

DISCUSSION: ‘THE SCIENTIFIC MODEL OF CAUSALITY’

*Michael E. Sobel**

1. INTRODUCTION

Heckman advocates an approach to causal inference that draws upon structural modeling of the outcome(s) of interest (which he calls scientific), and he contrasts this approach sharply with that arising out of the statistical literature on experimentation. Drawing extensively on several previous papers—for example, Heckman (1997, 2000, 2001) and Heckman and Navarro-Lozano (2004)—Heckman goes even further here, arguing that the statistical literature on causal inference is incomplete because it does not attempt to model the process by which subjects are selected into treatments (or what statisticians have called the “treatment assignment mechanism”) and that this literature confounds the task of defining parameters with the tasks of identifying and estimating these parameters. I shall return to these points later.

But whereas Heckman distinguishes sharply between these approaches (and hence between certain literatures in economics and statistics), on balance I find the similarities in the approaches he discusses much more profound than the dissimilarities. To elaborate, while there has been and continues to be much philosophical

For financial support, I am grateful to the John D. and Catherine T. MacArthur Foundation. For helpful remarks, I am grateful to the members of the causal inference study group at Columbia University. Address all correspondence to Michael Sobel, 421 Fayerweather Hall, Columbia University, New York, NY, 10027 (or mes105@columbia.edu).

*Columbia University

disagreement about the nature of the causal relation, in both these literatures, there are strong similarities in the way that the word “cause” is used. In particular, a causal relation sustains a counterfactual conditional, while a noncausal relation need not do so. Second, causal effects are allowed to be heterogeneous across units, a point that both statisticians and Heckman have emphasized. Third, dovetailing with this shared perspective on the nature of the causal relation, the potential outcomes notation invented by Neyman (1923), widely used since by statisticians working on experimental design (see for example, the textbooks by Cox [1958] and Kempthorne [1952]), is now standard notation in both these literatures. This notation, adopted also by Heckman during the latter 1980s and since in a number of his papers (and here) nicely captures the idea that a causal relationship sustains a counterfactual conditional statement. To be sure, this idea can be represented in other ways (Robins and Greenland 2000), but the potential outcomes notation is very easy to work with and easy to use without leading oneself astray. The importance of good notation cannot be emphasized strongly enough. As Whitehead ([1911], 1958:39) pointed out, “By relieving the brain of all unnecessary work, a good notation sets it free to concentrate on more advanced problems, and in effect increases the power of the race.” I would also propose that the use of a common notation encourages investigators to think similarly about a problem.

The incorporation of Neyman’s notation into the modern literature on causal inference is due to Rubin (1974, 1977, 1978, 1980), who, using this notation, saw the applicability of the work from the statistical literature on experimental design to observational studies and gave explicit consideration to the key role of the treatment assignment mechanism in causal inference, thereby extending this work to observational studies. To be sure, previous workers in statistics and economics (and elsewhere) understood well in a less formal way the problems of making causal inferences in observational studies where respondents selected themselves into treatment groups, as evidenced, for example, by Cochran’s work on matching and Heckman’s work on sample selection bias. But Rubin’s work was a critical breakthrough. The introduction of a suitable notation allowed in principle the clarification and formalization of the problem of making causal inferences with non-experimental data. Further, the use of this notation allowed the tasks of defining etimands to be separated

from the tasks of identifying and estimating these. This separation enabled Rubin not only to pinpoint the key role of the treatment assignment mechanism but to state precise conditions under which this mechanism was “ignorable” or not. These conditions have been used by subsequent workers (including Heckman) to evaluate and clarify existing procedures for causal inference (for example, instrumental variables) and to develop new methods for estimating causal effects (for example, by creating matched samples using propensity scores).

And fourth, although Heckman criticizes the “treatment effects” literature for modeling the effects of causes, as opposed to modeling the causes of effects, the majority of his paper also focuses on modeling the effects of an intervention (cause) on an outcome of interest.

That said, some of the problems involved in making causal inferences about agents who are not (or cannot in practice be) subjected by an investigator to one or another treatment of interest will be somewhat different than those that typically arise when a treatment is applied or not to a plot of land in an agricultural experiment. In the latter case, where a randomized experiment can be conducted, the treatment assignment mechanism is essentially a (possibly biased) coin toss (or a coin toss within distinguishable types of plots). Causal inference is typically more straightforward in this case. However, in observational studies, individuals typically sort themselves into treatment groups and how they do so may also be of independent interest. Economists often argue that individuals make choices by behaving as if they are maximizing expected utility. When the utility associated with making particular choices is also related to the outcome under consideration, this may create problems when the agent uses more information than that available to the economist (or the economist simply does not know the treatment assignment mechanism). In this case, consideration of the available information may be inadequate for a “sufficient” description of the process by which agents allocate themselves to treatment groups. Here, the assignment mechanism cannot be treated as a coin toss (net of the available information). Thus, methods based on this premise are inadequate for this case. In other social and behavioral sciences, where notions of decision making may take on a different flavor, the above may or may not be problematic. In any event, the point is that even when the same notion of causation is under

consideration, some variation by discipline in approaches to causal inference should be expected.

In addition, although assumptions are always present, investigators differ in the extent to which they are comfortable using these to make inferences. Heckman advocates herein what he has elsewhere (Heckman 2000) called a structural approach to estimating causal parameters. This model based approach can be very powerful and can be used as a basis for generating inferences that are sometimes much stronger and broader in scope than those typically made in the statistical literature Heckman criticizes. But the structural approach also typically features stronger assumptions. As Heckman (2000) documents, frustration with the seemingly arbitrary nature of the assumptions (for example, exclusion restrictions) used to identify structural models has led several generations of economists to eschew structural modeling in favor of other approaches, recently including experimentation (Heckman is quite critical of this “natural experiment movement;” for a nice treatment of the issues see Rosenzweig and Wolpin (2000)). As before, even when the same notion of causation is held, some differences in approaches to causal inference should be expected.

However, if one takes the view that statistical procedures should be tailored to address questions of interest to various constituencies (for example, different groups of scientists and policymakers), such differences should be regarded as both natural and desirable. Accordingly, I aim primarily to give a balanced overview of the issues.

To make my discussion as useful as possible for *Sociological Methodology* readers, I sometimes elaborate on material covered by Heckman in this or previous papers. Second (and primarily for the same reason), my remarks are organized around the following four themes in Heckman’s paper: 1) the nature of the causal relation, 2) definitions of causal estimands, 3) policy evaluation and forecasting, and 4) the identification and estimation of causal effects. Although I do not find simultaneous equation models (and more generally, structural equation models) in their current form very useful for causal inference, I do not take up this subject here, in large measure because a thorough treatment would require substantially increasing the length of an already long discussion. For some previous and related discussions of causality in

simultaneous equation models, the reader might also wish to consult Strotz and Wold (1960), Fisher (1970), and Sobel (1990).

To improve readability, in most instances, I include here the equations referred to, even if some of these already appear in Heckman's article. Whenever possible, I use notation similar or identical to Heckman's. In some instances, in order to retain consistency of style, minor deviations have been necessary.

2. THE CAUSAL RELATION

Heckman argues (page 2, this volume) that "Science is all about constructing models of the causes of effects" (vs. studying the effects of causes). He also argues that the notion of causality involves manipulating one or more variables and comparing the outcomes of these manipulations. Economists perform these (hypothetical) manipulations using models. And since models are mental constructs, Heckman concludes that causality resides in the mind. In addition, as different people think different thoughts and therefore construct different models, causes are "relative" (to use the language of Collingwood ([1940], 1972). That is, with equal legitimacy, different investigators may identify different factors as causes, while ignoring others. This is part of what Heckman refers to as the provisional nature of causal knowledge. The implications of this are further discussed in Sobel (1995), which readers may also wish to consult for more background on the causal relation.

Modeling the causes of effects is certainly an important scientific activity. But it should also be understood that many "scientific" questions are not causal. For example, NASA recently crashed a probe from the Deep Impact spacecraft into comet Tempell with the objective of learning more about the structure and composition of cometary nuclei. See Bunge (1979) for a discussion of the many kinds of noncausal questions that are of scientific interest. Further, studying the effects of causes is an important scientific activity: figuring out the causes of global warming would be considerably less important if the effects of global warming were inconsequential.

Next, the nature of the causal relation has consumed the attention of philosophers since well before Hume, without resolution. Regularity theories are one attempt to explicate the nature of

causation. Here the cause (or causes of the effect) is (are) usually thought of as some set of necessary and/or sufficient antecedents for the effect. Regularity theories are both general and (typically) deterministic. In addition, contemporary regularity theories usually require that a causal statement sustain a counterfactual conditional. The foregoing ideas may be expressed mathematically using functions, as in an “all-causes” model:

$$y(s) = g_s(x, u), \quad (1)$$

where $y(s)$ is an outcome of interest depending on state s , g_s is a function of x , a vector of observables, and u , a vector of unobservables. By varying the components of g_s , effects of the arguments (which may be state specific) can be defined under suitable conditions. In other instances, attention focuses on the results of manipulating the states s , which might, for example, index a set of treatments that are to be applied.

The antecedents above are also held to have causal priority over the effect. Explicating the nature of this priority has proven to be a difficult task. Most philosophers hold to the view that there is more to causal priority than mere temporal order. Thus, some sequences are to be regarded as causal while others are not. Although manipulating a cause is one way to establish the priority of the cause over the effect, in many theories of causation, including regularity theories, manipulability is not regarded as essential.

On the other hand, manipulability theories of causation emphasize the ability of a human agent to manipulate a cause: “A cause is an event or state of things which it is in our power to produce or prevent, and by producing or preventing which we can produce or prevent that whose cause it is said to be” (Collingwood [1940], 1972: p. 296–97). Whereas regularity theories are theories about the causes of effects, manipulability theories are theories about the effects of causes. Such theories correspond more closely with the way an experimentalist thinks of causation. These theories are also readily combined with singular theories of the causal relation.

At first glance, it appears that manipulability theories are inherently at odds with regularity theories. Thus, it might be argued that the modern literatures on causal inference in both statistics and econometrics, because these literatures are typically concerned with

the identification and estimation of the effect(s) of particular causes—for example, the effects of a policy intervention, as in Heckman's paper—are, from the scientific standpoint advocated by Heckman, misdirected. Importantly, this is not the case, and manipulability theories can be reconciled with regularity theories by noting that a manipulated cause is simply one component of (1) with u unknown (some components of u may be unknown and others simply not observed). To illustrate this point using (1), suppose the treatment variable s is 0 (no treatment) or 1 (treatment), $x \in \Omega_x$ and $u \in \Omega_u = \Omega_{0u} \cup \Omega_{1u}$, with $\Omega_{0u} \cap \Omega_{1u} = \emptyset$. Suppose that for $(x, u) \in \Omega_x \times \Omega_{0u}$, $g_0(x, u) = g_1(x, u) = 0$, while for $(x, u) \in \Omega_x \times \Omega_{1u}$, $g_0(x, u) = 0$, $g_1(x, u) = 1$. In a manipulability theory, the variable s is singled out for attention. If the investigator knows (s, x, u) and the functions g_s , the effect of s varies over x and u in a known (to the investigator) way. In practice, the investigator observes only s and x , in which case the effect of s (which the investigator may not be able to identify from data) varies over x in an apparently nondeterministic manner.

The relativity of causation (part of what Heckman calls the provisional nature of causal knowledge) is also easily illustrated. In (1) the treatment variable s may be singled out for attention and manipulated. The other arguments (x and u) remain in the causal background. A different investigator might identify one or more components of $x \equiv (x_1, x_2)$ as the cause and rewrite (1) as $h_{x_1}(x_2, u, s)$.

When the cause can be manipulated, each unit in a research study can receive any of the various levels of the cause (even though in practice a given unit is observed only at one level). In his presentation of what has come to be known in the statistical community as "Rubin's model for causal inference," Holland (1986), like Heckman, also emphasizes the importance of manipulating the cause, going so far as to coin the (unfortunate) phrase, "No causation without manipulation." Whereas Holland appears to insist on the actual ability of an investigator to manipulate the cause(s), many others, including Heckman, have argued that it is the idea of manipulating the cause, even if this can only be done hypothetically, that is key in defining causal relationships. If this point of view is taken (but not if not), Holland appears to be conflating the distinct problems of defining and identifying causal effects.

But perhaps the most controversial aspect of Heckman's brief treatment of causality is his claim that "causality is in the mind." This claim stems from (a) the fact that causal effects are defined as changes in outcomes when variables in a model are (hypothetically) manipulated and (b) the view that models are mental constructs made up by the scientist, "not empirical statements or descriptions of actual worlds" (page 3). While Heckman's conclusion is consistent with (a) and (b), and Heckman is certainly free to define causality in this fashion, I do not believe that most scientists (or philosophers) would subscribe to this view, and were they to do so, they would presumably have little further interest in causality (as science typically purports to be concerned with the real world).

In this vein, models (be they mathematical or of some other sort) are often constructed by scientists to represent causal processes (causal mechanisms) believed to be operating in the actual world (not just the mind). To be sure, the models have to be imagined and in this sense, our notion of the causal process(es) at play comes from the mind, but the processes (which we may or may not accurately model) are also believed to reside in the actual world. That is, the causal relation is typically held to describe a relation that is believed to exist in the real world.

3. DEFINING CAUSAL ESTIMANDS

In the statistical literature on causal inference, as in Heckman, assumptions (A-1) and (A-2) are typically made; Rubin (1980) has called this the stable unit treatment value assumption (SUTVA). When these assumptions are not made, the problem of defining causal estimands is more difficult, as is the problem of making inferences about these. In addition to Heckman, several others have worked on this problem (Halloran and Struchiner 1995; Sobel 2001, 2003). But this is fertile ground for social scientists, where interference due to social interactions and other constraints are the norm. Nevertheless, following Heckman, I shall hereafter assume SUTVA holds.

With (A-1) and (A-2) in hand, the response of unit ω to level s of the cause may be written as $Y_s(\omega)$; for the purposes at hand, assume that each unit can take on every level of the cause. Individual (unit) causal effects are then defined as an intra-unit, between-treatment

comparison $h(Y_s(\omega), Y_{s'}(\omega))$. Because each study unit can actually be observed only under one treatment, it is not possible to observe unit causal effects. Holland (1986) refers to this fact as “the fundamental problem of causal inference.”

Heckman focuses attention on three estimands in this paper, the average causal effect (ACE), the effect of treatment on the treated (TT), and the marginal treatment effect (MTE). Although the MTE can be useful for understanding other estimators, I do not discuss it further herein, as I believe sociologists will usually be more interested in the other two estimands. The local average treatment effect (LATE) will be discussed subsequently.

Let S denote a set of treatments of interest. The average causal effect of treatment s versus s' ($ACE(s, s')$) is defined as

$$E(Y_s - Y_{s'}), \quad (2)$$

in which $h(Y_s(\omega), Y_{s'}(\omega)) = (Y_s(\omega) - Y_{s'}(\omega))$. The ACE can also be defined conditionally on covariates W ; as in Heckman, this is denoted $ACE(s, s' | W)$, and when it is obvious which treatments are being compared (as in the case where there is just one treatment compared to no treatment) simply $ACE(W)$, or ACE in the case where there are no covariates. Since, following Heckman, $Y_s(\omega)$ is defined as the outcome of unit ω when treatment s is received, herein the ACE is the average difference when all units receive treatment s as versus s' . It is also (by virtue of assumptions (A-1) and (A-2)), the effect of receiving treatment s versus s' for a randomly selected person from the population.

The average effect of treatment s versus s' on the treated ($TT(s, s')$) is another parameter of longstanding interest:

$$E((Y_s - Y_{s'})|D = s), \quad (3)$$

where D is the random variable denoting which treatment in S is actually received. Thus, $TT(s, s')$ is the average effect of treatment s versus s' for those units that actually take up treatment s .

To round out the discussion, I also want to consider a parameter that has received a great deal of attention from biostatisticians and the public health community, the so called “intent to treat” estimand ($ITT(s, s')$). For all $s \in S$, we define $\tilde{Y}_s(\omega)$ as the outcome

of unit ω when assigned to treatment s . (The treatment to which a subject is assigned may differ from the treatment received because subjects will not always take up the treatment to which they are assigned; thus $\tilde{Y}_s(\omega) \neq Y_s(\omega)$ in general.) $\text{ITT}(s, s')$ is then defined by (2) with \tilde{Y}_s and $\tilde{Y}_{s'}$ replacing Y_s and $Y_{s'}$, respectively. Note that in the case where all subjects would take up their assignments, for any possible assignment, $\tilde{Y}_s(\omega) = Y_s(\omega)$ and $\text{ITT}(s, s') = \text{ACE}(s, s')$.

There has been some controversy over which of the parameters above are of greatest interest. I take the view that it all depends on the problem at hand, the goals of the scientist(s) analyzing the data and the purposes of the person(s) making policy on the basis of the analysis. Some examples where one or more of the parameters above are of interest follow.

For policies with universal coverage and universal participation, the ACE is the obvious parameter of interest. For example, consider the effect of a specific currency devaluation ($s = 0$ if no devaluation, 1 otherwise) on household spending. Here $\tilde{Y}_s(\omega) = Y_s(\omega)$ for all ω , implying $\text{ACE} = \text{ITT}$. If the devaluation is implemented, $\text{ACE} = \text{TT}$ as well (as every unit takes up the treatment).

For policies with universal coverage that do not require participation, some units may not take up the treatment. Because nonparticipating units will not obtain the benefits of participation, it might be argued that knowing the average effect of treatment for these units is irrelevant, suggesting TT is the parameter of interest. However, the untreated might take up treatment in the future if they believed the treatment were effective (for them). Thus, we might wish to also know the TUT (effect of treatment on the untreated). Alternatively, policymakers might want to know the effect for the nonparticipating units, for if this is deemed substantial, they will then want to make efforts to obtain the participation of such units. They will then also want to know the ACE, which is a weighted average of the TT and the effect of treatment on the untreated (TUT).

But some might argue instead that the effect that should be of interest is the effect of offering the program. For example, consider the case of a new contraceptive method. Whereas some scientists may be more interested in the ACE (or possibly the TT), which measures more directly the clinical effectiveness of the contraceptive, policymakers considering whether or not to widely distribute the contraceptive in a developing country are more concerned with the cost and the efficacy

of the contraceptive in the field (where some people do not follow instructions). Consequently, they are more interested in the ITT.

Finally, it is worth noting that if receipt of treatment is independent of the potential outcomes, given a set of known covariates W (including the case of no covariates), $TT(W) = TUT(W) = ACE(W)$.

Heckman also discusses a number of outcome measures that may be of interest to social scientists and economists but which are not discussed in the statistical literature he criticizes, where the outcomes $Y_s(\omega)$ are typically straightforward measures of the status of a unit—for example, the income of a family under treatment s or the survival time of a subject after surgery. In particular, Heckman considers outcomes $V(Y_s(\omega))$ where V is some function of the outcome—for example, the utility of $Y_s(\omega)$ to individual ω (or to a policymaker) under policy s . He then uses these to define various parameters comparing the benefit (welfare) associated with alternative policies. Although mathematically nothing new is involved here, this is useful, especially because it is possible that $E(V(Y_s) - V(Y_{s'})) \leq 0$ when (2) > 0 , for example. Thus, if V were to measure a social planner's utility, the planner would not wish to choose policy s over s' even though the average causal effect is greater than 0. Choosing a policy is often not this simple, however; for some interesting recent work that applies decision theory to the problem of treatment choice, see Manski (2000, 2004).

The estimands above are differences between means. Because the integral is a linear operator, these estimands only require knowledge of the marginal distributions $F(y_s)$ and $F(y_{s'})$ of potential outcomes. Under some circumstances (discussed later), these distributions can be identified.

Heckman also discusses a number of other estimands $h(Y_s, Y'_s)$ of substantive interest that depend on the joint distribution $F(y_s, y'_s)$ of $(Y_s, Y_{s'})$. However, the fundamental problem of causal inference precludes the simultaneous observation of $Y_s(\omega)$ and $Y'_s(\omega)$, implying that it is not possible to know more than the marginal distributions. And while knowledge of the marginal distributions imposes some constraints on the joint distribution, these constraints often do not allow much useful information on the joint to be extracted (for example, if the marginals are normal with known means and variances, this is consistent with non-normal joint distributions, as well as a bivariate normal with any correlation between -1 and 1). Thus,

much stronger assumptions will be required to point identify and estimate parameters depending on the joint distribution of potential outcomes than parameters depending only on the marginal distribution of the potential outcomes (for an example of this, see Carneiro, Hansen and Heckman 2003). Since the data impose few constraints (as discussed above) and the joint distribution of potential outcomes is not even an explicit auxiliary consideration in any substantive theory I can think of, the possibility that mathematical assumptions made primarily for the sake of convenience or tractability may be in large measure generating the “empirical” results seems especially strong here; sensitivity analyses should be a must.

4. POLICY EVALUATION AND FORECASTING

Drawing upon themes explicated at greater length in Heckman (2000, 2001) and several subsequent papers, Heckman emphasizes the value of the “scientific approach” (as exemplified by structural models) for policy evaluation and forecasting. He distinguishes three problems: (1) evaluating policies that have been implemented, (2) extrapolation of these to new environments, and (3) forecasting the effects of policies that have not been implemented to new environments. Heckman uses a structural equation model of the form $\phi(X(\omega), U(\omega))$ to examine this problem, writing the expectation of the observed outcome Y in the historical population, conditional on X as

$$E_H(Y|X = x) = \int_U \phi(x, u) dF_H(u|x), \quad (4)$$

where $F_H(u | x)$ is the conditional distribution of U given $X = x$ in the historical population. For problem 2, we want to know $E_T(Y | X = x)$. It is clear from equation (4) that this problem is easily solved if the distribution $F_T(u | X = x)$ is known in the new environment (target), assuming also the invariance of ϕ and the condition that the support of (X, U) in the target population is contained in the support of (X, U) in the historical population. Of course, the assumptions and information needed to solve this problem are very strong. The third problem can be dealt with in a similar fashion (see the appendix to Heckman’s paper), although it is more complicated.

As Heckman points out, the statistical literature on causal inference has focused on estimating the impact of policies in a given environment and problems 2 and 3 have not received much explicit attention. But certainly problem 2 is easy to address within the usual "treatment effect" framework and perhaps this is why it has not been addressed explicitly; problem 3 I discuss momentarily.

I now proceed to discuss problem 2 within a "treatment effects" framework for several reasons. First, I want the reader to understand that the "treatment effects" framework and the "scientific" framework, despite apparent differences, often yield very similar answers to real questions. In particular, that must be the case when we see that the answers actually rest on similar assumptions, once these are explicated. Second, in comparing Heckman's structural approach with the alternative I explicate below, I believe that some researchers who might need to address this problem in their future substantive work may find it easier to think about this problem from the "treatment effects" perspective.

For the sake of concreteness, consider the problem of extrapolating the historical ACE $E_H(Y_1 - Y_0)$ to a new population T. The obvious thing to do is to think of a set of covariates Z such that the historical and target ACEs are identical, and to average the historical ACE over the marginal distribution of Z in the target population.

More formally (assuming Y_1 and Y_0 are real valued scalars), the conditional ACE in the target population is

$$E_T((Y_1 - Y_0)|Z = z) = \int_R y_1 dF_T(y_1|z) - \int_R y_0 dF_T(y_0|z). \quad (5)$$

Knowledge of the target distributions $F_T(y_0 | z)$ and $F_T(y_1 | z)$ is sufficient to determine the value of equation (5); of course, the problem is that target distributions are unknown and it might be very difficult to specify them. The simplest thing is to assume the historical and target distributions are the same (where these are both defined) $F_H(y_s | z) = F_T(y_s | z)$ for $s = 0, 1$. Alternatively, in this case, we might just as well assume the weaker condition $E_T((Y_1 - Y_0) | Z = z) = E_H((Y_1 - Y_0) | Z = z)$. Either of these is an invariance assumption and should not be lightly made. But this characterization

of the problem seems to be one that is intuitively easy to understand, and this should allow an investigator to think reasonably about the necessary components of Z . Continuing, the target ACE is then $E_T(Y_1 - Y_0) = \int_R E_H((Y_1 - Y_0) | z) dF_T(z)$. Of course, to average the integrand over the target distribution, it must be defined for all values (up to a set of probability measure 0) that Z takes on in the target population. This will be the case, for example, when the supports of (Y_s, Z) in the target population are contained in the supports of (Y_s, Z) in the historical population.

Now suppose Y_s is, as in Heckman, the invariant (over the historical and target population) structural equation; $Y_s = \phi_s(Z, U_s)$ for $s = 0, 1$, then

$$E_T(Y_s | Z = z) = \int_R \phi_s(z, u_s) dF_T(u_s | z). \quad (6)$$

Analogous to the case above, if the target distributions $F_T(u_0 | z)$ and $F_T(u_1 | z)$ are known, the value of (6) is known. If these are assumed to be identical to their historical counterparts, this implies $F_H(y_s | z) = F_T(y_s | z)$ for $s = 0, 1$; if some other assumption is made, this cannot be the case. Note also that $F_H(y_s | z) = F_T(y_s | z)$ for $s = 0, 1$ does not imply invariance of the structural equation model nor the conditional distributions of U_s . As before, the ACE is obtained by averaging over the marginal distribution of Z in the target population.

As for the second point, I at least find it easier to think about the distributions $F_T(y_s | z)$ (or the conditional ACE) than to think about invariant structural equations and the conditional distributions of the unobservables. (Of course, this does not invalidate a structural approach.)

Heckman is also very critical of the “treatment effects literature” for its failure to deal with problem P3, and he briefly (see some of Heckman’s more recent work with Vytlacil for a more detailed treatment) considers this problem here, suggesting that treatments be viewed as a bundle of characteristics. The relationship between these characteristics (as versus just the treatments themselves) and the response (possibly with covariates) can then be modeled, and the relationship transported to the new environment, as per problem P2.

This idea is obvious (although its implementation can be difficult), which makes one wonder why statisticians have not addressed this topic. In that regard, several points are in order.

First, for more than 75 years, statisticians and applied workers have been using factorial experiments in conjunction with Fisher's analysis of variance (and more generally, response surface methodology) to both identify and estimate the effects of the factors (characteristics) comprising the treatment on the response, and to extrapolate these to conditions not actually experienced. A simple example is a partial factorial design, where higher order interactions are assumed to be 0, allowing extrapolation to combinations of the components not actually observed.

The solution above to problem P3 will be inadequate when the effects of the factors vary by covariates whose distributions are different in the historical and target population. In this case, it would be necessary to estimate the effects conditionally and then average over the distribution of these in the target population, as above. Conceptually, this is straightforward. Practically, the problem is to know what covariates to use and the relationship between the effects of the factors and the covariates. In the simplest case, where the investigator really does know what covariates to use and the covariates take on only a few levels, it may not be necessary to introduce (possibly arbitrary) modeling assumptions about the relationship between the effects of the factors and the covariates to make headway. But when there are many covariates and/or several continuous covariates, such assumptions become necessary.

There are two matters that make for additional complexity and the need for yet more assumptions. In observational studies, treatment assignment may not be ignorable. If it is ignorable, given known covariates, one can (in theory) proceed as above. If not, other avenues must be considered to achieve identification of parameters of interest.

Finally and perhaps critically, in contrast to the case in the experimental design literature, in most observational studies and social experiments, the number of characteristics an investigator would like to consider may far exceed the number of treatment groups. This will make for identification problems and point identification may end up resting on a number of assumptions that are

difficult to substantively justify. For example, consider the case of an observational study where it is reasonable to assume treatment assignment is ignorable (without covariates) and the average effects do not depend on covariates whose distributions differ in the historical and target populations. In this case, transporting the relationship between the response and its components to the new environment is simple, once the relationship is determined. Suppose now there are 5 components, each having 2 values (i.e., there are 32 combinations of component values); to identify all the effects in the most general case, 32 treatment groups are needed. Even if many of the higher order interactions disappear, identification problems will remain if there are few treatment groups, as in the usual case. This may be the primary reason that the “treatment effects” literature has not explicitly unbundled the components of interventions and attempted to address problem P3 in its full generality. That said, it is unfortunate that social experiments are not usually designed to facilitate understanding the relationship between the components and the effect.

5. IDENTIFYING AND ESTIMATING CAUSAL EFFECTS

5.1. *Background*

Since the invention of randomization (generally attributed to Fisher [1925]), statisticians have emphasized the importance of study design for the estimation of causal effects. In a completely randomized experiment—assuming random sampling from the population of interest and (A-1) and (A-2)—the outcomes of subjects assigned to receive treatment s are a random sample from the distribution of \tilde{Y}_s ; thus, this distribution can be consistently estimated from the data collected in the experiment.

Consequently, as previously noted, comparisons of potential outcomes that only require knowledge of the marginal distribution of outcomes can be made in randomized experiments. For example, statisticians have tested whether or not the outcome under treatment s is stochastically higher than the outcome under treatment s' . Another example is the ITT. Letting s denote treatment and s' the control treatment, statisticians have long known that when data are collected using randomized experiments, the difference between the treatment group mean and the control group mean on the outcome is an unbiased estimate of the ITT.

Under complete randomization the set of potential outcomes

$$(\{\tilde{Y}_s\}_{s \in S}) \perp\!\!\!\perp A, \quad (7)$$

where A is the treatment assignment variable and the notation is used to denote statistical independence. (Note that A refers to the treatment assigned, which may not be the treatment actually received.) Letting \tilde{Y} denote the observed response, under (7), $E(\tilde{Y} | A = s) = E(\tilde{Y}_s | A = s) = E(\tilde{Y}_s)$; thus, the observable conditional expectations identify the parameter $ITT(s, s')$.

The completely randomized experiment is a special case of the conditionally randomized experiment in which subjects are first grouped according to a set W of pretreatment covariates, and a completely randomized experiment is then conducted within the groups. Under conditional randomization, treatment assignment is "ignorable" given the covariates W :

$$(\{\tilde{Y}_s\}_{s \in S}) \perp\!\!\!\perp A | W. \quad (8)$$

Consequently, $ITT(s, s' | W)$ is identified from the observable conditional expectations:

$$E(\tilde{Y}_s - \tilde{Y}_{s'} | W) = E(\tilde{Y} | A = s, W) - E(\tilde{Y} | A = s', W). \quad (9)$$

Rubin (1977, 1978) saw that the conditionally randomized study provides a means to bridge the gap between experimental and observational studies. In observational studies, it is often not reasonable to believe the ignorability assumption:

$$(\{Y_s\}_{s \in S}) \perp\!\!\!\perp D. \quad (10)$$

However, if covariates W can be found that determine the treatment receipt process (in the sense that given these covariates, receipt of treatment does not depend on the potential outcomes), treatment assignment is ignorable, given the covariates (Barnow, Cain and Goldberger 1980 dubbed this "selection on observables"):¹

¹Editor's Note: This sentence is misprinted. The latter part of this sentence should read as follows: "... treatment assignment is ignorable, given the covariates (Barnow, Cain and Goldberger 1980). Heckman dubbed this 'selection on observables':" After this page was typeset and finalized, it was discovered that the character string "). Heckman" was inadvertently omitted.

$$(\{Y_s\}_{s \in S}) \parallel D | W. \quad (11)$$

Under (11), the conditional means for treatments s and s' identify $\text{ACE}(s, s' | W)$:

$$E(Y_s - Y_{s'} | W) = E(Y | D = s, W) - E(Y | D = s', W). \quad (12)$$

The intuition behind (11) is straightforward and readily lends itself to use by empirical investigators. Within levels of W , treatment receipt is decided (in the binary case) by the toss of a (possibly biased) coin. If the parameter (2) is of interest, as versus (12), this is obtained by averaging over the marginal distribution of W . If the average effect is the same for all values of W , it is not necessary to know the distribution of W . Otherwise, it must be possible to estimate this distribution from the data or the distribution must be known; in practice, it may be that neither of these conditions is attainable.

By way of contrast, despite a longstanding interest in making causal statements, until more recently economists were less interested in experimental data than statisticians. In part, this is due to the fact that economists are interested in many questions that are not particularly amenable to experimentation.

Economists have also long recognized that human agents make choices and they use theories of rational decision making to characterize the manner in which agents choose among alternatives. That is, economists attempt to carefully consider one set of mechanisms that individuals might use to allocate themselves to treatments (how agents choose D). Further, this allocation process is often of intrinsic interest to economists.

Heckman characterizes the statistical literature as incomplete, in part because statisticians do not model the allocation process. An example of this is adjusting for covariates using regression analysis, long advocated by statisticians. Here modeling the conditional expectation $E(Y | W, D)$ alone leads to an estimate of (12). If interest resides solely in estimating (12) when (11) holds, there is no need to model the allocation process. But even when (11) holds, especially in an observational study where W may be a large vector, statisticians will often advocate modeling the allocation process to reduce the dimensionality of the estimation problem, a subject to which I shall return.

Nevertheless, the focus in the statistical literature is primarily on obtaining the best possible estimate of the causal parameter of interest.

From this point of view, all else being equal, given the choice between a randomized experiment and an observational study where units select their own treatment, the experiment is typically preferred (especially when ITT is the parameter of interest and/or ACE is the parameter of interest and subjects comply with their assignments (that is, for all ω and for all $s \in S$, subject ω takes up assignment s when assigned to s).

In general, the way in which units are allocated in the experiment will not reflect the real-world allocation mechanism where human agents are making choices, as studied by economists. As such, the opportunity to learn about this mechanism (at least from the experimental study) is given up. This is the price we pay to ensure (7) or (8).

In the observational study, however, we cannot be certain that all relevant covariates have been taken into account. If (11) holds and the regression function is modeled correctly, we can learn about both the allocation process and the causal parameter(s) of interest. But if one or more covariates have not been taken into account and (11) is assumed, credible estimates of causal parameters may not be obtained. As Heckman points out, individuals making decisions may have relevant information that is not accessible to the investigator and therefore such information cannot be included in the investigator's model of the agent's choice. In economic models of behavior, agents use this "hidden" information in computing the expected utility of different choices. The agent then makes the choice that maximizes expected utility. Since it is not unreasonable to suppose that utility is a monotone function of many of the types of outcomes (for example, earnings) studied by economists, in such circumstances (11) will in general not be satisfied for the set of pretreatment covariates accessible to the investigator, and (12) will then not hold. In this case, if (11) is (correctly) not assumed for a given set of available covariates W , credible estimates might be obtained using other methods—for example, fixed effects models (including differences in differences), control functions, instrumental variables. But if the assumptions underlying the use of these alternatives are incorrect in the application under consideration, then as before, credible estimates may not be obtained.

5.2. *Matching, Control Functions, and Instrumental Variables*

These are three approaches to estimating causal parameters. Interestingly, although the rationale and assumptions needed to

justify these approaches differ, the propensity score (discussed below) figures prominently in all three.

In observational studies where it is believed that (11) holds, there still remains the problem of estimating $E(Y | W, D)$. When W is a high-dimensional vector and/or several components have many values, it may be difficult to specify the form of this function correctly, which can lead to faulty inferences. Matching will also be problematic in this case.

Let $S = \{0,1\}$. In a key paper, Rosenbaum and Rubin (1983) showed that when (11) holds and

$$0 < \Pr(D = 1 | W) < 1, \quad (13)$$

then

$$(\{Y_s\}_{s \in S} || \underline{D}) | P(W), \quad (14)$$

$$0 < \Pr(D = 1 | P(W)) < 1, \quad (15)$$

where $P(W) = \Pr(D = 1 | W)$ is the “so called” propensity score. Imbens (2000) generalizes the notion of a propensity score to the case of finitely many treatments. Imai and van Dyk (2004) extend the notion of a propensity score to the more general case where D may take on infinitely many values.

As a consequence of (15),

$$E(Y_1 - Y_0 | P(W)) = E(Y | D = 1, P(W)) - E(Y | D = 0, P(W)). \quad (16)$$

Equation (16) provides the mathematical justification for matching on the one-dimensional propensity score (as opposed to the multidimensional vector W , which may well be sparse), in which observations with the same values of $P(W)$ —one with $D = 1$, the other with $D = 0$ —are randomly paired, their difference providing an unbiased estimate of (16). An unbiased estimate of the parameter (2) can then be formed by taking the appropriate weighted average. Equation (16) can also be used to justify a related method called subclassification and to justify covariance adjustment using only D and $P(W)$ (as versus D and W). Other parameters (for example, TT) can also be estimated using these

methods. At this point, there is a large statistical literature on matching and related methods. The interested reader might wish to consult Smith (1997) for a sociological application and Imbens (2004) for a nice overview of estimating average treatment effects under the assumption (11).

The beauty of matching is explained quite nicely by Heckman (page 65, this volume) and in Heckman and Navarro-Lozano (2004:33): matching “does not require separability of outcome or choice equations into observable and unobservable components, exogeneity of conditioning variables, exclusion restrictions or adoption of specific functional forms of outcome equations.” Other methods of estimating causal effects, such as instrumental variables, fixed effects, and control functions, normally require one or more assumptions of the form above.

Nevertheless, Heckman is quite critical of matching on the propensity score. First, the method breaks down if $P(W) = 0$ or 1 for one or more values of W . In practice, even in less extreme cases, an investigator may encounter the case where the estimated $P(W)$ is close to, for example, 1 and there are no “good” matches from the control group. When such data are excluded, as is often the case, the causal parameter that is actually estimated (an average effect on a common support) may be of less interest. Second, when $P(W)$ is unknown (the typical case) and it is estimated nonparametrically, the dimensionality problem is simply transferred to this estimation problem.

Heckman also argues that it is often difficult to justify the use of (11) for some conditioning set W . According to him, this situation is exacerbated by the absence of an explicit model of treatment choice. Finally, he states that (11) is quite strong substantively, implying $MTE(W) = ACE(W) = TT(W)$.

Of course, it can be argued that (14) may hold even if (11) does not. But it is difficult to think of substantive situations where we would want to argue that (14) holds and hence that (16) holds but (11) does not. We should note also that (12) may hold even if (11) does not hold, and that (16) can hold even if (14) does not. However, as above, it is difficult to think of instances where we would want to argue that one of the weaker conditions holds, but the stronger does not. Thus, I do not consider it worthwhile to further entertain arguments of this nature.

Heckman questions the value of assumption (11) in social contexts. He suggests that when agents have hunches about the values of the potential outcomes, and treatment choice is based on those hunches, assumption (11) will not hold. While often true, there may nevertheless be situations where an investigator knows and measures the covariates on which the agents' decisions are based, in which case (11) holds. See also Imbens (2004) for less trivial examples.

When investigators do not think carefully about the treatment assignment process in observational studies, they are likely to omit important covariates from consideration. That said, it is not the statistician's job to substantively justify a particular model of choice. Nor would it be correct to suggest that statisticians are ignorant of, or do not stress the importance of understanding the treatment assignment mechanism. Indeed, going back to Fisher (quoted in Cochran 1965) statisticians have long acknowledged the importance of having a good theory of the treatment assignment mechanism; see also Rosenbaum (2002, ch. 1), who pays a great deal of attention to this matter.) Rosenbaum and others (see Rosenbaum [2002] for further citations) have also studied the consequences due to the failure to adjust for relevant omitted covariates.

Nevertheless, even when an investigator pays very close attention to the treatment assignment mechanism, a covariate (set of covariates) known to be relevant may be missing from the data and/or some relevant covariates are unknown to the investigator. This will be the case in some instances where treatment assignment is the result of an economic agent behaving rationally and in other instances where some other process describes the allocation to treatment groups. Unfortunately, assumption (11) is not directly testable, though it may be possible, by introducing auxiliary assumptions, to test this indirectly. Heckman's Tables 2 and 3 simply demonstrate what they should: if the assumptions underlying the use of matching are incorrect and the assumptions underlying Heckman's particular example of the use of control functions are correct, the observable parameters that also equal TT and ACE in the case where matching hold are now biased for TT and ACE. When it is suspected that (11) does not hold, an investigator can attempt to conduct sensitivity analyses (as statisticians have long advocated), construct bounds on the parameter(s) of interest—for example, Manski (1990) and Robins (1989)—or use some other approach—for example, fixed effects,

instrumental variables, control functions—to estimate the causal parameter of interest.

Following Heckman, I now examine the method of control functions, expositing the additively separable case also considered by him. He assumes (his equations 22a–22c)

$$V = \mu_V(W) + U_V, \quad E(U_V|W) = 0, \quad (17)$$

$$Y_s = \mu_s(X) + U_s, \quad E(U_s|X) = 0, \quad (18)$$

where $s = 0$ or 1 and $D = 1$ if and only if $V > 0$.

The observable conditional expectations ($Y = Y_1$ if $D = 1$, Y_0 if $D = 0$) are (using 18)

$$E(Y|X, Z, D = s) = \mu_s(X) + E(U_s|X, Z, D). \quad (19)$$

Under assumption (18), when (11) holds (with $(X, Z) = W$), $E(Y | X, Z, D = s) = E(Y_s | X, Z) = \mu_s(X)$. Note that the first equality follows from (11) and the second from the additional assumption (18); that is, the additional assumption (18) is not needed to justify matching on the propensity score. In the method of control functions, however, assumption (11) is not made and the components $E(U_s | X, Z, D = s)$ are modeled. Note that $E(U_1 | X, Z, D = 1) = E(U_1 | X, Z, V > 0) = E(U_1 | X, Z, U_V > -\mu_V(Z))$ by virtue of assumption (17); similarly, $E(U_0 | X, Z, D = 0) = E(U_0 | X, Z, V \leq 0)$. Thus, under (17) and (18), it might seem that the method of control functions is more general than matching. But modeling $E(U_s | X, Z, D = s)$ will require additional assumptions—for example, Heckman's assumption (C-1): $(U_1, U_0, U_V) \perp\!\!\!\perp (X, Z)$. Assumption (C-1) implies $(U_1, U_0) \perp\!\!\!\perp (X, Z) | U_V$, so that $E(U_s | X, Z, D = s)$ depends on X, Z only through the propensity score $P(X, Z)$. As in matching, a problem involving high dimensionality is now reduced to a one-dimensional problem through the use of the propensity score. It is worth noting that assumption [C-1] does not imply (11). Nor does (11) imply (C-1). Thus, even if (17) and (18) hold, it is not the case that “the control function approach is more general than the matching approach” (page 73, this volume). (Heckman points out that assumption (C-1) is not essential. Nevertheless, if this assumption is removed, others will

have to be made.) The two approaches simply make different assumptions and will thus be useful in different circumstances.

One other point should be made. Heckman notes: “Without invoking parametric assumptions, the method of control functions requires an exclusion restriction (a variable in Z that is not in X) to achieve nonparametric identification.” But he is far less critical of these assumptions (and others noted above) than he is of those required to justify matching and the use of instrumental variables. In that vein, Vella (1998, p. 131) points out the sensitivity to parametric assumptions of Heckman’s original work: “As estimation relies heavily on the normality assumption, the estimates are inconsistent if normality fails.” Vella (1998, p. 135) also notes that the exclusion restriction is “controversial” and he argues that many theoretical economic models of behavior, including the Roy model discussed by Heckman, explicitly impose $Z = X$.

Using instrumental variables is another way to estimate treatment effects in observational studies, and it makes assumptions that are different than those made in matching or the method of control functions. Social scientists have long used instrumental variables to estimate treatment effects when treatment choice is “endogenous.” Traditionally, the technique is explicated as follows. Consider the regression

$$Y = \mu(X) + \tau D + \varepsilon, \quad (20)$$

where $D = 1$ if the treatment is received, 0 otherwise, τ is the desired treatment effect, and $E(\varepsilon | X) = 0$. The problem here is that D is correlated with ε , so in general $E(\varepsilon | X, D) \neq 0$ (equivalently, $E(Y | X, D) = \mu(X) + \tau D + E(\varepsilon | X, D)$). However, if a variable Z can be obtained that is associated with Y only through D , i.e., Z does not directly affect the outcome, $E(\varepsilon | X, Z) = E(\varepsilon | X) = 0$, in which case $E(Y | X, Z) = \mu(X) + \tau E(D | X, Z)$. Consequently (assuming $E(D | X, Z = 1) - E(D | X, Z = 0) \neq 0$),

$$\tau = \frac{E(Y|X, Z = 1) - E(Y|X, Z = 0)}{E(D|X, Z = 1) - E(D|X, Z = 0)}. \quad (21)$$

From a causal standpoint, the formulation above is quite vague. Heckman has helped to clarify the literature on instrumental variables. Angrist, Imbens, and Rubin (1996) is another paper that I

find useful, and the approach taken there is somewhat different than Heckman's. Thus, I briefly exposit this approach and subsequently tie it to the exposition in Heckman; see also Vytlačil (2002).

I will focus on several parameters discussed by Heckman ($ACE(X)$), the local average treatment effect (hereafter $LATE(X)$), $TT(X)$, and I will also briefly discuss $ITT(X)$. Following Heckman, Z is the instrumental variable. It will also be taken to be binary, as in Angrist et al. (1996). (See Angrist and Imbens [1995] for some generalizations of the setup considered herein.) Let Z (previously denoted A) denote the treatment to which a subject is assigned (0 if assigned to the control group, 1 if assigned to the treatment group). Let $D(\omega)$ denote the observed choice of unit (ω) and let $D_z(\omega)$ denote the choice unit ω makes when assigned to treatment $z \in \{0, 1\}$. Similarly, let $Y_{(z,D_z)}(\omega)$ denote the response of unit ω when that unit is assigned to treatment z and chooses outcome $D_z(\omega)$. (Previously, $Y_{(z,D_z)}(\omega)$ was denoted $\tilde{Y}_z(\omega)$.) Let $Y_{zs}(\omega)$ denote the outcome of unit ω when that unit is assigned to treatment z and "takes up" treatment s , for $z = 0, 1$, $s = 0, 1$. Note that for each assignment, individuals take up only one treatment; nevertheless, as above, potential outcomes assuming they had taken up the treatment they did not take up can be defined.

To begin, it is useful to formalize the exclusion restriction—that is, the idea that the instrumental variable only affects the outcome by affecting D . This is the assumption (Holland 1988)

$$Y_{(0,s)}(\omega) = Y_{(1,s)}(\omega) \quad (22)$$

for $s = 0, 1$ and all ω . Consequently, the potential outcomes may be written as $Y_s(\omega)$. The exclusion restriction is very strong, and it can be quite difficult to find instruments that satisfy this assumption.

The problem with estimating the effect of D (conditional on the covariates X) on the outcome is that (11) will not generally hold, because D is "endogenous"; thus, in general, $E(Y | D = s, X) \neq E(Y_s | X)$. However, if (8) holds (with Z in place of A), as would be the case in a randomized experiment,

$$E(Y|Z = 1, X) - E(Y|Z = 0, X) = E(Y_{1,D_1} - Y_{0,D_0}|X), \quad (23)$$

that is, $ITT(X)$ is the numerator of the IV estimand (21). (Recall the previous discussion, which suggests that at least in some instances,

$ITT(X)$ and/or ITT may be the parameter(s) of greatest interest to a policymaker.)

Continuing, $ITT(X)$ may be broken down into the following four components:

$$E(Y_{1,D_1} - Y_{0,D_0} | X) = EE((Y_{1,D_1} - Y_{0,D_0}) | D_0, D_1, X), \quad (24)$$

where $(D_0, D_1) = (0, 0)$ or $(0, 1)$ or $(1, 0)$ or $(1, 1)$. By virtue of the exclusion restriction (22), units who always take up the treatment ($D_0(\omega) = D_1(\omega) = 1$), hereafter called “always takers,” or never take up the treatment ($D_0(\omega) = D_1(\omega) = 0$), hereafter called “never takers,” contribute nothing to (24). Angrist et al. (1996) call subjects with $D_1 = 1, D_0 = 0$ compliers and subjects with $D_1 = 0, D_0 = 1$ defiers; only these two types of units contribute to (24) under the exclusion restriction.

Angrist et al. (1996) also assume there are no defiers (the monotonicity assumption), in which case

$$ITT(X) = E((Y_{1,D_1} - Y_{0,D_0}) | D_0=0, D_1=1, X) \Pr(D_0=0, D_1=1 | X). \quad (25)$$

Dividing $ITT(X)$ by the compliance probability (assuming this is greater than 0) gives the parameter $LATE(X)$, the average treatment effect for the compliers (at X). The compliance probability $\Pr(D_1 = 1, D_0 = 0 | X) > 0$ may also be written (under the assumptions here) as $E((D_1 - D_0) | X)$. But this is equal to $E(D | X, Z = 1) - E(D | X, Z = 0)$ when treatment assignment (Z) is ignorable, given X , as here. Thus, under the assumptions above, $LATE(X) = IV(X)$. Note also that the compliance probability may be written as $\Pr(D_1 = 1 | X) - \Pr(D_0 = 1 | X) = P(X, 1) - P(X, 0)$, which makes the connection with the propensity score evident.

The parameter $LATE(X)$ (or $LATE$ when there are no covariates X) will not always have policy implications of interest. To begin, the compliers constitute a latent subpopulation. So, even if we wanted to administer the treatment only to the compliers and it was politically feasible to do so, it is not possible to identify these individuals (in practice, we could model the probability of being a complier and administer the program to those deemed “most likely” to be compliers). Second, when the compliers are a “small” fraction of the population, it may be difficult to argue that the results are of great

interest. For example, the question addressed by Angrist et al. (1996) is the excess civilian mortality (between 1974 and 1983) resulting from service in the Vietnam War (not the excess mortality among compliers). For men born in 1950, the compliers constitute only 15.9 percent of the population; technically, *LATE* only applies to this fraction of the population. In some applications, however, even if the compliers are a small fraction of the population, *LATE* (or *LATE(X)*) is nevertheless a parameter of great interest. This would be the case when it could be argued that the noncompliers, had they complied, would experience the same benefits as the compliers. I return to this subject momentarily. Third, Heckman (1997) has also pointed out that *LATE* (*LATE(X)*) is an unusual parameter, insofar as its very definition depends on the instrumental variable chosen. Thus, in some cases, *LATE(X)* and/or *LATE* may identify a parameter with policy relevance (as when *Z* represents assignment under a particular policy of interest), and in other cases it may not. For further discussion of *LATE* and other possible parameters of interest, see the discussion following Angrist et al. (1996) and Heckman (1997).

Although the parameters *LATE* and *LATE(X)* may not always be of great substantive interest, the methodological point is that the meaning of the IV estimand has been clarified (which has great substantive implications). In particular, a basis is provided that makes it very easy to ask if *IV(X)* identifies other parameters of possibly greater interest, such as *TT(X)* and *ACE(X)*.

To see this, consider the parameter *TT(X)*, which conditions on receipt of treatment ($D = 1$). The units receiving treatment are the compliers in the treatment group and the always takers (still assuming there are no defiers). It follows from the foregoing results that $IV(X) \neq TT(X)$ in general, and that $IV(X) = TT(X)$ if and only if the average effect of receiving treatment for the always takers (assuming the probability of being an always taker is greater than 0) and compliers is the same. Put this way, an analyst can ask whether the equality of treatment effects across these two groups is a reasonable assumption to make. If the analyst suspects, for example, that the always takers know that (even after conditioning on *X*) they will benefit by taking up the treatment (or have higher gains than others by so doing), he or she will not want to assume equality across groups and hence that $IV(X) = TT(X)$.

It is also easy to see that there is one important case where $IV(X)$ must equal $TT(X)$. If the treatment cannot be obtained in the control group, as in many social programs, it is not possible to be an always taker. In this case, $LATE(X) = TT(X)$ (without it being necessary to assume that the average effects of receiving treatment are the same for compliers and always takers), hence $IV(X) = TT(X)$.

Similarly, if the average effect of D on the response is the same for compliers, always takers, and never takers, $IV(X) = LATE(X) = TT(X) = ACE(X)$. If it is not possible to be an always taker, $LATE(X) = TT(X)$ (as above) and $LATE(X) = ACE(X)$ (hence $IV(X) = ACE(X)$) when it is assumed that the average effects of receiving treatment are identical for never takers and compliers. In cases where it is impossible to be a never taker (programs with universal coverage and participation), $LATE(X) = ACE(X)$ if it is assumed that the average effects of D on Y are identical for always takers and compliers.

In the case where the unit effects of D on Y are the same for all ω , the average effects of receiving treatment must be the same for all units, hence all groups, implying $IV = TT = ACE$. Of course, the assumption of constant effect is quite strong and not likely to be substantively reasonable in most social science applications.

Finally, if the probability of being a defier is nonzero, in general $IV(X) \neq LATE(X)$; but in the special case where the average effect of receiving treatment for compliers and defiers is the same, $IV(X) = LATE(X)$. Angrist et al. (1996) also discuss the consequences of violating the exclusion restriction, and there is some literature on estimating complier average causal effects in the absence of this restriction (for example, see, Jo [2002]).

Heckman approaches this subject somewhat differently. He imposes the additively separable model (18) on the potential outcomes. He then writes the observed outcome Y in terms of the potential outcomes as

$$Y = \mu_0(X) + (\mu_1(X) - \mu_0(X) + U_1 - U_0)D + U_0, \quad (26)$$

expresses the parameters $TT(X)$ and $ACE(X)$ in terms of (26), and states identifiability conditions in terms of D , U_0 , and U_1 .

The assumption of a constant effect holds ($Y_1 - Y_0$ is the same for all units) if $U_0 = U_1 \equiv U$. In this case, the instrumental variable Z

needs to satisfy the condition $E(U | X, Z) = E(U | X) = 0$ (equivalently, under (18) $E(Y_s | X, Z) = E(Y_s | X)$ for $s = 0, 1$). As above, a sufficient condition for this is $Y_s \perp\!\!\!\perp Z | X$ for $s = 0, 1$, and as above, assuming $P(X, 1) - P(X, 0) \neq 0$, $IV(X) = LATE(X) = TT(X) = ACE(X)$.

When the constant effect assumption fails but $E(U_1 - U_0 | X, D = 1) = 0$, Heckman (1997) shows that $TT(X) = ACE(X)$. A sufficient condition for this is

$$(U_1 - U_0) \perp\!\!\!\perp D | X \quad (27)$$

(or more generally $(Y_1 - Y_0)D | X$). Though weaker than the condition (11), which was not presumed to hold, Heckman points out that the sufficient condition above is nevertheless quite strong, requiring that receipt of treatment not depend, given X , on gains anticipated by subjects. That is, in general, we should not expect $TT(X) = ACE(X)$. We can also see this from the results above, where it was established that if there are no defiers, and the average effect of D on the response is identical for compliers and always takers, $LATE(X) = TT(X) = ACE(X)$. Similarly, if there are defiers and the average effect of receiving treatment is identical for defiers and compliers, and for compliers and always takers, $LATE(X) = TT(X) = ACE(X)$.

Heckman also gives general conditions under which $IV(X) = TT(X)$ and $IV(X) = ACE(X)$. As in the simpler case above, and for the same reasons, Heckman argues that these conditions are quite strong. Again, this argument seems most compelling when the analyst does not have access to data that the decision maker is using to make his decision and this information is predictive of the potential outcomes. For further details, the reader may consult Heckman (1997) or his paper in this volume.

6. CONCLUSION

Heckman argues for the use of an approach to causal inference in which structural models play a central role. It is worth remembering that these models are often powerful in part because they make strong assumptions. When these assumptions are correct, powerful (and correct) inferences may be obtained. Such inferences are likely to be stronger than those that would be made by advocates of randomized

experiments. For example, using a structural model in an observational study, we might learn about the treatment assignment mechanism and various average effects, and we might extrapolate the results to a new policy in a new environment. But when the assumptions are arbitrarily invoked in applications or require the use of knowledge that the investigator does not have, as seems often the case, so are the inferences derived from such modeling exercises. Thus, an investigator might well prefer to stick with simple estimators from randomized experiments, whenever possible. In such a case (presuming the experiment did not get botched and subjects complied with experimental protocols), the investigator can have greater confidence in his or her estimates of parameters such as ITT and ACE, for example.

But I do not want to argue that structural modeling is not useful, nor do I want to suggest that methodologists should bear complete responsibility for the use of the tools they have fashioned. To my mind, both structural modeling and approaches that feature weaker assumptions have their place, and in some circumstances, one will be more appropriate than the other. Which approach is more reasonable in a particular case will often depend on the feasibility of conducting a randomized study, what we can actually say about the reasonableness of invoking various assumptions, as well as the question facing the investigator (which might be dictated by a third party, such as a policy-maker). An investigator's tastes and preferences may also come into play. A cautious and risk-averse investigator may care primarily about being right, even if this limits the conclusions he or she draws, whereas another investigator who wants (or is required) to address a bigger question may have (or need to have) a greater tolerance for uncertainty about the validity of his or her conclusions.

In his introductory section, Heckman claims to make two major points: (1) that "causality is a property of a model of hypotheticals" (page 2), and (2) that statisticians have conflated the distinct tasks of defining parameters of interest, identification, and estimation. I have already discussed the first point. I conclude with a discussion of the second. With respect to this point, Heckman writes (page 5): "This emphasis on randomization or its surrogates (like matching) rules out a variety of alternative channels of identification of counterfactuals from population or sample data. It has practical consequences because of the conflation of step one with steps two and three in Table 1. Since randomization is used to define the parameters of

interest, this practice sometimes leads to the confusion that randomization is the only way—or at least the best way—to identify causal parameters from real data.”

Heckman appears to be arguing here that statisticians are putting the cart before the horse by focusing interest on average causal effects that do not depend on the joint distribution of potential outcomes and emphasizing identification conditions in observational studies that parallel random assignment, thus justifying estimation methods such as matching and even randomization itself. While it is impossible to assess such a claim, it is worth noting that average causal effects such as the ACE and ITT have been of great interest in public health, for example, for many years. These parameters can and have been used to address policy questions that are of great interest. Recall also that both the potential outcomes notation and the ACE (Neyman 1923) preceded randomization.

Of course, Heckman is certainly correct to note that there are interesting estimands that depend on the joint distribution and that here, randomization is of considerably less help. In addition, as he and many others have pointed out, when it is impossible for the investigator to obtain a sufficiently rich set of covariates to condition on, other methods of identifying, and estimating causal effects (including the usual effects that do not depend on joint distributions of potential outcomes) must be used.

But Heckman goes much further, arguing that statisticians have confounded the tasks of defining, identifying, and estimating causal parameters and, as above, even use randomization to define parameters of interest. By and large (except for some minor quibbles one might have about the way some authors have defined *LATE*), I would argue the opposite. One of the key contributions that statisticians have made is to unconfound these issues, paving the way for (1) the assessment of conditions under which valid causal inferences are permitted and (2) the development of appropriate methods for making valid causal inferences.

Consider the claim that randomization is used to define causal parameters of interest. In the introduction, I stressed the importance of good notation. By using the potential outcomes notation, statisticians (recall Neyman 1923 and later Rubin) were able to define causal estimands that mirrored their thinking on the counterfactual nature of the causal relation and that were different from the usual descriptive (observable) parameters.

Once such estimands have been defined, it can then be asked under what conditions various observable parameters are equal to (identify) these estimands. Randomization is a device for assigning subjects to treatments that makes the ignorability assumptions (conditions) (8) and/or (11) plausible. When these conditions hold, various observable parameters also equal the causal estimands. These conditions may also be met when randomization has not been used. This demonstrates the logical independence between the ignorability conditions and randomization. And clearly, these conditions are also logically independent of the definitions of causal estimands such as the ITT, ACE, and TT. (Readers might also want to look directly at the definition of these parameters and note that no mention of randomization is made.)

More generally, defining causal estimands independently of the conditions that must be met in order to identify them allows for the development of appropriate procedures (including randomization, matching, IV, control functions, etc.) for identifying (and then estimating) the causal parameters. This is the approach taken in both the “treatment effects” literature and recent econometric literatures, and it is also the approach that Heckman takes. It is a big step forward.

Another way to see the utility of making the definitions of causal effects logically independent of the conditions needed to identify them is to consider the usual approach to regression analysis (or structural equation models) which is typically taken (both in the past and often even now) by many social scientists. The parameters of a regression are certainly interpretable in a descriptive sense, but social scientists often impart a causal interpretation to one or more (often to all) parameters, which are typically interpreted as “effects” in this counterfactual sense (see Sobel [1990] for more on this point). Justifications for such interpretations have included the notion that the model is well specified and/or that important confounders have been controlled and/or that the causal ordering is correct. All of these justifications are extra-mathematical and virtually impossible to evaluate, insofar as a target (i.e., a well-defined estimand) has not even been defined. Using an appropriate notation allows the researcher to clearly define the estimand of interest independently of the regression parameter(s), enabling the analyst to give conditions under which the regression parameter(s) actually identify the target(s) of interest.

Although I disagree with him on this point and a number of others, Heckman, in conjunction with his collaborators, has made

useful contributions to the literature on causal inference. I hope the next generation of researchers will cooperate and incorporate the various literatures on causal inference, including the statistical and econometric literatures, under one umbrella. Science will be better served when this is the case.

REFERENCES

- Angrist, Joshua D., and Guido W. Imbens. 1995. "Two Stage Least squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90:431–42.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables" (with discussion). *Journal of the American Statistical Association* 91:444–72.
- Barnow, Bert S., Cain, Glenn C., and Arthur S. Goldberger. 1980. "Issues in the Analysis of Selectivity Bias." Pp. 43–59 in *Evaluation Studies Review Annual, 5*, edited by E. Stromsdorfer and G. Farkas. Beverly Hills: Sage.
- Bunge, Mario. 1979. *Causality and Modern Science*. 3d ed. New York: Dover.
- Carneiro, Pedro, Hansen, Karsten T., and James J. Heckman. 2003 "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44:361–432.
- Cochran, William G. 1965. "The Planning of Observational Studies of Human Populations." *Journal of the Royal Statistical Society, Series. A*, 128:234–55.
- Collingwood, Robin G. 1940: 1972. *An Essay on Metaphysics*. Chicago, IL.: Henrey Regnery Company.
- Cox, David R. 1958. *The Planning of Experiments*. New York: Wiley.
- Fisher, Franklin M. 1970. "A Correspondence Principle for Simultaneous Equation Models." *Econometrica* 38:73–92.
- Fisher, Ronald A. 1925. *Statistical Methods for Research Workers*. Edinburgh, Scotland: Olive and Boyd.
- Halloran, M. E., and C. J. Struchiner. 1995. "Causal Inference in Infectious Diseases." *Epidemiology*, 6:142–51.
- Heckman, James J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32:441–62.
- . 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics* 115:45–97.
- . 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109:673–748.

- Heckman, James J., and Salvador Navarro-Lozano. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics* 86:30–57.
- Holland, Paul W. 1986. "Statistics and Causal Inference" (with discussion). *Journal of the American Statistical Association* 81:941–70.
- . 1988. "Causal Inference, Path Analysis, and Recursive Structural Equations Models." (with discussion). Pp. 449–493 in *Sociological Methodology*, edited by C. C. Clogg. Washington, D.C: American Sociological Association.
- Imbens, Guido W. 2000. "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika* 87:706–10.
- . 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *Review of Economics and Statistics* 86:4–29.
- Imai, Kosuke, and David A. van Dyk. 2004. "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association* 99:854–66.
- Jo, Booil. 2002. "Estimation of Intervention Effects with Noncompliance: Alternative Model Specifications" (with discussion). *Journal of Educational and Behavioral Statistics* 27:385–420.
- Kemphorne, Oscar. 1952. *The Design and Analysis of Experiments*. New York: Wiley.
- Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Review Papers and Proceedings* 80:319–23.
- . 2000. "Identification Problems and Decisions Under Ambiguity: Empirical Analysis of Treatment Response and Normative Choice of Treatment Choice." *Journal of Econometrics* 95:415–42.
- . 2004. "Statistical Treatment Rules for Heterogeneous Populations." *Econometrica* 72:1221–46.
- Neyman, Jerzy S. 1923: 1990. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9" (with discussion). *Statistical Science* 4:465–80.
- Pearl, Judea. 2000. *Causality*. Cambridge, England: Cambridge University Press.
- Robins, James M. 1989. "The Analysis of Randomized and Non-Randomized AIDS Trials Using a New Approach to Causal Inference in Longitudinal Studies." Pp. 113–59 in *Health Service Research Methodology: A Focus on AIDS*, edited by Lee Sechrest, Howard Freeman, and Albert Mulley. Washington, DC: U.S. Public Health Service, National Center for Health Services Research.
- Robins, James M., and Sander Greenland. 2000. "Comment on 'Causal Inference without Counterfactuals,' by A. Philip Dawid." *Journal of the American Statistical Association* 95:431–35.
- Rosenbaum, Paul R. 2002. *Observational Studies*. 2d ed. New York: Springer.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41–55.
- Rosenzweig, Mark R., and Kenneth I. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38:827–874.

- Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688–701.
- . 1977. "Assignment to Treatment Groups on the Basis of a Covariate." *Journal of Educational Statistics* 2:1–26.
- . 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics* 6:34–58.
- . 1980. "Comment on 'Randomization Analysis of Experimental Data: The Fisher Randomization Test,' by D. Basu." *Journal of the American Statistical Association* 75:591–93.
- Smith, Herbert L. 1997. "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies." Pp. 325–53 in *Sociological Methodology*, vol. 27, edited by Adrian E. Raftery. Boston, MA: Blackwell Publishing.
- Sobel, Michael E. 1990. "Effect Analysis and Causation in Linear Structural Equation Models." *Psychometrika* 55:495–515.
- . 1995. "Causal Inference in the Social and Behavioral Sciences." Pp. 1–38 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by G. Arminger, C. C. Clogg, and M. E. Sobel. New York: Plenum Press.
- . 2001. "Spatial Concentration and Social Stratification. Does the Clustering of Disadvantage 'Beget' Bad Outcomes?" Forthcoming in *Poverty Traps*, edited by S. Bowles, S. N. Durlauf, and K. Hoff. New York: Russel Sage Foundation.
- . 2003. "What Do Randomized Studies of Housing Mobility Demonstrate: Causal Inference in the Face of Interference." Unpublished manuscript, Columbia University.
- Strotz, Robert H., and Herman O. A. Wold. 1960. "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis (Part 1)." *Econometrica* 28:417–27.
- Vella, Francis. 1998. "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources* 33: 127–169.
- Vytlačil, Edward. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica* 70:331–41.
- Whitehead, Alfred N. [1911] 1958. *An Introduction to Mathematics*. New York: Oxford University Press.

