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

Natalia Lazzati †

September 05, 2010

Abstract

This paper studies (nonparametric) partial identification of treatment response with social interactions. It imposes minimal monotone structure to the model, and derives sharp restrictions on the distribution of potential outcomes via monotone comparative statics. Under three sets of conditions, we identify sharp distributional bounds on the potential outcomes given observable data. The endpoints of the identified areas are in terms of stochastic dominance. Most of the behavioral assumptions we use are economically driven, as reflected in the various examples we provide. We apply our results to the analysis of crimes in the state of New York, showing the monotone conditions we enforce provide quite valuable information.

JEL codes: C31, D71

Key words and phrases: Treatment effects, social interactions, nonparametric bounds, monotonicity, first order stochastic dominance.

*The author thanks Rabah Amir and Keisuke Hirano for all their valuable suggestions. She also thanks the other members of her dissertation committee Stanley Reynolds, Mark Walker and John Wooders for all their help. Comments by Amilcar Menichini are also acknowledged.

†Natalia Lazzati, University of Arizona, *e-mail:* nlazzati@email.arizona.edu.

1 Introduction

This paper studies partial identification of treatment response in models that exhibit endogenous social interdependences. During the last three decades, social interactions have become an essential component of the economic analysis. In this study, when we say social interactions we mean interdependences between individual decisions (or achievements) that are not fully mediated by ordinary markets. Activities subject to strong social pressure include crimes, schooling, infection diseases, addictions and fertility decisions. Newer models of industrial organization, such as network goods and two-sided markets, display this feature as well.[1] 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. Manski (2010) is one of the few exceptions, and a direct predecessor of this article.

Models of endogenous social interactions often start with the outline of a system of structural equations that describe the outcome of each individual as a function of a vector of observable covariates and the outcomes of the other members of the same group. The simultaneous solution to the system of equations (when it exists) is the predicted outcome or behavior of the group members.[2] Most of the research in econometrics is concerned with the identification of the structural equations. In the treatment response literature the latter is often obtained via exclusion restrictions, e.g. assuming the outcome functions are statistically independent of realized treatments. Following Manski (2010), the purpose here is quite different. Our objective is partial identification of the outcome distributions under alternative treatments.[3] The underlying approach is to impose minimal monotone structure to the model to derive clear restrictions on the predicted distributions in terms of stochastic dominance by means of monotone comparative statics. We next provide a detailed summary of our results, coupled with the related literature.

Previous studies have required specific functional forms on the system of structural equations. Brock and Durlauf (2001, 2007), among others, propose a specific model and address point and

[1] Manski (2000) describes many applications of social interactions in economics and reviews part of the empirical literature.

[2] We will elaborate on the possibility of multiple solutions (or equilibria) later.

[3] Manski (2003) offers a nice analysis of partial identification of probability distributions. Tamer (2009) provides an updated review of partial identification in economics.

partial (or set) identification of a structural parameter that captures the strength of the social interactions.[4] These methods work when the proposed parametric forms approximate well the true structural functions [see Varian (1982, 1984) for an early exposition of this issue]. The functional forms used in practice are (in general) not theoretically driven. We use an alternative nonparametric method that requires no ad hoc specifications of functional forms. Founded on Manski (1990, 1997, 2010) our approach combines economically meaningful restrictions on the primitives (i.e., the structural equations) with data to derive bounds for the counterfactual distributions, that is, the outcome distributions that would have been experienced by group members who did not receive specific treatments, had these treatments been applied to the groups.

The first two conditions we impose are as follows: the outcome of each individual increases with the outcomes of the others; and it varies monotonically with the treatment to be received by the group. 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. Molinari and Rosen (2008) take advantage of similar properties to show how the analysis of Aradillas-Lopez and Tamer (2008) on the identification power of equilibrium in games can be extended to supermodular games. We make use of these constraints for a different purpose.

Using well-known results in economic theory, we show the last two assumptions 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 makes certain the system has a minimal and a maximal solution, i.e., the model is always consistent.[5] 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)]. This paper provides a precise econometric framework that translates

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

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

the monotone comparative statics result into testable implications on the distribution of potential outcomes, i.e., we identify sharp distributional bounds on the potential outcomes given observable data.

Manski (1997) introduces a monotone shape restriction to study identification in settings where a person's outcome varies only with his own treatment, i.e., SUTVA.[6] Manski (2010) generalizes that analysis to settings with exogenous social interactions, that is, models where an individual's outcome varies not only with his own treatment but also with the treatments of the other members of his reference group. The result described in the previous paragraph extends Manski (1997, 2010) to settings with endogenous interdependences. This last feature calls for a deeply different methodology to tackle questions that are similar a priori, i.e., a game theoretic approach.

The last two conditions restrict the response function of each group member; they do not involve any cross-group restriction on response. The third condition we impose allows us to use the empirical evidence of some groups to learn about the potential outcome of alternative ones. Standard econometric analysis of simultaneity combines shape conditions on the treatment response with exclusion restrictions. For example, many studies have established identification results by assuming the outcome functions are statistically independent of realized treatments. In certain contexts, this assumption is hard to justify. Manski and Pepper (2000, 2008) weaken that restriction by proposing a monotone constraint that fits better many applications of interest. They refer to this assumption as monotone treatment selection (MTS).[7] We extend their findings to nonparametric interactions-based models. From a theoretical standpoint our propositions require the comparison of equilibrium outcomes for different groups, and are based on the methodology proposed by Amir (2008) to contrast Nash equilibria of different games.

All previous results are developed in a tractable econometric framework, which is an additional contribution of the present study. We typify groups through the relevant features of the group members, and relate the distribution of types of groups to the underlying process of group formation. An extra value of this set-up is that it allows the researcher to specify individual and social aspects

[6]In the treatment response literature SUTVA stands for Stable Unit Treatment Value Assumption.

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

of behavior simultaneously, observing the natural hierarchy of individuals and groups.[8] Following Manski (2010), we distinguish two alternative social interaction models. The first setting handles small groups, where each member has a distinctive role, e.g. men and women in married couples. In this case the interest is on the joint outcome distribution that would occur if all the groups were to receive a specific treatment. The notion of treatment we use is quite general, as each possible treatment may indicate, for instance, a different policy for members that have distinct roles. The second setting is appropriate to study large neighborhoods with anonymous interactions, that is, models where each individual assigns an identical weight to the outcomes of every other member of the same group. In the latter case, the attention is on individual treatment response generated by members of the groups.

The rest of the paper is organized as follows. Section 2 presents an initial example that motivates the rest of the paper. Sections 3 and 4 study identification of social interaction models for small and large groups, respectively. Section 5 describes a simple empirical application of our results to crime rates and social interactions. Section 6 discusses a relevant extension of the methodology, and Section 7 concludes. We collect all the proofs in Section 8.

2 An Initial Example

This section studies a model similar to Becker (1991) and Becker and Murphy (2000). The example has a dual purpose: it highlights the distinctive features of the interactions-based models that justify our approach; and it helps us to contrast the classical (econometric) identification objective of previous analysis 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 (d_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

[8] Manski (2007) shows how to predict counterfactual discrete choice behavior when the behavioral model partially identifies the choice probabilities; to this end he models discrete choice as treatment response. To typify groups in the population we use an approach similar to the one developed in that paper.

his utility considering as given the behavior of the others. His demand for the good takes the form

$$d_j = f_j \left[p, (1/|G|) \sum_{i \in G} d_i \right], \; j \in G \tag{1}$$

where p is a potential price for the service and $(1/|G|) \sum_{i \in G} d_i$ 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 d_j hardly affect the average of the choices.[9] An equilibrium in this market is a solution to the system of equations (1)

$$\mathbf{d}(p) \equiv [d_j(p), \; j \in G]. \tag{2}$$

In each neighborhood G that has a branch of the popular restaurant, transactions take place at some realized price z. The empirical evidence consists of vectors of prices and individual demands $[(z^m, \mathbf{d}^m), m \in M]$, where m indicates a particular neighborhood and M is the set of all neighborhoods in the sample. The analyst objective is to combine the empirical evidence with prior information to learn about the distribution of individual demands that would occur (in the study population) if the groups were to receive a specific treatment p, i.e., $P[d(p)]$.

The econometric literature refers to system (1) as the structural equations. As Manski (2010) notices, system (1) is tautological in the absence of shape restrictions or distributional assumptions. Without further constraints, it might have multiple solutions or no solution. Tamer (2003) has named cases of multiple solutions as incomplete models. Models with no solution are said to be inconsistent or incoherent.

Previous analysis have posed specific functional forms on the system of structural equations

$$d_j = 1 \left[\alpha_1 + (\alpha_2/|G|) \sum_{i \in G} d_i + \varepsilon_j \geq p \right], \; j \in G \tag{3}$$

where $1(.)$ is the standard indicator function, and ε_j is a random term that differentiates individuals. Brock and Durlauf (2001, 2007), among others, address point and partial (or set) identification of the structural parameter α_2 that captures the strength of the social interactions.[10] Awareness of

[9] This assumption is sustained in Section 4 below. It simplifies the theoretical analysis without changing the nature of the results in any fundamental way.

[10] This example deviates from Brock and Durlauf (2001, 2007) in many respects. It is, neverthelesss, the simplest possible representation to clarify how our approach differs from the classical one.

the parameters and the distribution of latent variables $(\varepsilon_j,\ j \in G)$ imply knowledge of the demand functions, and this information permits counterfactual predictions of individual demands for different prices. If the parameters are just partially identified, or the equilibrium is not unique, then the predictions may comprise a set of possible outcome distributions. These procedures only work when the maintained parametric forms approximate the true demand functions well. As we describe next, we use a nonparametric approach that combines economically meaningful assumptions on (1) with data to derive distributional bounds on the potential outcomes for treatments (i.e., prices in this example) that may differ from the realized ones.

In this application there are two assumptions on the system of structural equations that can be easily justified: individual j's demand increases with the decisions of the other members of his group (M0); and individual j's demand is a downward sloping function of the market price (M1). 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 study in this example.[11]

Using existing results in economic theory we show that the last two assumptions imply clear restrictions on the demands at equilibrium [see Tarski's Fixed Point Theorem and Milgrom and Roberts' Theorem, in Section 8 (Proofs)]. First, for each given neighborhood, the system of structural equations has a smallest and a largest solution. Second, the extremal equilibrium demands decrease in the potential treatment p.[12] The challenge is to provide a precise framework so that the comparative statics result translate into testable implications on the distribution of potential outcomes, or demands, $P\left[d\left(p\right)\right]$.

In this paper we assume that either the solution set is a singleton, or that neighborhoods always

[11] The analyst observes realized prices and quantities in different neighborhoods, and uses the two described assumptions to make inference on 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. [See Manski (1997) for an alternative explanation of why the well-known lack of identification problem of demand functions does not hold here.]

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

coordinate in one of the extremal equilibria. This constraint requires a tailored justification in each specific application. For instance, studying crime rates and social interactions, Glaeser, Sacerdote and Scheinkman (1996) assume their data is generated from up to seven different distributions, they argue that the observed differences across cities can not be simply explained through multiplicity and use a framework that relies on a unique solution. Studying network externalities in the ACH banking industry, Ackerberg and Gowrisankaran (2006) do not find evidence of multiple equilibria. In addition, the models discussed in this article often allow the Pareto ranking of the equilibrium outcomes. Several studies claim that when the largest solution Pareto dominates the other ones, it may be reasonable to believe that people will coordinate in that particular outcome [see, e.g., Gowrisankaran and Stavins (2004)]. The same logic applies to the lowest equilibrium when it is the Pareto preferred.

The application of our results to this example confirms that M0 and M1 have identifying power. Condition M0 guarantees the model is consistent, and M1 leads to a sharp identification region for $P[d(p)]$ in terms of stochastic dominance. That is, we identify two extreme distributions that are function of the observable data, so that any distribution in between cannot be rejected as the true one. The endpoints of the identified area usually differ, so $P[d(p)]$ is (almost always) just partially identified.

The two discussed conditions assume nothing about the process of treatment selection, i.e., price determination. However, in the present analysis it would be reasonable to think that the owner of the restaurant sets higher prices to those groups with stronger demands (M2). (The assumption sustained in this study is weaker than M2.) This paper shows that M0 and M2 also lead to a sharp identification region for $P[d(p)]$ in terms of stochastic dominance. This inference approach is quite different from the previous one (based on M0 and M1), as it relies on the use of realized treatments as monotone instrumental variables.

The next analysis formalizes and extends all previous ideas. Sections 3 and 4 establish regions of identification for the distribution of potential outcomes in frameworks where groups are small and large, respectively. The restaurant example is closer to the second setting.

3 Model of Social Interactions where Identities Matter

This section addresses identification of treatment response in situations where groups are small (S), and each group member has a distinctive role. Models of decisions of married couples, soccer teams and pairs of patients and doctors fit in this setting.

3.1 Econometric Framework and the Analyst's Problem

In this section 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 for an arbitrary l in L. Each group is composed of a set of $|L|$ individuals, one of each class. If we were modeling partner decisions, then classes may refer to gender and a group could be defined as either a dating or a married couple (see, for instance, Examples 1 and 2 below).

Let $t \in T$ indicate a potential treatment to be received by a group G in the universe of groups $\mathcal{U}$, and let $\mathbf{y}(t) \equiv [y_{j_l}(t), j_l \in G] \in Y^{|L|}$ denote the related vector of outcomes for the members of the group. The vector of potential outcomes $\mathbf{y}(t)$ solves the system of structural equations

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

where $y_{j_l}(t)$ is the outcome of individual j_l in G, and $\mathbf{y}_{-j_l}(t) \equiv [y_{j_m}(t), j_m \in G, m \neq l]$ denotes the vector of outcomes by the other members of the same group. In (4) the outcome of each individual depends on both the treatment received by the group, and the outcomes by the other members of the same neighborhood. This last effect is the source of endogeneity.

This model conditions social interactions to occur within (and not across) reference groups. In addition, we assume group membership is known to the econometrician. The model is quite general as regards to the treatment. For instance, it allows a treatment t to indicate a distinct policy for individuals that belong to different classes. The discussion that follows elaborates on the econometric aspects of our set-up.

The distributional assumptions incorporate two dimensions. We first typify groups through the outcome functions of their members, and then relate the distribution of types of groups to the underlying process of group formation. Before formalizing this approach, we describe an example that captures the main insights.

Example 1 *Let us consider a population of six people, three of which are boys and three are girls, i.e.,* $J \equiv (J_1, J_2)$ *where* $J_1 = (1_1, 2_1, 3_1)$ *is the set of boys and* $J_2 = (1_2, 2_2, 3_2)$ *denotes the set of girls. A group is defined as a dating couple.*

This application describes a scenario of a tobacco prevention program and smoking decisions. Let each member of a dating 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. Therefore $T = \{0, 1\}$, $Y = \{0, 1\}$ *and* $f_{j_l} : T \times Y \to Y$ *with* $j = 1, 2, 3$ *and* $l = 1, 2$.

In this paper we categorize individuals and couples in terms of their outcome functions, that is, in terms of the way they react to the treatment and partner's decisions. According to this criterion, we have here 2^4 *(or* 16*) possible kinds of people—for each of the four possible configurations of treatments and partner's decisions there are two possible decisions to make—and thereby* 16^2 *(or* 256*) types of dating couples.*

Assume the population of interest has outcome functions as the ones described in Table 1 below. 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 that conform the columns of the two matrices. As in (4), the first argument indicates the value of the treatment and the second one specifies the smoking decision of each person's partner. For instance, the one in square brackets means that the second boy would smoke if he does not receive information about the risks of smoking and his girlfriend smokes.*

Table 1: Population Outcome Functions

Each row indicates the outcomes of a specific individual and each column specifies a pair of treatment and partner's decision.

$$
\text{Boys} \qquad \begin{array}{c} \\ y_{1_1} \\ y_{2_1} \\ y_{3_1} \end{array} \left[\begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ 0 & 1 & 0 & 0 \\ 0 & [\,1\,] & 0 & 0 \\ 0 & 1 & 0 & 1 \end{array} \right] \qquad \text{Girls} \qquad \begin{array}{c} \\ y_{1_2} \\ y_{2_2} \\ y_{3_2} \end{array} \left[\begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{array} \right] .
$$

The population described in Table 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 fraction of couples that will actually belong to each type will depend not only on the smoking preferences of the population, but also on their dating inclinations. Let us assume, for instance, that these individuals prefer to date rather than remaining alone, and to share their time with other individuals with similar smoking tastes. In this case two out of the three couples will have members that will smoke if and only if they are uninformed and their partners smoke, and the other one will have members that will smoke if and only if their partners smoke irrespective of the treatment. The other two types of groups entail members with different outcome functions, and have no chance to be formed given the assumed dating preferences.*

Let us formalize the idea in Example 1. 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. For instance, in Example 1, the outcome function of the first and the second boy are equal to each other but differ from that of the third one. Here we assume the cardinality of T and Y is at most countable.[13] 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. When we say a group is of type k we mean its members have structural equations $\mathbf{f}_k \equiv [f_{kl}(.), l \in L]$, 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 J. 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 the population. 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

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

couples, we should then expect a high proportion of spouses in the population with very similar outcome 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 could be described by $\boldsymbol{\pi} = (2/3, 1/3, 0, ..., 0)$ where $\boldsymbol{\pi}$ has 256 elements.

In this study, group formation is exogenous to the model in the sense that potential treatments do not affect the distribution of types of groups, $\boldsymbol{\pi}$.[14] Nevertheless, realized treatments may provide rich information about the type of groups that could have generated the data. The informational content of realized treatments is nicely captured by the restaurant example in Section 2. In that case, it is reasonable to think that potential prices do not affect the composition of the neighborhoods. However, the owner of the restaurant may set larger prices in those areas that have higher willingness to pay. Then realized prices could help the analyst to infer some relevant features of the neighborhoods under study.

From the last remark it follows that $[(\mathbf{f}_k, \pi_k), k \in K]$ characterizes the universe of groups, $\mathcal{U}$, in the study population. Groups of the population have observable realized treatments z^m and outcomes $\mathbf{y}^m \equiv \left[y_1^m(z^m), y_2^m(z^m), ..., y_{|L|}^m(z^m)\right]$. Let $[(z^m, \mathbf{y}^m), m \in M]$ denote the empirical evidence. The outcomes that would have been experienced under other treatments are counterfactual. The analyst's problem is to combine the empirical evidence with prior information (i.e., some plausible assumptions) to learn about the joint outcome distribution that would occur if the groups were to receive a specific 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 an $|L|$-dimensional random vector.

Example 2 *A model as the one described can be used to study retirement decisions of husbands and wives. Here J_1 may indicate 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 either the income tax or the cost of Social Security benefits. Many studies postulate that endogenous interactions

[14] Manski (2010) describes a similar condition by saying that reference groups are treatment-invariant. He adds that under the latter condition groups of reference are necessarily non-manipulable, in the sense that the social planner cannot use the treatments in T to change a person's group of influence.

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 specific monotone shape restrictions that are naturally satisfied in many social-interactions models (e.g. in Example 2).

3.2 Identification Region for $P\left[\mathbf{y}\left(t\right)\right]$

Sub-sections 3.2.1, 3.2.2 and 3.2.3 impose three alternative conditions on the primitives (i.e., the structural equations) to derive distributional bounds for $P\left[\mathbf{y}\left(t\right)\right]$ in terms of stochastic dominance. Sub-section 3.2.4 briefly discusses an estimation strategy for the identified endpoints.

3.2.1 Consistency of the Model

This sub-section takes advantage of the main results in Manski (1990) and Tarski (1955). The framework in Sub-section 3.1 only conditions social interactions to occur within the reference groups. Without further restrictions, the system of structural equations (4) might have multiple solutions or no solution. The next assumption imposes a shape restriction on $(\mathbf{f}_k, k \in K)$. This single condition precludes the possibility of inconsistency.

M0.S. Let $Y \subseteq \mathbb{R}$ be compact.[15] Let $\mathbf{y}, \mathbf{y}' \in Y^{|L|-1}$. Then

$$\mathbf{y} \geq \mathbf{y}' \Longrightarrow f_{kl}\left(t, \mathbf{y}\right) \geq f_{kl}\left(t, \mathbf{y}'\right) \tag{5}$$

$\forall t \in T$ and $\forall l \in L$. This condition holds for all $k \in K$.

Condition M0.S requires the outcome of each individual to increase with the vector of outcomes by the other members of its group, i.e., reinforcing endogenous interactions. It also demands the set of feasible outcomes to be a compact subset of the real line. For further reference, the minimum and maximum of Y are denoted $\underline{Y}$ and $\overline{Y}$respectively.

[15] In Sub-section 3.1 we already assumed Y is countable. M0.S imposes further structure on the outcome set. While the restrictions in assumption M0.S are required by the theoretical methodology we use, the former condition is just imposed to simplify our exposition.

Within the broad range of models that satisfy the last condition we distinguish two kind of settings. The first one encompasses the 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. In Example 2, the hypothesis that spouses obtain greater pleasure from retirement if they retire together is consistent with the basic restriction maintained in this class of games. The second kind includes models with positive externalities, in which the interactions occur at the level of payoffs, e.g. peer effects in the classroom [see Graham (2008) and Hahn and Hirano (2009), among others]. In this second case the system of structural equations (4) should be understood as agents' payoffs or achievements.

The next lemma uses assumption M0.S to state consistency, and describes the set of solutions when the model is incomplete. In Lemma 3 smallest and largest mean coordinatewise smallest and largest.

Lemma 3 *Assume M0.S holds, then the system of structural equations* $\mathbf{f}_k$ *has a smallest and a largest solution for all* $t \in T$ *and for all* $k \in K$.

The system of structural equations leads to a 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. Assumption M0.S ensures this function is increasing, so that the proof of Lemma 3 follows directly from Tarski's Fixed Point Theorem [see Section 8 (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. Exceptions to this approach are Ackerberg and Gowrisankaran (2006) who establish existence by invoking existing results in the literature of supermodular games, as we do here.

Let the symbol $\phi(t, k)$ denote the set of solutions (or fixed points) to system $\mathbf{f}_k$ for a potential treatment t. Lemma 3 states that $\phi(t, k)$ is non-empty, and has a smallest and a largest vector for all $(t, k) \in T \times K$. Our main propositions take for granted that either the system of structural equations has a unique solution or that the equilibrium selection rule is such that one of the extremal equilibria is always chosen.

Several sufficient conditions on the primitives (i.e., the structural equations) would rule out the possibility of multiple equilibria, e.g. uniqueness would hold if system $\mathbf{f}_k$ were a contraction. However, as these sufficient conditions cannot (in general) be justified on economic grounds, we prefer to state the assumption directly on the solution set.

U. One of the next conditions is satisfied (i) the solution set is a singleton; or (ii) the selection rule is such that either the smallest or the largest element of the solution set is always chosen. This condition holds for all the groups.

As we mentioned earlier, the latter constraint needs to be justified on a case-by-case scenario. To this end the distinction between the two kind of models that satisfy condition M0.S turns out to be quite important. If the endogenous interactions occur at the level of actions or strategies, then coordination among the members of a group could be a plausible justification for U(ii) under specific circumstances (we elaborate on this point in the restaurant example, in Section 2). If the endogenous interactions occur at the level of payoffs or achievements, then coordination cannot be used to justify U(ii). In models with positive externalities, U(i) is frequently (explicitly or implicitly) invoked. Much of the empirical analysis assumes linear structural functions, and uniqueness follows by a rank condition on the coefficient matrix.

The emphasis of the literature on the extremal elements of the solution set relies (at least partially) on a simple observation: in these models the minimal and the maximal solutions are (almost always) stable with respect to alternative adaptive dynamics, and therefore robust under occasional perturbation [see, e.g., Brock and Durlauf (2001), Echenique (2002) and Milgrom and Roberts (1990)].

Under assumption M0.S the model is consistent, adding U guarantees it is also complete. It is readily verified that when M0.S and U hold then the probability that the joint vector of potential outcomes falls in a specified set $B \subseteq \mathbb{R}^{|L|}$ is given by

$$P\left[\mathbf{y}\left(t\right) \in B\right] = \textstyle\sum_{k \in K} 1\left[\mathbf{y}_k\left(t\right) \in B\right] \pi_k \tag{6}$$

where $\mathbf{y}_k\left(t\right)$ is either the unique vector of $\phi\left(t,k\right)$ or the extreme one that is selected.

The observation set $\left[\left(z^m, \mathbf{y}^m\right), m \in M\right]$ may entail several different realized treatments. Let $\pi_{k|z}$ indicate the fraction of groups in $\mathcal{U}$ which are of type k conditional on z, and let $P\left(z\right)$ denote the distribution of realized treatments across groups. Here we make the convention that $\pi_{k|z} \equiv 0$ 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

$$P(\mathbf{y} \in B) = \textstyle\sum_{s \in T} P(\mathbf{y} \in B | z = s) P(z = s) \tag{7}$$

where $P(\mathbf{y} \in B | z = s) = \sum_{k \in K} 1[\mathbf{y}_k(s) \in B] \pi_{k|z=s}$. Then $P(\mathbf{y})$ is a mixture of the distributions of realized outcomes for different realized treatments, with $P(z)$ 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 problem of identification is captured by the next identity

$$P[\mathbf{y}(t)] = P[\mathbf{y}(t) | z = t] P(z = t) + P[\mathbf{y}(t) | z \neq t] P(z \neq t). \tag{8}$$

Given the sustained assumptions, the equality (8) follows by the Law of Total Probability.[16] The empirical evidence reveals $P[\mathbf{y}(t) | z = t] = P(\mathbf{y} | z = t)$, $P(z = t)$ and $P(z \neq t)$.

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., $P[\mathbf{y}(t) | z \neq t]$. However, by assumption M0.S, the set of possible vectors of outcomes is bounded from below and above 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 distributional bounds for the counterfactual distribution $P[\mathbf{y}(t) | z \neq t]$. Manski (1990) elaborates on this simple route for inference.

Let $\triangle_{Y^{|L|}}$ denote the set of multivariate distribution functions that are consistent with the nature of $Y^{|L|}$, and let the Greek letter H stand for the set of distributions that are consistent with the initial assumptions given the empirical evidence, i.e., the identification region for $P[\mathbf{y}(t)]$. The previous analysis suggests that $P[\mathbf{y}(t)]$ lies in the identification region indicated in the following proposition, which is expressed in terms of stochastic dominance (st) [see Section 8 (Proofs) for the typical multivariate characterization of the standard partial order of stochastic dominance].

[16] Manski (2003) uses this kind of decomposition to explain the anatomy of the problem of inference in alternative settings, e.g. the problem posed by missing data.

Proposition 4 *Assume M0.S and U hold. Then,*

$$H_M \{P[\mathbf{y}(t)]\} = \left\{ \delta \in \triangle_{Y^{|L|}} : \begin{array}{c} P(\mathbf{y}\,|z=t)\,P(z=t) + P\left(\overline{Y}^{|L|}\right) P(z \neq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y}\,|z=t)\,P(z=t) + P\left(\underline{Y}^{|L|}\right) P(z \neq t) \end{array} \right\} \tag{9}$$

for all $t \in T$. These bounds are sharp.

The next two sub-sections address the power of identification of two extra monotone restrictions, namely, monotone treatment response and monotone treatment selection.

3.2.2 Identification Using a Monotone Treatment Response Assumption

This sub-section builds on both Manski (1997, 2010) and Milgrom and Roberts (1990). The analysis that follows assumes the structural equations are weakly increasing in potential treatments. Opposite results apply if we reverse the signs of these effects.

M1.S. Let T be a partially ordered set. Let $t, t' \in T$. Then

$$t \geq t' \Longrightarrow f_{kl}(t, \mathbf{y}) \geq f_{kl}(t', \mathbf{y}) \tag{10}$$

$\forall \mathbf{y} \in Y^{|L|-1}$ and $\forall l \in L$. This condition holds for all $k \in K$.

Manski (1997) introduces a monotone shape restriction to study identification in settings where a person's outcome varies only with his own treatment, i.e., SUTVA.[17] Manski (2010) extends that analysis to settings with exogenous social interactions. In both studies the monotone shape restriction is imposed directly on the outcome functions. In this article the primitives are the structural equations, so M1.S is defined on system $\mathbf{f}_k$, and the assumptions sustained by Manski (1997, 2010) are in line with the statement in the next lemma.

Lemma 5 *Assume M0.S and M1.S hold. Then the least and the greatest solution vectors in $\phi(t,k)$ increase in t, for all $t \in T$ and for all $k \in K$.*

[17] In the treatment response literature SUTVA stands for Stable Unit Treatment Value Assumption.

The last result is well-known in the literature of supermodular games [see, e.g., Milgrom and Roberts (1990), Theorem 6]. The intuition behind the comparative statics is nicely captured by Example 2. If the income tax increases for the spouses, then the extra benefit of working an additional year decreases for both members of the married couple, and they have incentives to retire earlier (this is the dual version of M1.S). This first effect is reinforced then by the positive endogenous interactions: if one of the spouses brings forward the retirement decision, then the other one has an additional stimulus to do so (M0.S). As the direct effect and the indirect one act in the same direction, then the overall effect of the policy change can be predicted at the extremal equilibria. We next elaborate on the identification power of introducing condition M1.S.

If in addition to M0.S and M1.S, U is also satisfied, then Lemma 5 implies that $\mathbf{y}_k(t)$ increases in t. Assume these conditions hold and consider a type-k group that has empirical evidence $(z^m, \mathbf{y}^m)$. If $z^m \leq t$ then $\mathbf{y}^m$ is a sharp lower bound for $\mathbf{y}_k(t)$. Otherwise, the empirical evidence is uninformative, and $\underline{Y}^{|L|}$ constitutes the sharp lower bound. Alternatively, if $z^m \geq t$ 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 the sustained assumptions do not entail cross-restrictions (and k was arbitrarily chosen), then this analysis extends to all groups in $\mathcal{U}$. These observations justify the next identification region for $P[\mathbf{y}(t)]$ in terms of stochastic dominance.

Proposition 6 *Assume M0.S, M1.S and U hold. Then,*

$$H_{MTR}\{P[\mathbf{y}(t)]\} = \left\{ \delta \in \triangle_{Y^{|L|}} : \begin{array}{c} P(\mathbf{y}\,|z \geq t)\,P(z \geq t) + P\left(\overline{Y}^{|L|}\right)P(z \ngeq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y}\,|z \leq t)\,P(z \leq t) + P\left(\underline{Y}^{|L|}\right)P(z \nleq t) \end{array} \right\} \tag{11}$$

for all $t \in T$. These bounds are sharp.

The multivariate standard stochastic order is closed with respect to marginalization. Therefore, Proposition 6 also implies sharp bounds for the marginal distributions of potential outcomes of the 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 simple observation.[18] (We write L/S for the set of classes in L different from those in S.)

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

Corollary 7 *Assume M0.S, M1.S and U hold. Then, for all $t \in T$,*

$$H_{MTR}\left\{P\left[\mathbf{y}_S\left(\mathbf{t}\right)\right]\right\} = \left\{ \delta \in \triangle_{Y^{|S|}} : \begin{array}{c} P\left(\mathbf{y}_S | z \geq t\right) P\left(z \geq t\right) + P\left(\overline{Y}^{|S|}\right) P\left(z \ngeq t\right) \\ \geq_{st} \delta \geq_{st} \\ P\left(\mathbf{y}_S | z \leq t\right) P\left(z \leq t\right) + P\left(\underline{Y}^{|S|}\right) P\left(z \nleq t\right) \end{array} \right\} \quad (12)$$

where $P\left(\mathbf{y}_S | \,.\right) = E_{\mathbf{y}_{L/S}}\left[P\left(\mathbf{y} \mid .\right)\right]$. These bounds are sharp.

Proposition 6 can also be used to derive bounds for any increasing function g from the potential outcomes $\mathbf{y}\left(t\right)$ to $\mathbb{R}^n$ (with $n \leq |L|$), e.g. the mean of the vector of potential outcomes $E\left[\mathbf{y}\left(t\right)\right]$ [see Müller and Stoyan (2002), Theorem 3.3.11, p. 94]. Manski (1997, 2010) elaborates on this characterization for cases where the outcome of interest has only one dimension.

The elements of T need not be ordered. Therefore, in some circumstances the identified bounds in (11) may not improve the ones in (9). The next sub-section introduces a third monotone restriction that allows making counterfactual predictions for a subset of groups in the population using the empirical evidence of an alternative subset of groups.

3.2.3 Identification Using a Monotone Treatment Selection Assumption

This sub-section takes advantage of Amir (2008) and Manski and Pepper (2000, 2008). Many studies have established identification results by assuming the outcome functions are statistically independent of realized treatments, i.e., $P\left[\mathbf{y}\left(t\right) | z = t\right] = P\left[\mathbf{y}\left(t\right)\right]$. In our case this condition would be plausible if an explicit randomization mechanism had been used to assign treatments to the groups, so that $\pi_{k|z} = \pi_k$. This assumption is hard to justify in many circumstances. Manski and Pepper (2000, 2008) weaken that restriction by proposing a monotone constraint that fits better many applications of interest. Their constraint requires the outcome functions to be monotone in realized treatments. They refer to this assumption as monotone treatment selection (MTS). The extension of their finding to interactions-based models requires the comparison of equilibrium outcomes for groups that are of different types. To this end we use some of the results derived by Amir (2008).

In models of social interactions the MTS condition can be interpreted as suggesting that groups that have larger realized treatments have potential outcomes that are statistically larger. The next analysis introduces a restriction on the primitives (i.e., the structural equations) that leads to the latter result. Prior to describe condition M2.S we define a (natural) partial order on $(\mathbf{f}_k, k \in K)$.

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

According to the last definition, $\mathbf{f}_k$ is larger than $\mathbf{f}_{k'}$ if the former function is pointwise greater than the latter one. Let $\mathbf{f}$ denote a random vector of functions with support $(\mathbf{f}_k, k \in K)$, and define $P(\mathbf{f} = \mathbf{f}_k \mid z) \equiv \pi_{k|z}$ for all $k \in K$. The next condition formalizes the main assumption in the present sub-section.

M2.S. Let T be a partially ordered set. Let $s, s' \in T$. Then,

$$s \geq s' \Longrightarrow P(\mathbf{f} \mid z = s) \geq_{st} P\left(\mathbf{f} \mid z = s'\right). \tag{13}$$

Condition M2.S simply states that the fraction of groups with larger structural equations increases with the realized treatments. Depending on the context, an alternative interpretation would be that groups that select higher treatments have stochastically weakly larger structural functions than those that select lower ones. Manski and Pepper (2000, 2008) introduce the MTS condition to study identification in settings where a person's outcome varies only with his own treatment, i.e., SUTVA. Their sustained assumption is similar to the statement in the next lemma.

Lemma 9 *Assume M0.S, M2.S and U hold. Let* $s, s' \in T$. *Then,*

$$s \geq s' \Longrightarrow P[\mathbf{y}(t) \mid z = s] \geq_{st} P\left[\mathbf{y}(t) \mid z = s'\right] \tag{14}$$

for all $t \in T$.

The last result states that groups that select larger treatments have potential outcomes that are weakly larger. As we mentioned earlier, this lemma requires the comparison of equilibrium solution sets for groups that are of different types. Lemma 21, in Section 8 (Proofs), shows that if M0.S is satisfied and $\mathbf{f}_k \geq \mathbf{f}_{k'}$ then the least and the greatest solution vectors in $\phi(t, k)$ are larger than the corresponding ones in $\phi(t, k')$. This statement is based on a clever observation of Milgrom and Roberts' Theorem, made by Amir (2008). The proof of Lemma 9 is based on the previous result and the characterization of standard stochastic dominance in terms of the expectation of increasing functions. Although the outcome of Lemma 9 is intuitive, its proof is quite involved.

It is immediate to notice that M2.S will have identification power. The empirical evidence reveals $P[\mathbf{y}(t) \mid z = t] = P(\mathbf{y} \mid z = t)$. If M0.S, M2.S and U hold, then $P(\mathbf{y} \mid z = t)$ is a lower bound for $P[\mathbf{y}(t) \mid z \geq t]$, and it is an upper bound for $P[\mathbf{y}(t) \mid z \leq t]$ as well. The last two observations, formalized in Corollary 22, in Section 8 (Proofs), are direct implications of Lemma 9. The next proposition uses these remarks (and the natural bounds for the vector of potential outcomes) to state sharp distributional bounds for $P[\mathbf{y}(t)]$.

Proposition 10 *Assume M0.S, M2.S and U hold. Then,*

$$H_{MTS}\{P[\mathbf{y}(t)]\} = \left\{ \delta \in \triangle_{Y^{|L|}} : \begin{array}{c} P(\mathbf{y} \mid z = t) P(z \leq t) + P\left(\overline{Y}^{|L|}\right) P(z \nleq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y} \mid z = t) P(z \geq t) + P\left(\underline{Y}^{|L|}\right) P(z \ngeq t) \end{array} \right\} \tag{15}$$

for all $t \in T$. These bounds are sharp.

The bounds in Proposition 10 are neither tighter nor looser than the ones in Proposition 6, they are just different. A more precise region of identification for $P[\mathbf{y}(t)]$ is obtained when the analyst is confident about M0.S/M1.S and M0.S/M2.S so that he can combine both results [see Manski and Pepper (2000, 2008) for further details].

Here again, Proposition 10 implies sharp bounds for the marginal distributions of potential outcomes of the 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 11 (without proof) formalizes this claim.

Corollary 11 *Assume M0.S, M2.S and U hold. Then, for all $t \in T$,*

$$H_{MTS}\{P[\mathbf{y}_S(t)]\} = \left\{ \delta \in \triangle_{Y^{|S|}} : \begin{array}{c} P(\mathbf{y}_S \mid z = t) P(z \leq t) + P\left(\overline{Y}^{|S|}\right) P(z \nleq t) \\ \geq_{st} \delta \geq_{st} \\ P(\mathbf{y}_S \mid z = t) P(z \geq t) + P\left(\underline{Y}^{|S|}\right) P(z \ngeq t) \end{array} \right\} \tag{16}$$

where $P(\mathbf{y}_S \mid .) = E_{\mathbf{y}_{L/S}}[P(\mathbf{y} \mid .)]$. These bounds are sharp.

As Manski and Pepper (2000) state, although assumptions M0.S/M1.S and M0.S/M2.S are not individually refutable, the combined restriction can be shown false given the data. 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.

3.2.4 Estimation from Sample Data

The bounds identified in Propositions 4, 6 and 10 can be consistently estimated from sample data. The analyst just needs to substitute all the empirical distributions with their sample analogs. The techniques developed by Imbens and Manski (2004) and Stoye (2009) for partially identified parameters, could be useful here to construct confidence intervals for the estimated distributional bounds.

In this analysis data at the group level is needed because we take the research question to be inference about the joint distribution of potential outcomes. If the analyst's problem were inference on the marginal distributions for a specific class $l \in L$, then a random sample of individuals that fit in l would be sufficient (with treatment information at the group level).

4 Model of Social Interactions where Identities Don't Matter

This section addresses identification of treatment response in situations where groups are large (L), and social interactions are anonymous. Models of crime decisions, schooling, infectious diseases and addictions fit in this setting.

4.1 Econometric Framework and the Analyst's Problem

In this section groups are composed of a large (but finite) number of individuals of the population J, and we allow them to differ with respect to the number of members. Social interactions are global or anonymous, in the sense that each individual assigns an identical weight to the outcome of every other member of its group of reference.

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

$$y_j(t) = f_j \{t, P[y(t)]\}, j \in G \tag{17}$$

where $P[y(t)]$ is the distribution of individual outcomes induced by $\mathbf{y}(t)$, i.e., $P[y(t) \in B] = \sum_{j \in G} 1[y_j(t) \in B](1/|G|)$ for all set $B \subset \mathbb{R}$. Here $|G|$ is assumed to be large enough that changes in any individual outcome hardly affect $P[y(t)]$.

The distributional assumptions are similar to the ones in the previous section, so we omit a thoughtful justification. We say a group is of type k if its members have structural equations $\mathbf{f}_k \equiv [f_{kn}(.), n \in N_k]$, where $f_{kn} : T \times \triangle_Y \to Y$ and N_k denotes the set of members for a type-k group. Then the type of a group specifies how its members react to the treatments and outcome distributions by all members of the same neighborhood. The set of possible types of groups is denoted again by K.

The relative amount of agents of distinct types in J, i.e., with different outcome functions, and the underlying social and economical incentives of group formation define the distribution of types of groups in the universe $\mathcal{U}$. 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. As in this section groups might differ with respect to the number of members, here π_k reflects 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 the population.

Groups have observable vectors of realized treatment z^m and outcomes $\mathbf{y}^m$. The objective of the analysis is to gain knowledge of the outcome distribution (at the individual level) that would occur if all groups were to receive a treatment t, i.e., $P[y(t)]$. To this end we will use the empirical evidence, $[(z^m, \mathbf{y}^m), m \in M]$, taking advantage of some prior information, i.e., monotone restrictions.

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 all members of a certain city.*

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)].

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

4.2 Identification Region for $P\left[y\left(t\right)\right]$

The first sub-section introduces a monotone shape restriction that guarantees consistency. The following two address the identification power of imposing some additional structure via monotone comparative statics. We end the analysis with a brief discussion of sample estimation.

4.2.1 Consistency of the Model

The first assumption imposes a monotone shape restriction on $(\mathbf{f}_k, k \in K)$. As in Sub-section 3.2.1 this single condition guarantees coherency.

M0.L. Let $Y \subseteq \mathbb{R}$ be compact. Let $\mathbf{y}, \mathbf{y}' \in Y^{|N_k|}$. Then

$$P\left(y\right) \geq_{st} P\left(y'\right) \Longrightarrow f_{kn}\left[t, P\left(y\right)\right] \geq f_{kn}\left[t, P\left(y'\right)\right] \tag{18}$$

$\forall n \in N_k$ and $\forall t \in T$. This condition holds for all $k \in K$.

Condition M0.L requires the outcome of each individual to increase with the distribution of outcomes by all members in the same group. As in this setting interactions are anonymous, M0.L is clearly weaker than condition M0.S in Sub-section 3.2.1. In Example 12 this constraint is carefully justified by Sah (1991). Let us fix police per-capita at some initial level; he argues that if crime participation rate increases, then the probability that a person who commits a crime will be punished decreases thereby increasing his incentives to become a criminal.

The next lemma uses M0.L to state consistency.

Lemma 13 *Assume M0.L holds, then the system of structural equations* $\mathbf{f}_k$ *has a smallest and a largest solution for all* $t \in T$ *and for all* $k \in K$.

The proof of Lemma 13 follows again by Tarski's Fixed Point Theorem. Most of the social interactions models for large groups introduce M0.L as the distinctive feature of the phenomenon of interest [see, e.g., Becker (1991) and Glaeser Sacerdote, and Scheinkman (1996)]. In this kind of models coherence is obtained without further restrictions.

Let $\varphi(t,k)$ denote the set of solutions to system $\mathbf{f}_k$ for a potential treatment $t \in T$. Since we take the analyst's objective to be inference about the outcome distribution $P[y(t)]$, we find it convenient to let the elements of $\varphi(t,k)$ be the distribution functions $P[y_k(t)]$ induced by the solution vectors $\mathbf{y}_k(t)$. Lemma 13 states that $\varphi(t,k)$ is non-empty, and has a least and a greatest distribution. (Here least and greatest should be understood with respect to the partial order of standard stochastic dominance.) As in Sub-section 3.2, the main propositions in this sub-section take for granted that either the system of equations $\mathbf{f}_k$ has a unique solution, or that one of the extremal distributions is always selected.

Assumptions M0.L and U together ensure the model is both coherent and complete. Throughout we write $P[y_k(t)]$ for the unique element of $\varphi(t,k)$, or the extremal one that is always selected. If M0.L and U are satisfied, then the probability that the potential outcome falls in a specified set $B \subseteq \mathbb{R}$ is given by

$$P[y(t) \in B] = \textstyle\sum_{k\in K} P[y_k(t) \in B]\,\pi_k. \tag{19}$$

and it is linear in π_k.

The observation set $[(z^m, \mathbf{y}^m), m \in M]$ may comprise different realized treatments. Let $\pi_{k|z}$ indicate the proportion of individuals in J that belong to groups which are of type k conditional on the realized treatment z. We let $P(z)$ 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

$$P(y \in B) = \textstyle\sum_{\mathbf{s}\in T} P[y(t) \in B|z = s]\,P(z = s) \tag{20}$$

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

The research objective is inference about the outcome distribution $P[y(t)]$, which describes individual treatment response across groups for a potential treatment t. The difficulties for addressing point identification are captured by the next equality

$$P[y(t)] = P[y(t)\,|z = t]\,P(z = t) + P[y(t)\,|z \neq t]\,P(z \neq t) \tag{21}$$

where $P[y(t)\,|z = t]$ denotes the distribution of potential outcomes for those individuals whose groups have received treatments equal to the potential one, and $P[y(t)\,|z \neq t]$ indicates the distribution of potential outcomes for the other members of the population. Given the sustained assumptions, the equality holds by the Law of Total Probability.

The empirical evidence reveals $P[y(t)|z=t] = P(y|z=t)$, $P(z=t)$ and $P(z \neq t)$. Data alone are uninformative about $P[y(t)|z \neq t]$. However, the degenerate distributions $P(\underline{Y})$ and $P(\overline{Y})$ are sharp lower and upper bounds for this counterfactual distribution. This claim follows directly from condition M0.L.

The previous analysis leads to the next identification region for $P[y(t)]$.

Proposition 14 *Assume M0.S and U hold. Then,*

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

for all $t \in T$. These bounds are sharp.

The endpoints of (22) coincide with one another only in exceptional cases, thus $P[y(t)]$ is almost always just partially identified. The next two sub-sections address the power of identification of two extra monotone restrictions that shed extra light on the counterfactual distribution $P[y(t)|z \neq t]$.

4.2.2 Identification Using a Monotone Treatment Response Assumption

The next condition is the analogue to condition M1.S is Sub-section 3.2.2.

M1.L. Let T be a partially ordered set. Let $t, t' \in T$. Then

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

$\forall n \in N_k$ and $\forall P(y) \in \Delta_Y$. This condition holds for all $k \in K$.

Assumption M1.L has direct implications in terms of comparative statics with respect to the extreme distributions of the solution set $\varphi(t,k)$, as stated in the next lemma.

Lemma 15 *Assume M0.L and M1.L hold. Then the smallest and the largest distributions in $\varphi(t,k)$ increase in t with respect to the standard stochastic order, $\forall (t,k) \in T \times K$.*

It is readily apparent that M1.L will have identification power. If in addition to M0.L and M1.L, U is also satisfied, then Lemma 15 implies that $P[y_k(t)]$ is stochastically increasing in t, i.e., if $t \geq t'$ then $P[y_k(t)] \geq_{st} P[y_k(t')]$. Assume these conditions hold and consider a group of type k that has empirical evidence $(z^m, \mathbf{y}^m)$. Let $P(y^m)$ denote the outcome distribution induced by the vector $\mathbf{y}^m$. If $z^m \leq t$ then $P(y^m)$ is a sharp lower bound for $P[y_k(t)]$. Otherwise, the empirical evidence is uninformative, and $P(\underline{Y})$ constitutes the sharp lower bound. On the other hand, if $z^m \geq t$ then $P(y^m)$ is a sharp upper bound for $P[y_k(t)]$. Otherwise, the sharp upper bound is just the degenerate distribution $P(\overline{Y})$. Since the sustained assumptions do not entail cross-restrictions, this analysis extends to all groups in $\mathcal{U}$ and justifies the next distributional bounds for $P[y(t)]$.

Proposition 16 *Assume M0.L, M1.L and U hold. Then,*

$$H_{MTR}\{P[y(t)]\} = \left\{ \delta \in \triangle_Y : \begin{array}{c} P(y \mid z \geq t)\, P(z \geq t) + P(\overline{Y})\, P(z \ngeq t) \\ \geq_{st} \delta \geq_{st} \\ P(y \mid z \leq t)\, P(z \leq t) + P(\underline{Y})\, P(z \nleq t) \end{array} \right\} \tag{24}$$

for all $t \in T$. These bounds are sharp.

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 allows 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 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) elaborates on this characterization.

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 3.2 because it is not immediate that the multivariate standard stochastic order has clear predictions for all the existing definitions.

4.2.3 Identification Using a Monotone Treatment Selection Assumption

This sub-section provides sufficient conditions on the primitives (i.e., the structural equations) to validate the use of realized treatments as monotone instrumental variables in the sense of outcome monotonicity. This analysis is harder than the one of Sub-section 3.2.3 because it requires the comparison of equilibrium sets for groups that differ in the number of members. Since the solution sets, $\varphi(t,k)$, comprise distribution functions we use the standard stochastic order to contrast them.

Let us define

$$F_k[y \in B \mid t, P(y)] \equiv \sum\nolimits_{n \in N_k} 1\{f_{kn}[t, P(y)] \in B\}(1/|N_k|) \tag{25}$$

for all set $B \subseteq \mathbb{R}$. Then F_k indicates the probability that the individual outcomes in a type-k group lie in a set $B \subseteq \mathbb{R}$, for any given treatment $t \in T$ and any initial distribution $P(y) \in \Delta_Y$ that its members may face. We next introduce a partial order on $(F_k, k \in K)$ similar to the one introduced by Definition 8.

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

According to the last description, 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 $P(F = F_k \mid z) \equiv \pi_{k|z}$ for all $k \in K$. The next condition is the analogue to condition M2.S in Sub-section 3.2.3.

M2.L. Let T be a partially ordered set. Let $s, s' \in T$. Then,

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

Condition (26) states that the proportion of individuals that belong to groups with weakly higher outcome distributions increases with the realized treatments. The motivation for this constraint is well captured by Example 12: in the study of crime rates it would be reasonable to think that the public authority invests more (per-capita) on the criminal apprehension system of those cities where people's willingness to commit crimes is believed to be higher.

The last constraint has direct implications for the conditional distributions of potential outcomes, as stated in next lemma.

Lemma 18 *Assume M0.L, M2.L and U hold. Let $s, s' \in T$. Then,*

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

for all $t \in T$.

It is again readily apparent that M2.L will have identification power. The empirical evidence reveals $P[y(t) \mid z=t] = P(y \mid z=t)$. Corollary 25, in Section 8 (Proofs), shows that $P(y \mid z=t)$ is a lower bound for $P[y(t) \mid z \geq t]$ and an upper bound for $P[y(t) \mid z \leq t]$. The latter results derive directly from the previous lemma. These remarks (and the natural bounds for the individual outcomes) justify the next identification region for $P[y(t)]$ in terms of stochastic dominance.

Proposition 19 *Assume M0.L, M2.L and U hold. Then,*

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

for all $t \in T$. These bounds are sharp.

As we mentioned earlier, a more precise region of identification for $P[y(t)]$ is obtained when the analyst is confident about M0.L/M1.L and M0.L/M2.L so that he can combine both results.

4.2.4 Estimation from Sample Data

As in Sub-section 3.2.4, the distributional bounds identified in Propositions 14, 16 and 19 can be consistently estimated from finite-sample data. The analyst just needs to substitute all the empirical distributions with the sample analogs. The main difference between Sub-sections 3.2.4 and 4.2.4 is that in the latter sampling groups is not needed. The estimations can be performed by sampling individuals [see, e.g., Brock and Durlauf (2001)].

5 Application: Crime Rates and Social Interactions

This section applies our results to the analysis of crime rates and social interactions. 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 be a criminal is his perceived probability of punishment. Subsequent work emphasizes the importance of positive social interactions in motivating criminal behavior [Glaeser, Sacerdote and Scheinkman (1996) review this literature].

The framework in Section 4 seems appropriate 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 level of the city, 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 certain city is, the lower the probability of being punished. His argument justifies M0.L. 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 M1.L holds here as well. As we explained in Sub-section 4.2.3, it could also be reasonable to think that the public authority invests 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 M2.L.

The aim of the analysis that follows is to learn about the distribution of crimes in the state of New York, if all its cities were to be assigned a certain 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

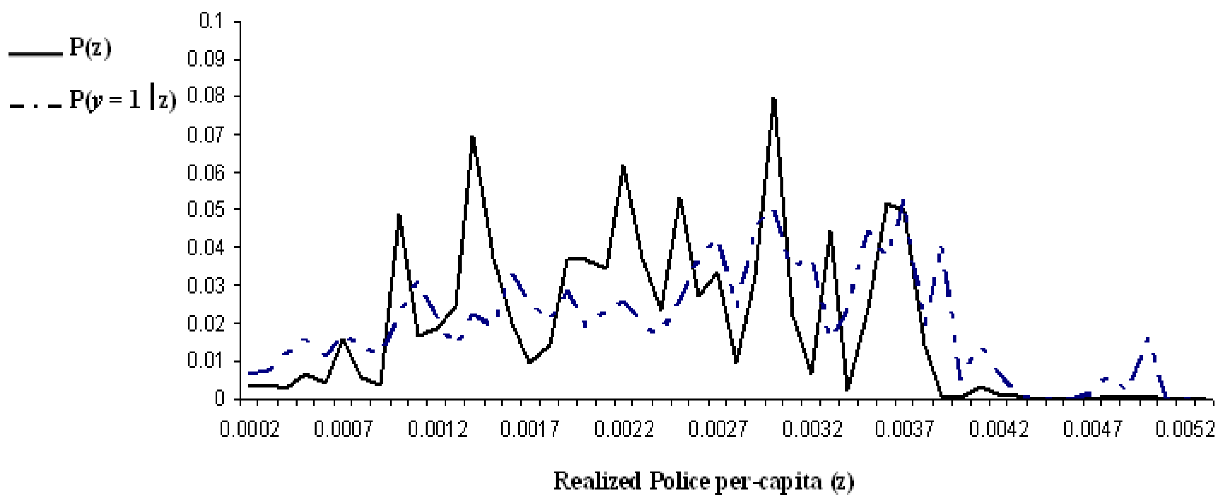
Our data source is the Uniform Crime Reporting (UCR) program of the FBI, for the year 2008 [similar data are used by Glaeser et. al. (1996)]. The UCR program informs crimes reported (and verified), and classifies them in two groups. The first group, violent crimes, includes murder and manslaughter, forcible rape, robbery and aggravated assault. The second one, property crimes, comprises burglary, larceny-theft, motor vehicle theft and arson. We focus on the latter offenses because Glaeser et. al. (1996) provide evidence that social interactions are stronger there.

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 New York from the sample, as its features (e.g. population) are markedly different from the characteristics of other ones. Our data covers 40% of the remaining population in the state of New York. To perform the analysis we discretized the level of police per-capita (i.e., number of police in city i/population in city i) in multiples of 0.0001.

Figure 1 displays the data for the 99.6% of the sampled population—to display the data we eliminated a few observations with extremely high levels of police per-capita for expositional ease, these observations were taken into account for the estimation purpose. For levels of police per-capita between 0 and 0.0052, the probabilities $P(z)$ and $P(y = 1 \mid z)$ indicate the fraction of individuals in the sample that received those treatments and the proportion of individuals that committed a crime during the year 2008 conditional on realized treatments, respectively.

Since a large part of the sampled population—specifically, 94.32%—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 1: Realized Police per-capita and Criminal Activity



Source: FBI, UCR program.

5.2 Findings

In this application, the outcome of interest is binary, i.e., $Y = \{0, 1\}$. Therefore, the fraction of people that commits a crime at a given level of police per-capita t, $P[y(t) = 1]$, fully describes the potential outcome distribution of criminal activity in the state of New York for the specified treatment.

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 we report the lower and the upper bounds for $P[y(t)=1]$ under three sets of monotone restrictions. These bounds are expressed in percentage points. First we assume only M0.L, then M0.L/M1.L and finally M0.L/M1.L/M2.L. This analysis implicitly assumes that each individual has committed at most one crime during the year under study. Table 2 displays the results.

Table 2: Lower and Upper Bounds for $P[y(t)=1]$ in %

Police per-capita	M0.L		M0.L/M1.L		M0.L/M1.L/M2.L	
$t=0.001$	00.01	95.22	02.84	90.78	02.84	90.78
$t=0.002$	00.07	96.36	02.16	62.96	02.16	62.85
$t=0.003$	00.40	92.44	01.19	24.92	01.19	24.92
$t=0.004$	00.00	99.93	00.02	4.06	00.02	01.64

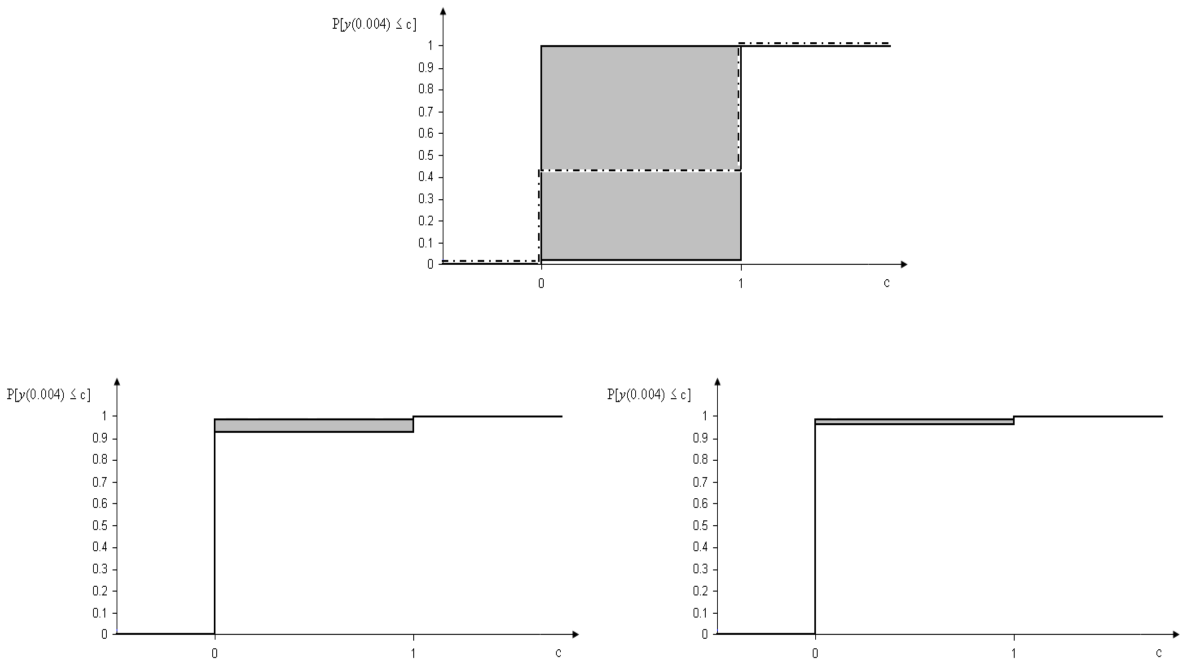
Source: FBI, UCR program.

The three sets of restrictions in Table 2 are clearly nested. As a consequence, the informational content of the estimations increases as we move from the left to the right along each row. In this analysis, the informational content of each pair of values increases when the absolute value of their difference gets smaller. We next highlight the main findings.

We can clearly see that M0.L 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 M1.L noticeably improves all the predictions. For instance, for $t=0.004$, M0.S alone predicts that $P[y(t)=1]\in[0,99.93]$, while the introduction of M1.S reduces the interval to $P[y(t)=1]\in[0.02,4.06]$. The reason is that a very small fraction of the population has realized treatment $z=0.004$—specifically, 0.67%—and then the empirical evidence alone is ineffective to identify the potential outcome. However, by adding M1.S all the data are used in the estimation: part of the observations shed light on the upper bound, other segment contributes to identify the lower bound and the remaining observations play a role in the two extremal distributions. For this treatment level, M2.L provides significant extra information as the upper bound for $P[y(t)=1]$ decreases from 4.06 to 1.64, while the lower bound remains unchanged.

Figure 2 illustrates the latter observations. The grey areas indicate the identification regions for the potential outcome cumulative distribution functions (CDF's) under the three sets of assumptions. The dashed line on the first picture represents one of the alternative CDF's that are consistent with M0.L given the empirical evidence. As we observed before, the identification regions become much tighter as we add monotone restrictions to the model.

Figure 2: Bounds for the Potential Outcome CDF's for $t = 0.004$



On the top M0.L, on the bottom-left M0.L/M1.L and on the bottom-right M0.L/M1.L/M2.L. Source: FBI, UCR program.

6 A Potential Extension to Incomplete Models

The main propositions in Sections 3 and 4 take for granted that either the system of structural equations has a unique solution, or that the equilibrium selection rule is such that one of the extremal equilibria is always chosen. Although many situations of interest may fulfill one of these conditions, relaxing them is a natural direction for further research.

As Echenique and Komunjer (2009) notice, when the model is incomplete (i.e., it has multiple solutions) the empirical evidence has a mixture distribution.[19] Each group's observations derive

[19] The framework of Echenique and Komunjer (2009) follows Jovanovic (1989).

from an equilibrium selection rule, and the analyst does not know (in general) which equilibrium was actually chosen. Assuming a fairly general equilibrium selection device, they show that the comparative statics has testable implications on sufficiently small and large quantiles of the outcome distribution. Although their analysis does not allow for counterfactual predictions—the sufficiently small and large quantiles are not identified—the framework they use seems appropriate to extend the previous results to incomplete models. The main difficulty will be to separate the distributions of the extremal equilibria from the other ones in the empirical evidence. The statistical literature on "the mixing problem" may shed light on this issue.

7 Concluding Remarks

The paper extends Manski (1990, 1997, 2010) and Manski and Pepper (2000, 2008) to treatment response models with endogenous social interactions. This last feature introduces new difficulties into the analysis, and calls for a deeply different approach to tackle questions that are similar a priori. To this end, we make use of the theoretical contributions of Amir (2008), Milgrom and Roberts (1990) and Tarski (1955), most of which relate to the theory of supermodular games. Our approach to partial identification is nonparametric, and it is based on mild monotone restrictions on the primitives, i.e., the structural equations. 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 within these bounds. The nonparametric bounds for the counterfactual outcome distributions are sharp. By applying our results to the study of crimes and social interactions in the state of New York, we show that they can also be very informative.

We provide a tractable econometric framework, which is an additional contribution of the present study. This framework typifies groups through the relevant features of the group members, thereby observing the natural hierarchy of individuals and groups. The extension of our results to incomplete models is part of the agenda for future research.

8 Proofs

In an attempt to make this paper self-contained we provide three theorems that are invoked in the subsequent proofs.

Tarski's Fixed Point Theorem *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 *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 (greatest) fixed point of f is increasing in s on S [Milgrom and Roberts (1990)].*

The characterization of first order stochastic dominance (FOSD), often called standard stochastic dominance, requires the definition of 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 *Let $X, X' \in \mathbb{R}^n$ be two random vectors. The next conditions are equivalent*

(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.*

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

Proof of Lemma 3: Consider the next mapping

$$\begin{aligned} M_{t,k}: \quad & Y^{|L|} && \to Y^{|L|} \\ & \left(y_1, y_2, ..., y_{|L|}\right) && \to \left(f_{k1}, f_{k2}, ..., f_{k|L|}\right). \end{aligned} \tag{29}$$

It is immediate to notice that the range of $M_{t,k}$ is as given. By construction, the solution set to the system of structural equations $\mathbf{f}_k$, $\phi(t,k)$, coincides with the set of fixed points of $M_{t,k}$. Then the proof of Lemma 3 reduces to show that the set of fixed points of $M_{t,k}$ is non-empty and has a least and a greatest element.

Here $(Y^{|L|}, \geq)$ is a complete lattice for the ordering $\mathbf{y} \geq \mathbf{y'}$ if $y_l \geq y'_l$ for all $l \in L$, i.e., the coordinatewise (partial) order. By assumption M0.S, $M_{t,k}$ is monotone increasing. Hence, Lemma 3 follows by Tarski's Fixed Point Theorem. *Q.E.D.*

Proof of Proposition 4: We first show

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

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

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

Under M0.S and U, the first equality holds by the Law of Total Probability. The empirical evidence reveals $P\left[\mathbf{y}\left(t\right) \in U | z=t\right] = P\left(\mathbf{y} \in U | z=t\right)$. The inequality is true because $\underline{Y}^{|L|}$ is a lower bound for any possible realization of the vector of potential outcomes. Since U was arbitrarily selected, our first claim follows by FOSD's Theorem(i).

The lower bound is sharp because, given the empirical evidence, the prior information is consistent with $P\left[\mathbf{y}\left(t\right)\left|z \neq t\right.\right] = P\left(\underline{Y}^{|L|}\right)$.

The proof for the upper bound is quite similar, so we omit it. *Q.E.D.*

Proof of Lemma 5: Lemma 3 shows that if M0.S holds then $\phi\left(t,k\right)$ is non-empty and has a least and a greatest element. Condition M1.S ensures the system of structural equations $\left[f_{kl}\left(.\right), l \in L\right]$ increases in t for any fixed $k \in K$. Then $M_{t,k}$, as defined in (29), increases in t on T and Lemma 5 follows by Milgrom and Roberts' Theorem. *Q.E.D.*

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

Corollary 20 *Assume M0.S, M1.S and U are satisfied. Then* $P\left[\mathbf{y}\left(t\right)\left|z \leq t\right.\right] \geq_{st} P\left(\mathbf{y}\left|z \leq t\right.\right)$ *and* $P\left(\mathbf{y}\left|z \geq t\right.\right) \geq_{st} P\left[\mathbf{y}\left(t\right)\left|z \geq t\right.\right]$.

Proof of Corollary 20: To prove the first claim let $U \subset \mathbb{R}^{|L|}$ be an upper set, and let us consider the next three steps

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

Under M0.S and U, the two equalities hold by the Law of Total Probability. If we add M1.S, the inequality follows by Lemma 5, as it implies that $\mathbf{y}_k\left(t\right) \geq \mathbf{y}_k\left(s\right)$ for all $t \geq s$. Since U was arbitrarily selected, the first claim follows by FOSD's Theorem(i).

The proof for the second claim is similar, so we omit it. *Q.E.D.*

Proof of Proposition 6: We first show

$$P\left[\mathbf{y}\left(t\right)\right] \geq_{st} P\left(\mathbf{y} \mid z \leq t\right) P\left(z \leq t\right) + P\left(\underline{Y}^{|L|}\right) P\left(z \nleq t\right). \tag{31}$$

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

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

Under M0.S and U, the first equality holds by the Law of Total Probability. The empirical evidence reveals $P\left(\mathbf{y} \in U | z \leq t\right)$, and the first inequality follows by Corollary 20 and FOSD's Theorem(i). The second inequality is true as $\underline{Y}^{|L|}$ is a lower bound for any possible realization of the vector of potential outcomes. Since U was arbitrarily selected, the claim follows by FOSD's Theorem(i).

The lower bound is sharp because, given the empirical evidence, the prior information is consistent with both $P\left[\mathbf{y}\left(t\right) | z \leq t\right] = P\left(\mathbf{y} \mid z \leq t\right)$ and $P\left[\mathbf{y}\left(t\right) | z \nleq t\right] = P\left(\underline{Y}^{|L|}\right)$.

The proof for the upper bound is quite similar, so we omit it. *Q.E.D.*

The proof of Lemma 9 requires an additional result. Lemma 21 is based on Amir (2008).

Lemma 21 *Assume M0.S holds. Let $k, k' \in K$. If $\mathbf{f}_k \geq \mathbf{f}_{k'}$ according to Def. 8, then the least (greatest) vector in $\phi\left(t, k\right)$ is larger than the least (greatest) vector in $\phi\left(t, k'\right)$ for all $t \in T$.*

Proof of Lemma 21: Lemma 3 shows that if M0.S holds then $\phi(t,k)$ and $\phi(t,k')$ are non-empty and 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. The result follows by Milgrom and Roberts' Theorem, because $\phi(t,k)$ and $\phi(t,k')$ are the sets of fixed points of $M_{t,k}$ and $M_{t,k'}$ respectively. *Q.E.D.*

Proof of Lemma 9: Let $s,s'\in T$, with $s\geq s'$, and let us assume that M0.S and U hold. Fix an upper set $U\subset\mathbb{R}^{|L|}$. By definition,

$$P[\mathbf{y}(t)\in U|z=s]=\textstyle\sum_{k\in K}1[\mathbf{y}_k(t)\in U]\,\pi_{k|z=s}. \tag{32}$$

Let us define $g(\mathbf{f}_k,t,U)\equiv 1[\mathbf{y}_k(t)\in U]$. By Lemma 21, $g(\mathbf{f}_k,t,U)\geq g(\mathbf{f}_{k'},t,U)$ if $\mathbf{f}_k\geq\mathbf{f}_{k'}$. Substituting $1[\mathbf{y}_k(t)\in U]$ by $g(\mathbf{f}_k,t,U)$ in (32) we get

$$P[\mathbf{y}(t)\in U|z=s]=\textstyle\sum_{k\in K}g(\mathbf{f}_k,t,U)\,\pi_{k|z=s}. \tag{33}$$

Then $P[\mathbf{y}(t)\in U|z=s]$ is the expectation of an increasing function of $\mathbf{f}$ conditional on $z=s$. By M2.S, we know that $P(\mathbf{f}\mid z=s)\geq_{st}P(\mathbf{f}\mid z=s')$. Then, by FOSD's Theorem(ii), we get that $P[\mathbf{y}(t)\in U|z=s]\geq P[\mathbf{y}(t)\in U|z=s']$. The result follows by FOSD's Theorem(i) as U was arbitrarily selected. *Q.E.D.*

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

Corollary 22 *Assume M0.S, M2.S and U are satisfied. Then* $P[\mathbf{y}(t)|z\geq t]\geq_{st}P(\mathbf{y}\,|z=t)$ *and* $P(\mathbf{y}\,|z=t)\geq_{st}P[\mathbf{y}(t)|z\leq t]$, *for all* $t\in T$.

Proof of Corollary 22: To prove the first claim let $U\subset\mathbb{R}^{|L|}$ be an upper set, and let us consider the next four steps

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

Under M0.S and U, the first and the third equalities hold by the Law of Total Probability. The second one is true as $\sum_{s\in T} 1(s \geq t) P(z = s|z \geq t) = 1$. If we add M2.S, the inequality follows by Lemma 9. Since U was arbitrarily selected, the first claim holds by FOSD's Theorem(i). The proof for the second claim is similar, so we omit it. *Q.E.D.*

Proof of Proposition 10: We first show

$$P[\mathbf{y}(t)] \geq_{st} P(\mathbf{y}\,|t = z) P(z \geq t) + P\left(\underline{Y}^{|L|}\right) P(z \ngeq t). \tag{34}$$

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

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

Under M0.S and U, the first equality holds by the Law of Total Probability. The empirical evidence reveals $P(\mathbf{y} \in U|z = t)$. Under the sustained assumptions, the first inequality follows by Corollary 22, and the last one holds as $\underline{Y}^{|L|}$ is a lower bound for any possible realization of the vector of potential outcomes. Since U was arbitrarily selected, our first claim follows by FOSD's Theorem(i).

The lower bound is sharp because, given the empirical evidence, the prior information is consistent with both $P[\mathbf{y}(t)\,|z \geq t] = P(\mathbf{y} \mid z = t)$ and $P[\mathbf{y}(t)\,|z \ngeq t] = P\left(\underline{Y}^{|L|}\right)$.

The proof for the upper bound is quite similar, so we omit it. *Q.E.D.*

Proof of Lemma 13: Consider the next mapping

$$\begin{aligned}
R_{t,k}: \quad & Y^{|N_k|} && \to Y^{|N_k|} \\
& \left(y_1, y_2, ..., y_{|N_k|}\right) && \to \left(f_{k1}, f_{k2}, ..., f_{k|N_k|}\right).
\end{aligned} \tag{35}$$

It is immediate to notice that the range of $R_{t,k}$ is as given. By construction, the solution set to the system of structural equations $\mathbf{f}_k$ coincides with the set of fixed points of $R_{t,k}$. Then the proof of Lemma 3 reduces to show that the set of fixed points of $R_{t,k}$ is non-empty and has a least and a greatest element.

Here $(Y^{|N_k|}, \geq)$ is a complete lattice for the ordering $\mathbf{y} \geq \mathbf{y}'$ if $y_n \geq y'_n$ for all $n \in N_k$, i.e., the coordinatewise (partial) order. By assumption M0.L, $R_{t,k}$ is monotone increasing. To see why, notice that if $\mathbf{y} \geq \mathbf{y}'$ then the induced distributions satisfy $P[y(t)] \geq_{st} P[y'(t)]$ and (by M0.L) the structural functions are higher at $\mathbf{y}$ than at $\mathbf{y}'$. Hence, the claim follows by Tarski's Fixed Point Theorem. *Q.E.D.*

Proof of Proposition 14: This proof is almost identical to the one of Proposition 4, thus we omit it. *Q.E.D.*

Proof of Lemma 15: Lemma 13 shows that if M0.L holds then $\varphi(k,t)$ is non-empty and has a least and a greatest element. Condition M1.L ensures the system of structural equations $[f_{kn}(.), n \in N_k]$ increases in t for any fixed $k \in K$. Then $R_{t,k}$, as defined in (35), increases in t on T, and Lemma 5 follows by Milgrom and Roberts' Theorem. *Q.E.D.*

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

Corollary 23 *Assume M0.L, M1.L and U are satisfied. Then* $P[y(t)|z \leq t] \geq_{st} P(y|z \leq t)$ *and* $P(y|z \geq t) \geq_{st} P[y(t)|z \geq t]$.

Proof of Corollary 23: To prove the first claim let $U \subset \mathbb{R}$ be an upper set, and let us consider the next three steps

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

Under M0.L and U, the two equalities hold by the Law of Total Probability. If we add M1.L, the inequality follows by Lemma 15 and FOSD's Theorem(i), as they imply that $P[y_k(t) \in U] \geq P[y_k(s) \in U]$ for all $t \geq s$. Since U was arbitrarily selected, the first claim follows again by FOSD's Theorem(i).

The proof for the second claim is similar, so we omit it. *Q.E.D.*

Proof of Proposition 16: This proof is almost identical to the one of Proposition 6, thus we omit it. *Q.E.D.*

The proof of Lemma 18 requires an additional result. Lemma 24 is based on Amir (2008).

Lemma 24 *Assume M0.L holds. Let $k, k' \in K$. If $F_k \geq F_{k'}$ according to Def. 17, then the least (greatest) distribution in $\varphi(t,k)$ is stochastically larger than the least (greatest) distribution in $\varphi(t,k')$ for all $t \in T$.*

Proof of Lemma 24: Lemma 13 shows that if M0.L holds then $\varphi(t,k)$ and $\varphi(t,k')$ are non-empty and have a least and a greatest element. The infimum fixed point of the mapping (35), which is the smallest distribution of the solution set $\varphi(t,k)$, is also a fixed point of $F_k[t, P(y)]$, as defined in (25). In addition, the smallest fixed point of the mapping (35) can be obtained by a successive iteration procedure starting at $(y_{kn}, n \in N_k) = \underline{Y}^{|N_k|}$. The same remarks hold for a type-k' group. If $F_k \geq F_{k'}$, then in each iteration the induced outcome distribution will be larger for the type-k group as compared to the type-k'. It follows that the least element in $\varphi(t,k)$ must be larger than the least element in $\varphi(t,k')$. A similar argument applies to the greatest distributions of the solution sets. *Q.E.D.*

Proof of Lemma 18: Let $s, s' \in T$, with $s \geq s'$, and let us assume that M0.L and U hold. Fix an upper set $U \subset \mathbb{R}$. By definition,

$$P[y(t) \in U \mid z = s] = \textstyle\sum_{k \in K} P[y_k(t) \in U] \, \pi_{k|z=s}. \tag{36}$$

Let us define $g(F_k, t, U) \equiv P[y_k(t) \in U]$. By Lemma 24, we get that $g(F_k, t, U) \geq g(F_{k'}, t, U)$ if $F_k \geq F_{k'}$. Substituting $P[y_k(t) \in U]$ by $g(F_k, t, U)$ in (36) we obtain

$$P[y(t) \in U \mid z = s] = \textstyle\sum_{k \in K} g(F_k, t, U) \, \pi_{k|z=s}. \tag{37}$$

Then $P[y(t) \in U \mid z = s]$ is the conditional expectation of a function that is increasing in F. By M2.L and FOSD's Theorem(ii) it follows that $P[y(t) \in U \mid z = s] \geq P[y(t) \in U \mid z = s']$. The result follows by FOSD's Theorem(i) as U was arbitrarily selected. *Q.E.D.*

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

Corollary 25 *Assume M0.L, M1.L and U are satisfied. Then $P[y(t) | z \geq t] \geq_{st} P(y|z=t)$ and $P(y|z=t) \geq_{st} P[y(t) | z \leq t]$.*

Proof of Corollary 25: To prove the first claim let $U \subset \mathbb{R}$ be an upper set, and let us consider the next four steps

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

Under M0.L and U, the first and the last equalities hold by the Law of Total Probability. If we add M2.L, the inequality follows by Lemma 18. The second equality is true as

$$\textstyle\sum_{s\in T} 1\left(s\geq t\right) P\left(z=s \mid z\geq t\right) = 1. \tag{38}$$

Since U was arbitrarily selected, the first claim is true by FOSD's Theorem(i). The proof for the second claim is similar, so we omit it. *Q.E.D.*

Proofs of Proposition 19: The proof of this proposition derives from Corollary 25. It is almost identical to the one of Proposition 6, thus we omit it. *Q.E.D.*

References

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

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

[3] Aradillas-Lopez, A., and E. Tamer (2008): "The Identification Power of Equilibrium in Simple Games,"*Journal of Business & Economic Statistics*, 26, 261-283.

[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] Becker, G., and K. Murphy (2000): "Social Economics," The Belknap Press of Harvard University Press.

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

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

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

[10] Echenique, F., and I. Komunjer (2009): "Testing Models with Multiple Equilibria by Quantile Methods," *Econometrica*, 77, 1281-1297.

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

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

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

[14] Hahn, J., and K. Hirano (2009): "Design of Randomized Experiments to Measure Social Interaction Effects," forthcoming in *Economics Letters*.

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

[16] Jovanovic, B. (1989): "Observable Implications of Models with Multiple Equilibria," *Econometrica*, 57, 1431-1437.

[17] Koopmans, T. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17, 125–144.

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

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

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

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

[22] Manski, C. (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14, 115-136.

[23] Manski, C. (2003): "Partial Identification of Probability Distributions," Springer Series in Statistics.

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

[25] Manski, C. (2010): "Identification of Treatment Response with Social Interactions," mimeo.

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

[27] Manski, C., and J. Pepper (2008): "More on Monotone Instrumental Variables," *The Econometric Journal*, 12, 200-216.

[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," Social Dynamics, Ed. by Durlauf, S., and H. Young, MIT Press, 45-82.

[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] Morris, S., and H. Shin (2000): "Global Games: Theory and Applications," mimeo.

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

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

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

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

[36] Sutton, J. (1999): "Technology and Market Structure," MIT Press.

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

[38] Tamer, E. (2009): "Partial Identification in Econometrics," mimeo.

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

[40] Topa, G. (2001): "Social Interactions, Local Spillovers, and Unemployment," *Review of Economic Studies*, 68, 261-295.

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

[42] Varian, H. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 99-110.

[43] Varian, H. (1984): "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579-597.

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