The Econometrics Journal

Econometrics Journal (2013), volume **16**, pp. S1–S23.
doi: 10.1111/j.1368-423X.2012.00368.x

Identification of treatment response with social interactions

CHARLES F. MANSKI[†]

[†]*Department of Economics and Institute for Policy Research, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 USA*

E-mail: cfmanski@northwestern.edu

First version received: February 2011; final version accepted: November 2011

Summary This paper studies identification of potential outcome distributions when treatment response may have social interactions. Defining a person's treatment response to be a function of the entire vector of treatments received by the population, I study identification when nonparametric shape restrictions and distributional assumptions are placed on response functions. An early key result is that the traditional assumption of individualistic treatment response is a polar case within the broad class of *constant treatment response* (CTR) assumptions, the other pole being unrestricted interactions. Important non-polar cases are interactions within reference groups and anonymous interactions. I first study identification under Assumption CTR alone. I then strengthen this Assumption to semi-monotone response. I next discuss derivation of these assumptions from models of endogenous interactions. Finally, I combine Assumption CTR with statistical independence of potential outcomes from realized *effective treatments*. The findings both extend and delimit the classical analysis of randomized experiments.

Keywords: *Analysis of treatment response, Partial identification, social networks.*

1. INTRODUCTION

This paper studies identification of treatment response in settings with social interactions, where personal outcomes may vary with the treatment of others. Social interactions are common within households, schools, workplaces and communities. Yet research on treatment response has mainly assumed that a person's outcome may vary only with his own treatment, not with those of other members of the population. Cox (1958) called this 'no interference between units'. Rubin (1978) called it the Stable Unit Treatment Value Assumption. I call it *individualistic treatment response* (ITR), to mark it as an assumption that restricts the form of treatment response functions.

The present analysis extends my earlier work on identification with individualistic response, reported in Manski (1990, 1997, 2003), Manski and Pepper (2000, 2009), and elsewhere. Here, as there, I ask what can be learned about outcomes under potential treatments when data on realized treatments and outcomes are combined with assumptions on treatment response. I emphasize relatively weak assumptions that may be credible in applications and, hence, primarily report findings of partial rather than point identification.

The concerns of this paper differ from those of previous research on identification of social interactions. Econometrics has long studied identification of structural models of endogenous interactions, which suppose that individual outcomes vary with the outcomes of other members

of the population. Research on this subject began with classical analysis of linear simultaneous equations and has evolved through the recent literature on identification of linear-in-means models (Manski, 1993) and discrete choice models (Tamer, 2003, and Brock and Durlauf, 2007). See Blume et al. (2011) for a review of much of the modern literature. From the perspective of models of endogeneous interactions, treatment response is the reduced-form solution to a structural system. Section 4 of the present paper elaborates on this matter.

1.1. Basic concepts and notation

To set the stage, I now specify basic concepts and notation that will be used throughout the paper. This requires a modest but essential extension of the setup used in my earlier work on identification of treatment response. A clear and concise formal language is essential to the analysis.

When response is assumed to be individualistic, each member j of population J has a response function $y_j(\cdot)$: $T \to Y$ mapping the mutually exclusive and exhaustive potential treatments $t \in T$ into outcomes $y_j(t) \in Y$. Person j has an observable realized treatment $z_j \in T$ and realized outcome $y_j \equiv y_j(z)$. Suppose that the cardinality of T is at most countable. This enables analysis that uses only elementary probability theory.

Let J be a probability space (J, Ω, P). Then observation of $(y_j, z_j; j \in J)$ reveals $P(y, z)$, the joint distribution of realized outcomes and treatments. A common research objective has been to learn about the outcome distribution $P[y(t)]$ that would occur if all persons were to receive a specified treatment t. Interest in $P[y(t)]$ is often motivated by a decision problem in which a planner chooses between the realized treatments and a policy that mandates treatment t. Then the planner wants to compare $P[y(t)]$ with $P(y)$.

Now remove Assumption ITR, so each person's outcome may vary with the treatments received by all members of the population. To express this, one extends the domain of the response function from T to the Cartesian product of T across the population; that is $T^J \equiv \times_{k \in J} T$. The response function becomes $y_j(\cdot)$: $T^J \to Y$, mapping treatment vectors $t^J \in T^J$ into outcomes $y_j(t^J) \in Y$. Here $t^J \equiv (t_k, k \in J)$ denotes a potential treatment vector specifying the treatment to be received by every member of the population. Person j has observable realized treatment $z_j \in T$ and realized outcome $y_j \equiv y_j(z^J)$, where $z^J \equiv (z_k, k \in J)$.

I will take the research objective to be inference on the outcome distribution $P[y(t^J)]$ that would occur if the population were to receive any potential treatment vector t^J. Interest in $P[y(t^J)]$ may be motivated by a decision problem in which a planner chooses between the realized treatments z^J and a policy that mandates treatment vector t^J. Then the planner wants to compare $P[y(t^J)]$ with $P(y)$. Instances of such planning problems are studied in Graham (2011) and Manski (2009, 2010).

In my earlier work studying prediction when all persons receive a common treatment, I have let t denote the specified common treatment. Here I let t be the random variable generated by t^J. Thus, $P(y, z, t)$ is the empirical distribution of $(y_j, z_j, t_j; j \in J)$. I will use τ rather than t to denote a specific element of T.

1.2. Identification of potential outcome distributions

Comparison of the setup with and without Assumption ITR makes plain that identification without the assumption presents a much more severe challenge than with it. Given Assumption

ITR and no further assumptions, the Law of Total Probability shows that $H\{P[y(t^J)]\}$, the identification region for $P[y(t^J)]$, is the set of distributions $[P(y|z = t)P(z = t) + \delta P(z \neq t), \delta \in \Delta_Y]$, where Δ_Y denotes the space of all probability distributions on Y. This region is a proper subset of Δ_Y if and only if $P(z = t) > 0$, which occurs when a positive fraction of the population receive the same realized and potential treatment. I have previously reported this simple result in Manski (2003, Chapter 7) and elsewhere for the case when t^J assigns a common treatment to all persons. Section 2 below extends it to the general case where t^J is a vector of treatments that may vary across the population.

Without Assumption ITR or another assumption restricting social interactions, $H\{P[y(t^J)]\}$ is the singleton $P(y)$ when $z^J = t^J$ and is the set Δ_Y of all distributions whenever $z^J \neq t^J$. Thus, the empirical evidence alone is uninformative about $P[y(t^J)]$ when t^J has any counterfactual component. Partial or point identification of $P[y(t^J)]$ may become feasible when the empirical evidence is combined with assumptions that restrict the shape of the response functions $[y_j(\cdot), j \in J]$ and/or the distribution $P[y(\cdot), z]$ of response functions and realized treatments. The resulting form of $H\{P[y(t^J)]\}$ depends on the assumptions imposed and the treatment vector t^J under consideration.

1.3. Organization of the paper

This paper is entirely general with respect to the potential treatment vector, but focuses on particular classes of assumptions. Sections 2 and 3 study two shape restrictions on response functions, *constant treatment response* (CTR) and *semi-monotone treatment response* (SMTR).

Assumption CTR posits that a person's outcome remains constant when t^J varies within specified subsets of T^J. I refer to these subsets of T^J as the person's *effective treatments*. Leading cases are assumptions asserting that interactions may occur within but not across known reference groups. Then a person's outcome remains constant when treatment varies outside his reference group. Assumption ITR is the special case where each person is his own reference group.

Assumption SMTR states that set T is partially ordered and that outcomes vary monotonically across ordered pairs of treatment vectors. Important subcases are *reinforcing* and *opposing* interactions. A reinforcing interaction occurs when a person's outcome increases both with the value of his own treatment and with the values of the treatments received by others in the reference group. An opposing interaction occurs when a person's outcome increases with the value of his own treatment but decreases with the values of the treatments received by others.

In the analysis of Sections 2 and 3, a response function is a primitive that maps treatment vectors into outcomes. Section 4 shows how Assumptions CTR and SMTR may be derived from models of endogenous interactions. The primitive in such a model is a system of structural equations that take the outcome of each person to be a function of the treatment vector and of the outcomes of other members of the population. The response functions $[y_j(\cdot), j \in J]$ are a derived concept.

Section 5 combines Assumption CTR with the distributional assumption that potential outcomes are statistically independent of realized effective treatments. I show that $P[y(t^J)]$ is point-identified if and only if every effective treatment that occurs with positive empirical probability in t^J also occurs with positive empirical probability in z^J. This requirement is transparent under Assumption ITR, but is more subtle with treatment interactions. I show that it generically fails to hold in two settings where interactions are global in nature.

Moreover, random assignment of treatments loses its classical identifying power in these settings.

2. CONSTANT TREATMENT RESPONSE

Constant-response assumptions assert that treatment response does not vary over specified sets of treatment vectors. Section 2.1 poses the assumption in abstraction and establishes its identifying power. Section 2.2 describes the leading case of interactions within reference groups. Section 2.3 specializes further to anonymous, distributional and functional interactions. Section 2.4 briefly discusses estimation with data on a random sample of the population.

It will be evident that constant-response assumptions have only limited identifying power. Nevertheless, they are highly important to analysis of treatment response. They are basic assumptions that provide a foundation on which further assumptions may be placed.

2.1. The assumption in abstraction

Consider person j. Let $c_j(\cdot)$: $T^J \to C_j$ be a function mapping treatment vectors onto a set C_j. A constant-response assumption asserts:

ASSUMPTION CTR.

$$c_j(t^J) = c_j(s^J) \Rightarrow y_j(t^J) = y_j(s^J). \tag{2.1}$$

Thus, j experiences the same outcome for all treatment vectors that form a level set of $c_j(\cdot)$. With this in mind, I shall refer to C_j as the set of *effective treatments* for person j.

The present definition of Assumption CTR generalizes one given in Manski and Pepper (2009), who named the concept in an individualistic-response context considering treatments with multiple components. There we defined CTR as an exclusion restriction asserting that a person's outcome remains constant when some treatment components are altered, holding the other components fixed. We did not, however, study the identifying power of the assumption.

Suppose that one observes $[c_j(\cdot), y_j, z_j; j \in J]$; thus, function $c_j(\cdot)$ is an observed covariate. Consider inference on $y_j(t^J)$. The researcher can infer $y_j(t^J)$ if and only if $c_j(z^J) = c_j(t^J)$. When this event occurs, z^J and t^J are effectively the same treatment from the perspective of person j, yielding the same outcome $y_j(t^J) = y_j(z^J) = y_j$. When $c_j(z^J) \neq c_j(t^J)$, Assumption CTR and observation of y_j do not reveal $y_j(t^J)$.

Now consider identification of $P[y(t^J)]$. By the Law of Total Probability,

$$\begin{aligned} P\,[y(t^J)] = {} & P\,[y(t^J)|c\,(z^J) = c(t^J)] \cdot P\,[c(z^J) = c(t^J)] \\ & + P\,[y(t^J)|c\,(z^J) \neq c\,(t^J)]\; P[c(z^J) \neq c(t^J)]. \end{aligned} \tag{2.2}$$

Here $P[c(z^J) = c(t^J)]$ is the fraction of the population for whom $[c(z^J) = c(t^J)]$, and $P[y(t^J)|c(z^J) = c(t^J)]$ is the distribution of outcomes conditional on this event. Observation of realized treatments reveals $P[c(z^J) = c(t^J)]$ and $P[c(z^J) \neq c(t^J)]$. Assumption CTR implies that $P[y(t^J)|c(z^J) = c(t^J)] = P[y|c(z^J) = c(t^J)]$. Observation of realized treatments and outcomes reveals $P[y|c(z^J) = c(t^J)]$ when $P[c(z^J) = c(t^J)] > 0$. The empirical evidence and Assumption CTR are uninformative about the counterfactual outcome distribution $P[y(t^J)|c(z^J) \neq c(t^J)]$. Hence, we have

PROPOSITION CTR. *Given Assumption CTR, the identification region for* $P[y(t^J)]$ *is*

$$H\{P[y(t^J)]\} = \{P[y|c(z^J) = c(t^J)] \cdot P[c(z^J) = c(t^J)] + \delta P[c(z^J) \neq c(t^J)], \delta \in \Delta_Y\}. \tag{2.3}$$

Observe that the size of $H\{P[y(t^J)]\}$ varies inversely with $P[c(z^J) = c(t^J)]$. The region is the singleton $P(y)$ when $P[c(z^J) = c(t^J)] = 1$. It expands as $P[c(z^J) = c(t^J)]$ decreases, and becomes uninformative when $P[c(z^J) = c(t^J)] = 0$.

2.2. *Interactions within reference groups*

2.2.1. Concepts and notation. It is common in applications to assume that each member of the population has a known reference group, with interactions occurring within but not across groups. A person's reference group may be assumed to be the members of his family, neighbourhood, school, workplace, or some other group, depending on the context. One might, for example, assume that treatment interactions may occur within but not across neighbourhoods.

Let $G(j) \subset J$ denote the reference group of person j, let $T^{G(j)} \equiv \times_{k \in G(j)} T$, and let $t^{G(j)} \equiv [t_k, k \in G(j)]$ be the sub-vector of t^J specifying the treatments assigned to the members of the group. For $j \in J$ and $t^J \in T^J$, let $C_j = T^{G(j)}$ and $c_j(t^J) = t^{G(j)}$. Then an effective treatment for person j is the sub-vector of treatments in his reference group. A person's outcome remains constant when treatments outside the group are altered, holding fixed the treatments of persons in the group.

As defined here, reference groups are person-specific, treatment-invariant and non-manipulable. Person-specific means that person k may be a member of person j's group but not vice versa. It is often assumed that reference groups are symmetric, with person k being a member of j's group if and only if j belongs to k's group. However, symmetry is not descriptive of all interactions. Asymmetry is expressed graphically in social network analysis when a directed path either directly or indirectly connects person k to j, but no directed path connects j to k.

While the notation $G(j)$ makes the reference group person-specific, it does not permit the group to be treatment-specific. I could expand the notation to $G(j, t_j)$, letting the group vary with person j's own potential treatment, or even to $G(j, t^J)$, letting it vary with the entire potential treatment vector. However, I will reserve the term *reference group* for cases in which the group is the same, whatever the treatment vector may be. The general idea of Assumption CTR covers cases in which the persons who interact vary across treatments, but I will not refer to these cases as interactions within reference groups.

Given that reference groups are treatment-invariant, they necessarily are non-manipulable. That is, a planner cannot use the treatments in T to change a person's reference group.

2.2.2. Analysis. Consider inference on $y_j(t^J)$. The researcher knows the value of $y_j(t^J)$ if and only if $z^{G(j)} = t^{G(j)}$. Applying (2.3), the identification region for $P[y(t^J)]$ is

$$H\{P[y(t^J)]\} = [P(y|z^G = t^G) \cdot P(z^G = t^G) + \delta P(z^G \neq t^G), \delta \in \Delta_Y]. \tag{2.4}$$

Two polar cases of interactions within reference groups are unrestricted interactions, where reference groups are the entire population, and individualistic treatment response, where reference groups are single persons. In the former case, $G(j) = J$ for all $j \in J$. Then (2.4) becomes

$$H\{P[y(t^J)]\} = [P(y|z^J = t^J) \cdot P(z^J = t^J) + \delta P(z^J \neq t^J), \delta \in \Delta_Y]. \tag{2.5}$$

All persons face the same realized treatment vector z^J. Hence, $P(z^J = t^J) = 1$ if $z^J = t^J$ and $P(z^J = t^J) = 0$ if $z^J \neq t^J$. Thus, $H\{P[y(t^J)]\} = P(y)$ if $z^J = t^J$ and $H\{P[y(t^J)]\} = \Delta_Y$ if $z^J \neq t^J$. This shows that observation of realized treatments and outcomes per se is uninformative about the outcome distribution with a counterfactual treatment vector.

When response is individualistic, $G(j) = j$ for all $j \in J$. Then (2.4) becomes

$$H\{P[y(t^J)]\} = [P(y|z = t) \cdot P(z = t) + \delta \cdot P(z \neq t), \delta \in \Delta_Y]. \tag{2.6}$$

Result (2.6) extends my earlier work on identification with individualistic treatment response. I have earlier reported (2.6) for the special case in which the potential treatment vector t^J assigns the same treatment to all members of the population; see, for example, Manski (2003, Chapter 7). Then the treatment t on the right-hand side of (2.6) is the common treatment, say τ, and $t^J = (\tau, \tau, \ldots, \tau)$. Now (2.6) holds in the general case where t^J may be any treatment vector, possibly assigning different treatments to different persons.

The size of region (2.6) varies inversely with the magnitude of $P(z = t)$; that is, with the fraction of the population who have the same realized and potential treatments. Point-identification occurs if and only if $P(z = t) = 1$, which requires that $z^J = t^J$ if J is a countable population and permits deviation of z^J from t^J only on a negligible set of persons when J is a continuum. Region (2.6) grows smoothly from the singleton $P(y)$ to the entire space Δ_Y as $P(z = t)$ decreases from 1 to 0. This contrasts sharply with the unrestricted-interaction region (2.5), which equals Δ_Y whenever $P(z = t) < 1$.

2.3. Anonymous, distributional and functional interactions

2.3.1. Anonymous interactions. Region (2.4) characterized identification under the sole assumption that interactions occur within reference groups. The assumption that an interaction is *anonymous* goes further by asserting that the outcome of person j is invariant with respect to permutations of the treatments received by other members of his group. This further assumption is empty when the reference group contains only one person other than j, but is meaningful when the reference group is larger.

Consider, for example, vaccination of some children in a community. When considering illness from an infectious disease, one might think it credible to take each child's reference group to be the set of children who attend the same school. One might additionally think it credible to assume that each child's illness outcome may depend on his own vaccination treatment and on the number of children vaccinated in his school, but not on the identities of the other schoolmates vaccinated.

Formally, let $G(j)/j$ denote the reference group exclusive of person j himself and let $\pi[t^{G(j)/j}]$ denote the set of permutations of treatment vector $t^{G(j)/j}$. If $G(j)/j$ is empty, define $t^{G(j)/j}$ to be empty as well. Applying (2.3), the identification region is

$$\begin{aligned} H\{P[y(t^J)]\} = \{&[P(y|z = t, z^{G/} \in \pi(t^{G/})] \cdot P[z = t, z^{G/} \in \pi(t^{G/})] \\ &+ \delta \cdot P[z \neq t \text{ or } z^{G/} \notin \pi(t^{G/})], \delta \in \Delta_Y\}. \end{aligned} \tag{2.7}$$

This region is a subset of the region (2.4) obtained when it was assumed only that interactions occur within reference groups. Here the researcher knows the value of $y_j(t^J)$ when the event $[z_j = t_j, z^{G(j)/j} \in \pi(t^{G(j)/j})]$ occurs. Previously, $y_j(t^J)$ was known when $z^{G(j)} = t^{G(j)}$. The latter event implies the former one.

2.3.2. Distributional interactions. A *distributional* interaction strengthens an anonymous one by supposing that treatment response is invariant with respect to the size of the reference group and to permutations of the treatments received by other members of the group. Thus, the outcome of person j may vary only with his own treatment and with the empirical distribution of treatments among other members of the group.

Let Δ_T denote the space of all distributions on T. For $t^J \in T^J$, let $Q(t^{G(j)/j})$ be the empirical distribution of the treatments in $t^{G(j)/j}$. Thus, for $\tau \in T$, $Q(t^{G(j)/j} = \tau)$ is the fraction of the persons in $G(j)/j$ who receive treatment τ when t^J is the potential treatment vector. If group $G(j)/j$ is empty, define $Q(t^{G(j)/j}) = \phi$, where ϕ denotes the empty set. Then the identification region is

$$H\{P[y(t^J)]\} = \{[P(y|z = t,\, Q(z^{G/}) = Q(t^{G/})] \cdot P[z = t,\, Q(z^{G/}) = Q(t^{G/})] + \delta \cdot P(z \neq t \text{ or } Q(z^{G/}) \neq Q(t^{G/}), \delta \in \Delta_Y\}. \tag{2.8}$$

This region is a subset of the region (2.7) obtained when it was assumed only that interactions are anonymous. Here the researcher knows the value of $y_j(t^J)$ when the event $[z_j = t_j, Q(z^{G(j)/j}) = Q(t^{G(j)/j})]$ occurs. Previously, $y_j(t^J)$ was known when $[z_j = t_j, z^{G(j)/j} \in \pi(t^{G(j)/j})]$. The latter event implies the former one.

2.3.3. Functional interactions. Applied researchers often assume not only that interactions are distributional but also that $Q(t^{G(j)/j})$ affects outcomes solely through some functional of the distribution, say $F(t^{G(j)/j})$. A leading case is the *mean interaction*, where treatments are real-valued and $F(t^{G(j)/j}) = E(t^{G(j)/j})$, the empirical mean of the treatments in $t^{G(j)/j}$. A mean interaction is equivalent to a distributional interaction when set T has two treatments. It is a stronger assumption when there are more than two.

Another case of interest is the *supremum interaction*, where treatments are ordered and $F(t^{G(j)/j}) = \sup(t^{G(j)/j})$. Suppose that a treatment is information communicated within a reference group. Suppose that information treatments are ordered, with $\tau > \tau'$ meaning that a person with treatment τ receives all of the information in τ', plus some more. Then communication within the group ensures that person j effectively receives treatment $\sup(t^{G(j)})$.

Whatever functional F may be, the identification region is

$$H\{P[y(t^J)]\} = \{[P(y|z = t,\, F(z^{G/}) = F(t^{G/})] \cdot P[z = t,\, F(z^{G/}) = F(t^{G/})] + \delta \cdot P[z \neq t \text{ or } F(z^{G/}) \neq F(t^{G/})], \delta \in \Delta_Y\}. \tag{2.9}$$

This region is a subset of the region (2.8) obtained when it was assumed only that interactions are distributional. Here the researcher knows the value of $y_j(t^J)$ when the event $[z_j = t_j, F(z^{G(j)/j}) = F(t^{G(j)/j})]$ occurs. Previously, $y_j(t^J)$ was known when $[z_j = t_j, Q(z^{G(j)/j}) = Q(t^{G(j)/j})]$. The latter event implies the former one.

2.4. Estimation with data on a random sample of the population

Although this paper is about identification, I would be remiss to entirely ignore estimation with sample data. When analysing identification, I assume that one observes $[c_j(z^J), c_j(t^J), y_j]$ for every member of the population. Now suppose that one draws a random sample of N persons, say J_N, and observes $[c_j(z^J), c_j(t^J), y_j]$ only for $j \in J_N$. Let P_N denote the empirical distribution of J_N.

Then one may consistently estimate identification region (2.3) by its sample analogue

$$H\{P_N[y(t^J)]\} \equiv \{P_N[y|c(z^J) = c(t^J)] \cdot P_N[c(z^J) = c(t^J)] + \delta \cdot P_N[c(z^J) \neq c(t^J)], \delta \in \Delta_Y\}. \tag{2.10}$$

Statistical inference on parameters of $P[y(t^J)]$ may be performed using the methods developed over the past decade for settings with partial identification. For example, the method of Imbens and Manski (2004) may be used to compute confidence intervals for the mean outcome $E[y(t^J)]$.

The only subtlety of estimation with sample data is that application of (2.10) requires observation of sample members' realized and potential *effective* treatments, not just their own treatments. Excepting the special case of individualistic treatment response, the effective treatments of sample members generically depend on the treatments received by non-sample members. One must observe the treatments of all persons in a sample member's reference group even though some of these persons may not themselves be sample members. However, one does not need to observe the outcomes realized by non-sample members.

Observation of the treatments received by non-sampled persons in reference groups is realistic in some applied settings. Realized treatments for the entire population may be set by known regulations, may be observable prices, or may be recorded in accessible administrative databases. When population treatment data are not available in these ways, a survey researcher might ask sample members to report the treatments received by their reference groups. For example, if reference groups are families, a sample member might be asked not only to report her own treatment but also those received by other family members.

3. SEMI-MONOTONE TREATMENT RESPONSE

The constant-response assumptions considered in Section 2 were nested. Functional interactions strengthen distributional interactions, which strengthen anonymous interactions, which in turn strengthen interactions within a reference group. The various identification regions presented above were correspondingly nested sets. However, even the strongest of these assumptions has only limited identifying power.

Smaller identification regions emerge if the assumption that response is constant within level sets of $c(\cdot)$ is combined with the assumption that response is semi-monotone across level sets. Section 3.1 poses the assumption in abstraction and establishes its identifying power. Sections 3.2 and 3.3 present the polar cases of reinforcing and opposing interactions. As a further case study, Section 3.4 introduces monotone metric interactions. Section 3.5 uses vaccination as treatment for infectious disease to illustrate application of Assumptions CTR and SMTR.

3.1. The assumption in abstraction

Suppose that some constant-response assumption has been imposed. Considering person j, let the set C_j of effective treatments be partially ordered. Thus, given a pair of distinct values $(c, c') \in C_j \times C_j$, either $c \prec c'$ or $c \succ c'$ or (c, c') are unordered, in which case I write $c \not\gtrless c'$. Let the outcome space Y be a subset of the real line. Let t^J and s^J be two potential treatment vectors. The assumption of semi-monotone response asserts

ASSUMPTION SMTR.

$$c_j(t^J) \geq c_j(s^J) \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.1}$$

This assumption encompasses Assumption CTR, as the equality $c_j(t^J) = c_j(s^J)$ is equivalent to the two inequalities $c_j(t^J) \geq c_j(s^J)$ and $c_j(t^J) \leq c_j(s^J)$.

Considering individualistic response, Manski (1997), Proposition S1 showed that observation of realized treatments and outcomes combined with assumption SMTR yields a sharp bound on any parameter of the outcome distribution that respects stochastic dominance. It is straightforward to extend the argument to settings with social interactions.

Consider the outcome of person j when the treatment vector is t^J. Let $y_0 \equiv \inf Y$ and $y_1 \equiv \sup Y$ be the logical lower and upper bounds on outcomes. Combining the empirical evidence with assumption SMTR yields this sharp bound on $y_j(t^J)$:

$$\begin{aligned} &c_j(t^J) < c_j(z^J) \Rightarrow y_0 \leq y_j(t^J) \leq y_j \\ &c_j(t^J) = c_j(z^J) \Rightarrow y_j(t^J) = y_j \\ &c_j(t^J) > c_j(z^J) \Rightarrow y_j \leq y_j(t^J) \leq y_1 \\ &c_j(t^J) \varnothing c_j(z^J) \Rightarrow y_0 \leq y_j(t_J) \leq y_1. \end{aligned} \tag{3.2}$$

Let $y_{jL}(t^J)$ and $y_{jU}(t^J)$ denote the lower and upper bounds on $y_j(t^J)$ stated in (3.2). Given that (3.2) holds for all $j \in J$, the population distribution of $y_{jU}(t^J)$ stochastically dominates that of $y(t^J)$, which in turn dominates that of $y_{jL}(t^J)$. Given that (3.2) exhausts the available information, we have

PROPOSITION SMTR. *Given Assumption SMTR, the identification region for* $P[y(t^J)]$ *is*

$$H\{P[y(t^J)]\} = \{\delta \in \Delta_Y \colon P[y_U(t^J)] \geq_{sd} \delta \geq_{sd} P[y_L(t^J)]\}. \tag{3.3}$$

Here $\geq_{sd}$ denotes the weak stochastic dominance relationship.

Let D be any parameter of the outcome distribution that respects stochastic dominance. For example, D may be a quantile or the mean of an increasing function of the outcome. Region (3.3) immediately yields this sharp bound on $D[y(t^J)]$:

$$D[y_L(t^J)] \leq D[y(t^J)] \leq D[y_U(t^J)]. \tag{3.4}$$

Considering individualistic response, Manski (1997), Corollaries S1.1–S1.3 gave the explicit form of bound (3.4) for various D-parameters. The extensions to settings with social interactions are immediate. In particular, the result for the mean outcome $E[y(t^J)]$ is

$$\begin{aligned} y_0 \cdot P[c(t^J) < c(z^J) \cup c(t^J) \varnothing c(z^J)] \; + \; E[y | c(t^J) \geq c(z^J)] \cdot P[c(t^J) \geq c(z^J)] \leq E[y(t^J)] \\ \leq y_1 \cdot P[c(t^J) > c(z^J) \cup c(t^J) \varnothing c(z^J)] \; + \; E[y | c(t^J) \leq c(z^J)] \cdot P[c(t^J) \leq c(z^J)]. \end{aligned} \tag{3.5}$$

3.2. Reinforcing interactions

I defined reinforcing interactions verbally in the Introduction. Formally, let T be partially ordered. Let j have reference group $G(j)$ and let $T^{G(j)}$ inherit the partial ordering on T. That is, given two treatment vectors t^J and s^J, let $c_j(t^J) \geq c_j(s^J)$ mean that $[t_k \geq s_k$, all $k \in G(j)]$. A reinforcing

interaction occurs when

$$[t_k \geq s_k, \text{all } k \in G(j)] \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.6}$$

When (3.6) holds, the response function increases with the treatment that person j receives and with the treatments of other members of the reference group. Thus, the treatments received by others reinforce a person's own treatment.

Consider, for example, vaccination against an infectious disease. Vaccination of person j may reduce the chance that this person will become ill, and vaccination of other persons may also reduce his probability of illness, reinforcing the effect of own vaccination. Or consider provision of tutoring to a class of students. Tutoring student j may increase his achievement, and tutoring other students in the class may help him achieve as well.

3.2.1. Reinforcing distributional interactions. The definition of a reinforcing interaction stated in (15) orders treatment vectors only when every member of the reference group of person j receives at least as large a treatment with $t^{G(j)}$ as with $s^{G(j)}$. Suppose that the social interaction is distributional. Then we may strengthen the idea of a reinforcing interaction by letting $c_j(t^J) \geq c_j(s^J)$ mean that $[t_j \geq s_j, Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})]$. A reinforcing distributional interaction occurs when

$$[t_j \geq s_j, Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})] \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.7}$$

The event $[t_k \geq s_k, \text{all } k \in G(j)]$ implies the event $[t_j \geq s_j, Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})]$. Hence, a reinforcing distributional interaction orders all treatment pairs that are ordered by a reinforcing interaction, and possibly more. It follows that the present identification region for $P[y(t^J)]$ is a subset of the one obtained when the interaction is only assumed reinforcing.

When person j's reference group is large, the stochastic dominance inequality $Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})$ appearing in (3.6) is approximately the same as $Q(t^{G(j)}) \geq_{sd} Q(s^{G(j)})$, which includes j in the group distribution. The latter inequality is simpler to use in some applications.

3.2.2. Reinforcing D-interactions. A yet smaller identification region results when a distributional interaction is assumed to be a functional interaction, where the functional is a parameter D that respects stochastic dominance. Now take $c(t^J) \geq c(s^J)$ to mean that $[t_j \geq s_j, D(t^{G(j)/j}) \geq D(s^{G(j)/j})]$. A reinforcing D-interaction occurs when

$$[t_j \geq s_j, D(t^{G(j)/j}) \geq D(s^{G(j)/j})] \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.8}$$

The event $[t_j \geq s_j, Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})]$ implies the event $[t_j \geq s_j, D(t^{G(j)/j}) \geq D(s^{G(j)/j})]$. Hence, a reinforcing D-interaction orders all treatment pairs that are ordered by a reinforcing distributional interaction, and possibly more. Therefore, the present identification region for $P[y(t^J)]$ is a subset of the one obtained with a reinforcing distributional interaction.

3.3. Opposing interactions

An opposing interaction reverses the direction of the inequality relating a person's outcome to the treatments received by other members of his reference group. An opposing interaction occurs when

$$[t_j \geq s_j, \{t_k \leq s_k, k \in G(j)/j\}] \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.9}$$

When (3.9) holds, the response function increases with the treatment that person j receives and decreases with the treatments of other members of the reference group. Thus, the treatments received by others act in opposition to a person's own treatment.

Consider, for example, training that provides occupation-specific human capital. Training person j may increase the chance that this person finds employment in the occupation. Training other persons increases the supply of trained labour and, hence, may decrease the probability that person j finds employment.

Opposing distributional and D-interactions are defined in the obvious way. The former occurs when

$$[t_j \geq s_j, Q(s^{G(j)/j}) \geq_{sd} Q(t^{G(j)/j})] \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.10}$$

The latter occurs when

$$[t_j \geq s_j, D(s^{G(j)/j}) \geq D(t^{G(j)/j})] \Rightarrow y_j(t^J) \geq y_j(s^J). \tag{3.11}$$

3.4. Monotone metric interactions

Whereas much empirical research assumes that social interactions are anonymous, some studies position persons spatially on social networks and suppose that the strength of interaction between two persons decreases with the distance between them. Distance may be purely geographic—a person may interact most closely with members of his household and successively less closely with residents of his block, neighbourhood, city and nation. Or the metric measuring social distance may jointly consider geography, occupation, political views, religion and other attributes.

The ordinal essence of this idea may be formalized as a case of semi-monotone treatment response. To begin, partially order the members $G(j)$ of person j's reference group in terms of their distance from j. Thus, list j as the first member of the group, with distance zero from himself. Next list the member of $G(j)/j$ closest to j, and so on. If multiple persons are equidistant from j, they are unordered relative to one another. Let $G^*(j)$ denote the partially ordered version of $G(j)$.

Now compare response to certain permutations of a specified treatment vector. In particular, consider re-allocations in which two group members who differ in distance from j exchange treatments. Whereas the person closer to j originally was to receive the smaller of their two treatments, the exchange makes this person receive the larger treatment. I will say that a *monotone metric interaction* occurs if any such re-allocation weakly increases the outcome experienced by j.

Formally, let $s^{G^*(j)}$ be a specified vector of reference-group treatments. Let $c_j(t^J) \geq c_j(s^J)$ mean that $t^{G^*(j)}$ is a permutation of $s^{G^*(j)}$ that exchanges the treatments of two ordered group members, say k and m, with $k < m$, $s_k < s_m$, $t_k = s_m$ and $t_m = s_k$. A monotone metric interaction occurs if $y_j(t^J) \geq y_j(s^J)$.

3.4.1. Example: incentives for clean-burning fuels. Consider a geographic region subject to air pollution created by residential burning of fossil fuels for heating and cooking. Let treatments be taxes or regulations that provide incentives for use of clean-burning fuels. Let the outcomes of interest be the health status of the residents of the region.

In this context, one may find it credible to assume that treatment interactions are both reinforcing and monotone metric. Reinforcing means that region-wide strengthening of the incentives for use of clean fuels improves the health status of all persons in the region.

Monotone metric means that the health status of person j improves with a reallocation of heterogeneous incentives within the region, strengthening them for persons who live close to j and correspondingly weakening them for persons who live far from j.

The assumption of a reinforcing interaction has relatively transparent credibility. A region-wide strengthening of incentives should induce all residents to use cleaner fuels and, hence, reduce pollution region wide. Supposing that health status decreases with exposure to pollution, the result should be a region-wide increase in health status.

Assessment of the monotone-metric assumption is more subtle. The envisioned reallocation of incentives should induce use of cleaner fuels by residents who live close to j and dirtier fuels by those who live far from j. Suppose that pollution decays with distance from the source of the burning. Suppose also that persons who live at different distances from j respond similarly to incentives. Then the reallocation should yield a net reduction in the exposure of person j to pollution.

3.5. Vaccination against infectious disease

This section uses a simple scenario of vaccination against infectious disease to illustrate some of the findings of Sections 2 and 3. Let $T = \{0, 1\}$, with $(\tau = 1)$ denoting vaccination and $(\tau = 0)$ no vaccination. Let the outcome of interest be a binary measure of health status, with y $= 1$ if a person remains in good health and $y = 0$ if he becomes ill with the disease. Then sufficient statistics for the distribution $P(y, z)$ of realized treatments and outcomes are $P_{11} \equiv P(y = 1|z = 1)$, $P_{10} \equiv P(y = 1|z = 0)$ and $p \equiv P(z = 1)$. The realized probability of good health is $P(y = 1) = pP_{11} + (1 - p)P_{10}$.

Consider a potential treatment vector t^J that increases the population rate of vaccination from p to some $q > p$. In particular, t^J sets $t_j = 1$ for all persons with $z_j = 1$ and for some of those with $z_j = 0$.

The objective is to learn $P[y(t^J) = 1]$. One may interpret $P[y(t^J) = 1]$ retrospectively as the population rate of good health that would have occurred if vaccination had been performed for all persons who were actually vaccinated and for a specified subset of those who were not. Or one may interpret $P[y(t^J) = 1]$ prospectively as the health rate that will occur if treatment vector t^J is applied to a new population that is identical in composition to the study population.

The identification region for $P[y(t^J) = 1]$ depends on the maintained assumptions. I first assume that treatment is individualistic and then add the assumption of monotone treatment response, in the sense that vaccination never lowers health status and may improve it. I next consider a reinforcing interaction within the population as a whole.

3.5.1. Individualistic response. Suppose that a person's health status depends only on his own treatment. This assumption is not credible when considering an infectious disease, but I begin with it to provide contrast with the findings when social interactions are considered. The identification region under Assumption ITR was given in (2.6). With a binary outcome, (2.6) becomes the interval

$$H\{P[y(t^J) = 1]\} = [P(y = 1|z = t) \cdot P(z = t), P(y = 1|z = \mathrm{t}) \cdot P(z = t) + P(z \neq t)]. \quad (3.12)$$

Consider the fraction $P(z = t)$ of the population whose realized and potential treatments coincide. This includes the group of size p who realize treatment 1, all of whom would continue to receive it under t^J. It also includes the group of size $1 - q$ which realizes treatment 0 and would

continue to receive it under t^J. Hence, $P(z = t) = p + 1 - q$. Correspondingly, $P(z \neq t) = q - p$. Observe that $P(z \neq t)$ is the width of the interval on the right-hand side of (3.12).

Consider $P(y = 1|z = t)$, the probability of good health in the group with $(z = t)$. It is the case that

$$\begin{aligned} P(y = 1|z = t) &= P(y = 1|z = t = 1) \cdot P(z = 1|z = t) \; + \; P(y = 1|z = t = 0) \cdot P(z = 0|z = t) \\ &= P_{11}[p/(p + 1 - q)] + P(y = 1|z = \mathrm{t} = 0) \cdot [(1 - q)/(p + 1 - q)]. \end{aligned} \tag{3.13}$$

The first equality applies the Law of Total Probability. Our derivation of P(z = t) shows that $P(z = 1|z = t) = p/(p + 1 - q)$ and $P(z = 0|z = t) = (1 - q)/(p + 1 - q)$. We have $P(y = 1|z = t = 1) = P_{11}$ because $z = 1 \Rightarrow t = 1$. We have not yet encountered $P(y = 1|z = t = 0)$, the probability of good health in the group who realized treatment 0 and who would continue to receive 0 under t^J. This conditional probability is revealed by the empirical evidence once t^J is specified. Hence, all quantities on the right-hand side of (3.13) are known.

3.5.2. Monotone-individualistic response. Continue to suppose that a person's health status depends only on his own treatment. Also suppose that treatment response is monotone in the sense that $y_j(1) \geq y_j(0)$ for all $j \in J$. This is credible in settings where vaccines do not have adverse side effects. Then vaccination never makes a person worse off and may improve his health status.

The identification region is given by (3.5), which reduces in the present case to

$$H\{P[y(t^J) = 1]\} = [P(y = 1|t \geq z) \cdot P(t \geq z), P(t > z) + P(y = 1|t \leq z) \cdot P(t \leq z)]. \tag{3.14}$$

The inequality $t^J \geq z^J$ holds in this illustration. Hence, $P(t \geq z) = 1$, $P(t > z) = q - p$ and $P(t \leq z) = P(t = z) = p + 1 - q$. Moreover, $P(y = 1|t \geq z) = P(y = 1)$ and $P(y = 1|t \leq z) = P(y = 1|t = z)$, whose value was derived in (3.13). The result is

$$H\{P[y(t^J) = 1]\} = [P(y = 1), q - p + P(y = 1|t = z) \cdot (p + 1 - q)]. \tag{3.15}$$

The lower bound is larger than the one obtained using Assumption ITR alone. The upper bound is the same as with Assumption ITR alone.

3.5.3. Reinforcing interactions. Now suppose that a person's health status may depend on the entire population vector of vaccination treatments. In the absence of any restrictions, $H\{P[y(t^J) = 1]\}$ is the [0, 1] interval. However, it is reasonable to assume that interactions are reinforcing.

Application of (3.5) in the present setting gives

$$\begin{aligned} H\{P\,[y(t^J) = 1]\} = [&P(y = 1|t^J \geq z^J) \cdot P(t^J > z^J),\, P(t^J > z^J \cup t^J \varnothing z^J) \\ &+ P(y = 1|t^J \leq z^J) \cdot P(t^J \leq z^J)]. \end{aligned} \tag{3.16}$$

We have $t^J > z^J$ by design. Hence, (3.16) reduces to

$$H\{P[y(t^J) = 1]\} = [P(y = 1),\, 1]. \tag{3.17}$$

The lower bound is the same as with the assumption of monotone-individualistic response. The upper bound is 1 because a reinforcing interaction permits the possibility that increasing the vaccination rate completely eliminates disease transmission.

4. DERIVATION OF ASSUMPTIONS CTR AND SMTR FROM MODELS OF ENDOGENOUS INTERACTIONS

I have thus far viewed a response function as a primitive concept mapping population treatments into personal outcomes. Hence, I posed Assumptions CTR and SMTR as direct restrictions on this function. Researchers often model the social mechanism mapping treatments into outcomes. Economists relate outcomes to choices made by members of the population and suppose that these choices express individual optimizing behaviour and the equilibria of games. Epidemiologists studying treatment of infectious diseases study models of infection and contagion.

From the perspective of such models, response functions are not primitives but rather are quantities whose properties stem from the mechanism under study. Hence, Assumptions CTR and SMTR should be derived rather than posed directly. In this section, I consider Assumptions CTR and SMTR from the viewpoint of econometric models of endogenous interactions.

4.1. Concepts and notation

The primitive in a model of endogenous interactions is a system of *structural equations* that takes the outcome of each person to be a function of population treatments and outcomes. Formally, one supposes that the potential outcome vector $y^J(t^J) \equiv [y_j(t^J),\ j \in J]$ solves the structural equations

$$y_j(t^J) = f_j[t_j, t^{J/j}, y^{J/j}(t^J)], j \in J. \tag{4.1}$$

Here $t^{J/j} \equiv (t_k, k \in J, k \neq j)$ and $y^{J/j}(t^J) \equiv [y_k(t^J), k \in J, k \neq j]$ are the treatment and outcome vectors for the population exclusive of person j.

The *structural function* $f_j(\cdot)$ permits $y_j(t^J)$ to be determined by j's own treatment as well as by the treatments and outcomes of other members of the population. The term *endogenous interaction* describes $y^{J/j}(t^J)$ as an argument of $f_j(\cdot)$. If $y^{J/j}(t^J)$ were not an argument, $f_j(\cdot)$ would simply be the person's response function. The presence of $y^{J/j}(t^J)$ makes (4.1) a system of simultaneous equations.

The term *exogenous interaction* describes $t^{J/j}$ as an argument of $f_j(\cdot)$. It is important to differentiate between $t^{J/j}$ as an argument of $f_j(\cdot)$ and as an argument of the response function $y_j(\cdot)$. In a *pure* endogenous interactions model, $t^{J/j}$ is not an argument of $f_j(\cdot)$. Yet $t^{J/j}$ may still affect response through the system of simultaneous equations.

An outcome vector $y^J(t^J)$ that solves (4.1) is said to be a *reduced form* of the structural equations. A model is *complete* if (4.1) has a unique solution for all feasible structural functions. A model is incomplete if (4.1) may have multiple solutions or no solutions. Incomplete models are not abnormal. Structural equations with multiple solutions may describe games with multiple equilibria. Those with no solutions may describe games with no equilibria. See, for example, Brock and Durlauf (2001) and Tamer (2003).

Here are two examples of endogenous interactions models.

EXAMPLE 4.1. Consider illness from an infectious disease. Let the outcome of interest measure health status. Let the treatment be vaccination status. Illness may vary with a person's own vaccination status, the status of others (an exogenous interaction) and the illness outcomes of others (an endogenous interaction).

EXAMPLE 4.2. Consider labour supply in a population of husband-wife couples. Let the outcome of interest be hours worked. Let the treatment be a person's market wage. One may think it reasonable to assume that labour-supply interactions occur only within couples, not between them. Within each couple, a person's labour supply may vary with his or her own wage, the wage of the spouse (an exogenous interaction) and the spouse's labour supply (an endogenous interaction).

Sections 4.2 and 4.3 relate identification of structural functions and response functions in abstraction. Section 4.4 considers derivation of Assumption CTR from models of endogenous interactions, focusing on the determination of reference groups. Section 4.5 considers derivation of Assumption SMTR from such models, focusing on reinforcing interactions.

4.2. Identification of structural functions and response functions

Econometricians have long studied identification of structural functions. Observation of realized treatments and outcomes reveals that

$$y_j = f_j(z_j, z^{J/j}, y^{J/j}), \quad j \in J. \tag{4.2}$$

Thus, the empirical evidence pins down one point on the structural function of each population member. Econometricians have combined this evidence with restrictions on f^J, the objective being identification of features of these structural functions. Classical econometric analysis of structural equations begins by assuming that each f_j is a linear function of its arguments and then adds further assumptions to achieve point identification of the entire vector f^J. See Goldberger (1991) for a textbook exposition. Recent analysis of so-called linear-in-means models proceeds similarly, with particular attention to separation of exogenous and endogenous interactions. See Manski (1993).

Our concern is identification of response functions, not structural functions. From this perspective, it is not important to determine the mechanism through which population treatments affect personal outcomes. A model of endogenous interactions is useful if credible assumptions on $f^J(\cdot)$ imply restrictions on $y^J(\cdot)$. Our particular concern is identification of $P[y(t^J)]$, the empirical distribution of $y^J(t^J)$.

While inference on structural functions has been the dominant theme of econometric research, econometricians have occasionally observed that the objective may be to infer response functions rather than structural functions. Arthur Goldberger put it this way in his ET Interview (Kiefer and Goldberger, 1989, p. 150): 'Well, that's one position, that the entire content in a structural model is simply in the restrictions, if any, that it implies on the reduced form—that's true. That gives priority to the reduced form'.

The relationship between identification of structural functions and response functions is straightforward when the structural functions are linear in treatments and outcomes. Then solution of the structural equations shows that response functions are linear in treatments. The parameters of response functions are many-to-one functions of the parameters of the structural functions. Hence, identification of response functions is a simpler objective than identification of structural functions.

Outside of linear models, the relationship between identification of structural functions and reduced forms is largely an open question. This question is much too broad for a comprehensive analysis here, but I will make a small start. Section 4.3 calls attention to the fact that the

relationship between structural and response functions differs qualitatively in complete and incomplete models.

4.3. Complete and incomplete models

Given a complete model, identification of $P[y(t^J)]$ is logically no more difficult than identification of f^J, and may be easier. With an incomplete model, identification of $P[y(t^J)]$ may be more difficult than identification of f^J. I explain here. In what follows, Φ denotes the identification region for f^J.

Suppose first that the model is complete; thus, (4.1) has a unique solution for each element of Φ. For each $f^J \in \Phi$, Let $y^J(t^J, f^J)$ denote this solution. Then the identification region for $y^J(t^J)$ is $[y^J(t^J, f^J), f^J \in \Phi]$. The cardinality of this set cannot be larger than that of Φ, and it may be smaller. In particular, a model that point-identifies f^J necessarily point-identifies $y^J(t^J)$. Knowledge of $y^J(t^J)$ implies knowledge of $P[y(t^J)]$. Hence, identification of $P[y(t^J)]$ is logically no more difficult than identification of f^J.

Next suppose that the model is incomplete, with at least one solution to (4.1) for every feasible value of f^J and multiple solutions for some values. For each $f^J \in \Phi$, let $\Upsilon(t^J, f^J)$ denote the set of solutions to the structural equations. Then the identification region for $y^J(t^J)$ is $\{\Upsilon(t^J, f^J), f^J \in \Phi\}$. In general, the cardinality of this set may be larger or smaller than that of Φ. It necessarily is larger when the model point-identifies f^J. Then f^J is known, but $\Upsilon(t^J, f^J)$ contains multiple elements. Hence, $H\{P[y(t^J)]\}$ may contain multiple elements.

Finally, consider an incomplete model having no solution to (4.1) for some $f^J \in \Phi$. There are two ways to interpret non-existence of a solution. One might interpret it to mean that the value of f^J under consideration is not feasible. Then one should eliminate this value from Φ. This done, non-existence of a solution logically cannot occur.

Alternatively, one might interpret non-existence to mean that the endogenous-interactions model is silent on $y^J(t^J)$. Then the model has no identifying power for $P[y(t^J)]$. This interpretation is reasonable in analysis of games, where a finding that no equilibrium exists implies that the specified equilibrium concept makes no prediction about the actions chosen by players.

4.4. Structural and response reference groups

With the above as background, we may ask what specific shape restrictions on structural functions imply about the shape of response functions. This section considers how assumptions specifying structural reference groups determine response groups. The next considers the implications of structural monotonicity for response monotonicity.

Researchers modeling endogenous interactions regularly assume that interactions occur within known reference groups. Let $F(j)$ denote the structural reference group of person j. Then a model may assume that

$$y_j(t^J) = f_j[t_j, t^{F(j)/j}, y^{F(j)/j}(t^J)], \quad j \in J. \tag{4.3}$$

There exists no universal relationship between the structural reference group $F(j)$ and the response group $G(j)$ defined in Section 2. However, having specified the structural groups of an endogenous-interactions model, one can derive the implied response groups. I give three illustrative polar cases here. I suppose throughout that the endogenous-interactions model is complete.

4.4.1. Symmetric Structural Groups. Let F be a group of persons and suppose that membership in structural group F is symmetric. That is, $F(j) = F$ for all $j \in F$. Then the structural equations pertaining to persons in F are

$$y_j(t^F) = f_j[t_j, t^{F/j}, y^{F/j}(t^F)], j \in F. \tag{4.4}$$

Completeness of the model implies that these equations have a unique solution $y_j(t^F)$, $j \in F$. Hence, the members of F share the same response reference group, namely $G(j) = F$ for all $j \in F$.

4.4.2. Recursive Structural Groups. Let population J be an ordered set of persons, indexed by the positive integers. Let the structural equations have the form

$$y_1(t^J) = f_1(t_1), \tag{4.5a}$$

$$y_j(t^J) = f_j[t_j, t_{j-1}, y_{j-1}(t^J)], \quad j \in (2, \ldots). \tag{4.5b}$$

Thus, the structural reference group for person j is $F(j) = (j-1, j)$. Recursively solving the equations shows that the response group for j is $G(j) = (1, \ldots, j)$.

4.4.3. Partly Responsive Structural Groups. Let the population contain two types of persons. A person is *responsive* to treatment if the value of his structural function may vary with his own treatment, all else equal. A person is *unresponsive* to treatment otherwise. Suppose that no one is responsive to the treatments received by others. Let R denote the sub-population of responsive persons. Thus, the structural equations are

$$y_j(t^J) = f_j[t_j, y^{J/j}(t_J)], \quad j \in R, \tag{4.6a}$$

$$y_j(t^J) = f_j[y^{J/j}(t^J)], \quad j \notin R. \tag{4.6b}$$

Thus, the structural group for person j is $F(j) = J$ if $j \in R$ and $F(j) = J/j$ if $j \notin R$. Solving the equations shows that, for all members of the population, outcomes may vary only with the treatments of responsive persons. Hence, $G(j) = \mathrm{R}$ for all $j \in J$.

Observe how the relationship between structural and response reference groups differs across these cases. Structural and response groups are identical in a population with symmetric structural groups. Structural groups are smaller than response ones when structural groups are recursive, the structural group of person j being $(j-1, j)$ and the response group being $(1, \ldots, j)$. Structural groups are larger than response ones when structural groups are partly responsive. The former group is either J or J/j, but the latter is R.

4.5. Reinforcing structural and response-function interactions

A researcher posing an endogenous-interactions model may assume that structural functions are monotone in their arguments. In work that builds on an earlier draft of the present paper, Lazzati (2010) studies the implications of such assumptions for response functions. I summarize her findings here.

Lazzati supposes that structural reference groups are symmetric. Hence, structural and response groups are identical. She also supposes that the range space of outcomes is compact. In this setting, she considers two forms of monotonicity of the structural functions.

First, suppose that endogenous interactions are reinforcing. Formally, consider persons in a symmetric structural group F. For all $j \in F$ and $t^F \in T^F$, assume that $f_j(t_j, t^{F/j}, \cdot)$ weakly increases in its final argument, $y^{F/j}(t^F)$. Lazzati brings to bear Tarski's fixed-point theorem to show that the structural equations have at least one solution. If they have multiple solutions, there exist smallest and largest solutions, whose values may depend on t^F.

Next, suppose that the structural functions are monotone in own treatment and also that exogenous interactions are reinforcing. Formally, for all $j \in F$ and $y^{F/j}(t^F) \in Y^{F/j}$, assume that $f_j[\cdot, \cdot, y^{F/j}(t^F)]$ weakly increases in $(t_j, t^{F/j})$. Combining this monotonicity assumption with the earlier one, she shows that the smallest and largest solutions to the structural equations are weakly increasing functions of $(t_j, t^{F/j})$.

When the endogenous-interactions model is complete, these two results imply that response functions satisfy Assumption SMTR, with $t^F \geq s^F \Rightarrow y_j(t^F) \geq y_j(s^F)$ for all $j \in F$. However, this conclusion does not necessarily hold when the model is incomplete. The potential problem is that, when the structural equations have multiple solutions, the social mechanism at work in the population may possibly select a smaller solution under treatment vector t^F than under s^F. Recognizing this possibility, Lazzati introduces the further assumption that the mechanism always selects either the smallest or the largest solution to the equations. Then Assumption SMTR holds even when the endogenous-interactions model is incomplete. A caveat is that the credibility of the further assumption may be difficult to assess in applications.

5. STATISTICAL INDEPENDENCE OF POTENTIAL OUTCOMES AND REALIZED EFFECTIVE TREATMENTS

Assumptions CTR and SMTR restrict the shape of individual response functions, without constraining the distribution of response across the population. Research under Assumption ITR regularly joins shape restrictions on response functions with distributional assumptions. Similarly, studies of models of endogenous interactions pose shape restrictions and distributional assumptions on structural functions.

A classical union of shape restrictions and distributional assumptions combines Assumption ITR with the assumption that potential outcomes are statistically independent of realized treatments. The statistical independence assumption has high credibility when realized treatments are randomly assigned. The pair of assumptions transparently yields point identification of potential outcome distributions, provided only that realized treatments equal potential treatments for a positive fraction of the population.

In this section I generalize the classical derivation to settings with social interactions. I combine Assumption CTR with the assumption that potential outcomes are statistically independent of realized effective treatments. I show that this pair of assumptions point identifies potential outcome distributions, provided that realized effective treatments equal potential effective treatments for positive fractions of certain subpopulations. Although the present derivation directly generalizes the one under Assumption ITR, the finding presented here is not as positive. The requirement that realized effective treatments equal potential effective treatments for positive fractions of certain subpopulations regularly fails to hold when social interactions are global in nature.

Section 5.1 presents the analysis. Section 5.2 interprets the finding. Section 5.3 draws cautionary implications for the identifying power of random assignment.

5.1. Analysis

As prelude, recall the classical argument. Suppose that the objective is to learn $P[y(\tau)]$ for some $\tau \in T$. One pairs Assumption ITR with the statistical-independence Assumption $P[y(\tau)] = P[y(\tau)|z = \tau]$. Assumption ITR implies that $P[y(\tau)|z = \tau] = P(y|z = \tau)$. Observation of realized treatments and outcomes reveals $P(y|z = \tau)$ if and only if $P(z = \tau) > 0$. Hence, the assumption that event $[z = \tau]$ is statistically independent of $y(\tau)$ point-identifies $P[y(\tau)]$ if and only if $P(z = \tau) > 0$. If $P(z = \tau) = 0$, the empirical evidence and assumption are uninformative about $P[y(\tau)]$.

Now pose any version of Assumption CTR. To generalize the classical argument, I first decompose the population into a set of distinct *effective-treatment types*. I will say that persons i and j have the same type if there exists a permutation operator π_{ij}: $T^J \to T^J$ such that $c_i(t^J) = c_j[\pi_{ij}(t^J)]$ for all $t^J \in T^J$. Suppose, for example, that i and j both have reference groups of size N. Then $c_i(t^J)$ and $c_j(t^J)$ are both subvectors of t^J of length N. A permutation of t^J transforms $c_i(t^J)$ into $c_j(t^J)$.

To enable use of elementary probability theory, I will suppose that the population is composed of a finite set M of types, each type having finitely many potential effective treatments. Let J_m denote the sub-population of type m. Let C_m be the common set of effective treatments for persons of type m.

For a given t^J and $\gamma \in C_m$, let $J_{m\gamma} \equiv [j \in J_m: c_j(t^J) = \gamma]$. In words, $J_{m\gamma}$ is the group of persons of type m who have effective treatment γ when the potential treatment vector is t^J. Outcomes in groups with zero probability mass do not affect outcome distribution $P[y(t^J)]$. Hence, in what follows, it suffices to consider groups with $P(J_{m\gamma}) > 0$.

Now assume statistical independence (SI) of potential outcomes and realized effective treatments. Formally, the assumption is

ASSUMPTION SI. *For each group $J_{m\gamma}$ with $P(J_{m\gamma}) > 0$,*

$$P[y(t^J)|J_{m\gamma}] = P[y(t^J)|J_{m\gamma}, c(z^J) = \gamma]. \tag{5.1}$$

Assumption CTR implies that $P[y(t^J)|J_{m\gamma}, c(z^J) = \gamma] = P[y|J_{m\gamma}, c(z^J) = \gamma]$. Observation of realized treatments and outcomes reveals $P[y|J_{m\gamma}, c(z^J) = \gamma]$ if and only if $P[c(z^J) = \gamma|J_{m\gamma}] > 0$. Hence, Assumption SI point-identifies $P[y(t^J)|J_{m\gamma}]$ if and only if $P[c(z^J) = \gamma|J_{m\gamma}] > 0$. If $P[c(z^J) = \gamma|J_{m\gamma}] = 0$, the empirical evidence and assumption are uninformative about $P[y(t^J)|J_{m\gamma}]$.

It remains to aggregate across groups. The Law of Total Probability gives

$$P[y(t^J)] = \sum_{(m\in M, \gamma\in C_m)} P[y(t^J)|J_{m\gamma}] \cdot P(J_{m\gamma}). \tag{5.2}$$

Hence, we have

PROPOSITION SI. *Given Assumption SI, the identification region for $P[y(t^J)]$ is*

$$H\{P[y(t^J)]\} = \left\{ \sum_{(m\in M, \gamma\in C_m: P[c(z^J)=\gamma|J_{m\gamma}]>0)} P[y|J_{m\gamma}, c(z^J) = \gamma] \cdot P(J_{m\gamma}) + \delta \cdot \sum_{(m\in M, \gamma\in C_m: P[c(z^J)=\gamma|J_{m\gamma}]=0)} P(J_{m\gamma}), \delta \in \Delta_Y \right\}. \tag{5.3}$$

5.2. Identifying power

Proposition SI shows that Assumption SI point-identifies $P[y(t^J)]$ if and only if $P[c(z^J) = \gamma | J_{m\gamma}] > 0$ for all $m \in M$ and $\gamma \in C_m$ such that $P(J_{m\gamma}) > 0$. Thus, every relevant conjectured effective treatment must be on the support of the distribution of realized effective treatments. The magnitudes of the probabilities $\{P[c(z^J) = \gamma | J_{m\gamma}], m \in M, \gamma \in C_m\}$ are immaterial. All that matters is that they be positive.

This requirement is transparent under Assumption ITR. Then everyone has the same type and the common set of effective treatments is $C = T$. Let J_τ denote the sub-population who would receive treatment τ under potential treatment vector t^J. The requirement for point identification of $P[y(t^J)]$ is $P(z = \tau | J_\tau) > 0$ for all $\tau \in T$ such that $P(J_\tau) > 0$. This support condition generically holds if realized treatments are assigned randomly with *ex ante* assignment probabilities $\varphi(\tau) > 0$, all $\tau \in T$. Then familiar arguments using laws of large numbers show that $P(z = \tau | J_\tau) \cong \varphi(\tau)$ for all τ such that $P(J_\tau) > 0$.

The support condition is considerably more subtle with treatment interactions. Suppose that persons of type m have reference groups of size S. Then the set of effective treatments is $C_m = T^S$ and $P[c(z^J) = \gamma | J_{m\gamma}] = P[z^{G(\cdot)} = \gamma | J_{m\gamma}]$. Thus, equalizing the realized and conjectured effective treatments of one person $j \in J_m$ necessitates fixing the realized treatments of all members of his reference group $G(j)$. For each k $\in G(j)$, let $D(k) \equiv [i \in J\colon k \in G(i)]$ denote the subset of J who list k as a member of their groups. Equalizing the realized and conjectured effective treatments of person j constrains the realized effective treatments of the entire class of persons $[D(k), k \in G(j)]$. This phenomenon can make it difficult to satisfy the support condition when reference groups are large or when there exist persons who belong to many reference groups.

I give two illustrations below. In both cases, interactions are *global* in the sense that the class of persons $[D(k), k \in G(j)]$ comprises the entirety of J_m.

5.2.1. Groups with leaders and followers. Let type m consist of persons having reference groups of size $N + L$. In these groups, membership is symmetric for N persons. That is, if the group of person i contains j, then the group of j contains i. Membership is asymmetric for L persons. These persons are in all groups but are not themselves type-m. I will call the L persons *leaders* and the N persons *followers*.

For example, the N persons in a group may be family members, perhaps husbands and wives. The L persons may be public figures, perhaps celebrities or opinion leaders. The treatments received by public figures may affect the outcomes of all families. The treatments received by family members have no impacts outside of the family.

The effective treatment of a person of type m is a subvector of t^J of length $N + L$. Let $\Lambda(m) \subset J$ denote the leaders of type m. Let $\gamma = (\tau^N, \tau^{\Lambda(m)})$ denote a situation in which followers receive the N treatments τ^N and leaders receive the L treatments $\tau^{\Lambda(m)}$. Then

$$P[c(z^J) = \gamma | J_{m\gamma}] = P[(z^N, z^{\Lambda(m)}) = (\tau^N, \tau^{\Lambda(m)}) | J_{m\gamma}] = P(z^N = \tau^N | J_{m\gamma}) \cdot 1[z^{\Lambda(m)} = \tau^{\Lambda(m)}]. \tag{5.4}$$

Thus, $P[c(z^J) = \gamma | J_{m\gamma}] > 0$ if and only if $z^{\Lambda(m)} = \tau^{\Lambda(m)}$ and $P(z^N = \tau^N | J_{m\gamma}) > 0$. In words, realized and potential treatments must coincide for all leaders. Moreover, they must coincide for all followers in a positive fraction of the groups of type m.

5.2.2. Population-wide distributional interactions. Suppose that the population contains one type of person. The common reference group is the entire population, and interactions are

distributional. This may be a reasonable idealization of some vaccination scenarios. One may think it credible to assume that each person's health status varies with his own vaccination and with the population rate of vaccination.

In this setting, $c_j(t^J) = [t_j, Q(t^J)]$ and $c_j(z^J) = [z_j, Q(z^J)]$ for all $j \in J$. The feasible values of γ are the pairs $[\tau, \mathrm{Q(t^J)}]$, $\tau \in T$. Fixing τ and letting J_τ be the subpopulation who would receive τ under potential treatment vector t^J, we have

$$P[c(z^J) = \gamma | J_{m\gamma}] = P\{[z, Q(z^J)] = [\tau, Q(t^J)] | J_\tau\} = P(z = \tau | J_\tau) \cdot 1[Q(z^J) = Q(t^J)]. \quad (5.5)$$

Thus, $P[c(z^J) = \gamma | J_{m\gamma}] > 0$ if and only if $Q(z^J) = Q(t^J)$ and $P(z = \tau | J_\tau) > 0$. The realized and potential population distributions of treatments must coincide. Moreover, a positive fraction of the persons in sub-population J_τ must receive realized treatment τ.

5.3. Random assignment of realized treatments

In research making Assumption ITR, Assumption SI is often motivated by knowledge that realized treatments were randomly assigned to the population. Random assignment may also motivate Assumption SI in the presence of interactions. However, in contrast to the situation with individualistic response, random assignment may not have identifying power. The two above illustrations demonstrate the difficulty.

Consider a random assignment process that independently assigns persons to treatments, with *ex ante* probability distribution φ on T. Suppose that φ is non-degenerate, placing positive mass on at least two elements of T. Then random assignment does not yield a determinate vector z^J of realized treatments. Instead, it yields an *ex ante* probability distribution for z^J.

Consider groups with leaders and followers. We showed above that the empirical evidence and Assumption SI can be informative about $P[y(t^J)|J_{m\gamma}]$ only if $z^{\Lambda(m)} = \tau^{\Lambda(m)}$. With random assignment, the *ex ante* probability that $z^{\Lambda(m)} = t^{\Lambda(m)}$ is $\Pi_{\,j\,\in\Lambda(m)}\, \varphi(t_j)$. This probability is less than one. Hence, random assignment yields positive *ex ante* probability that $z^{\Lambda(m)} \neq t^{\Lambda(m)}$.

Consider a population-wide distributional interaction. We showed above that the empirical evidence and Assumption SI can be informative about $P[y(t^J)|J_{m\gamma}]$ only if $Q(z^J) = Q(t^J)$. Using random assignment, this equality occurs with *ex ante* probability less than one. In the limit case of an uncountably large population, $Q(z^J) = \varphi$.

These negative findings do not appear in classical analysis of random assignment, which assumes individualistic response. Nor do they appear in the scattered efforts that researchers have made to study random assignment in settings with social interactions, such as Hudgens and Halloran (2008). These authors, and others they cite, assume that the population partitions into a large number of symmetric reference groups, each of finite size. Supposing that interactions may occur within groups but not across groups, they extend the classical analysis of random assignment. The feasibility of this extension is unsurprising, as Assumption ITR holds when the population is defined to be a collection of groups rather than persons.

6. CONCLUSION

This paper has studied identification of potential outcome distributions when treatment response may have social interactions. Defining a person's treatment response to be a function of the entire vector of treatments received by the population, I studied identification when non-parametric

shape restrictions and distributional assumptions are placed on response functions. An early key result was that the traditional assumption of individualistic treatment response is a polar case within the broad class of *constant treatment response* (CTR) assumptions, the other pole being unrestricted interactions. Important non-polar cases are interactions within reference groups and anonymous interactions. I first studied identification under Assumption CTR alone. I then strengthened this assumption to semi-monotone response. I next discussed derivation of these assumptions from models of endogenous interactions. Finally, I combined Assumption CTR with statistical independence of potential outcomes from realized effective treatments. Three propositions expressed the basic results of the paper, with special cases and illustrations fleshing them out.

I believe that these contributions provide a secure foundation for much further work. Many treatment-response assumptions beyond CTR, SMTR and SI warrant attention. Sustained study of the use of models of social mechanisms to derive restrictions on response functions would also be welcome.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation grant SES-0911181. I have benefitted from the opportunity to present this work as the Econometrics Journal Lecture at the 21st Annual $(EC)^2$ conference, at an invited session of the 2010 North American Winter Meeting of the Econometric Society, and in seminars at NYU, Northwestern, Nuffield College, UCL, UCLA, the University of Munich, and the University of Virginia. I am grateful for comments from Mike Fu, Guido Imbens, John Pepper and two referees.

REFERENCES

Blume, L., W. Brock, S. Durlauf and Y. Ioannides (2011). Identification of social interactions. In J. Benhabib, A. Bisin and M. Jackson (Eds.), *Handbook of Social Economics*, 854–964. Amsterdam: North-Holland.

Brock, W. and S. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*, 235–60.

Brock, W. and S. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*, 52–75.

Cox, D. (1958). *Planning of Experiments*. New York, NY: Wiley.

Goldberger, A. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Graham, B. (2011). Econometric methods for the analysis of assignment problems in the presence of complimentary and social spillovers. In J. Benhabib, A. Bisin and M. Jackson (eds.), *Handbook of Social Economics*, 965–1052. Amsterdam: North-Holland.

Hudgens, M. and E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*, 832–42.

Imbens, G. and C. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*, 1845–57.

Kiefer, N. and A. Goldberger (1989). The ET interview: Arthur S. Goldberger. *Econometric Theory 5*, 133–60.

Lazzati, N. (2010). Treatment response with social interactions: partial identification via monotone comparative statics. Working paper, Department of Economics, University of Michigan.

Manski, C. (1990). Nonparametric bounds on treatment effects. *American Economic Review Papers and Proceedings 80*, 319–23.

Manski, C. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies 60*, 531–42.

Manski, C. (1997). Monotone treatment response. *Econometrica 65*, 1311–34.

Manski, C. (2003). *Partial Identification of Probability Distributions.* New York, NY: Springer-Verlag.

Manski, C. (2009). Diversified treatment under ambiguity. *International Economic Review 50*, 1013–41.

Manski, C. (2010). Vaccination with partial knowledge of external effectiveness. *Proceedings of the National Academy of Sciences 107*, 3953–60.

Manski, C. and J. Pepper (2000). Monotone instrumental variables: with an Application to the Returns to Schooling. *Econometrica 68*, 997–1010.

Manski, C. and J. Pepper (2009). More on monotone instrumental variables. *The Econometrics Journal 12*, S200–S216.

Rubin, D. (1978). Bayesian inference for causal effects: the role of randomization. *Annals of Statistics 6*, 34–58.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*, 147–67.