Abstract

This paper studies identification of potential outcome distributions when treatment response may have social interactions. Defining a person’s treatment response to be a function of the entire vector of treatments received by the population, I study identification when 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 first study identification under assumption CTR alone and when this shape restriction is strengthened to semi-monotone response. I then consider distributional assumptions using instrumental variables, with focus on statistical independence assumptions. Finally, I discuss the use of models of endogenous social interactions to derive restrictions on response 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, the University of Munich, and the University of Virginia. I am grateful for comments from Mike Fu, Guido Imbens, and John Pepper.
1. Introduction

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

The present analysis extends my earlier work on identification with individualistic response, reported in Manski (1990, 1997, 2003), Manski and Pepper (2000), 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 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 $y_j(\cdot): T \rightarrow Y$ 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 $y_j = y_j(z)$. Suppose that the cardinality of $T$ is at most countable. This enables analysis that uses only elementary probability theory.

Let $J$ be a probability space $(J, \Omega, P)$. 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 learn about the outcome distribution $P[y(t)]$ that would occur if all persons were to receive a specified treatment $t$. Interest in $P[y(t)]$ is often motivated by a decision problem in which a planner chooses between the realized treatments and a policy that mandates treatment $t$. Then the planner wants to compare $P[y(t)]$ with $P(y)$.

Now remove assumption ITR, so each person’s outcome may vary with the treatments received by all members of the population. To express this, one extends the domain of the response function from $T$ to the Cartesian product of $T$ across the population; that is, $T' = \times_{k \in J} T$. The response function becomes $y_j(\cdot): T' \rightarrow Y$, mapping treatment vectors $t' \in T'$ into outcomes $y_j(t') \in Y$. Here $t' = (t_k, k \in J)$ 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 $y_j = y_j(z')$, where $z' = (z_k, k \in J)$.

I will take the research objective to be inference on the outcome distribution $P[y(t')]$ that would occur if the population were to receive any potential treatment vector $t'$. Interest in $P[y(t')]$ may be motivated by a decision problem in which a planner chooses between the realized treatments $z'$ and a policy that mandates treatment vector $t'$. Then the planner wants to compare $P[y(t')]$ with $P(y)$. Instances of such planning problems are studied in Graham, Imbens, and Ridder (2009) and Manski (2009, 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 $t'$. 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 $\tau$ rather than $t$ to denote a specific element of $T$.

Identification of Potential Outcome Distributions

Comparison of the setup with and without assumption ITR makes plain that identification without the assumption presents a much more severe challenge than with it. Given assumption ITR and no further
assumptions, the Law of Total Probability shows that $H\{P[y(t')]\}$, the identification region for $P[y(t')]$, is the set of distributions $[P(y|z = t)P(z = t) + \delta P(z \neq t), \delta \in \Delta_y]$, where $\Delta_y$ denotes the space of all probability distributions on $Y$. This region is a proper subset of $\Delta_y$ 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 $t'$ assigns a common treatment to all persons. Section 2.1 below extends it to the general case where $t'$ is a vector of treatments that may vary across the population.

Without assumption ITR or another assumption restricting social interactions, $H\{P[y(t')]\}$ is the singleton $P(y)$ when $z' = t'$ and is the set $\Delta_y$ of all distributions whenever $z' \neq t'$. Thus, the empirical evidence alone is uninformative about $P[y(t')]$ when $t'$ has any counterfactual component. Partial or point identification of $P[y(t')]$ may become feasible when the empirical evidence is combined with assumptions that restrict the shape of the response functions $[y(\cdot), j \in J]$ and/or the distribution $P[x, y(\cdot), z]$ of covariates, response functions, and realized treatments. The resulting form of $H\{P[y(t')]\}$ depends on the assumptions imposed and the treatment vector $t'$ under consideration.

Organization of the Paper

This paper is entirely general with respect to the potential treatment vector, but focuses on particular classes of assumptions. Section 2 studies two shape restrictions on response functions, constant treatment response (CTR) and semi-monotone treatment response (SMTR). These assumptions partially identify potential outcome distributions.

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

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

Section 3 combines shape restrictions with distributional assumptions that use instrumental variables. In research under assumption ITR, it has been common to assume that treatment response is statistically independent of an instrumental variable. It has been particularly common to use the realized treatment as the instrument. I extend the analysis to assumption CTR, using effective treatments as the instrumental variable.

I show that $P[y(t')]$ is point-identified if and only if every effective treatment that occurs with positive empirical probability in $t'$ also occurs with positive empirical probability in $z'$. This requirement is transparent under assumption ITR, but is more subtle with treatment interactions. I show that it generically fails to hold in two scenarios where interactions are global rather than local in nature. Moreover, random assignment of treatments loses its classical identifying power in these scenarios.

Section 4 discusses assumptions framed as models of endogenous interactions. In the analysis of Sections 2 and 3, a response function is a primitive that maps treatment vectors into outcomes. The primitive in a model of endogenous interactions 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 $[y(j), j \in J]$ 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 $P[y(t')]$, not identification of structural equations. A model of endogenous
interactions may have identifying power for $P[y(t')]$ if the specified structural equations imply restrictions on the reduced form. I distinguish complete models, which have unique reduced forms, from incomplete models, which do not. Point identification of a complete model implies point identification of its reduced form, but this logic does not hold for incomplete models.

Although this paper is about identification, I would be remiss to entirely ignore estimation with sample data. At the end of Section 2, I suppose that one poses 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.

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 $[y_{j}(·), j \in J]$. 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 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 provide a foundation on which further assumptions may be placed.

2.1.1. The Assumption in Abstraction

Consider person $j$. Let $c_j(t) \in T \rightarrow C_j$ be a function mapping treatment vectors onto a set $C_j$. A constant-response assumption asserts:

**Assumption CTR:**

$\forall j \in J, (1) \quad c_j(t') = c_j(s') \implies y_j(t') = y_j(s')$.

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

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

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

Now consider identification of \( P[y(t')] \). By the Law of Total Probability,

\[
(2) \quad P[y(t')] = P[y(t')|c(z') = c(t')]P[c(z') = c(t')] + P[y(t')|c(z') \neq c(t')]P[c(z') \neq c(t')].
\]

Here \( P[c(z') = c(t')] \) is the fraction of the population for whom \( c(z') = c(t') \), and \( P[y(t')|c(z') = c(t')] \) is the distribution of outcomes conditional on this event. Observation of realized treatments reveals \( P[c(z') = c(t')] \) and \( P[c(z') \neq c(t')] \). Assumption CTR implies that \( P[y(t')|c(z') = c(t')] = P[y|c(z') = c(t')] \). Observation of realized treatments and outcomes reveals \( P[y|c(z') = c(t')] \) when \( P[c(z') = c(t')] > 0 \). The empirical evidence and assumption CTR are uninformative about the counterfactual outcome distribution \( P[y(t')|c(z') \neq c(t')] \).

Hence, we have

\[
\text{Proposition CTR: Given Assumption CTR, the identification region for } P[y(t')] \text{ is}
\]

\[
(3) \quad H\{P[y(t')]\} = \{P[y|c(z') = c(t')]P[c(z') = c(t')] + \delta \cdot P[c(z') \neq c(t')], \delta \in \Delta_y \}. \quad \square
\]

Observe that the size of \( H\{P[y(t')]\} \) varies inversely with \( P[c(z') = c(t')] \). The region is the singleton \( P(y) \) when \( P[c(z') = c(t')] = 1 \). It expands as \( P[c(z') = c(t')] \) decreases, and becomes uninformative when \( P[c(z') = c(t')] = 0 \).
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. A person’s reference group may be 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.

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

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

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

Given that reference groups are treatment-invariant, they necessarily are non-manipulable. That is, a planner cannot use the treatments in T to change a person’s reference group. 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 \( y_j(t') \). The researcher knows the value of \( y_j(t') \) if and only if \( z^{(j)} = t' \).

Applying (3), the identification region for \( P[y(t')] \) is

\[
H\{P[y(t')]\} = \{P(y|z^{(j)} = t') P(z^{(j)} = t') + \delta \cdot P(z^{(j)} \neq t') \, | \, \delta \in \Delta_v\}.
\]

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

\[
H\{P[y(t')]\} = \{P(y|z = t') P(z = t') + \delta \cdot P(z \neq t'), \delta \in \Delta_v\}.
\]

All persons face the same realized treatment vector \( z' \). Hence, \( P(z' = t') = 1 \) if \( z' = t' \) and \( P(z' = t') = 0 \) if \( z' \neq t' \). Thus, \( H\{P[y(t')]\} = P(y) \) if \( z' = t' \) and \( H\{P[y(t')]\} = \Delta_v \) if \( z' \neq t' \). 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

\[
H\{P[y(t')]\} = \{P(y|z = t) P(z = t) + \delta \cdot P(z \neq t), \delta \in \Delta_v\}.
\]
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 $t'$ 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 $\tau$, and $t' = (\tau, \tau, \ldots, \tau)$. Now (6) holds in the general case where $t'$ 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 $z' = t'$ if $J$ is a countable population and permits deviation of $z'$ from $t'$ only on a negligible set of persons when $J$ is a continuum. Region (6) grows smoothly from the singleton $P(y)$ to the entire space $\Delta_\nu$ as $P(z = t)$ decreases from 1 to 0. This contrasts sharply with the unrestricted-interaction region (5), which equals $\Delta_\nu$ 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 empirical 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 \( C_j = T \times \Delta_j \), where \( \Delta_j \) 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 \( t' \in T_j \), let \( c_j(t') = [t_j, Q(t^{G(j)})] \), where \( Q(t^{G(j)}) \) is the within-group distribution of the treatments in \( t^{G(j)} \). That is, for \( \tau \in T \), \( Q(t^{G(j)} = \tau) \) is the fraction of the persons in \( G(j)/j \) who receive treatment \( \tau \) when \( t' \) 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 \( Q(t^{G(j)}) = 0 \), where \( 0 \) denotes the empty set.

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

\[
(7) \quad H \{ P[y(t')] \} = \{ P[y|z = t, Q(z^{G(j)}) = Q(t^{G(j)})] \cdot P[z = t, Q(z^{G(j)}) = Q(t^{G(j)})] + \delta \cdot P(z \neq t \text{ or } Q(z^{G(j)} \neq Q(t^{G(j)}), \delta \in \Delta_j} \}
\]

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 \( y_j(t') \) when the event \( [z_j = t_j, Q(z^{G(j)}) = Q(t^{G(j)})] \) occurs. Previously, \( y_j(t') \) 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 \( Q(t^{G(j)}) \) affects outcomes solely through some functional of the distribution, say \( F(t^{G(j)}) \). A leading case is the mean interaction, where treatments are real-valued and \( F(t^{G(j)}) = E(t^{G(j)}) \), the within-group mean of the treatments in \( t^{G(j)} \). 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 \( F_0(t^{G(j)}) = \sup(t^{G(j)}) \). Suppose that a treatment is information communicated within a reference group. Suppose that information treatments are ordered, with \( \tau > \tau' \) meaning that a person with treatment \( \tau \) receives
all of the information in $\tau'$, plus some more. Then communication within the group ensures that person $j$ effectively receives treatment $\sup(t^{(0)})$.

Whatever functional $F$ may be, let $C_j = T \times \Phi$, where $\Phi$ is the range space for $F$. Let $c_j(t') = [t, F(t^{(0)})]$. Then (3) becomes

\[ (8) \quad H\{P[y(t')]\} = \frac{\{[P(y|z = t, F(z^{(0)}) = F(t^{(0)})]P[z = t, F(z^{(0)}) = F(t^{(0)})] + \delta \cdot P(z \neq t \text{ or } F(z^{(0)}) \neq F(t^{(0)}), \delta \in \Delta_t}\}. \]

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 $y_j(t')$ when the event $[z_j = t, F(z^{(0)}) = F(t^{(0)})]$ occurs. Previously, $y_j(t')$ was known when $[z_j = t, Q(z^{(0)}) = Q(t^{(0)})]$. 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 response strengthens functional interactions, which strengthens distributional interactions, which in turn strengthens interactions within a reference group. The various identification regions presented above were correspondingly nested sets. However, even the assumption of individualistic 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 $c \sim c'$. Let the outcome space $Y$ be a subset of the real line. Let $t'$ and $s'$ be two potential treatment vectors. The assumption of semi-monotone response asserts

**Assumption SMTR:**

\[
\begin{align*}
(9) \quad & c_j(t') \geq c_j(s') \implies y_j(t') \geq y_j(s') \quad \square
\end{align*}
\]

This assumption encompasses assumption CTR, as the equality $c_j(t') = c_j(s')$ is equivalent to the two inequalities $c_j(t') \geq c_j(s')$ and $c_j(t') \leq c_j(s')$.

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

Consider the outcome of person $j$ when the treatment vector is $t'$. Let $y_{\min} \equiv \inf Y$ and $y_{\max} \equiv \sup Y$ be the logical lower and upper bounds on outcomes. Combining the empirical evidence with assumption SMTR yields this sharp bound on $y_j(t')$:

\[
\begin{align*}
(10) \quad & c_j(t') < c_j(z') \implies y_{\min} \leq y_j(t') \leq y_{\max} \\
& c_j(t') = c_j(z') \implies y_j(t') = y_j
\end{align*}
\]
\(c(t') > c(z') \rightarrow y_j \leq y_j(t') \leq y_i\)
\(c(t') \vartriangleleft c(z') \rightarrow y_0 \leq y_i(t') \leq y_i.\)

Let \(y_{jL}(t')\) and \(y_{jU}(t')\) denote the lower and upper bounds on \(y_j(t')\) stated in (10). Given that (10) holds for all \(j \in J\), the population distribution of \(y_{jL}(t')\) stochastically dominates that of \(y(t')\), which in turn dominates that of \(y_{jU}(t')\). Given that (10) exhausts the available information, we have

**Proposition SMTR**: Given assumption SMTR, the identification region for \(P[y(t')]\) is

\[(11) \quad H\{P[y(t')]\} = \{\delta \in \Delta_y : P[y_{jL}(t')] \succeq_{sd} \delta \succeq_{sd} P[y_{jU}(t')]\}. \]

Here \(\succeq_{sd}\) 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 \(D[y(t')]\):

\[(12) \quad D[y_{jL}(t')] \leq D[y(t')] \leq D[y_{jU}(t')]\].

Considering individualistic 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 \(E[y(t')]\) is

\[(13) \quad y_i \cdot P[c(t') < c(z') \cup c(t') \vartriangleleft c(z')] + E[y|c(t') \succeq c(z')] \cdot P[c(t') \succeq c(z')] \leq E[y(t')] \leq y_i \cdot P[c(t') > c(z') \cup c(t') \vartriangleleft c(z')] + E[y|c(t') \succeq c(z')] \cdot P[c(t') \succeq c(z')].\]
2.2.2. Reinforcing Interactions

I defined reinforcing interactions verbally in the Introduction. Formally, let $T$ be partially ordered. Let $j$ have reference group $G(j)$ and let $T^{G(j)}$ inherit the partial ordering on $T$. That is, given two treatment vectors $t'$ and $s'$, let $c_j(t') \geq c_j(s')$ mean that $[t_k \geq s_k, \text{all } k \in G(j)]$. A reinforcing interaction occurs when

$$[t_k \geq s_k, \text{all } k \in G(j)] \rightarrow y_j(t') \geq y_j(s').$$

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.

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

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 $c_j(t') \geq c_j(s')$ mean that $[t_j \geq s_j, Q(t^{G(j)}) \geq_{sd} Q(s^{G(j)})]$. A reinforcing distributional interaction occurs when
(15) \[ t_j \geq s_j, \ Q(t^{G(j)}) \succeq_{sd} Q(s^{G(j)}) \] \rightarrow y_j(t) \geq y_j(s).

The event \([t_k \geq s_k, \text{ all } k \in G(j)]\) implies the event \([t_j \geq s_j, \ Q(t^{G(j)}) \succeq_{sd} Q(s^{G(j)})]\). 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 \(P[y(t')]\) 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 \(Q(t^{G(j)}) \succeq_{sd} Q(s^{G(j)})\) appearing in (15) is approximately the same as \(Q(t^{G(j)}) \succeq_{sd} Q(s^{G(j)})\), which includes \(j\) in the group distribution. The latter inequality is simpler to use in some applications.

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 \(c(t') \succeq c(s')\) to mean that \([t_j \geq s_j, \ D(t^{G(j)}) \succeq D(s^{G(j)})]\). A reinforcing D-interaction occurs when

(16) \[ t_j \geq s_j, \ D(t^{G(j)}) \succeq D(s^{G(j)}) \] \rightarrow y_j(t') \geq y_j(s').

The event \([t_j \geq s_j, \ Q(t^{G(j)}) \succeq_{sd} Q(s^{G(j)})]\) implies the event \([t_j \geq s_j, \ D(t^{G(j)}) \succeq D(s^{G(j)})]\). 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 \(P[y(t')]\) 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

\[
[t_j \geq s_j, \{t_k \leq s_k, k \in G(j)\}] \land y_j(t) \leq y_j(s).
\]

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.

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

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

\[
[t_j \geq s_j, Q(s^{(t)}) \geq Q(s^{(t')})] \land y_j(t) \geq y_j(s).
\]

The latter occurs when

\[
[t_j \geq s_j, D(s^{(t)}) \geq D(s^{(t')})] \land y_j(t) \geq y_j(s).
\]
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 $P_{11} = P(y = 1|z = 1)$, $P_{10} = P(y = 1|z = 0)$, and $p = P(z = 1)$. The realized probability of good health is $P(y = 1) = pP_{11} + (1 - p)P_{10}$.

Consider a potential treatment vector $t'$ that increases the population rate of vaccination from $p$ to some $q > p$. In particular, $t'$ sets $t = 1$ for all persons with $z = 1$ and for some of those with $z = 0$.

The objective is to learn $P[y(t') = 1]$. One may interpret $P[y(t') = 1]$ 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 $P[y(t') = 1]$ prospectively as the health rate that will occur if treatment vector $t'$ is applied to a new population that is identical in composition to the study population.

The identification region for $P[y(t') = 1]$ depends on the maintained assumptions. I first assume that treatment is individualistic and then add the assumption of monotone treatment response, in the sense that vaccination never lowers health status and may improve it. I next consider a reinforcing interaction within the population as a whole.

2.3.1. Individualistic Response

Suppose that a person’s health status depends only on his own treatment. This assumption is not credible when considering an infectious disease, but I begin with it to provide contrast with the findings when
social interactions are considered. The identification region under assumption ITR was given in (6). With a binary outcome, (6) becomes the interval

\[ H\{y(t') = 1\} = \{P(y = 1|z = t)P(z = t), P(y = 1|z = t)P(z = t) + P(z = t)\}. \]

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 \(t'\).

It also includes the group of size \(1 - q\) who realize treatment 0 and would continue to receive it under \(t'\). 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

\[ P(y = 1|z = t) = P(y = 1|z = t, z = 1)P(z = 1|z = t) + P(y = 1|z = t, z = 0)P(z = 0|z = t) \]

\[ = P_{11}\frac{p}{(p + 1 - q)} + P(y = 1|z = t, z = 0)\frac{(1 - q)\frac{(p + 1 - q)}{p + 1 - q}}{t = 1}. \]

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 \(t'\). This conditional probability is revealed by the empirical evidence once \(t'\) is specified. Hence, all quantities on the right-hand side of (21) are known.

2.3.2. Monotone-Individualistic Response

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

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

$$(22) \quad \mathbb{H}\{P[y(t') = 1]\} = [P(y = 1 \mid t \geq z) \cdot P(t \geq z), P(t > z) + P(y = 1 \mid t \leq z) \cdot P(t \leq z)].$$

The inequality $t' \geq z'$ 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

$$(23) \quad \mathbb{H}\{P[y(t') = 1]\} = [P(y = 1), q - p + P(y = 1 \mid t = z) \cdot (p + 1 - q)].$$

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. Reinforcing Interactions

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

Application of (13) in the present setting gives
(24) \[ H\{P[y(t') = 1]\} = [P(y = 1 | t' \geq z') P(t' \geq z'), P(t' > z' \cup t' \varnothing z') + P(y = 1 | t' \leq z') P(t' \leq z')]. \]

We have \( t' > z' \) by design. Hence, (24) reduces to

(25) \[ H\{P[y(t') = 1]\} = [P(y = 1), 1]. \]

The lower bound is the same as with the assumption of monotone-individualistic response. The upper bound is 1 because a reinforcing interaction permits the possibility that increasing the vaccination rate 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 \( t' \), one observes \([c_j(z'), c_j(t'), y_j; j \in J]\). This enables computation of the identification regions for \( P[y(t')] \) under assumptions CTR and SMTR, given in (3) and (11) respectively.

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

\[
(3') \quad H_{P_N}[y(t')] = \{P_N[y(c(z') = c(t'))] P_N[c(z') = c(t')] + \delta P_N[c(z') \neq c(t')], \delta \in \Delta_y\}, \]

\[
(11') \quad H_{P_N}[y(t')] = \{\delta \in \Delta_y; P_N[y_{t'}(t')] \geq_{sd} \delta \geq_{sd} P_N[y_{t'}(t')]\}, \]

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

This argument requires observation of sample members’ realized effective treatments $[c_j(z'), j \in J_n]$, not just their realized own treatments $(z_r, j \in J_n)$. 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 one assumes a functional interaction, then it suffices to observe $z_j$ and $F(z_j^{G(j)})$. 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 generally does not enable consistent estimation of identification regions (3) and (11). With social interactions, observation of $(z_r, j \in J_n)$ generally does not reveal $[c_j(z'), j \in J_n]$. Hence, one cannot determine how $c_j(z')$ is related to $c_j(t')$ for $j \in J_n$.

3. Distributional Assumptions Using Instrumental Variables

This section combines the shape restrictions of Section 2 with distributional assumptions using instrumental variables. In research assuming individualistic response, an instrumental variable (IV) is a specified function $v = v(x, z)$ of the observed covariates $x$ and realized treatments $z$. Assumptions typically
restrict how the conditional response distributions $P[y(w, v)]$ may vary with $v$, where $w = w(x, z)$ is another function of $(x, z)$. When studying treatment with social interactions, I will let $v = v(x', z')$ and $w = w(x', z')$. To simplify the exposition, I will suppress $w$ until Section 3.2, where it is necessary to make it partially explicit.

Various distributional assumptions may merit consideration. In research making 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 treatment with social interactions, the objective being inference on $P[y(t')]$ for a specified $t' \in T'$. Then one may assume that $P[y(t')|v] = P[y(t')]$ or $E[y(t')|v] = E[y(t')]$.

Section 3.1 considers assumption SI in abstraction. The analysis is a simple extension of my earlier work assuming individualistic response (Manski, 1990; 2003, Chapter 2 and Section 7.4). I again use vaccination to illustrate. In research making 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 3.2 extends the argument to inference on $P[y(t')]$ under assumption CTR, using the realized effective treatment $c(z')$ as the instrumental variable.

3.1. Statistical Independence

3.1.1. 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 $P[y(t')|v = v]$, where $v \in V$, the support of $v$. One simply needs to condition each reference to $P$ on the event $[v = v]$ and repeat the derivations. Let $H\{P[y(t')|v = v]\}$ denote the resulting identification region. The identification region for the collection of distributions $\{P[y(t')|v = v], v \in V\}$ is
the Cartesian product $\times_{v \in V} H\{P[y(t^i)|v = v]\}$. This result holds 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

**Assumption SI:**

\begin{equation}
(26) \quad P[y(t^i)|v] = P[y(t^i)].
\end{equation}

Then $P[y(t^i)]$ must lie within the intersection of the identification regions $H\{P[y(t^i)|v = v]\}$, $v \in V$. Moreover, every distribution in this intersection is a feasible value of $P[y(t^i)]$. Hence, we have

**Proposition SI:** Given Assumption SI, the identification region for $P[y(t^i)]$ is

\begin{equation}
(27) \quad H\{P[y(t^i)]\} = \bigcap_{v \in V} H\{P[y(t^i)|v = v]\}.
\end{equation}

The assumptions used to derive this region may be jointly testable. The empirical evidence may reveal that the region is empty. If so, then some assumption is incorrect. The testability of assumption SI contrasts with assumptions CTR and SMTR. Those assumptions are not refutable.

3.1.2. Vaccination

To illustrate result (27), 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 $t'$ that equalizes the vaccination rates of the two groups at an intermediate level $q$. In particular, $t'$ 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 $P[y(t') = 1]$.

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

$$P[y(t') = 1] = P[y(t') = 1 \mid v = 0]P(v = 0) + P[y(t') = 1 \mid v = 1]P(v = 1).$$

Application of (25) to the group with $v = 0$ shows that $H\{P[y(t') = 1 \mid v = 0]\} = [P(y = 1 \mid v = 0), 1]$. A derivation analogous to that yielding (25) shows that $H\{P[y(t') = 1 \mid v = 1]\} = [0, P(y = 1 \mid v = 1)]$. The joint identification region for $P[y(t') = 1 \mid v = 0]$ and $P[y(t') = 1 \mid v = 1]$ is the Cartesian product of the marginal regions. Hence, the identification region for $P[y(t') = 1]$ is

$$H\{P[y(t') = 1]\} = [P(y = 1 \mid v = 0)P(v = 0) + P(y = 1 \mid v = 1)P(v = 1)].$$

The lower bound occurs if the change in treatment from $z'$ to $t'$ 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 interaction is combined with assumption SI, namely $P[y(t') = 1 \mid v = 0] = P[y(t') = 1 \mid v = 1]$. Then $H\{P[y(t') = 1]\}$ is the intersection of the
identification regions obtained above for $P[y(t^*) = 1 \mid v = 0]$ and $P[y(t^*) = 1 \mid v = 1]$. Thus,

$$H\{P[y(t^*) = 1]\} = [P(y = 1 \mid v = 0), P(y = 1 \mid v = 1)].$$

Inspection of region (30) shows that the pair of assumptions used to derive the region are jointly testable. Suppose the empirical evidence reveals that $P(y = 1 \mid v = 0) > P(y = 1 \mid v = 1)$. Then region (30) is empty. It follows that at least one of the assumptions is incorrect.

When region (30) is non-empty, it is natural to ask whether the maintained assumptions are credible. The assumption of a reinforcing interaction seems sensible enough, but it may be difficult to assess assumption SI. The fact that $t^*$ equalizes the vaccination rates of the two groups is 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 $v$ is a valid instrumental variable.

3.2. Using Realized Effective Treatments as Instrumental Variables

In research making Assumption ITR, it is common to let $v = z$ and assume that $P[y(\tau)] = P[y(\tau) \mid z]$. Assumption ITR implies that $P[y(\tau) \mid z = \tau] = P(y \mid z = \tau)$. Observation of realized treatments and outcomes reveals $P(y \mid z = \tau)$ if and only if $P(z = \tau) > 0$. Hence, the assumption that $z$ is statistically independent of $y(\tau)$ point-identifies $P[y(\tau)]$ if and only if $P(z = \tau) > 0$. If $P(z = \tau) = 0$, the empirical evidence and assumption are uninformative about $P[y(\tau)]$.

This reasoning extends to analysis of treatment response under assumption CTR, the instrumental variable now being the realized effective treatment $c(z')$. Section 3.2.1 derives the extended result in
abstraction. Section 3.2.2 considers its identifying power. Section 3.2.3 draws cautionary implications for the identifying power of random assignment.

3.2.1. The Assumption in Abstraction

As prelude, I decompose the population into a set of distinct effective-treatment types. I will say that persons \( i \) and \( j \) have the same type if there exists a permutation operator \( \pi_j : T^i \rightarrow T^j \) such that \( c(t^i) = c[\pi(t^i)] \) for all \( t^i \in T^i \). Suppose, for example, that \( i \) and \( j \) both have reference groups of size \( N \). Then \( c(t^i) \) and \( c(t^j) \) are both subvectors of \( t^i \) of length \( N \). A permutation of \( t^i \) transforms \( c(t^i) \) into \( c(t^j) \).

Let the population be composed of a finite set \( M \) of types. Let \( e(j) \in M \) denote the type of person \( j \). Let \( C_m \) be the common set of effective treatments for persons of type \( m \). For a given \( t^i \) and \( i \in C_m \), let \( J_{my} = \{ j \in J : e(j) = m, c(t^i) = \gamma \} \). In words, \( J_{my} \) is the group of persons of type \( m \) who have effective treatment \( \gamma \) when the potential treatment vector is \( t^i \). Outcomes in groups with zero probability mass do not affect outcome distribution \( P[y(t^i)] \). Hence, in what follows, I only consider groups with \( P(J_{my}) > 0 \).

Now assume statistical independence with effective treatments as instruments:

**Assumption SI-ET**: For each group \( J_{my} \) with \( P(J_{my}) > 0 \),

\[
(31) \quad P[y(t^i)|J_{my}] = P[y(t^i)|J_{my}, c(z')].
\]

Assumption CTR implies that \( P[y(t^i)|J_{my}, c(z') = \gamma] = P[y|J_{my}, c(z') = \gamma] \). Observation of realized treatments and outcomes reveals \( P[y|J_{my}, c(z') = \gamma] \) if and only if \( P[c(z') = \gamma|J_{my}] > 0 \). Hence, the assumption that \( c(z') \) is statistically independent of \( y(t^i) \) point-identifies \( P[y(t^i)|J_{my}] \) if and only if \( P[c(z') = \gamma|J_{my}] > 0 \). If \( P[c(z') = \gamma|J_{my}] = 0 \), the empirical evidence and assumption are uninformative about \( P[y(t^i)|J_{my}] \).
It remains to aggregate across groups. The Law of Total Probability gives

$$P[y(t')] = \sum_{(m \in M, \gamma \in C_m)} P[y(t')|J_m] \cdot P(J_m).$$

Hence, we have

**Proposition SI-ET:** Given assumption SI-ET, the identification region for $P[y(t')]$ is

$$H\{P[y(t')]\} = \{ \sum_{(m \in M, \gamma \in C_m; P[c(z) = \gamma|J_m] > 0)} P[y|J_m, c(z') = \gamma] \cdot P(J_m) \ + \ \delta \sum_{(m \in M, \gamma \in C_m; P[c(z) = \gamma|J_m] = 0)} P(J_m), \delta \in \Delta \}. \quad \square$$

### 3.2.2. Identifying Power

Given assumption SI-ET, $P[c(z') = \gamma|J_m] > 0$ is the essential requirement for identification of $P[y(t')|J_m]$. This requirement is transparent under assumption ITR, where it reduces to $P(z = \tau) > 0$. It is more subtle with treatment interactions, and it often fails to hold. I provide two illustrations here.

**Groups with Leaders and Followers**

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

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

The effective treatment of a person of type m is a subvector of \( t' \) of length \( N + L \). Let \( \Lambda(m) \subset J \) denote the leaders of type m. Let \( \gamma = (\tau^N, \tau^{\Lambda(m)}) \) denote a situation in which followers receive the N treatments \( \tau^N \) and leaders receive the L treatments \( \tau^{\Lambda(m)} \). Then

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

Thus, \( P[c(z') = \gamma | J_{m\tau}] > 0 \) if and only if two conditions hold. First, realized and conjectured treatments must coincide for all leaders. Second, they must coincide for all followers in a positive fraction of the groups of type m.

Population-wide Distributional Interactions

Suppose that the population contains one type of person. The common reference group is the entire population, and interactions are distributional. This may, perhaps, be a reasonable idealization of some vaccination scenarios. One may think it credible to assume that each person’s health status varies with his own vaccination and with the population rate of vaccination.

In this setting, \( c_j(t') = [t_j, Q(t')] \) and \( c_j(z') = [z_j, Q(z')] \) for all \( j \in J \). The feasible values of \( \gamma \) are the pairs \( [\tau, Q(t')] \), \( \tau \in T \). Fixing \( \tau \) and letting \( J_\tau \) be the subpopulation who would receive \( \tau \) under potential treatment vector \( t' \), we have

\[
P[c(z') = \gamma | J_{m\tau}] = P[z, Q(z')] = [\tau, Q(t')] | J_{\tau}] = P(z = \tau | J_{\tau}) \cdot 1[Q(z') = Q(t')].
\]
Thus, $P[c(z') = \gamma | J_{m}] > 0$ if and only if two conditions hold. First, the realized and conjectured population distributions of treatments must coincide. Second, a positive fraction of the persons in subpopulation $J$, must receive realized treatment $\tau$.

3.2.3. Random Assignment of Realized Treatments

In research making assumption ITR, use of $z$ as the instrumental variable is often motivated by knowledge that realized treatments were randomly assigned to a large study population. Random assignment may also motivate assumption SI-ET. However, in contrast to the situation with individualistic response, random assignment may not have identifying power. The two above illustrations demonstrate the difficulty.

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

Consider groups with leaders and followers. We showed above that the empirical evidence and assumption SI-ET can be informative about $P[y(t') | J_{m}]$ only if $z^{A(m)} = t^{A(m)}$. With random assignment, the ex ante probability that $z^{A(m)} = t^{A(m)}$ is $\prod_{j \in A(m)} \pi(t)$. This probability is less than one. Hence, random assignment yields positive ex ante probability that $z^{A(m)} \neq t^{A(m)}$.

Consider a population-wide distributional interaction. We showed above that the empirical evidence and assumption SI-ET can be informative about $P[y(t') | J_{m}]$ only if $Q(z') = Q(t')$. Using random assignment, this equality occurs with ex ante probability less than one. In the limit case of an uncountably large population, $Q(z') = \pi$.

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

The illustrative scenarios considered here—groups with leaders and followers, and population-wide distributional interactions—have a very different structure. These scenarios exhibit global social interactions rather than local ones. We have demonstrated that, when interactions are global, random assignment does not retain its classical identifying power.

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 response functions. It may also be productive to 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 discuss use of endogenous-interaction models to identify the distribution of treatment response.
4.1. Identification of Structural Functions and Reduced Forms

Considered in abstraction, a model of endogenous interactions supposes that the potential outcome vector \( y(t^*) = [y_j(t^*), j \in J] \) solves the \textit{structural equations}

\[
y_j(t^*) = f_j[t_j, t_j^*, y_j(t^*)], \quad j \in J.
\]

Here \( t_j^* = (t_k, k \in J, k \neq j) \) and \( y_j^*(t^*) = [y_k(t^*), k \in J, k \neq j] \) are the treatment and outcome vectors for the population exclusive of person \( j \). The structural function \( f_j(\cdot) \) permits \( y_j(t^*) \) to be determined by \( j \)'s own treatment as well as by the treatments and outcomes of other members of the population. The term \textit{exogenous} interaction describes \( t_j^* \) as an argument of \( f_j(\cdot) \), while \textit{endogenous} interaction describes \( y_j^*(t^*) \). If \( y_j^*(t^*) \) were not an argument, \( f_j(\cdot) \) would simply be the person’s response function. The presence of \( y_j^*(t^*) \) as an argument of \( f_j(\cdot) \) makes (36) a system of simultaneous equations.

An outcome vector \( y_j(t^*) \) that solves (36) is said to be a \textit{reduced form} of the structural equations. A model is \textit{complete} if (36) 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).

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

\[
y_j = f_j(z_j, z_j^*, y_j^*), \quad j \in J.
\]
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 $f^j$; see, for example, Goldberger (1991). Work on linear models of social interactions does likewise (e.g., Manski, 1993). Semiparametric and nonparametric research explores identification under weaker assumptions.

Our concern is identification of potential outcome distributions. Thus, we 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 $y^j(t')$. Our particular concern is $P[y(t')]$, the empirical distribution of $y^j(t')$.

It has occasionally been observed, but often neglected, 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.”

The relationship between identification of structural functions and reduced forms is straightforward in analysis of linear models. There, the parameters of the reduced form are many-to-one functions of the parameters of the structural function. Hence, identification of the reduced form is a simpler objective than identification of the structural functions.

Outside of linear models, the relationship between identification of structural functions and reduced forms is largely an open question. This question is much too broad for a comprehensive analysis here, but I will make a small start. Section 4.2 calls attention to the fact that the relationship between structural functions and reduced form differs qualitatively in complete and incomplete models.
4.2. Complete and Incomplete Models

Given a complete model, identification of $P[y(t')]$ is logically no more difficult than identification of $f'$, and may be easier. With an incomplete model, identification of $P[y(t')]$ may be more difficult than identification of $f'$. I explain here. In what follows, $\Phi$ denotes the identification region for $f'$.

Suppose first that the model is complete; thus, (36) has a unique solution for each element of $\Phi$. For each $f' \in \Phi$, let $y'(t', f')$ denote this solution. Then the identification region for $y'(t')$ is $[y'(t', f'), f' \in \Phi]$. The cardinality of this set cannot be larger than that of $\Phi$, and it may be smaller. Thus, a model that point-identifies $f'$ necessarily point-identifies $y'(t')$. Knowledge of $y'(t')$ implies knowledge of $P[y(t')]$. Hence, identification of $P[y(t')]$ is logically no more difficult than identification of $f'$.

Next suppose that the model is incomplete, with at least one solution to (36) for every feasible value of $f'$ and multiple solutions for some values. For each $f' \in \Phi$, let $Y(t', f')$ denote the set of solutions to (38). Then the identification region for $y'(t')$ is $\{Y(t', f'), f' \in \Phi\}$. In general, the cardinality of this set may be larger or smaller than that of $\Phi$. However, it necessarily is larger when the model point-identifies $f'$. Then $f'$ is known, but $Y(t', f')$ contains multiple elements. Hence, $H\{P[y(t')]\}$ may contain multiple elements.

Finally, consider an incomplete model having no solution to (36) for some $f' \in \Phi$. There are two distinct ways to interpret non-existence of a solution. One might interpret it to mean that the value of $f'$ under consideration is not feasible. Then one should eliminate this value from $\Phi$. 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 $y'(t')$. Then the model has no identifying power for $P[y(t')]$. 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.
5. Conclusion

This paper has studied identification of potential outcome distributions when treatment response may have social interactions. Defining a person’s treatment response to be a function of the entire vector of treatments received by the population, I analyzed identification when shape restrictions and distributional assumptions are placed on response functions. I first posed assumption CTR alone and next strengthened it to semi-monotone response. I then posed statistical independence assumptions using instrumental variables, with focus on the use of effective treatments as instruments. Four propositions expressed the basic findings, with many special cases and illustrations fleshing them out.

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


