cemmap
centre for microdata methods and practice

Identification of treatment response with social interactions

Charles F. Manski

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP01/10

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

IDENTIFICATION OF TREATMENT RESPONSE WITH SOCIAL INTERACTIONS

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

First Public Draft: October 7, 2009
This Draft: April 3, 2010

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 distributional 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. However, I find that random assignment of realized treatments generically has no identifying power when some persons are *leaders* who may affect outcomes throughout the population. Finally, I consider use of models of endogenous social interactions to derive restrictions on response functions. I emphasize that identification of potential outcome distributions differs from the longstanding econometric concern with identification of structural functions.

This research was supported in part by National Science Foundation grant SES-0911181. I have benefitted from the opportunity to present this work at an invited session of the 2010 North American Winter Meeting of the Econometric Society and in seminars at NYU, Northwestern, UCL, and the University of Munich. I am grateful for comments from Mike Fu and Guido Imbens.

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. Some researchers have called this "no interference between units" or the Stable Unit Treatment Value Assumption. I have called 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), 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 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. A clear and concise formal 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 outcomes, and realized treatments. A common research objective has been

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 $y_j(\cdot)$: $T^J \to Y$, 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 a specified feasible 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). Instances of such planning problems are studied in Graham, Imbens, and Ridder (2009) and Manski (2010).

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

The notation introduced above suffices to cover many but not all interesting problems of treatment response with social interactions. It covers cases in which a planner assigns treatments to individual

members of the population, perhaps medical or educational or economic treatments, who then interact with one another to produce outcomes. It also covers cases in which a planner allocates persons to groups within which interactions occur; for example, the treatments in T may be school classrooms, hospital wards, or prisons. Perhaps the primary restriction of the notation is that it portrays a static planning problem. The notation does not cover dynamic problems, where initial treatment assignments are followed by the realization of initial outcomes, then assignment of further treatments, and so on.

Identification of $P[y(t^J)]$

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 $[P(y \mid z = t)P(z = t) + \delta P(z \neq t), \delta \in \Delta_Y]$, 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.

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 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. The size and shape of $H\{P[y(t^J)]\}$ depends on the assumptions imposed and the treatment vector t^J under consideration.

Innumerable assumptions and potential treatment vectors may be of interest. This paper is entirely general with respect to the potential treatment vector. However, I cannot constructively analyze all possible assumptions. To make progress, I consider some simple nonparametric assumptions that may be credible in applications. Section 2 studies two shape restrictions on response functions, *constant treatment response* (CTR) and *semi-monotone treatment response* (SMTR). Section 3 combines these shape restrictions with statistical-independence assumptions that use instrumental variables. Section 4 considers identification using models of endogenous interactions. Section 5 concludes, calling attention to some of the many potentially fruitful directions for future research. The remainder of this Introduction gives a more detailed overview of what lies ahead.

Restrictions on the Shape of Response Functions

Section 2 begins with the assumption of constant treatment response, which states 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*. 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.

As defined here, reference groups are person-specific but non manipulable. Person-specific means that person k may be a member of person j's group but not vice versa. Non-manipulability means that a planner cannot use the treatments in T to alter a person's reference group. Prediction of outcomes when treatments can manipulate group composition is not discussed in this paper.

An important subclass of interactions within reference groups assumes that interactions are *distributional*. 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 *mean interaction*, applicable when treatments are real-valued, assumes that a person's outcome may vary with the mean treatment of others in the group. A *supremum interaction*, applicable when treatments are ordered, assumes that a person's outcome may vary with the largest treatment in the 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 (1997) 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.

Although this paper is about identification, I would be remiss to entirely ignore estimation with sample data. To conclude Section 2, I suppose that one poses a version of assumption CTR and observes a random sample of the population. I show that the identification region for any potential outcome distribution may be consistently estimated if one observes the realized outcome and effective treatment of each sample member. However, it generally does not suffice to observe the realized outcome and own treatment for each member.

Looking beyond this paper, the reader should be aware that assumptions CTR and SMTR address the existence and sign of treatment interactions but do not restrict magnitudes. One may want to relax the binary idea that person k either is or is not a member of person j's reference group. Instead, one might a priori bound the extent to which k's treatment may affect j's outcome. Or one may find it credible to strengthen assumption SMTR by posing concavity, convexity, or other second-order shape restrictions. A possible starting point for future analysis is the Manski (1997) study of concave-monotonicity under assumption ITR.

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 independent; that is, $P[y(\tau)|v] = P[y(\tau)]$.

The statistical-independence assumption extends 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 $P[y(t^J)|v] = P[y(t^J)]$. I first characterize the identifying power of this assumption 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 $P[y(\tau)]$ if and only if $P(z = \tau) > 0$. I extend the argument underlying this result to inference under assumption CTR, using effective treatments as the instrumental variable. 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.

I apply the analysis to settings in which the population consists of *leaders* and *followers*. The treatments received by leaders may affect the outcomes of all members of the population, but those received by followers may affect only the person receiving treatment. An important negative finding is that random assignment of realized treatments generically has no identifying power when the population contains leaders.

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 compare the identifying power of complete and incomplete models. Analysis of the linear-in-means model of interactions in large groups illustrates that point identification of a complete model is not necessary for point identification of potential outcome distributions.

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. Section 2.4 discusses estimation of identification regions from random sample data.

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 distributional 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 $c_j(\cdot)$: $T^J \to C_j$ be a 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 $c_j(\cdot)$. With this in mind, I shall refer to C_j as the set of *effective treatments* for person j.

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,

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

Here $P[c(z^J) = c(t^J)]$ is the fraction of the population for whom $[c(z^J) = c(t^J)]$, and $P[y(t^J) \mid c(z^J) = c(t^J)]$ 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 $P[y(t^J) \mid c(z^J) = c(t^J)] = P[y \mid c(z^J) = c(t^J)]$. Observation of realized treatments and outcomes reveals $P[y \mid c(z^J) = c(t^J)]$ when $P[c(z^J) = c(t^J)] > 0$. The empirical evidence and assumption CTR are uninformative about the counterfactual outcome distribution $P[y(t^J) \mid c(z^J) \neq c(t^J)]$. Hence, the identification region for $P[y(t^J)]$ is

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

Observe that the size of $\mathrm{H}\{P[y(t^J)]\}$ 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

Concepts and Notation

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 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. At the extreme, the reference group for person j might be the entire population while that of person k might be this person alone. I will later (see Section 3.3) consider a strong form of asymmetry in which the population consists of *leaders* and *followers*. The treatments received by leaders may affect the outcomes of all members of the population, but those

received by followers may affect only the person receiving treatment.

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 who interact 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. The general idea of assumption CTR covers cases in which a planner can manipulate the group with whom a person interacts.

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 (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

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

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 (4) becomes

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

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, say τ, and $t^J = (\tau, \tau, \ldots, \tau)$. 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 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 (5), which equals Δ_Y whenever $P(z = t) < 1$.

2.1.3. Distributional Interactions

Region (4) characterized identification under the sole assumption that interactions occur within reference groups. Applied research often assumes that interactions are distributional. A *distributional*

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 distributional-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 $Q(t^{G(j)/j})$ 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 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 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 within-group 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 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 is 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, let $C_j = T \times \Phi$, where Φ is the range space for F. Let $c_j(t^J) = [t_j, F(t^{G(j)/j})]$. Then (3) becomes

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

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

This region is a subset of the region (7) 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.

2.2. 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 distributional 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 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 $c < c'$ or $c > c'$ or (c, c') are unordered, in which case I write $c \varnothing c'$. 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 that

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

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 treatment 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 $y_j(t^J)$:

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

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

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

where $\ge_{sd}$ 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 (1997), 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) $[t_k \ge s_k$, 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 Distributional 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 $s^{G(j)}$. 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

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

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 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 (15) is 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. I will use it in Section 2.3.

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

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

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.

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

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 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 permit distributional interactions and then specialize to reinforcing distributional interactions.

2.3.1. Individualistic Treatment 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 (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 ≠ t) = q − p. Observe that P(z ≠ 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

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

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 ⇒ 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

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

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, $P(y = 1 \mid t \geq z) = P(y = 1)$ and $P(y = 1 \mid t \leq z) = P(y = 1 \mid t = z)$, whose value was derived in (21). The result is

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

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. Distributional 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 (7), which becomes

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

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 Distributional 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 (15) holds. Suppose that the population is large, so $Q(t^{G(j)/j}) \cong Q(t^{G(j)})$ and $Q(s^{G(j)/j}) \cong Q(s^{G(j)})$. Using this approximation, (15) 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

$$(26)\quad H\{P[y(t^J) = 1]\} = [P[y = 1 \mid (t, q) \geq (z, p)]\cdot P[(t, q) \geq (z, p)],$$
$$P[(t, q) > (z, p) \cup (t, q) \oslash (z, p)] + P[y = 1 \mid (t, q) \leq (z, p)]\cdot P[(t, q) \leq (z, p)]].$$

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

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

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

2.4. Estimation of Identification Regions with Data on a Random Sample of the Population

The identification analysis of Sections 2.1 and 2.2 supposed that one observes $[c_j(\cdot), y_j, z_j; j \in J]$. Hence, for any potential treatment vector t^J, one observes $[c_j(z^J), c_j(t^J), y_j; j \in J]$. This enables computation of the identification regions for $P[y(t^J)]$ under assumptions CTR and SMTR, given in (3) and (11) respectively. To recall, these are

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

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

Now suppose that one does not observe $[c_j(z^J), c_j(t^J), y_j; j \in J]$. Instead, one draws a random sample of N persons, say J_N, and observes $[c_j(z^J), c_j(t^J), y_j; j \in J_N]$. Then one may estimate regions (3) and (11) by their sample analogs

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

(11′) $H\{P_N[y(t^J)]\} \equiv \{\delta \in \Delta_Y: P_N[y_U(t^J)] \geq_{sd} \delta \geq_{sd} P_N[y_L(t^J)]\},$

where P_N denotes the empirical distribution of J_N. If the population is uncountable, the laws of large numbers

for random sampling imply that each $H\{P_N[y(t^J)]\}$ converges in various senses to the corresponding $H\{P[y(t^J)]\}$ as $N \to \infty$. Thus, regions (3) and (11) may be estimated consistently.

This argument requires observation of sample members' realized effective treatments $[c_j(z^J), j \in J_N]$, not just their realized own treatments $(z_j, j \in J_N)$. Excepting the special case of individualistic treatment response, the effective treatments of sample members generically depend on the treatments received by non-sample members. If one assumes only that sample member $j \in J_N$ has reference group G(j), one must observe all of the treatments $[z_k, k \in G(j)]$. If assumes a functional interaction, then it suffices to observe z_j and $F(z^{G(j)/j})$. Importantly, one does not need to observe the outcomes realized by non-sample members.

Observation of the treatments received by non-sample members 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.

When a researcher only observes the own treatments of sample members, not their effective treatments, random sampling does not enable consistent estimation of identification regions (3) and (11). In the presence of social interactions, observation of $(z_j, j \in J_N)$ generically does not reveal $[c_j(z^J), j \in J_N]$. Hence, one cannot determine how $c_j(z^J)$ is related to $c_j(t^J)$ for $j \in J_N$.

Other sampling processes may enable consistent estimation in some settings. Suppose, for example, that the population partitions into uncountably many reference groups, with a finite upper bound on group size. Then random sampling of groups, with observation of the realized treatments and outcomes of all members of each sampled group, enables consistent estimation of potential outcome distributions.

Researchers sometimes face sampling processes with intermediate inferential problems, where effective treatments are partially observed. See Sojourner (2009) for analysis of treatment response in one such setting.

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 $v \equiv v(x^J, z^J)$ and $w \equiv w(x^J, z^J)$. To simplify the exposition, I will suppress w until Section 3.3, where it is necessary to make it partially 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, the objective being inference on $P[y(t^J)]$ for a specified $t^J \in T^J$. Then one may assume that $P[y(t^J)|v] = P[y(t^J)]$ or $E[y(t^J)|v] = E[y(t^J)]$.

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.

The discussion in Section 2.4 of estimation from sample data applies as well to the analysis of Section 3. Hence, I will not discuss estimation here.

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 $P[y(t^J) \mid v = \nu]$, 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 $H\{P[y(t^J) \mid v = \nu]\}$ denote the resulting identification region. The identification region for the collection of distributions $\{P[y(t^J) \mid v = \nu], \nu \in V\}$ is the Cartesian product $\times_{\nu \in V} H\{P[y(t^J) \mid v = \nu]\}$. 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 $P[y(t^J) \mid v] = P[y(t^J)]$. Then $P[y(t^J)]$ must lie within the intersection of the identification regions $H\{P[y(t^J) \mid v = \nu]\}$, $\nu \in V$. 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

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

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

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

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 persons with v = 1; thus, $P(z = 1 \mid v = 0) < P(z = 1 \mid v = 1)$. 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, $P(t = 1 \mid v = 0) = P(t = 1 \mid v = 1) = q$. The objective is to learn $P[y(t^J) = 1]$.

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

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

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

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

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 distributional interaction is combined with assumption SI, namely $P[y(t^J) = 1 \,|\, v = 0] = P[y(t^J) = 1 \,|\, v = 1]$. Then $\overline{H}\{P[y(t^J) = 1]\}$ is the intersection of the identification regions obtained above for $P[y(t^J) = 1 \,|\, v = 0]$ and $P[y(t^J) = 1 \,|\, v = 1]$. Thus,

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

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 distributional, 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 per se 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 v is a valid instrumental variable.

3.3. Using Realized Effective Treatments as Instrumental Variables

In research under assumption ITR, it is common to let v = z and assume that $P[y(\tau)] = P[y(\tau) \,|\, z]$ for

a specified $\tau \in T$. This version of assumption SI may be motivated by knowledge that realized treatments were randomly assigned within a large population. 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 $P(z = \tau) > 0$.

This reasoning extends directly to situations in which a large population partitions into a continuum of reference groups, each of bounded finite size. For example, the groups might be husband-wife pairs or they might be pairs of twins. Then assumption ITR holds when the population is defined to be the collection of groups rather than persons.

I show here that the reasoning extends further to general analysis of treatment response under assumption CTR, the instrumental variable now being the realized effective treatment $c(z^J)$. Section 3.3.1 derives the extended result in abstraction. Sections 3.3.2 applies it to interactions with leaders and followers. Section 3.3.3 develops implications for inference with random assignment of realized treatments.

3.3.1. Identification Using Assumptions CTR and SI

Let the previously suppressed conditioning covariates w include $c(t^J)$ and perhaps other functions of x^J. It is important to the present analysis to explicitly condition on $c(t^J)$, but I will continue to suppress other components of w to simplify the exposition. I will assume that $P[y(t^J) | c(t^J)] = P[y(t^J) | c(t^J), c(z^J)]$. This extension of assumption SI reduces to the familiar $P[y(\tau)] = P[y(\tau) | z]$ when $t^J = (\tau, \ldots, \tau)$ and response is individualistic.

Let C denote the set of all effective treatments, and let $C(t^J)$ be the subset of C that occur with positive empirical probability in t^J. Thus, $C \equiv \cup_{j \in J} C_j$ and $C(t^J) \equiv \{\gamma \in C: P[c(t^J) = \gamma] > 0\}$. To enable use of elementary probability theory, I will assume that C is countable.

Successively apply the Law of Total Probability, assumption SI, 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 $\gamma \in C(t^J)$. 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

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

Here $C_1(t^J) \equiv \{\gamma \in C(t^J)\colon P[c(z^J) = \gamma \,|\, c(t^J) = \gamma] > 0\}$ are the elements of $C(t^J)$ that occur with positive empirical probability in z^J, and $C_0(t^J) \equiv \{\gamma \in C(t^J)\colon P[c(z^J) = \gamma \,|\, c(t^J) = \gamma] = 0\}$ are the elements with zero empirical probability.

Point identification occurs if and only if $C_0(t^J)$ is empty. Then the final term in (34) disappears. Contrariwise, assumption SI and the empirical evidence are uninformative if $C_1(t^J)$ is empty. Then (34) reduces to $H\{P[y(t^J)]\} = \Delta_Y$.

An Alternative Derivation

Derivation of (34) did not use the general finding (28) expressing the identification region with assumption SI as the intersection of the regions across values of the instrumental variable. We can alternatively use (28) to obtain (34).

For each $\gamma \in C(t^J)$, application of (28) gives

$$(35) \quad H\{P[y(t^J)\,|\,c(t^J) = \gamma]\} = \bigcap_{\gamma' \in C} H\{P[y(t^J)\,|\,c(t^J) = \gamma,\, c(z^J) = \gamma']\}.$$

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 (35) 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. This result and the first equality in (33) imply (34).

3.3.2. Interactions with Leaders and Followers

It is illuminating to apply result (34) to settings where the population is composed of leaders and followers. Suppose that treatments received by leaders may affect the outcomes of all members of the population, but those received by followers may affect only the person receiving treatment. Assumption ITR is the polar case in which all persons are followers. An unrestricted interaction is the polar case in which all persons are leaders and no other assumptions are imposed.

Let $L \subset J$ denote the sub-population of leaders, and let its complement $M = J/L$ denote the followers. Then $c_j(t^J) = t^L$ and $c_j(z^J) = z^L$ for all $j \in L$, while $c_j(t^J) = (t^L, t_j)$ and $c_j(z^J) = (z^L, z_j)$ for all $j \in M$. Hence, assumption SI consists of the equation $P[y(t^J)\,|\,c(t^J) = t^L] = P[y(t^J)\,|\,c(t^J) = t^L, z^L]$ and the set of equations $P[y(t^J)\,|\,c(t^J) = (t^L, \tau)] = P[y(t^J)\,|\,c(t^J) = (t^L, \tau), (z^L, z)]$, $\tau \in T$. The former equation, which applies to leaders, holds tautologically because treatment vector z^L is constant across the population. The latter equations, which apply to followers, hold if M is a large sub-population whose realized treatments were randomly assigned.

In this setting, the set $C(t^J)$ contains the element t^L if $P(L) > 0$. For each $\tau \in T$, it contains the element

(t^L, τ) if $P(M) > 0$ and $P(t = \tau \mid M) > 0$. The identification region for $P[y(t^J)]$ depends critically on whether $z^L = t^L$. If this equality does not hold, then $C_1(t^J)$ is empty and (34) reduces to $H\{P[y(t^J)]\} = \Delta_Y$. Thus, equality of the potential and realized treatments of leaders is a necessary condition for inference.

If $z^L = t^L$, then $C_1(t^J)$ contains t^L if $P(L) > 0$. For each $\tau \in T$, it contains (t^L, τ) if $P(M, t = \tau) > 0$ and $P(z = \tau \mid M, t = \tau) > 0$. Hence, result (34) becomes

$$(36) \quad H\{P[y(t^J)]\} = \{P(y \mid L)\cdot P(L) + \sum_{\tau \in T_{M1}(t^J)} P(y \mid M, t = \tau, z = \tau)\cdot P(M, t = \tau) + \delta \cdot P[M, t \in T_{M0}(t^J)],\ \delta \in \Delta_Y\},$$

where $T_M(t^J) \equiv [\tau \in T: P(t = \tau \mid M) > 0]$, $T_{M1}(t^J) \equiv [\tau \in T_M(t^J): P(z = \tau \mid M, t = \tau) > 0]$ and $T_{M0}(t^J) \equiv [\tau \in T_M(t^J): P(z = \tau \mid M, t = \tau) = 0]$. Point identification occurs if and only if $T_{M0}(t^J)$ is empty. Then the final term in (36) disappears.

Distributional and Functional Interactions with a Large Sub-Population of Leaders

The above derivation assumed only that the population is composed of leaders and followers. Given further assumptions, equality of the potential and realized treatments of leaders may not be necessary for inference. It is instructive to consider the case in which a large sub-population of leaders generates distributional or functional interactions.

Consider a distributional interaction. Then $c_j(t^J) = [Q(t^L), t_j]$ and $c_j(z^J) = [Q(z^L), z_j]$ for all $j \in J$. Hence, assumption SI consists of the equations $P[y(t^J) \mid c(t^J) = (Q(t^L), \tau)] = P[y(t^J) \mid c(t^J) = (Q(t^L), \tau), (Q(z^L), z)]$, $\tau \in T$. The conditioning on $Q(z^L)$ holds tautologically because this treatment distribution is constant across the population. The conditioning on z holds if realized treatments were randomly assigned.

In this setting, $C(t^J) = [(Q(t^L), \tau); \tau \in T \text{ s. t. } P(t = \tau) > 0]$. The identification region for $P[y(t^J)]$ depends critically on whether $Q(z^L) = Q(t^L)$. If this equality does not hold, then set $C_1(t^J)$ is empty and (34) reduces to $H\{P[y(t^J)]\} = \Delta_Y$. Thus, equality of leaders' distributions of potential and realized treatments is

necessary for inference.

If $Q(z^L) = Q(t^L)$, then $C_1(t^J) = [(Q(t^L), \tau) \in C(t^J)$ s. t. $P(z = \tau \,|\, z = \tau) > 0]$. Result (34) becomes

$$(37) \quad H\{P[y(t^J)]\} = \{ \sum_{\tau \in T_1(t^J)} P(y \,|\, t = \tau, z = \tau) \cdot P(t = \tau) + \delta \cdot P[t \in T_0(t^J)], \ \delta \in \Delta_Y \},$$

where $T(t^J) \equiv [\tau \in T: P(t = \tau) > 0]$, $T_1(t^J) \equiv [\tau \in T(t^J): P(z = \tau \,|\, t = \tau) > 0]$ and $T_0(t^J) \equiv [\tau \in T(t^J): P(z = \tau \,|\, t = \tau) = 0]$. Point identification occurs if and only if $T_0(t^J)$ is empty. Then the final term in (37) disappears.

The argument is analogous with a functional interaction. One need only replace the distribution Q with the functional F throughout the above derivation. The functional equality $F(z^L) = F(t^L)$ is weaker than the distributional equality $Q(z^L) = Q(t^L)$. Hence, assertion of a functional interaction enables inference in more settings than does assertion of a distributional interaction.

3.3.3. Random Assignment of Realized Treatments

This above analysis of interactions with leaders and followers has important negative implications for inference when realized treatments are randomly assigned. Random assignment does not yield a determinate vector z^J of realized treatments. Instead, it yields an ex ante probability distribution for z^J. The simplest randomization mechanism independently draws realized treatments from a common distribution. Let the ex ante distribution on T produced by the randomization mechanism be denoted π. Then $\pi(\tau)$ is the probability that a person is assigned to treatment τ.

The first part of Section 3.3.2 showed that random assignment of treatments to leaders generically has no identifying power. We found that $z^L = t^L$ is necessary for inference on $P[y(t^J)]$. If the sub-population L of leaders has finite cardinality, then the ex ante probability that $z^L = t^L$ is $\prod_{j \in L} \pi(t_j)$. This probability is generically less than one, the only exceptions occurring when $(t_j = \tau, j \in L)$ for some $\tau \in T$ and the

randomization mechanism sets $\pi(\tau) = 1$. These degenerate exceptions aside, random assignment always yields positive ex ante probability that $z^L \neq t^L$, in which case $H\{P[y(t^J)]\} = \Delta_Y$. If L is has infinite cardinality, the ex ante probability that $z^L \neq t^L$ is generically one.

The analysis assuming a distributional interaction showed that random assignment to leaders has identifying power only if π and $Q(t^L)$ coincide. We found that $Q(z^L) = Q(t^L)$ is necessary for inference on $P[y(t^J)]$. The assumption of a large sub-population of leaders implies that $Q(z^L) = \pi$. Hence, no inference is possible if $\pi \neq Q(t^L)$ and result (37) holds if $\pi = Q(t^L)$.

These negative findings do not appear in the classical analysis of random assignment, which assumes that response is individualistic. The reason is that there are no leaders under assumption ITR, only followers. Hence, random assignment point-identifies $P[y(t^J)]$ if $\pi(\tau) > 0$ for all τ such that $P(t = \tau) > 0$.

The conclusions drawn here send a cautionary message about the identifying power of random assignment. It has become commonplace to assert that random assignment enables prediction of potential outcome distributions under counterfactual treatments. This assertion is well-grounded under assumption ITR, but not necessarily so when treatments have social interactions. The nature of the interaction matters.

At one pole are situations in which a large population partitions into a continuum of reference groups, each of bounded finite size. Assumption ITR holds when the population is defined to be the collection of groups rather than persons, so the classical analysis of random assignment pertains. At the other pole are settings in which the population consists of leaders and followers, and no other assumptions are imposed. Here random assignment generically has no identifying power.

There are many intermediate settings where random assignment may have some identifying power but less than under assumption ITR. Section 3.3.2 showed this to be the case when leaders generate distributional or functional interactions. There is much scope for exploration of other intermediate cases.

4. Models of Endogenous Social Interactions

Sections 2 and 3 examined identification of potential outcome distributions when shape restrictions and distributional assumptions are placed directly on the response functions $[y_j(\cdot), j \in J]$. The assumptions studied there—constant treatment response, semi-monotonicity, and statistical independence—are relatively transparent. They often are easy to assess even though they approach treatment response as a black box.

It may also be productive to open the black box, model the process transforming treatments into outcomes, and use the model to derive restrictions on response functions. Econometricians have long studied models of endogenous social interactions, which posit that each person's outcome is a *structural function* of population treatments and outcomes. The econometric literature has focused on identification of the structural functions per se. Here I consider use of endogenous-interaction models to identify the distribution of treatment response. Section 4.1 poses the problem in abstraction.

4.1. The Problem in Abstraction

A model of endogenous interactions supposes that the potential outcome vector $y^J(t^J) \equiv [y_j(t^J), j \in J]$ solves the *structural equations*

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

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. 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 *exogenous* interaction describes $t^{J/j}$ as an argument of $f_j(\cdot)$, while *endogenous* interaction describes $y^{J/j}(t^J)$.

If $y^{J/j}(t^J)$ were not an argument, $f_j(\cdot)$ would simply be the person's response function. The presence of $y^{J/j}(t^J)$ as an argument of $f_j(\cdot)$ makes (38) a system of simultaneous equations.

Example 1: Consider labor 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 labor-supply interactions occur only within couples, not between them. Within each couple, 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). Formally, this gives the model

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

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

where (k1, k2) is the ordered pair of persons in couple k. □

Example 2: Consider 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). Formally, this gives the model

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

Here x_j denotes the metropolitan area within which person j resides, while $P(t|x_j)$ and $P[y(t^J)|x_j]$ are the

distributions of treatments and outcomes in this area.

Identification of Structural Functions and Reduced Forms

An outcome vector $y^J(t^J)$ that solves (38) is said to be a *reduced form* of the structural equations. A model is *complete* if (38) has a unique solution for all feasible structural functions. A model is incomplete if (38) 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. A researcher may reasonably pose such models. See, for example, Brock and Durlauf (2001) and Tamer (2003).

Research in econometrics has long been concerned with identification of structural functions. Observation of realized treatments and outcomes reveals that

$$(41) \quad y_j = f_j(z_j, z^{J/j}, y^{J/j}), \qquad j \in J.$$

Thus, the empirical evidence pins down one point on the structural function of each population member. Econometricians have studied identification when this evidence is combined with shape restrictions and distributional assumptions on f^J. Classical analysis of linear structural equations combines several strong assumptions to achieve point identification of f^J. Recent work on control functions and nonparametric instrumental variables explores identification when the classical assumptions are weakened in various ways.

The concern of this paper is identification of potential outcome distributions. Thus, I want to use endogenous-interactions models to identify the reduced forms of structural equations, not to identify structural functions per se. A model has identifying power for the reduced form if the empirical evidence and the maintained assumptions on f^J imply restrictions on $y^J(t^J)$. Our particular concern is $P[y(t^J)]$, the empirical distribution of $y^J(t^J)$.

The relationship between identification of structural functions and reduced forms is straightforward in classical analysis of linear structural equations. There, the parameters of the reduced form are a specific many-to-one function of the parameters of the structural function. See, for example, Goldberger (1991). It has occasionally been observed that interest may center on the reduced form rather than the structural functions. 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."

Outside of the classical linear context, 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 can make a start. Section 4.2 distinguishes the identifying power of complete and incomplete models. Section 4.3 examines the linear-in-means model of interactions in large groups.

4.2. The Identifying Power of Complete and Incomplete Models

When considering the use of endogenous-interactions models to identify potential outcome distributions, one should be careful to distinguish 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, (38) 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. Thus, 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 (38) 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 (38). 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. However, 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. This suggests, but does not formally imply that $H\{P[y(t^J)]\}$ contains multiple elements.

Finally, consider an incomplete model having no solution to (38) for some $f^J \in \Phi$. There are at least two distinct 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 common in game theory, where a finding that no equilibrium exists is taken to mean that the specified equilibrium concept makes no prediction about the actions chosen by players.

4.3. The Linear-in-Means Model of Interactions in Large Groups

This section uses a simple complete model to demonstrate the difference between identification of structural functions and reduced forms. I consider the linear-in-means model of interactions in large groups, which I have previously studied from the perspective of identification of structural functions (Manski, 1993).

Let the population partition into a finite set of large reference groups characterized by values for a covariate x. Thus, all persons with the same value of x belong to the same reference group. Empirical researchers often assume the linear-in-means model

$$(42) \quad y_j(t^J) = \alpha + \beta_1 t_j + \beta_2 E(t|x_j) + \gamma E[y(t^J)|x_j] + u_j.$$

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

$$(43) \quad E[y(t^J)|x_j] = \alpha + (\beta_1 + \beta_2)E(t|x_j) + \gamma E[y(t^J)|x_j] + E(u|x_j).$$

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

$$(44) \quad E[y(t^J)|x_j] = \frac{\alpha}{1-\gamma} + \frac{\beta_1 + \beta_2}{1-\gamma} E(t|x_j) + \frac{E(u|x_j)}{1-\gamma}.$$

Insertion of the right-hand side of (44) into (42) yields the reduced form

$$(45) \quad y_j(t^J) = \frac{\alpha}{1-\gamma} + \beta_1 t_j + \frac{\gamma\beta_1 + \beta_2}{1-\gamma} E(t|x_j) + \frac{\gamma}{1-\gamma} E(u|x_j) + u_j.$$

Thus, the linear model of endogenous interactions posed in (42) implies that the response functions derived in (45) 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 (42) makes strong assumptions, 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 assumptions; see Manski (1993). However, our objective is identification of potential outcome distributions, not identification of structural parameters.

A simple result emerge if one imposes the mean-independence assumption $E(u|z, x) = 0$. With this assumption, evaluation of (45) with the realized treatments and outcomes yields the linear mean regression

$$(46) \quad E(y|z, x) = \frac{\alpha}{1 - \gamma} + \beta_1 z + \frac{\gamma\beta_1 + \beta_2}{1 - \gamma} E(z|x) \equiv \phi_0 + \phi_1 z + \phi_2 E(z|x),$$

where $\phi \equiv (\phi_0, \phi_1, \phi_2)$ 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 illustrates that point identification of a complete endogenous-interactions model is not necessary for point identification of potential outcome distributions.

5. Conclusion

This paper has developed a formal language for study of treatment response with social interactions, and has used it to obtain new findings on identification of potential outcome distributions.

The analysis of shape restrictions in Section 2 began by showing 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 distributional interactions. I showed 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.

Section 3 studied distributional assumptions using instrumental variables. I showed that findings obtained previously under assumption ITR extend when assumptions of statistical independence (SI) are posed in settings with social interactions. In particular, I extended the well-known result on identification using realized treatments as an instrumental variable, with realized effective treatments now being the instrument. However, I found that random assignment of realized treatments generically has no identifying power when some persons are leaders who may affect outcomes throughout the population.

Section 4 observed that models of endogenous social interactions may be used to derive restrictions on response functions. I emphasized that the use of such models to identify potential outcome distributions differs from the longstanding econometric concern with identification of structural functions. I distinguished the identifying power of complete and incomplete models.

Looking beyond this paper, I see enormous scope for further work in all of the directions initiated here. Many shape restrictions beyond assumptions CTR and SMTR warrant attention. I think it highly important to obtain a more complete understanding of the identifying power of random assignment in settings with social interactions. Study of the use of models of endogenous interactions to derive restrictions on response functions would be welcome.

References

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

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

Graham, B., G. Imbens, and G. Ridder (2009), "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis," National Bureau of Economic Research Working Paper W14860.

Kiefer, N. and A. Goldberger (1989), "The ET Interview: Arthur S. Goldberger," *Econometric Theory*, 5, 133-160.

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

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

Manski, C. (2010), "Vaccination with Partial Knowledge of External Effectiveness," *Proceedings of the National Academy of Sciences*, 107, 3953-3960.

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.

Sojourner, A. (2009), "Inference on Peer Effects with Missing Peer Data: Evidence from Project STAR," Carlson School of Management, University of Minnesota.

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