IDENTIFICATION OF TREATMENT RESPONSE WITH SOCIAL INTERACTIONS

Charles F. Manski
Department of Economics and Institute for Policy Research, Northwestern University

Draft: October 7, 2009

Abstract

This paper develops a formal language for study of treatment response with social interactions, and uses it to obtain new findings on identification of potential outcome distributions. Defining a person's treatment response to be a function of the entire vector of treatments received by the population, I study identification when shape restrictions and distributional assumptions are placed on response functions. An early key result is that the traditional assumption of *individualistic treatment response* (ITR) 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 show that established findings on identification under assumption ITR extend to assumption CTR. These include identification with assumption CTR alone and when this shape restriction is strengthened to semi-monotone response. I next study distributional assumptions using instrumental variables. Findings obtained previously under assumption ITR extend when assumptions of statistical independence (SI) are posed in settings with social interactions, but with two caveats. The extended version of assumption SI has no power to identify counterfactual outcome distributions when social interactions are unrestricted. When interactions are restricted, the extended assumption may not be credible even if treatments are randomly assigned. Finally, considering models of endogenous social interactions, I show that identification of structural equations differs from identification of outcome distributions under potential treatments. Analysis of familiar linear models illustrates this general point.

This research was supported in part by National Science Foundation grant SES-0911181. I am grateful for comments from Mike Fu.

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 arguably are ubiquitous within households, schools, workplaces, and communities. Yet research on treatment response commonly assumes that a person's outcome may vary only with his own treatment, not with those of other members of the population. Some researchers call this "no interference between units" or 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, 1997a, 2003), Manski and Pepper (2000), 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 nonparametric assumptions that may be credible in applications and, hence, primarily report findings of partial rather than point identification.

To set the stage, I next specify basic concepts and notation that will be used throughout the paper. This requires a modest but essential extension of the formal setup used in my earlier work. I have then and now found that a clear and concise language enormously simplifies analysis of treatment response.

Basic Concepts and Notation

When response is assumed to be individualistic, each member j of population J has observable covariates $x_j \in X$ and 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 $y_j \equiv y_j(z)$. Let J be a probability space (J, Ω, P). Then observation of $[(x_j, y_j, z_j), j \in J]$ reveals P(x, y, z), the joint distribution of covariates, realized treatments and outcomes. A common research objective is to combine this empirical evidence with credible assumptions 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, I extend 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 $y_j(\cdot)$: $T^J \rightarrow Y$, 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 $z^J \equiv (z_k, k \in J)$. Observation of $[(x_j, y_j, z_j), j \in J]$ again reveals $P(x, y, z)$.

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 a specified treatment vector t^J. I will not require that t^J assign a common treatment to all persons, nor that it assign treatments randomly. I will suppose that the cardinality of T is at most countable. This enables an analysis that uses only elementary probability theory, particularly the Law of Total Probability. 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 $P(y)$.

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

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 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 2.1 below extends it to the general case where t^J is a vector of treatments that may vary across the population.

I show in Section 2.1 that, without assumption ITR or another assumption restricting the social interaction, the identification region for $P[y(t^J)]$ is the set of distributions $[P(y|z^J = t^J)P(z^J = t^J) + \delta P(z^J \neq t^J), \delta \in \Delta_Y]$. This is the singleton P(y) when $z^J = t^J$ and is the set Δ_Y of all distributions when $z^J \neq t^J$. 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 form of the response functions $[y_j(\cdot), j \in J]$ and/or the distribution $P[x, y(\cdot), z]$ of covariates, response functions, and realized treatments. This paper determines the identifying power of some nonparametric assumptions that may be credible in applications, and also considers some familiar parametric models.

Section 2 studies two nonparametric shape restrictions on response functions, *constant treatment response* (CTR) and *semi-monotone treatment response* (SMTR). Section 3 combines these shape restrictions with distributional assumptions that use instrumental variables. Section 4 considers identification using models of endogenous interactions. Section 5 concludes. The remainder of this Introduction gives an overview of what lies ahead.

Restrictions on the Shape of Response Functions

Constant response is a class of assumptions stating that a person's outcome remains constant when t^J varies within specified subsets of T^J. This 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.

Considering treatment with social interactions, a leading subclass of constant-response assumptions assert 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. In applied work, a person's reference group is often assumed to be the members of his family, neighborhood, school, workplace, or some other group, depending on the context. One might, for example, assume that treatment interactions may occur within but not across neighborhoods. Assumption ITR is the special case where each person's reference group includes only himself.

An important subclass of interactions within reference groups assumes that interactions are *anonymous*. Here the outcome of a person may vary with his own treatment and with the distribution of treatments among others in his reference group, but does not vary with the size of the group or with permutations of the treatments received by other group members. A simple form of anonymous interaction is the assumption that a person's outcome varies only with his own treatment and with the mean treatment of others in the reference group.

Semi-monotone response is a class of assumptions stating that set T^J is partially ordered and that outcomes vary monotonically across ordered pairs of treatment vectors. This class of assumptions was introduced and studied in Manski (1997a) in the context of individualistic treatment response. There the set T was partially ordered and it was assumed that outcomes vary monotonically across ordered pairs of treatments. Extending the idea to settings with social interactions is straightforward.

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. 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 who are in contact with person j may also reduce his probability of illness, reinforcing the effect of own vaccination. I will use vaccination to illustrate findings on identification in Sections 2.3 and 3.2.

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. 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, but training other persons increases the supply of trained labor and, hence, may decrease the probability that person j finds employment.

Distributional Assumptions Using Instrumental Variables

Having studied the identifying power of shape restrictions alone in Section 2, Section 3 combines shape restrictions with distributional assumptions that use instrumental variables. In research under assumption ITR, empirical researchers often pose assumptions about the distribution of response. Particularly common are assumptions that use an instrumental variable $v \equiv v(x, z)$, where $v(\cdot, \cdot): X \times T \to V$ is a specified function of observed covariates and realized treatments. Taking the objective to be inference on $P[y(\tau)]$ for a specified $\tau \in T$, it is common to assume that $y(\tau)$ and v are statistically or mean independent; that is, $P[y(\tau)|v] = P[y(\tau)]$ or $E[y(\tau)|v] = E[y(\tau)]$.

These assumptions extend immediately to research on treatment with social interactions, where the objective is to infer $P[y(t^J)]$ for a specified $t^J \in T^J$. Here one may assume that $P[y(t^J)|v] = P[y(t^J)]$ or $E[y(t^J)|v] = E[y(t^J)]$. I first characterize the identifying power of these assumptions abstractly. The analysis extends my earlier work on identification when assumption ITR is combined with independence assumptions using instrumental variables (Manski, 1990; 2003, Chapter 2 and Section 7.4). I use a vaccination scenario to illustrate the findings.

I then consider the use of realized treatments as the instrumental variable. Given assumption ITR, it is well known that the assumption $P[y(\tau)|z] = P[y(\tau)]$ point identifies $P[y(\tau)]$. Moreover, this assumption has high credibility when realized treatments are randomly assigned. I show that the familiar identification argument extends to inference under assumption CTR, using appropriate person-specific functions of z^J as the instrumental variable.

These positive findings must be tempered by two caveats. First, the extended statistical independence assumption has no power to identify counterfactual outcome distributions when social interactions are unrestricted. Second, when interactions are restricted, the extended assumption may not be credible even if treatments are randomly assigned.

Models of Endogenous Social Interactions

In the analysis of Sections 2 and 3, response functions are primitives that map treatment vectors into outcomes. Section 4 considers identification with models of endogenous interactions. The primitive of 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, called the *reduced form* of the model.

A large body of econometric research has studied identification of structural equations. However, our objective is identification of $P[y(t^J)]$, not identification of structural equations. A model of endogenous interactions may have identifying power for $P[y(t^J)]$ if the specified structural equations imply restrictions on the reduced form. I use two familiar models to illustrate the broad idea. One is a linear simultaneous equations model of interaction between pairs of persons. The other is a linear-in-means model of anonymous interactions within large reference groups. Analysis of these simple models demonstrates that identification of $P[y(t^J)]$ is not equivalent to identification of structural equations. I also briefly discuss incomplete models, which do not have a unique reduced form.

2. Restrictions on the Shape of Response Functions

This section studies the identifying power of assumptions that restrict the shape of the treatment response functions $[y_j(\cdot), j \in J]$. I begin with constant treatment response in Section 2.1 and then add semi-monotone treatment response in Section 2.2. Section 2.3 uses vaccination against infectious disease to illustrate findings.

2.1. Constant Treatment Response

Constant-response assumptions assert that treatment response does not vary over specified sets of treatment vectors. Section 2.1.1 poses the assumption in abstraction and establishes its identifying power. Section 2.1.2 specializes to CTR assumptions that restrict social interactions to reference groups. Section 2.1.3 specializes further to anonymous interactions.

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 often have high credibility. As such, they provide a foundation on which further assumptions may be placed.

2.1.1. The Assumption in Abstraction

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

(1) $c_j(t^J) = c_j(s^J) \;\Rightarrow\; y_j(t^J) = y_j(s^J)$.

Thus, j experiences the same outcome for all treatment vectors that form a level set of $c_j(\cdot)$.

Consider inference on $y_j(t^J)$. The event $[c_j(z^J) = c_j(t^J)]$ may or may not occur. The researcher knows the value of $y_j(t^J)$ when this event occurs, because $y_j(t^J) = y_j(z^J) = y_j$. The researcher does not know $y_j(t^J)$ when the event does not occur.

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

$$(2) \qquad P[y(t^J)] = P[y(t^J)|c(z^J) = c(t^J)]\cdot P[c(z^J) = c(t^J)] + P[y(t^J)|c(z^J) \neq c(t^J)]\cdot P[c(z^J) \neq c(t^J)].$$

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, the identification region for $P[y(t^J)]$ is

$$(3) \qquad H\{P[y(t^J)]\} = \{P[y|c(z^J) = c(t^J)]\cdot P[c(z^J) = c(t^J)] + \delta\cdot P[c(z^J) \neq c(t^J)], \delta \in \Delta_Y\}.$$

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 $P[c(z^J) = c(t^J)] = 0$.

2.1.2. Interactions within Reference Groups

It is common in applications to assume that each member of the population has a known reference groups, with interactions occurring within but not across groups. 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 we have an exclusion restriction asserting that person j's outcome remains constant when treatments of persons outside the group are altered, holding fixed the treatments of persons within the group.

As defined here, reference groups are person-specific but not manipulable. Being person-specific permits person k to belong to person j's group but not vice versa. Such 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. At the extreme, the reference group for person j might be the entire population while that of person k might be this person alone.

Non-manipulability means the treatments T under consideration cannot alter a person's reference group. The present analysis can be extended to consider treatments that manipulate group composition. However, I will not pursue this here.

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 (3), the identification region for $P[y(t^J)]$ is

$$(4) \qquad H\{P[y(t^J)]\} \; = \; [P(y|z^G = t^G) \cdot P(z^G = t^G) + \delta \cdot P(z^G \neq t^G) \, , \, \delta \in \Delta_Y].$$

Two polar cases of interactions within reference groups are unrestricted interactions, where all reference groups are the entire population, and individualistic treatment response, where all reference groups are single persons. In the former case, $G(j) = J$ for all $j \in J$. Then (4) becomes

$$H\{P[y(t^J)]\} \;=\; [P(y|z^J = t^J)\cdot P(z^J = t^J) + \delta\cdot P(z^J \neq t^J),\ \delta \in \Delta_Y]. \tag{5}$$

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 result justifies the statement in the Introduction 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 (4) becomes

$$H\{P[y(t^J)]\} \;=\; [P(y|z = t)\cdot P(z = t) + \delta\cdot P(z \neq t),\ \delta \in \Delta_Y\}. \tag{6}$$

Result (6) extends my earlier work on identification with individualistic treatment response. I have earlier reported (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 (6) is the common treatment and $t^J = (t, t, \ldots, t)$. Now (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 (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. The identifying power of assumption ITR appears when $0 < P(z = t) < 1$. Region (6) grows 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 (5), which equals Δ_Y whenever $P(z = t) < 1$.

2.1.3. Anonymous Interactions

Region (4) characterized identification under the sole assumption that interactions occur within reference groups. Applied research often assumes that interactions are anonymous. An *anonymous* social interaction is one where the outcome of person j may vary with his own treatment and with the distribution of treatments among other members of the reference group, but is invariant with respect to the size of the group and permutations of the treatments received by other members of the group. The anonymous-interaction assumption is empty when a reference group contains one or two persons, 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 rate of vaccination in his school, but not on the identities of the specific other schoolmates vaccinated.

Formally, let $C_j = T \times \Delta_T$, where Δ_T is the space of all distributions on T. For $j \in J$ with $|G(j)| > 1$, let G(j)/j denote the reference group exclusive of person j himself. For $t^J \in T^J$, let $c_j(t^J) = [t_j, Q(t^{G(j)/j})]$, where $Q(t^{G(j)/j}) \equiv P[t^J | G(j)/j]$ is the within-group distribution of the treatments in $t^{G(j)/j}$. That is, for $\tau \in T$, $Q(t^{G(j)/j} = \tau)$ is the fraction of the persons in G(j)/j who would receive treatment τ when t^J is the potential treatment vector. For $j \in J$ with $|G(j)| = 1$, the set G(j)/j is empty. To formally cover this case, I define $Q(t^{G(j)/j}) = \emptyset$, where $\emptyset$ denotes the empty set.

With this definition of C_j and $c_j(\cdot)$, the abstract constant-response region (3) takes the form

$$(7)\quad H\{P[y(t^J)]\} = \{[P(y|z = t, Q(z^G) = Q(t^G)]\cdot P[z = t, Q(z^G) = Q(t^G)] + \delta\cdot P(z \neq t \text{ or } Q(z^G) \neq Q(t^G), \delta \in \Delta_Y\}.$$

This region is a subset of the region (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 $[z_j = t_j, Q(z^{G(j)/j}) = Q(t^{G(j)/j})]$ occurs. Previously, $y_j(t^J)$ was known when $z^{G(j)} = t^{G(j)}$. The latter event implies the former one.

Functional Interactions

Applied research often assumes not only that interactions are anonymous but also that $Q(t^{G(j)/j})$ affects outcomes solely through some functional of Q, say $F_Q(t^{G(j)/j})$. A leading case is the *mean interaction*, where treatments are real-valued and $F_Q(t^{G(j)/j}) = E_Q(t^{G(j)/j})$, the within-group mean of the treatments in $t^{G(j)/j}$. A mean interaction is equivalent to an anonymous interaction when set T has two treatments. It is a stronger assumption when there are more than two.

Another case of applied interest is the *supremum interaction*, where treatments are ordered and $F_Q(t^{G(j)/j}) = \sup(t^{G(j)/j})$. Suppose that a treatment is information that may be 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 all group members effectively receive treatment $\sup(t^G)$.

Whatever functional F may be, let $C_j = T \times \Phi$, where Φ is the range space for F. Let $c_j(t^J) = [t_j, F_Q(t^{G(j)/j})]$. Then (3) becomes

$$(8)\quad H\{P[y(t^J)]\} =$$

$$\{[P(y|z = t, F_Q(z^G) = F_Q(t^G)]\cdot P[z = t, F_Q(z^G) = F_Q(t^G)] + \delta\cdot P(z \neq t \text{ or } F_Q(z^G) \neq F_Q(t^G), \delta \in \Delta_Y\}.$$

This region is a subset of the region (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 $[z_j = t_j, F_Q(z^{G(j)/j}) = F_Q(t^{G(j)/j})]$ 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.

2.2. Semi-Monotone Treatment Response

The constant-response assumptions considered in Section 2.1 were nested. Individualistic treatment response weaky strengthens functional interactions, which weakly strengthens anonymous interactions, which in turn weakly strengthens interactions within a reference group. The various identification regions presented above were correspondingly nested sets. However, even the assumption of individualistic treatment response 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 2.2.1 poses the assumption in abstraction and establishes its identifying power. Sections 2.2.2 and 2.2.3 consider the important sub-cases of reinforcing and opposing interactions.

2.2.1. The Assumption in Abstraction

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

(9) $c_j(t^J) \geq c_j(s^J) \;\Rightarrow\; y_j(t^J) \geq y_j(s^J).$

This assumption strengthens 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 treatment response, Manski (1997a), 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 $y_j(t^J)$:

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

Let $y_{jL}(t^J)$ and $y_{jU}(t^J)$ denote the lower and upper bounds on $y_j(t^J)$ stated in (10). Given that (10) 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 (10) exhausts the available information, the identification region for $P[y(t^J)]$ is

$$
(11) \qquad H\{P[y(t^J)]\} = \{\delta \in \Delta_Y: P[y_U(t^J)] \geq_{sd} \delta \geq_{sd} P[y_L(t^J)]\},
$$

where $\geq_{sd}$ 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 (11) immediately yields this sharp bound on $D[y(t^J)]$:

(12) $D[y_L(t^J)] \le D[y(t^J)] \le D[y_U(t^J)]$.

Considering individualistic treatment response, Manski (1997a), Corollaries S1.1 – S1.3 gave the explicit form of bound (12) 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

(13) $$y_0 \cdot P[c(t^J) < c(z^J) \cup c(t^J) \oslash c(z^J)] + E[y \mid c(t^J) \ge c(z^J)] \cdot P[c(t^J) \ge c(z^J)] \le E[y(t^J)]$$
$$\le y_1 \cdot P[c(t^J) > c(z^J) \cup c(t^J) \oslash c(z^J)] + E[y \mid c(t^J) \le c(z^J)] \cdot P[c(t^J) \le c(z^J)].$$

2.2.2. Reinforcing Interactions

I defined reinforcing interactions verbally in the Introduction. Formally, let treatment set T be partially ordered. Let person 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 $c_j(t^J) \ge c_j(s^J)$ mean that $[t_k \ge s_k$, all $k \in G(j)]$. A reinforcing interaction occurs when

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

When (14) 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.

I earlier gave vaccination against an infectious disease as an example of an interaction that is credibly reinforcing. Another is provision of tutoring to students in a classroom. It is reasonable to think that tutoring a student weakly increases his achievement. It may also be reasonable to think that tutoring some students weakly increases the achievement of all students in the classroom.

Reinforcing Anonymous Interactions

The definition of a reinforcing interaction stated in (14) 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 $s^{G(j)}$. When the social interaction is anonymous, permutation of the treatments received by other members of the reference group does not affect the outcome of person j. Hence, let $c_j(t^J) \geq c_j(s^J)$ now mean that $[t_j \geq s_j, Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})]$. A reinforcing anonymous interaction occurs when

$$(15) \quad [t_j \geq s_j, Q(t^{G(j)/j}) \geq_{sd} Q(s^{G(j)/j})] \Rightarrow y_j(t^J) \geq y_j(s^J).$$

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 anonymous 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 (15) is well approximated by $Q(t^{G(j)}) \geq_{sd} Q(s^{G(j)})$, which includes j in the group distribution. The latter inequality is simpler to use in some applications. I will use it in Section 2.3.

Reinforcing D-Interactions

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

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

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

2.2.3. 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

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

When (17) 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. I earlier gave occupation-specific training as an example of an interaction that is credibly opposing.

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

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

The latter occurs when

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

2.3. Illustration: Vaccination Against Infectious Disease

I will use a simple scenario of vaccination against infectious disease to illustrate the findings of Sections 2.1 and 2.2. 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 probability 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 probability 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 permit mean interactions and then specialize to reinforcing mean interactions.

2.3.1. Individualistic Treatment Response

Suppose that a person's health status depends only on his own treatment. This assumption may not be 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 (6). With a binary outcome, (6) becomes the interval

$$(20) \qquad H\{P[y(t^J) = 1]\} \ = \ [P(y = 1 | z = t) \cdot P(z = t), \ \ P(y = 1 | z = t) \cdot P(z = t) + P(z \neq t)].$$

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 $1 - q$ who realize 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 (20).

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

$$\begin{aligned}(21) \qquad P(y = 1 | z = t) \ &= \ P(y = 1 | z = t, z = 1) \cdot P(z = 1 | z = t) \ + \ P(y = 1 | z = t, z = 0) \cdot P(z = 0 | z = t) \\ &= \ P_{11}[p/(p + 1 - q)] + P(y = 1 | z = t, z = 0) \cdot [(1 - q)/(p + 1 - q)].\end{aligned}$$

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 $P(y = 1 | z = t, z = 1) = P_{11}$ because $z = 1 \Rightarrow t = 1$. We have not yet encountered $P(y = 1 | z = t, z = 0)$, 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, which depends on t^J, is revealed by the empirical evidence once t^J is specified. Hence, all quantities on the right-hand side of (21) are known.

2.3.2. Monotone-Individualistic Treatment 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 (13), which reduces in the present case to

$$(22) \qquad H\{P[y(t^J) = 1]\} \;=\; [P(y = 1|t \geq z)\cdot P(t \geq z),\ P(t > z) + P(y = 1|t \leq z)\cdot P(t \leq z)].$$

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 (21). The result is

$$(23) \qquad H\{P[y(t^J) = 1]\} \;=\; [P(y = 1),\ q - p + P(y = 1|t = z)\cdot(p + 1 - q)].$$

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

2.3.3. Mean Interactions

Now suppose that a person's health status may depend on his own treatment and on the population vaccination rate. Then the identification region is given by (8), which becomes

$$(24)\quad H\{P[y(t^J)] = 1\} = [P(y = 1 | z = t, p = q)\cdot P(z = t, p = q),$$
$$P(y = 1 | z = t, p = q)\cdot P(z = t, p = q) + P(z \neq t \text{ or } p \neq q)].$$

By assumption, $q > p$ in this illustration. Hence, the identification region is [0, 1]. This result should be expected, because increasing the vaccination rate makes it counterfactual for the entire population.

2.3.4. Reinforcing Mean Interactions

Continue to suppose that a person's health status may depend on his own treatment and on the population vaccination rate. Also suppose that the interaction is reinforcing, with vaccination of an individual reducing his chance of infecting others. Then (16) holds. Suppose that the population is large, so $E_Q(t^{G(j)/j}) \cong E_Q(t^{G(j)}) = P(t = 1)$ and $E_Q(s^{G(j)/j}) \cong E_Q(s^{G(j)}) = P(s = 1)$. Using this approximation, (16) reduces to

$$(25)\quad t_j \geq s_j \cap P(t = 1) \geq P(s = 1) \Rightarrow y_j(t^J) \geq y_j(s^J).$$

The identification region is given by (13), which reduces in this case to

$$(26) \quad H\{P[y(t^J) = 1]\} = [P[y = 1 \,|\, (t, q) \geq (z, p)]\cdot P[(t, q) \geq (z, p)],$$

$$P[(t, q) > (z, p) \cup (t, q) \oslash (z, p)] + P[y = 1 \,|\, (t, q) \leq (z, p)]\cdot P[(t, q) \leq (z, p)].$$

Given that $t^J \geq z^J$ and $q > p$, $P[(t, q) > (z, p)] = 1$. Hence, (26) reduces to

$$(27) \quad H\{P[y(t^J) = 1]\} = [P(y = 1), 1].$$

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

3. Distributional Assumptions Using Instrumental Variables

This section combines the shape restrictions of Section 2 with distributional assumptions that use instrumental variables. In research assuming individualistic response, an instrumental variable (IV) is a specified function $v \equiv v(x, z)$ of the observed covariates x and realized treatments z. Assumptions typically restrict how the conditional response distributions $P[y(\cdot)|w, v]$ may vary with v, where $w \equiv w(x, z)$ is another function of (x, z). When studying treatment with social interactions, I will let $v \equiv v(x, z^J)$ and $w \equiv w(x, z^J)$. To simplify the exposition, I will suppress w until Section 3.3, where it is necessary to make it explicit.

Various distributional assumptions may merit consideration. In research under assumption ITR, where the objective is to infer $P[y(\tau)]$ for a specified $\tau \in T$, it has been common to assume statistical independence (SI) or mean independence (MI); that is, $P[y(\tau)|v] = P[y(\tau)]$ or $E[y(\tau)|v] = E[y(\tau)]$. These assumptions extend directly to research on treatment with social interactions, where the objective is to infer

$P[y(t^J)]$ for a specified $t^J \in T^J$. Then one may assume that $P[y(t^J)|v] = P[y(t^J)]$ or $E[y(t^J)|v] = E[y(t^J)]$. Research on *monotone instrumental variables* considers weaker assumptions that replace the above equalities with weak inequalities (Manski and Pepper, 2000). I will focus here on the traditional independence assumptions.

Section 3.1 considers these assumptions in abstraction. The analysis is a simple extension of my earlier work assuming individualistic response (Manski, 1990; 2003, Chapter 2 and Section 7.4). To illustrate, Section 3.2 considers a variation on the vaccination scenario of Section 2.3.

In research under assumption ITR, it has been common to point-identify $P[y(\tau)]$ by asserting assumption SI with the realized treatment z as the instrumental variable. Section 3.3 extends the argument to inference on $P[y(t^J)]$ under assumption CTR, using $c(z^J)$ as the instrumental variable. However, I caution that the extended assumption need not be credible even if treatments are randomly assigned. Random assignment implies that z^J and $y(t^J)$ are statistically independent, but it does not imply that $c(z^J)$ and $y(t^J)$ are independent. I give an example to illustrate.

3.1. The Assumption in Abstraction

To begin, observe that all of the findings obtained in Section 2 hold if one poses a shape restriction and considers identification of $P[y(t^J)|v = \nu]$, where $\nu \in V$, the support of v. One simply needs to condition every reference to P on the event $[v = \nu]$ and repeat the derivations. Let $H\{P[y(t^J)|v = \nu]\}$ denote the resulting identification region. The identification region for the collection of distributions $\{P[y(t^J)|v = \nu], \nu \in V\}$ is the Cartesian product $\times_{\nu \in V} H\{P[y(t^J)|v = \nu]\}$. These results hold because the shape restrictions of Section 2 operate separately on the response function of each member of the population. They restrict the distribution of response only through aggregation of their implications for individual response.

Now introduce the statistical-independence assumption $P[y(t^J)|v] = P[y(t^J)]$. Then $P[y(t^J)]$ must lie within the intersection of the identification regions $H\{P[y(t^J)|v = v]\}$, $v \in V$. Moreover, every distribution in this intersection is a feasible value of $P[y(t^J)]$. Hence, the identification region for $P[y(t^J)]$ is

$$(28) \quad H\{P[y(t^J)]\} = \bigcap_{v \in V} H\{P[y(t^J)|v = v]\}.$$

An analogous derivation holds for inference on $E[y(t^J)]$ under the mean-independence assumption $E[y(t^J)|v] = E[y(t^J)]$. In this case, the identification regions obtained in Section 2 are intervals of the generic form $[L_v(t^J), U_v(t^J)]$, $v \in V$. The region using assumption MI is the interval

$$(29) \quad H\{E[y(t^J)]\} = [\sup_{v \in V} L_v(t^J), \inf_{v \in V} U_v(t^J)].$$

The assumptions used to derive these identification regions may be jointly testable. The empirical evidence may reveal that the region in (28) or (29) is empty. If so, then some assumption is incorrect. When an identification region is non-empty, one cannot reject the maintained assumptions. Of course, a failure to reject the assumptions does not imply that they are correct.

3.2. Application to Vaccination

To illustrate, consider a vaccination scenario in which the population partitions into two reference groups. Persons with $v = 0$ belong to one group and those with $v = 1$ belong to the other. Treatment interactions may occur within but not across groups.

Suppose that the realized vaccination rate among persons with $v = 0$ is lower than among those with $v = 1$; thus, $P(z = 1|v = 0) < P(z = 1|v = 1)$. Consider a potential treatment vector t^J that equalizes the vaccination rates of the two groups at an intermediate level q. In particular, t^J sets $t_j = 1$ for all those with

($v_j = 0, z_j = 1$) and for some of those with ($v_j = 0, z_j = 0$). It sets $t_j = 0$ for all those with ($v_j = 1, z_j = 0$) and for some of those with ($v_j = 1, z_j = 1$). As a result, $P(t = 1|v = 0) = P(t = 1|v = 1) = q$. The objective is to learn $P[y(t^J) = 1]$.

First consider inference under the assumption of a reinforcing mean interaction. By the Law of Total Probability,

$$(30) \quad P[y(t^J) = 1] \; = \; P[y(t^J) = 1|v = 0]\cdot P(v = 0) \; + \; P[y(t^J) = 1|v = 1]\cdot P(v = 1).$$

Application of (27) to the group with $v = 0$ shows that $H\{P[y(t^J) = 1|v = 0]\} = [P(y = 1|v = 0), 1]$. A derivation analogous to that yielding (27) shows that $H\{P[y(t^J) = 1|v = 1]\} = [0, P(y = 1|v = 1)]$. The joint identification region for $P[y(t^J) = 1|v = 0]$ and $P[y(t^J) = 1|v = 1]$ is the Cartesian product of the marginal regions. Hence, the identification region for $P[y(t^J) = 1]$ is

$$(31) \quad H\{P[y(t^J) = 1]\} \; = \; [P(y = 1|v = 0)\cdot P(v = 0), P(v = 0) + P(y = 1|v = 1)\cdot P(v = 1)].$$

The lower bound occurs if the change in treatment from z^J to t^J has no positive health effect on those with $v = 0$ and a negative effect on everyone with $v = 1$. The upper bound occurs if the change makes everyone with $v = 0$ healthy and has no negative effect on those with $v = 1$.

Now consider inference when the assumption of a reinforcing mean interaction is combined with assumption SI, namely $P[y(t^J) = 1|v = 0] = P[y(t^J) = 1|v = 1]$. Then $H\{P[y(t^J) = 1]\}$ is the intersection of the reinforcing-mean identification regions obtained above for $P[y(t^J) = 1|v = 0]$ and $P[y(t^J) = 1|v = 1]$. Thus,

$$(32) \quad H\{P[y(t^J) = 1]\} \; = \; [P(y = 1|v = 0), P(y = 1|v = 1)].$$

Inspection of region (32) shows that the pair of assumptions used to derive the region are jointly testable. Suppose the empirical evidence reveals that $P(y = 1|v = 0) > P(y = 1|v = 1)$. Then region (32) is empty. It follows that at least one of the assumptions is incorrect.

When region (32) is non-empty, it is natural to ask whether the maintained assumptions are credible. The assumption of a reinforcing interaction seems sensible enough. It is less clear that the interaction is anonymous, as the structure of social networks may affect the transmission of disease.

It may be difficult to assess assumption SI. The fact that t^J equalizes the vaccination rates of the two groups may be suggestive, but it does not imply equal health outcomes in the two groups. The assumption may be credible if one somehow knows that members of the two groups have similar susceptibility to infection and that similar processes are used to assign treatments in the two groups. In the absence of such information, one may not be able to assess whether or not v is a valid instrumental variable.

3.3. Realized Treatments as Instrumental Variables

3.3.1. Combining Assumptions CTR and SI

In research assuming individualistic response, it is common to assume that $P[y(\tau)] = P[y(\tau)|z]$, $\tau \in T$. This version of assumption SI may be motivated by knowledge that realized treatments were randomly assigned, or it may be posed without clear knowledge of the assignment process. In any case, 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, taking the realized treatment z to be an instrumental variable that is statistically independent of $y(\tau)$ point-identifies $P[y(\tau)]$ if and only if $P(z = \tau) > 0$.

The above familiar derivation extends to settings with social interactions under assumption CTR. The appropriate extension of assumption SI is $P[y(t^J)|c(t^J)] = P[y(t^J)|c(t^J), c(z^J)]$. This takes the person-specific variable $c(z^J)$ to be the instrumental variable. It reduces to $P[y(\tau)] = P[y(\tau)|z]$ when response is individualistic and $t^J = (\tau, \ldots, \tau)$.

To begin the derivation, let $C \equiv \cup_{j \in J} C_j$ and let $C(t^J) \equiv \{\gamma \in C\text{: } P[c(t^J) = \gamma] > 0\}$. To enable use of elementary probability theory, this analysis assumes that C is countable. Successively apply the Law of Total Probability, the IV assumption, and assumption CTR to obtain

$$(33) \quad P[y(t^J)] = \sum_{\gamma \in C(t^J)} P[y(t^J)|c(t^J) = \gamma]P[c(t^J) = \gamma] = \sum_{\gamma \in C(t^J)} P[y(t^J)|c(t^J) = \gamma, c(z^J) = \gamma]P[c(t^J) = \gamma]$$

$$= \sum_{\gamma \in C(t^J)} P[y|c(t^J) = \gamma, c(z^J) = \gamma]P[c(t^J) = \gamma].$$

For each $\gamma \in C(t^J)$, observation of realized treatments and outcomes reveals $P[y|c(t^J) = \gamma, c(z^J) = \gamma]$ if and only if $P[c(t^J) = \gamma, c(z^J) = \gamma] > 0$. By construction, $P[c(t^J) = \gamma] > 0$ for all γ in $C(t^J)$. Hence, the empirical evidence reveals $P[y|c(t^J) = \gamma, c(z^J) = \gamma]$ if and only if $P[c(z^J) = \gamma|c(t^J) = \gamma] > 0$. It follows that the identification region for $P[y(t^J)]$ is

$$(34) \quad H\{P[y(t^J)]\} = \{ \sum_{\gamma \in C_1(t^J)} P[y|c(t^J) = \gamma, c(z^J) = \gamma]\cdot P[c(t^J) = \gamma] + \delta\cdot P[c(t^J) \in C_0(t^J)],\ \delta \in \Delta_Y\},$$

where $C_1(t^J) \equiv \{\gamma \in C(t^J)\text{: } P[c(z^J) = \gamma|c(t^J) = \gamma] > 0\}$ and $C_0(t^J) \equiv \{\gamma \in C(t^J)\text{: } P[c(z^J) = \gamma|c(t^J) = \gamma] = 0\}$.

Observe that $P[y(t^J)]$ is point-identified when $P[c(t^J) = \gamma] > 0 \Rightarrow P[c(z^J) = \gamma|c(t^J) = \gamma] > 0$. Then the set $C_0(t^J)$ is empty and $C_1(t^J) = C(t^J)$. Contrariwise, $P[y(t^J)]$ is entirely unknown when $P[c(t^J) = \gamma] > 0 \Rightarrow P[c(z^J) = \gamma|c(t^J) = \gamma] = 0$. Then $C_1(t^J)$ is empty and $C_0(t^J) = C(t^J)$.

Two Polar Cases

Region (34) simplifies in the polar case of individualistic response. Then $C = T$ and $C(t^J) = T(t^J) \equiv$

$[\tau \in T: P(t = \tau) > 0]$. Also $C_1(t^J) = T_1(t^J) \equiv [\tau \in T(t^J): P(z = \tau | t = \tau) > 0]$ and $C_0(t^J) = T_0(t^J) \equiv [\tau \in T(t^J): P(z = \tau | t = \tau) = 0]$. Hence, (34) becomes

$$(35) \quad H\{P[y(t^J)]\} = \{ \sum_{\tau \in T_1(t^J)} P(y | t = \tau, z = \tau)\cdot P(t = \tau) + \delta\cdot P[t \in T_0(t^J)], \ \delta \in \Delta_Y\}.$$

Region (34) also simplifies in the polar case of an unrestricted interaction. Here, $c_j(z^J) = z^J$ for all $j \in J$. Thus, $c(z^J)$ has a degenerate distribution, implying that assumption SI necessarily holds. Observe that $P[c(t^J) = \gamma] = 1$ if $\gamma = t^J$ and $P[c(t^J) = \gamma] = 0$ otherwise. Hence, $C(t^J)$ contains the single element t^J. We have $P[c(z^J) = t^J] = 1$ if $z^J = t^J$ and $P[c(z^J) = t^J] = 0$ otherwise. Hence, $H\{P[y(t^J)]\}$ is the singleton $P(y)$ if $z^J = t^J$ and is Δ_Y if $z^J \neq t^J$. This is the same result as holds using the empirical evidence alone. It shows, among other things, that random assignment of realized treatments has no power to identify counterfactual outcome distributions when social interactions are unrestricted.

An Alternative Derivation

Derivation of (34) did not use the general finding (28) expressing the identification region with an IV-assumption as the intersection of the regions across values of the instrument. We can alternatively use (28) to obtain (34). For each $\gamma \in C(t^J)$, application of (28) gives

$$(36) \quad H\{P[y(t^J) | c(t^J) = \gamma]\} = \bigcap_{\gamma' \in C} H\{P[y(t^J) | c(t^J) = \gamma, c(z^J) = \gamma']\}.$$

When $\gamma' = \gamma$, we have established that $H\{P[y(t^J) | c(t^J) = \gamma, c(z^J) = \gamma]\}$ is the singleton $P[y | c(t^J) = \gamma, c(z^J) = \gamma]$ if $P[c(z^J) = \gamma | c(t^J) = \gamma] > 0$ and is Δ_Y otherwise. When $\gamma' \neq \gamma$, $H\{P[y(t^J) | c(t^J) = \gamma, c(z^J) = \gamma']\} = \Delta_Y$. Hence, the intersection of regions in (36) is $P[y | c(t^J) = \gamma, c(z^J) = \gamma]$ if $P[c(z^J) = \gamma | c(t^J) = \gamma] > 0$ and is Δ_Y otherwise.

Comparison with Assumption CTR Alone

It is instructive to compare the present result with the one obtained in Section 2.1 using assumption CTR alone. When constant response is the only maintained assumption, observation of y_j and z^J reveals $y_j(t^J)$ if and only if $c_j(z^J) = c_j(t^J)$. When the empirical evidence does not reveal $y_j(t^J)$, it is uninformative about $P[y(t^J)]$. Hence, the identification region given in (3) enlarges with $P[c(z^J) \neq c(t^J)]$.

When constant response is combined with the IV assumption, the empirical evidence reveals $P[y(t^J)|c(t^J) = \gamma]$ if and only if $P[c(z^J) = \gamma|c(t^J) = \gamma] > 0$. Hence, region (34) enlarges with $P[c(t^J) \in C_0(t^J)]$. Distribution $P[y(t^J)]$ may be point-identified even if $P[c(z^J) \neq c(t^J)] = 1$.

Credibility of Assumption SI

I have thus far used assumption SI without questioning its credibility. Credibility has long been a concern in research under assumption ITR, where the assumption reduces to $P[y(\tau)] = P[y(\tau)|z]$. In that setting, the assumption has a firm foundation when treatments are randomly assigned but may not otherwise.

Credibility is yet more complex to assess when assumption ITR is weakened to CTR. Now the instrumental variable is not z but rather $c(z^J)$. Even if z and $y(t^J)$ are statistically independent, $c(z^J)$ and $y(t^J)$ may be dependent because the function $c(\cdot)$ is person-specific. Thus, assumption SI may not hold even if treatments are randomly assigned. I illustrate below.

3.3.2. Illustration: The Supremum Interaction with Information Treatments

Let treatments be ordered items of information, as considered in our discussion of supremum interactions in Section 2.1. Let the population partition in two sub-populations, J_0 and J_1. Members of J_0 have individualistic response, while those in J_1 share the common reference group J_1. Thus, members of J_0 know only the information that they receive directly, while members of J_1 share information. It follows that

$c_j(z^J) = z_j$ for $j \in J_0$, while $c_j(z^J) = \sup (z_k, k \in J_1)$ for $j \in J_1$.

Let $\tau_{max} \equiv \max T$ exist. Consider a potential treatment vector that gives every member of the population a specified treatment $\tau^* < \tau_{max}$. Then $c_j(t^J) = \tau^*$ for all $j \in J$, and assumption SI reduces to $P[y(\tau^*)] = P[y(\tau^*)|c(z^J)]$. Application of (34) shows that $P[y(\tau^*)] = P[y|c(z^J) = \tau^*]$, provided that $P[c(z^J) = \tau^*] > 0$.

The informative component of assumption SI is $P[y(\tau^*)] = P[y(\tau^*)|c(z^J) = \tau^*]$. Consider its credibility. The left-hand side is the distribution of outcomes when every member of the population has information τ^*. The right-hand side is this distribution within the sub-population who actually know τ^* and nothing more. Suppose that $P(z = \tau^*|J_0) > 0$ and $P(z = \tau_{max}|J_1) > 0$. Then $c_j(z^J) = \tau^*$ for those members of J_0 with $z = \tau$ and $c_j(z^J) = \tau_{max}$ for all $j \in J_1$. Thus, the group with $[c(z^J) = \tau^*]$ contains only certain members of J_0.

Given the above, Assumption SI is credible only if one has reason to think that the members of J_0 with realized treatment τ^* have the same distribution of response as the population in toto. It is not credible otherwise. The credibility issue arises whatever assignment process generates the realized treatments. If realized treatments are randomly assigned, the assumption is credible only if one has reason to think that sub-populations J_0 and J_1 have the same distribution of response.

4. Models of Endogenous Social Interactions

4.1. Basic Concepts and Notation

Sections 2 and 3 studied identification of outcome distributions when shape restrictions and distributional assumptions are placed directly on the response functions $[y_j(\cdot), j \in J]$. Models of endogenous social interactions begin with specification of structural equations of the general form

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

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. Function f_j 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 *exogenous* interaction describes $t^{J/j}$ as an argument of f_j, while *endogenous* interaction describes $y^{J/j}(t^J)$. This terminology follows Manski (1993).

For each value of t^J, (38) is a system of simultaneous equations that jointly restrict the population outcome vector $y^J(t^J) \equiv [y_j(t^J), j \in J]$. When there exists a unique $y^J(t^J)$ that solves (38) for each $t^J \in T^J$, this is the structural model's reduced form. The terms *structural equations* and *reduced form* stem from the classical econometrics literature on linear simultaneous equations models. Although an endogenous-interactions model specifies (38) for all $t^J \in T^J$, the econometrics literature has traditionally written the system of equations only for the realized treatment vector $t^J = z^J$, leaving implicit its extension to potential treatment vectors. When $t^J = z^J$, (38) becomes $y_j = f_j(z_j, z^{J/j}, y^{J/j})$, $j \in J$.

Equation (38) is tautological in the absence of shape restrictions or distributional assumptions on the functions $[f_j(\cdot), j \in J]$. Models of endogenous interactions pose such assumptions. When the maintained assumptions imply restrictions on the reduced form, they may have identifying power for $P[y(t^J)]$. It has been particularly common in empirical research to pose linear models of endogenous interactions. These models yield linear reduced forms.

I use two linear models to illustrate. Section 4.2 poses a linear simultaneous equations model of interactions between pairs of persons. Section 4.3 examines the linear-in-means model of interactions in large reference groups. These models are simple and familiar. However, previous analysis has typically taken the objective to be identification of their structural parameters. Here, in contrast, the objective is identification of outcome distributions under potential treatment vectors.

Section 4.4 briefly considers incomplete models of endogenous interactions. These are models in which the system of equations (38) does not have a unique solution. Such models may nonetheless have identifying power for outcome distributions.

Hoping to model endogenous interactions without having to solve systems of simultaneous equations, researchers have sometimes posed dynamic versions of equation (38), supposing that current outcomes are functions of past treatments and outcomes. I will not consider such models here, but think it important to caution the reader that dynamic models have their own subtleties. Indeed, they often are more complex than simultaneous models rather than less so. This has long been recognized in the econometrics literature on linear structural models, which distinguished early on between general dynamic models and the special case of recursive models. Some analysis of dynamic versions of the linear-in-means and other anonymous interactions models appears in Manski (1997b).

4.2. Interactions Between Pairs of Persons

Let the population partition into a set K of reference groups composed of pairs of persons. Formally, $J = [(k1, k2), k \in K]$, where (k1, k2) is the ordered pair of persons in group k. For each $t^J \in T^J$ and $k \in K$, the outcomes $[y_{k1}(t^J), y_{k2}(t^J)]$ solve the pair of simultaneous equations

(39a) $\quad y_{k1}(t^J) = f_{k1}[t_{k1}, t_{k2}, y_{k2}(t^J)],$

(39b) $\quad y_{k2}(t^J) = f_{k2}[t_{k1}, t_{k2}, y_{k1}(t^J)].$

A model of this type might be posed to study labor supply in a population of married couples. Let the ordered pairs be husbands and wives. 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 social interactions occur only within couples, not between them. Within couples, a person's labor supply may depend on his or her own wage, the wage of the spouse (an exogenous interaction), and the spouse's labor supply (an endogenous interaction).

4.2.1. The Linear Model

In the absence of further assumptions, the solution to equations (39) may be any function of (t_{k1}, t_{k2}). Thus, the endogenous-interactions model thus far restricts the response functions $[y_j(\cdot), j \in J]$ only by requiring that interactions occur within the reference groups K. Imposition of further assumptions may yield stronger restrictions on response functions.

Empirical researchers often assume the linear model

$$(40a) \quad y_{k1}(t^J) = \alpha_1 + \beta_{11}t_{k1} + \beta_{12}t_{k2} + \gamma_1 y_{k2}(t^J) + u_{k1},$$

$$(40b) \quad y_{k2}(t^J) = \alpha_2 + \beta_{21}t_{k1} + \beta_{22}t_{k2} + \gamma_2 y_{k1}(t^J) + u_{k2}.$$

Here $(\alpha_1, \ldots, \gamma_2)$ are parameters and (u_{k1}, u_{k2}) are person-specific variables. Unless $\gamma_1\gamma_2 = 1$, the implied reduced form is

$$(41a) \quad y_{k1}(t^J) = \frac{\alpha_1 + \gamma_1\alpha_2}{1 - \gamma_1\gamma_2} + \frac{\beta_{11} + \gamma_1\beta_{21}}{1 - \gamma_1\gamma_2} t_{k1} + \frac{\beta_{12} + \gamma_1\beta_{22}}{1 - \gamma_1\gamma_2} t_{k2} + \frac{u_{k1} + \gamma_1 u_{k2}}{1 - \gamma_1\gamma_2}$$

$$\equiv \varphi_{10} + \varphi_{11}t_{k1} + \varphi_{12}t_{k2} + e_{k1},$$

$$(41b) \quad y_{k2}(t^J) = \frac{\alpha_2 + \gamma_2\alpha_1}{1 - \gamma_1\gamma_2} + \frac{\beta_{21} + \gamma_2\beta_{11}}{1 - \gamma_1\gamma_2} t_{k1} + \frac{\beta_{22} + \gamma_2\beta_{12}}{1 - \gamma_1\gamma_2} t_{k2} + \frac{u_{k2} + \gamma_2 u_{k1}}{1 - \gamma_1\gamma_2}$$

$$\equiv \varphi_{20} + \varphi_{21}t_{k1} + \varphi_{22}t_{k2} + e_{k2},$$

where $\varphi \equiv (\varphi_{10}, \ldots, \varphi_{22})$ are composite parameters and (e_{k1}, e_{k2}) are composite variables.

The linear model of endogenous interactions posed in (40) implies that the response functions derived in (41) are linear in treatments. Slope parameters may vary within reference groups but not across them. For example, if applied to study the labor supply of couples, (41) permits husbands and wives to have different slope parameters, but it requires all husbands (and all wives) to have the same labor supply responses to wages.

Although model (40) makes strong assumptions that may lack foundation, it does not yet enable prediction of outcomes under potential treatment vectors. The reason is that the model does not yet restrict the person-specific variables (u_{k1}, u_{k2}), $k \in K$. Researchers have entertained various assumptions, with the objective of point identification of the structural parameters $(\alpha_1, \ldots, \gamma_2)$. However, our objective is identification of outcome distributions under potential treatment vectors, not identification of the structural parameters.

Leaving aside the question of credibility, a simple way to point-identify outcome distributions is to assume that $E(u_1, u_2 | z_1, z_2) = 0$. Given this mean-independence assumption, evaluation of (41) with the realized treatments and outcomes yields the linear mean regressions

(42a) $E(y_1 | z_1, z_2) = \varphi_{10} + \varphi_{11} z_1 + \varphi_{12} z_2,$

(42b) $E(y_2 | z_1, z_2) = \varphi_{20} + \varphi_{21} z_1 + \varphi_{22} z_2.$

The empirical evidence reveals $E(y_1 | z_1, z_2)$ and $E(y_2 | z_1, z_2)$ on the support of (z_1, z_2). Hence, the parameters φ are point-identified if the support of $(1, z_1, z_2)$ is not contained in a linear subspace of R^3.

From here, a short argument shows that $P[y(t^J)]$ is point-identified. Knowledge of φ, combined with observation of realized treatments and outcomes, implies knowledge of $[(e_{k1}, e_{k2}), k \in K]$. Knowledge of φ and $[(e_{k1}, e_{k2}), k \in K]$ implies knowledge of all of the response functions $[y_{k1}(\cdot), y_{k2}(\cdot), j \in J]$. This yields knowledge of $P[y(t^J)]$ for all $t^J \in T^J$.

Observe that point-identification of the six reduced-form parameters φ does not imply point-identification of the eight structural parameters $(\alpha_1, \ldots, \gamma_2)$. Indeed, the structural parameters are not point-identified under the assumptions maintained above. This illustrates the general point that identification of $P[y(t^J)]$ is not equivalent to identification of structural equations.

4.3. Anonymous Interactions in Large Reference Groups

Let the population partition into a set of reference groups characterized by values for a covariate x. Thus, all persons with the same value of x belong to the same reference group. Let each reference group be large, and suppose that anonymous treatment and outcome interactions may occur within each group. Thus, for each $t^J \in T^J$ and $x \in X$, outcome $y_j(t^J)$ solves the equation

$$(43) \quad y_j(t^J) = f_j\{t_j, P(t|x_j), P[y(t^J)|x_j]\},$$

where $P(t|x_j)$ and $P[y(t^J)|x_j]\}$ are the distributions of treatments and outcomes in the group who have covariate value x_j.

A model of this type might be posed to study illness from an infectious disease. Let reference groups be metropolitan areas. Let the outcome of interest measure health status. Let the treatment be binary, taking the value one if a person is vaccinated and zero otherwise. One may think it reasonable to assume that interactions occur only within metropolitan areas, not between them. Within an area, illness may depend on a person's own vaccination status, on the vaccination rate for the area population (an exogenous interaction), and on the illness rate for the area population (an endogenous interaction).

4.3.1. The Linear-in-Means Model

In the absence of further assumptions, the endogenous-interactions model restricts the response functions $[y_j(\cdot), j \in J]$ only by requiring that interactions be anonymous within the reference groups defined by values of x. Imposition of further assumptions may yield stronger restrictions on response functions.

Empirical researchers often assume the linear-in-means model

$$(44) \quad y_j(t^J) = \alpha + \beta_1 t_j + \beta_2 E(t|x_j) + \gamma E[y(t^J)|x_j] + u_j.$$

Here $(\alpha, \beta_1, \beta_2, \gamma)$ are parameters and u_j is a person-specific variable. Taking expectations of both sides, conditional on x_j, yields the equilibrium condition

$$(45) \quad E[y(t^J)|x_j] = \alpha + (\beta_1 + \beta_2)E(t|x_j) + \gamma E[y(t^J)|x_j] + E(u|x_j).$$

Unless $\gamma = 1$, the unique equilibrium value of $E[y(t^J)|x_j]$ is

$$(46) \quad E[y(t^J)|x_j] = \frac{\alpha}{1-\gamma} + \frac{\beta_1 + \beta_2}{1-\gamma} E(t|x_j) + \frac{E(u|x_j)}{1-\gamma}.$$

Insertion of the right-hand side of (46) into (44) yields the reduced form

$$(47) \quad y_j(t^J) = \frac{\alpha}{1-\gamma} + \beta_1 t_j + \frac{\gamma\beta_1 + \beta_2}{1-\gamma} E(t|x_j) + \frac{\gamma}{1-\gamma} E(u|x_j) + u_j.$$

Thus, the linear model of endogenous interactions posed in (44) implies that the response functions derived in (47) are linear in treatments, the slope parameters for own treatments and group-mean treatments being the same for all members of the population.

Although model (44) makes strong assumptions that may lack foundation, it does not yet enable prediction of outcomes under potential treatment vectors. The reason is that the model does not yet restrict the person-specific variables $(u_j, j \in J)$. Researchers have studied identification of the structural parameters under various assumption; see Manski (1993). However, our objective is identification of outcome distributions under potential treatment vectors, not identification of the structural parameters.

Leaving aside the question of credibility, a simple approach is to impose the mean-independence assumption $E(u|z, x) = 0$. With this assumption, evaluation of (47) with the realized treatments and outcomes yields the linear mean regression

$$(48) \quad E(y|z, x) = \frac{\alpha}{1-\gamma} + \beta_1 z + \frac{\gamma\beta_1 + \beta_2}{1-\gamma} E(z|x) \equiv \varphi_0 + \varphi_1 z + \varphi_2 E(z|x),$$

where $\varphi \equiv (\varphi_0, \varphi_1, \varphi_2)$ are composite parameters. The empirical evidence reveals $E(y|z, x)$ on the support of (z, x). Hence, the parameters φ are point-identified if the support of $[1, z, E(z|x)]$ is not contained in a linear subspace of R^3.

From here, a short argument shows that $P[y(t^J)]$ is point-identified. Knowledge of φ, combined with observation of realized treatments and outcomes, implies knowledge of $(u_j, j \in J)$. Knowledge of φ and $(u_j, j \in J)$ implies knowledge of all of the response functions $[y_j(\cdot), j \in J]$. This yields knowledge of $P[y(t^J)]$ for all $t^J \in T^J$.

Observe that point-identification of the three reduced-form parameters φ does not imply point-identification of the four structural parameters $(\alpha, \beta_1, \beta_2, \gamma)$. Structural parameter β_1 is point-identified under the assumptions maintained above, but $(\alpha, \beta_2, \gamma)$ are not. This again illustrates that identification of $P[y(t^J)]$ is not equivalent to identification of structural equations.

It is intriguing that both here and in Section 4.2, we found identification of $P[y(t^J)]$ to be a less difficult problem than identification of structural equations. It would be interesting to learn the extent to which this finding generalizes to other structural equation systems with unique solutions.

4.4. Incomplete Models of Endogenous Interactions

To conclude this discussion of endogenous-interactions models, I briefly consider inference when the structural equations do not have a unique solution. Tamer (2003) has described cases where (38) has multiple solutions as *incomplete* models. I will use this term more broadly to describe cases where (38) has multiple solutions or no solution. Incomplete models are not abnormal. Structural equations with multiple solutions describe endogenous interactions with multiple equilibria, while those with no solutions describe interactions with no equilibria. A researcher may reasonably pose such models.

Suppose that a system of structural equations has multiple solutions for a specified potential treatment vector t^J. Then the model does not imply a unique outcome distribution $P[y(t^J)]$. However, it may still restrict the feasible distributions. Thus, the model may have identifying power.

The situation differs when the structural equations have no solution. Then the model is silent on the outcome distribution. Hence, it has no identifying power.

To go beyond these general statements requires study of particular incomplete models. Consider, for example, binary response versions of the linear models examined in Sections 4.2 and 4.3. When outcomes are binary, researchers studying interactions between pairs of persons have posed the model

(49a) $y_{k1}(t^J) = 1[\alpha_1 + \beta_{11}t_{k1} + \beta_{12}t_{k2} + \gamma_1 y_{k2}(t^J) + u_{k1} > 0]$,

(49b) $y_{k2}(t^J) = 1[\alpha_2 + \beta_{21}t_{k1} + \beta_{22}t_{k2} + \gamma_2 y_{k1}(t^J) + u_{k2} > 0]$.

Researchers studying anonymous interactions in large reference groups have posed the model

$$(50)\quad y_j(t^J) = 1[\alpha + \beta_1 t_j + \beta_2 E(t|x_j) + \gamma E[y(t^J)|x_j] + u_j > 0].$$

Tamer (2003) has studied identification of the structural parameters of the former model, which has a unique equilibrium for some values of the parameter and person-specific variables, multiple equilibria for others, and no equilibrium for yet others. Brock and Durlauf (2001) have studied identification of the structural parameters of the latter model, which may have a unique or multiple equilibria. However, use of these models to identify outcome distributions under potential treatment vectors is a distinct question, which remains open.

5. Conclusion

This paper has developed a formal language for study of treatment response with social interactions, and has used it to derive a spectrum of new findings on identification. The basic idea is simple. I first defined individual treatment response to be a function of the entire vector of treatments received by the population. I then studied identification when shape restrictions and distributional assumptions are placed on the response functions.

An early key result was that the traditional assumption of individualistic response is a polar case within the broad class of constant response assumptions, the other pole being unrestricted interactions. Important non-polar cases are interactions within reference groups and anonymous interactions. I showed that findings on identification obtained earlier under assumption ITR extend to assumption CTR. These include identification with assumption CTR alone and when this shape restriction is strengthened to semi-monotone response.

I next studied distributional assumptions using instrumental variables. Here too, findings obtained previously under assumption ITR extend when assumptions of statistical or mean independence are invoked in settings with social interactions. However, these positive findings had to be tempered by two caveats. First, the extended version of assumption SI has no power to identify counterfactual outcome distributions when social interactions are unrestricted. Second, when interactions are restricted, the extended assumption may not be credible even if treatments are randomly assigned.

The final part of the paper considered models of endogenous social interactions. I emphasized that identification of structural equations differs from identification of outcome distributions under potential treatment vectors. Analysis of two familiar linear models illustrated this general point. I briefly discussed incomplete models and noted that researchers sometimes specify dynamic rather than simultaneous models of endogenous interactions. However, this paper has not sought to study the use of incomplete or dynamic models to identify counterfactual outcome distributions.

References

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

Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, 80, 319–323.

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

Manski, C. (1997a), "Monotone Treatment Response," *Econometrica*, 65, 1311–1334.

Manski, C. (1997b), "Identification of Anonymous Endogenous Interactions," in B. Arthur, S. Durlauf, and D. Lane (editors), *The Economy as an Evolving Complex System II*, Reading, Mass.: 369-384.

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

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.

Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147-167.