposition 4. □

Proof of Lemma 13: Consider the following mapping

$$\begin{aligned} N_{t,k}: \quad & \triangle_Y && \rightarrow \triangle_Y \\ & P(y) && \rightarrow P(y') \end{aligned} \tag{39}$$

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 13 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], $(\triangle_Y, \geq_{st})$ is a complete lattice. By assumption $A1'$,

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

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 Proposition 14: We first show

$$P[y(t)] \geq_{st} P(y|\tau = t) P(\tau = t) + P(\underline{Y}) P(\tau \neq t). \tag{40}$$

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

$$\begin{aligned} P[y(t) \in U] &= P[y(t) \in U|\tau = t] P(\tau = t) + P[y(t) \in U|\tau \neq t] P(\tau \neq t) \\ &\geq P(y \in U|\tau = t) P(\tau = t) + P(\underline{Y} \in U) P(\tau \neq t). \end{aligned} \tag{41}$$

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

Sharpness follows by the same argument as the one invoked in Proposition 4. □

Proof of Lemma 15: Lemma 13 shows that if $A1'$ holds then $\varphi(t,k)$ has a least and a greatest element, and $A2'$ ensures the system of equations $\mathbf{f}_k[t, P(y)]$ increases in t for any fixed $k \in K$. Then, for all $t > t'$,

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

for any upper level set U. It follows, by FOSD(i), that $N_{t,k}$, as defined in (39), 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 $A3'$(i) is then sufficient to our claim.

Assume condition $A3'$(ii) holds, and consider $\mathcal{A}(P^0(y), t, k)$. Since T is a chain then either

$$\begin{aligned} P^1(y \in U) &= (1/|G_k|) \sum_{l \in G_k} 1\left(f_{kl}\left[t, P^0(y)\right] \in U\right) \geq P^0(y \in U) \text{ or} \\ P^1(y \in U) &= (1/|G_k|) \sum_{l \in G_k} 1\left(f_{kl}\left[t, P^0(y)\right] \in U\right) \leq P^0(y \in U) \end{aligned}$$

for any upper level set U. Then, by FOSD(i), either $P^1(y) \geq_{st} P^0(y)$ or $P^1(y) \leq_{st} P^0(y)$. Assume the first case holds. (The other case can be shown in a similar way, thus we omit it.) It follows, by $A1'$, that the sequence $\{P^i\}_{i=0}^{\infty}$ increases with respect to FOSD. Being increasing and bounded, $\{P^i\}_{i=0}^{\infty}$ has a limit distribution $\widehat{P} = \sup\{P^i\}$. By the discrete nature of the potential outcomes, $\widehat{P}$ is a fixed point of $N_{t,k}$. By condition $A3'$(ii), $\widehat{P} = P[y_k(t)]$. By $A2$, if $t \geq t'$, then

$$(1/|G_k|) \sum_{l \in G_k} 1\left(f_{kl}\left[t, P^0(y)\right] \in U\right) \geq (1/|G_k|) \sum_{l \in G_k} 1\left(f_{kl}\left[t', P^0(y)\right] \in U\right).$$

It follows, by $A1$, that the sequence $\{P^i(y)\}_{i=0}^{\infty}$ is higher with t than with t'. Since monotonicity is preserved by taking limits, $P[y_k(t)] \geq_{st} P[y_k(t')]$.

The fact that $A3'$(iii) is sufficient to our claim holds by Echenique and Sabarwal (2003). □

The proof of Proposition 16 requires an intermediate result that relates to Lemma 15.

Corollary 23 *Assume $A1'$, $A2'$, and $A3'$ 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 of Corollary 23: Let $U \subset \mathbb{R}$ be an upper set, and consider the next steps

$$\begin{aligned} P[y(t) \in U|\tau \leq t] &= \textstyle\sum_{s \in T} \left\{\sum_{k \in K} P[y_k(t) \in U]\, \pi_{k|\tau=s}\right\} 1(s \leq t)\, P(\tau = s \mid \tau \leq t) \\ &\geq \textstyle\sum_{s \in T} \left\{\sum_{k \in K} P[y_k(s) \in U]\, \pi_{k|\tau=s}\right\} 1(s \leq t)\, P(\tau = s \mid \tau \leq t) \\ &= P(y \in U|\tau \leq t). \end{aligned} \tag{42}$$

Under $A1'$, $A2'$ and $A3'$, the inequality follows by Lemma 15, 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. □

Proof of Proposition 16: We next show

$$P[y(t)] \geq_{st} P(y|\tau \leq t) P(\tau \leq t) + P(\underline{Y}) P(\tau \nleq t). \tag{43}$$

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

$$\begin{aligned} P[y(t) \in U] &= P[y(t) \in U|\tau \leq t] P(\tau \leq t) + P[y(t) \in U|\tau \nleq t] P(\tau \nleq t) \\ &\geq P(y \in U|\tau \leq t) P(\tau \leq t) + P(\underline{Y} \in U) P(\tau \nleq t). \end{aligned}$$

Under $A1'$, $A2'$ and $A3'$, the inequality follows by Corollary 23 and FOSD(i), and because $\underline{Y}$ is a lower bound for any potential outcome. 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 the one invoked in Proposition 4. □

The proof of Lemma 18 requires an additional result.

Lemma 24 *Assume $A1'$ and $A5'$ hold. Let $k, k' \in K$. If $F_k \geq F_{k'}$ according to Def. 17, then $P[y_k(t)] \geq_{st} P[y_{k'}(t)]$ for all $t \in T$.*

Proof of Lemma 24: Lemma 13 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 (39). 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 $A5'$(i).

The fact that $A5'$(ii) is sufficient to our claim holds by Echenique and Sabarwal (2003). □

Proof of Lemma 18: Let $s, s' \in T$, with $s \geq s'$, and let us assume that $A1'$, $A4'$ and $A5'$ hold. Fix an upper set $U \subset \mathbb{R}$. By definition,

$$P[y(t) \in U \mid \tau = s] = \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=s}. \tag{44}$$

Recall $P(F = F_k \mid \tau) \equiv \pi_{k|\tau}$. By Lemma 24, $P[y_k(t) \in U] \geq P[y_{k'}(t) \in U]$ if $F_k \geq F_{k'}$. Then $P[y(t) \in U \mid \tau = s]$ is the expectation of an increasing function of F conditional on $\tau = s$. By $A4'$, we know that $P(F \mid \tau = s) \geq_{st} P(F \mid \tau = s')$. Then, by FOSD(ii), we get $P[y(t) \in U \mid \tau = s] \geq P[y(t) \in U \mid \tau = s']$. The result follows by FOSD(i) as U was arbitrarily selected. □

The proof of Proposition 19 requires an intermediate result that relates to Lemma 18.

Corollary 25 *Assume $A1'$, $A4'$ and $A5'$ 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 of Corollary 25: Let $U \subset \mathbb{R}$ be an upper set, and consider the next steps

$$\begin{aligned} P[y(t) | \tau \geq t] &= \sum_{s \in T} \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=s} \right\} 1(s \geq t) P(\tau = s \mid \tau \geq t) \\ &\geq \sum_{s \in T} \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=t} \right\} 1(s \geq t) P(\tau = s \mid \tau \geq t) \\ &= \left\{ \sum_{k \in K} P[y_k(t) \in U] \pi_{k|\tau=t} \right\} \sum_{s \in T} 1(s \geq t) P(\tau = s \mid \tau \geq t) \\ &= P(y \in U \mid \tau = t). \end{aligned} \tag{45}$$

Under $A1'$, $A4'$ and $A5'$, the inequality follows by Lemma 18. The third line is true as $P[y(t) \in U | \tau = t]$ is independent of s, and the last one 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. □

Proof of Proposition 19: We next show

$$P[y(t)] \geq_{st} P(y | \tau = t) P(\tau \geq t) + P(\underline{Y}) P(\tau \ngeq t). \tag{46}$$

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

$$\begin{aligned} P[y(t) \in U] &= P[y(t) \in U | \tau \geq t] P(\tau \geq t) + P[y(t) \in U | \tau \ngeq t] P(\tau \ngeq t) \\ &\geq P(y \in U | \tau = t) P(\tau \geq t) + P(\underline{Y} \in U) P(\tau \ngeq t). \end{aligned}$$

Under $A1'$, $A4'$ and $A5'$, the inequality follows by Corollary 25, and because $\underline{Y}$ is a lower bound for any potential outcome. Since U was arbitrarily selected, our first claim follows by FOSD(i). The proof for the upper bound is quite similar, so we omit it.

Sharpness follows by the same argument as the one invoked in Proposition 4. □

8 Appendix B

8.1 Estimators for the Extremal Bounds of $P[y(t) = 1]$

The parameter of interest is the fraction of people that would commit a crime in NY State at a level of police per capita t, i.e., $P[y(t) = 1]$. We observe criminal activity and realized levels of police per capita for a random sample of cities in NY 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 $\triangle(t)$ the distance between these bounds, i.e., $\triangle(t) \equiv P_u[y(t) = 1] - P_l[y(t) = 1]$. We distinguish estimators by using hats.

In what follows: τ_m is realized level of police per capita in city m; d_m is number of crimes in city m divided by number of people in city m; n_m is number of people in city m divided by 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 14, 16, and 19 to the analysis of crimes.

Proposition 26 *Assume $A1'$ holds. Then, for all $t \in T$,*

$$H\left\{P\left[y\left(t\right)=1\right]\right\} = \left\{\delta \in \triangle_Y : \begin{array}{c} P\left(y=1 \mid \tau=t\right) P\left(\tau=t\right) + P\left(\tau \neq t\right) \\ \geq_{st} \delta \geq_{st} \\ P\left(y=1 \mid \tau=t\right) P\left(\tau=t\right) \end{array}\right\}. \tag{47}$$

These bounds are sharp.

The estimators for the bounds are just the sample analogs of (47). That is, $\forall t \in T$,

$$\begin{aligned} \widehat{P}_u[y(t) &= 1] = \sum\nolimits_{m \in M} \left[d_m 1\left(\tau_m = t\right) + 1\left(\tau_m \neq t\right)\right] n_m \\ \widehat{P}_l[y(t) &= 1] = \sum\nolimits_{m \in M} d_m 1\left(\tau_m = t\right) n_m. \end{aligned}$$

(II) Positive Interactions and (Negative) Monotone Treatment Response

Proposition 27 *Assume A1′ and the duals of (A2′ and A3′) hold. Then, for all $t \in T$,*

$$H\left\{P\left[y\left(t\right)=1\right]\right\} = \left\{ \delta \in \triangle_Y : \begin{array}{c} P\left(y=1 \mid \tau \leq t\right) P\left(\tau \leq t\right) + P\left(\tau \nleq t\right) \\ \geq_{st} \delta \geq_{st} \\ P\left(y=1 \mid \tau \geq t\right) P\left(\tau \geq t\right) \end{array} \right\}. \tag{48}$$

These bounds are sharp.

The estimators for the bounds are just the sample analogs of (48). That is, $\forall t \in T$,

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

(III) Positive Endogenous Interactions and Monotone Treatment Selection

Proposition 28 *Assume A1′, A4′, and A5′ hold. Then, for all $t \in T$,*

$$H\left\{P\left[y\left(t\right)=1\right]\right\} = \left\{ \delta \in \triangle_Y : \begin{array}{c} P\left(y=1 \mid \tau = t\right) P\left(\tau \leq t\right) + P\left(\tau \nleq t\right) \\ \geq_{st} \delta \geq_{st} \\ P\left(y=1 \mid \tau = t\right) P\left(\tau \geq t\right) \end{array} \right\}. \tag{49}$$

These bounds are sharp.

The estimators for the bounds are just the sample analogs of (49). That is, $\forall t \in T$,

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

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

8.2 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 below 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 $\left(\widehat{\sigma}_l^2, \widehat{\sigma}_u^2, \widehat{\rho}\right)$ that converge to their population values uniformly in $P \in \Psi$.
(ii) For all $P \in \Psi$, $\underline{\sigma}^2 \leq \sigma_l^2 \leq \overline{\sigma}^2$ and $\underline{\sigma}^2 \leq \sigma_u^2 \leq \overline{\sigma}^2$ for some positive and finite $\underline{\sigma}^2$ and $\overline{\sigma}^2$, and $P_u[y(t) = 1] - P_l[y(t) = 1] \leq \overline{\triangle} < \infty$.
(iii) For all $\varepsilon > 0$, there are $\nu > 0$, K, and N_0 such that $N \geq N_0$ implies $\Pr\left(\sqrt{N}\left|\overline{\triangle} - \triangle\right| > K\triangle^{\nu}\right) < \varepsilon$ uniformly in $P \in \Psi$.

Stoye (2009) explains that Assumption 1(iii) requires $\widehat{\triangle}(t)$ to be super-efficient 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 $a_N \longrightarrow 0$, $a_N\sqrt{N} \to \infty$, and $\sqrt{N}\left|\overline{\triangle} - \triangle_N\right| \xrightarrow{p} 0$ for all sequences of distributions $P_N \subseteq \Psi$ with $\triangle_N \leq a_N$.

Though condition (iii′) is often hard to validate, Stoye (2009) provides a sufficient condition that clearly holds in our set-up.

Lemma 29 *Let Assumption 1(i) and (ii) hold, and assume*

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

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_α solves

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

Stoye (2009) shows uniform validity of CI_α under Assumption 1(i), (ii) and (iii′).[28]

References

[1] Ackerberg, D., and G. Gowrisankaran (2006): "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," *RAND J of Economics*, 37, 738-761.

[2] Amir, R. (2008): "Comparative Statics of Nash Equilibria," mimeo.

[3] Andrews, D., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119-157.

[4] Becker, G. (1968): "Crime and Punishment: An Economic Approach," *The Journal of Political Economy*, 76, 169-217.

[5] Becker, G. (1991): "A Note on Restaurant Pricing and other Examples of Social Influences on Price," *The Journal of Political Economy*, 99, 1109-1116.

[6] Blume, L., W. Brock, S. Durlauf, and Y. Ionnides (2010): "Identification of Social Interactions," Hanbook of Social Economics, J. Benhabib, A. Bisin, and M. Jackson, eds.

[7] Brock, W., and S. Durlauf (2001a): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235-260.

[28] 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.

[8] Brock, W., and S. Durlauf (2001b): "Interactions-Based Models," In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 5. North-Holland, 3297-3380.

[9] Brock, W., and S. Durlauf (2007): "Identification of Binary Choice Models with Social Interactions," *Journal of Econometrics*, 140, 52-75.

[10] De Paula, Á. (2009): "Inference in a Synchronization Game with Social Interactions," *Journal of Econometrics*, 148, 56-71.

[11] Echenique, F. (2002): "Comparative Statics by Adaptive Dynamics and the Correspondence Principle," *Econometrica*, 70, 833-844.

[12] Echenique, F. (2003): "Mixed Equilibria in Games of Strategic Complementarities," *Economic Theory*, 22, 33-44.

[13] Echenique, F., and T. Sabarwal (2003): "Strong Comparative Statics of Equilibria," *Games and Economic Behavior*, 42, 307–314.

[14] Glaeser, E., B. Sacerdote, and J. Scheinkman (1996): "Crime and Social Interactions," *The Quarterly Journal of Economics*, 111, 507-548.

[15] Gowrisankaran, G., and J. Stavins (2004): "Network Externalities and Technology Adoption: Lessons from Electronic Payments," *RAND Journal of Economics*, 35, 260-276.

[16] Graham, B. (2008): "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, 76, 643-660.

[17] Hahn, J., and K. Hirano (2009): "Design of Randomized Experiments to Measure Social Interaction Effects," *Economics Letters*, 106, 51-53.

[18] Imbens, G., and C. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1845-1857.

[19] Jia, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76, 1263-1316.

[20] Lee, S., O. Linton, and Y.-J. Whang (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77, 585–602.

[21] Manski, C. (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, 319-323.

[22] Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531-542.

[23] Manski, C. (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

[24] Manski, C. (2007): "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393-1410.

[25] Manski, C. (2011): "Identification of Treatment Response with Social Interactions," *The Econometrics Journal*, Forthcoming.

[26] Manski, C., and J. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997-1010.

[27] Matzkin, R. (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76, 945–978.

[28] Milgrom, P., and J. Roberts (1990): "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255-1277.

[29] Moffitt, R., (2001): "Policy Interventions, Low-Level Equilibria, and Social Interactions," In Social Dynamics, eds. S. Durlauf and P. Young, MIT Press.

[30] Molinari, F., and A. Rosen (2008): "The Identification Power of Equilibrium in Games: The Supermodular Case," *Journal of Business & Economic Statistics*, 26, 297-302.

[31] Mosler, K., and M. Scarsini (1991): "Some Theory of Stochastic Dominance," Stochastic Orders and Decisions under Risk. IML Lecture Notes. Monograph Series.

[32] Müller, A., and D. Stoyan (2002): "Comparison Methods for Stochastic Models and Risks," John Wiley & Sons, Inc.

[33] Rosen, A. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107-117.

[34] Sah, R. (1991): "Social Osmosis and Patterns of Crime," *Journal of Political Economy*, 99, 1972-1995.

[35] Shaikh, A., and E. Vytlacil (2011): "Partial Identification in Triangular Systems of Equations Binary Dependent Variables," *Econometrica*, 79, 949–955.

[36] Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299-1315.

[37] Stoye, J. (2010): "Partial identification of spread parameters," *Quantitative Economics*, 1, 3, 323–357.

[38] Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147-165.

[39] Tarski, A. (1955): "A Lattice Theoretical Fixpoint Theorem and its Applications," *Pacific Journal of Mathematics*, 5, 285-309.

[40] Topkis, D. (1979): "Equilibrium Points in Nonzero-sum n-person Submodular Games," *Journal of Control and Optimization*, 17, 773-787.

[41] Vives, X. (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305-321.

[42] Vives, X. (1999): "Oligopoly Pricing: Old Ideas and New Tools," MIT Press.