

Dialogue on Causality between James Heckman and Ian Shrier

James J. Heckman

jjh@uchicago.edu

*Center for the Economics of Human Development and Department of Economics
University of Chicago
1126 East 59th Street, Chicago, IL 60637, USA*

Ian Shrier

ian.shrier@mcgill.ca

*Centre for Clinical Epidemiology
Lady Davis Institute, Jewish General Hospital
3755 Cote Ste-Catherine Rd, Montreal, QC, Canada H3T 1E2*

1. General Perspective

Interviewer: What is your general approach to science and causality?

Heckman: The current literature on causality is filled with monologues of various participants hawking their wares, ignoring what others have to say. I have read the literature in statistics and computer science closely. It has had a big influence on many fields, including on my own, and often to their detriment. As an example, discussions about “preferred estimators” show its malign influence. Like much of the current “causal inference” literature, the emphasis on “preferred estimators” conflates estimation methods with conceptual definitions of causality.

I am an economist trained (at the college level) in physics and mathematics. I switched to economics out of interest in social and economic issues and in the belief that economics can be a science and can contribute to understanding and resolving important policy debates. The specificity of the policy problems addressed and the generality of the analytical framework used to address them attracted me to the field. Starting with my graduate training in economics at Princeton, I have also taken a strong interest in statistics as a tool for sharpening empirical investigations. I have published papers in statistics journals and symposia.

I have spent my entire professional life as an economist attempting to respect the high standards and the rigorous protocols of hard science in my research and in that of my students. I have sought to produce hard, verifiable (replicable), empirical evidence and to influence others around me to do the same.

I seek non-tautological, rigorously justified models derived from theory and verified on rigorously justified data. Measurement and theory together are the hallmarks of good

science. No serious scientist pretends to “let the data speak for themselves,” nor would they impose models on data that wouldn’t support them.

I have long practiced abductive inference (see, e.g., Josephson and Josephson, 1996), building and testing models using all sources of data (quantitative and qualitative), which may include censuses, cross-section surveys, experimental data, observer reports, newspaper accounts, interviews, ethnographic studies, and controlled and natural experiments. No particular source of data is uniquely privileged, although more objective, carefully collected and documented data are preferred if they are available. The goal is always to obtain evidence as free of personal belief as possible. If better documented sources are not available, other sources of data can be valuable even if they are less credible.

A rigorous scientific approach to any investigation requires admitting failure when it happens, revising models in the light of failures when they occur, extracting new implications of old models, and testing the revised models so derived on fresh samples if possible. The key to learning from data is being honest, determining and admitting multiple interpretations of the evidence if they appear. One should constantly play devil’s advocate and challenge one’s own work. The search for, care for, and consideration of alternative explanations is the key to rigorous empirical work.

A central feature of hard science and rigorous economics is that it addresses well-posed problems and presents qualified answers. It is not about seeking the estimand of a “preferred” estimator. It is about having well-posed scientific questions and separating anecdote and bias from hard evidence. It is a public activity that invites scrutiny and challenge and not about any particular statistical procedure.

2. On Mentorship

Interviewer: Who would you consider your most important mentors and why?

Heckman: My mentors in economics include three great minds from an earlier generation: Ragnar Frisch, who shared the first Nobel prize in economics and outlined the agenda of econometric policy evaluation (1930, published 2010; 1938); his student Trygve Haavelmo (1943; 1944) who formalized the notion of causality and simultaneous causality; and Jacob Marschak (1953) who further developed the study of causality in pursuit of answering well-posed policy problems.

The Cowles Commission at the University of Chicago (1939–1955), which supported Haavelmo for part of his career, was led, for a while, by Marschak. It developed the first rigorous framework for defining causality and making causal inferences. It defined and analyzed causality in simultaneous systems. It was sketched by Frisch (1930, published 2010), initially for linear models. These frameworks have been greatly extended beyond the early linear normal frameworks used by the Cowles pioneers, although some statisticians to this day continue to attack economics for using linear normal models and continue to claim that simultaneous causality is not possible, despite 80 years of research on the topic.

Two other economists, both still alive, taught me by example the possibility of using econometric models to answer the deep question of forecasting the demand for new goods never previously experienced: Richard Quandt (1958; 1966; 1976) and Daniel McFadden (1975).

3. On Causality

Interviewer: What do you consider the core issues in causality?

Heckman: Before one can meaningfully talk about causality one must define counterfactuals, which lie at its core. Counterfactuals that describe possible outcomes under different conditions are an expression of human imagination and creativity. Different groups judge the quality of the counterfactuals created by different standards. Counterfactuals are *imagined* outcomes: *mental* constructs defined according to some set of rules, which may be implicit.

There are no hard and fast rules for generating counterfactuals. Possibilities are only limited by the imaginations of their creators. Counterfactuals are products of the mind whose plausibility depends on how well they respect the rules invoked by their users. Scientific communities emerge to establish rules for creating counterfactuals from rigorous theory and to verify their construction. However standards vary across fields (see, e.g., Feynman, 1981 at <https://vimeo.com/118188988>).

Science doesn't stop with possibilities. It is about the verification *in objective data* of counterfactual predictions. Verification is the process of checking the ingredients, both theory and evidence, and holding them up for public scrutiny, including replicability. Public scrutiny is not a popularity contest, although it is often interpreted this way by networks of like-minded individuals who ratify each other and ignore outsiders (see, e.g., Carrell et al., 2022). At sufficient scale, such networks can prevent serious scrutiny of their core ideas.

While there may be many possible worlds, their plausibility depends on whether they rely on credible ingredients. Scientific research demands careful measurement and rigorous testing. It also asks that the counterfactuals generated are grounded in scientific principles previously established in the body of previous research. Science is a cumulative process that seeks consilience across studies.

Richard Feynman's book *The Character of Physical Law* (1965) is a superb popular discussion of how rigorous science works. He shows how abstract models of reality often have astounding accuracy in analyzing and predicting real phenomena. It is their power in making predictions in real data that creates their acceptance. Another famous physicist, Eugene Wigner, wrote an influential paper (1960) that captures his amazement and that of other scientists in how well abstract mathematical models predict real-world phenomena. Abstract models in physics follow the rules of physics up to a point but then may extend them. They are not empirical statements but—as Feynman and Wigner noted—they are often powerful in explaining empirical phenomena.

The acts of conceiving counterfactuals and their relationships—thought experiments—and the acts of estimating and testing the validity of these imagined relationships are fundamentally distinct. However, they are often confused, especially in fields without guiding abstract principles and cumulative knowledge. Purely statistical approaches often ignore the crucial point that science is an iterative process. Scientists build models, test them, adapt them to fresh data and/or on new data, and/or examine further implications of proposed new models to see if they hold up. It is by iteration and public argument that scholars learn from data and build models to explain them. This feature is often absent from many applications of “causal inference” in statistics.

There is no “correct” way to generate counterfactuals unless a set of generator rules is postulated. If counterfactuals claimed to be generated by such rules are in fact inconsistent with them, they are “incorrect” in terms of the announced rules. But there are no absolutes in this business—just agreement with prior knowledge up to the point of discovery. Arguments that counterfactuals are “nearest possible worlds” (see Lewis, 1973) flounder on the lack of any metric (or topology) for “closeness” or enumeration of the set of possible worlds.

When two (or more) counterfactuals are compared, a causal effect is obtained. Some counterfactuals may be rooted in fact, others in fancy. Thus, if we compare US history if the South had won the Civil War to US history if it had lost (as actually happened), we define a “causal parameter” (really a causal scenario) holding all else the same. Of course, the two counterfactuals differ greatly in terms of their anchor in fact. The first opens up many possibilities, none of which necessarily occurred, unless the two histories are identical. The second is a topic studied by historians. History itself is subject to controversy, as heated academic disputes attest. Our beliefs about the quality of causal effects are grounded in the quality of the counterfactuals underlying them.

3.1 Econometric Framework

Heckman: An elementary framework is helpful in understanding the econometric approach. Economic theory is grounded in abstract principles. Economic models are models of possible outcomes. They are thought experiments. Consider a simple linear model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U. \tag{1}$$

(X_1, X_2, U) are not necessarily random variables or generated by stochastic processes. These can also be constants when used in functional relationships; e.g., letting $(Y, X_1, X_2, U) = (y, x_1, x_2, u)$, we can write (1) as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u. \tag{2}$$

Equation (1) is an example of a function $g : (X_1, X_2, U) \rightarrow Y$. Equation (2) is an example of g evaluated at the specified values. Economists study how hypothetical variation in the

arguments (X_1, X_2, U) affect Y when the other arguments are fixed (held constant). This is a general functional relationship. One thought experiment imagines an increase in x_1 , holding x_2, u at fixed values. Of course in the real world, one may not be able to randomly vary X_1 without also varying X_2 and U . That is an entirely different matter.

In Equation (1), β_1 is the effect of a unit increase in X_1 , holding X_2 and U fixed. Each Y obtained by varying X_1 holding $(X_2 = x_2, U = u)$ is a possible (“potential”) outcome over the support of (1). Such relationships are at the core of economic theory and have been for over 130 years, and have been at the core of mathematics proper for centuries. Marshall (1890) referred to β_1 as the *ceteris paribus* effect of X_1 on Y . *Ceteris paribus* means “everything else the same.” Precisely the same types of qualifications are given when examining the pull of gravity on a feather.

Thought experiments often illuminate real world policy debates. Thus, if Y is the consumption of cigarettes and X_1 is the price of cigarettes, the *ceteris paribus* variation in X_1 is informative about the possible effect of a tax (which increases the price) on smoking. Hard science and rigorous economics are grounded in thought experiments like these. Einstein is famous for his use of thought experiments, and Feynman (1965) illustrates their usefulness. A variety of potential outcomes can be obtained by varying X_1, X_2 , and U in different ways.

I have deliberately used a linear model as my example of an abstract theoretical model. I distinguish it from a familiar linear regression model, as is usually taught in statistics. That model starts with a collection of random variables (Y, X_1, X_2, U) . Up to now, I have not had to specify whether Y, X_1, X_2, U are proper random variables or not.

Under normality of (X_1, X_2, U) (alternatively of (Y, X_1, X_2)) and taking conditional expectations of (1),

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + E(U|X = x_1, X_2 = x_2). \quad (3)$$

If $E(U|X = x_1, X = x_2) = 0$, one obtains

$$E(Y|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \quad (4)$$

Evaluating Equation (1) at the values $X_1 = x_1, X_2 = x_2, U = u$ gives Equation (2), which is very similar to but conceptually distinct from Equation (4). Setting U in (2) at the expected value of $U(u = 0)$, one obtains

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (5)$$

which is exactly Equation (4).

The fact that the right-hand sides of Equations (4) and (5) are mathematically identical obscures a crucial difference: Equations (1), (2), and (5) are the result of *thought experiments*. The analyst *conceptually sets* the values of X_1, X_2 and U . No data are in sight, although the thought experiment may have been motivated by empirical studies. No

properties of random variables need to be specified. In contrast, Equations (3) and (4) are generated by another type of thought experiment: by taking conditional expectations used to describe data. They are ingredients frequently used in statistical methods. In contrast, Equation (5) is a causal relationship that determines the value of outcome y . The values of x_1, x_2, u are *hypothetically* assigned to (X_1, X_2, U) . Equation (4) is defined by a statistical operation that might be used to estimate parameters from observed data over the support of the data. Equation (5) is defined without invoking any statistics.

Many statisticians work in fields in which there is no formal methodology for describing “setting” (or fixing) inputs like (x_1, x_2, u) . The Kolmogorov (1956) axioms do not need to be invoked in defining (5), although the laws describing mathematical functions are required (e.g., Knopp, 2016). Abstract models like (1), (2), and (5) are outside of the Kolmogorov framework. They are nonstochastic.

This observation helps explain the rise of interest in Pearl’s (2009) do-calculus in some quarters of statistics and social science. His “do operator” works by setting inputs in a formal structure like (1) and creating special rules outside of formal statistics to interpret operations like “fixing” or “setting.” However, his calculus is not needed to accomplish this task (see Pinto and Heckman, 2022).

Elsewhere, Heckman and Pinto (2015) formally extend conventional probability theory to include “fixing” or “setting” in the Kolmogorov system by introducing a new class of random variables. They develop an intuitive framework that enables analysts to express causal operations using only standard probability and statistical theory without the elaborate non-statistical rules used in the do-calculus.

Representations like Equation (1) arise from thought experiments. Building on our cigarette example, raising a price (X_1) likely reduces cigarette demand. Holding everything else constant, the law of demand suggests that $\beta_1 < 0$, i.e., as $X_1 \uparrow, Y \downarrow$ holding X_2, U fixed. A huge body of economic theory and evidence supports this prediction (see, e.g., Mas-Colell et al., 1995). One can *set* $X_1 = x_1$ and vary it to create different counterfactual worlds.

It may occur that in actual data, Equation (4) coincides with, or is very “close” to, Equation (5), but that is a separate issue. In the cigarette example, as in many other studies of demand, the correspondence between theory and evidence is quite close. In the language of econometrics, this would correspond to Equation (5) (or (1)) being identified (see, e.g., Matzkin, 2007 for a discussion of identification). The basic distinction between a thought experiment generating different values of y by varying x_1 , keeping x_2 (and u) constant, and conditional expectations ((3) and (4)) is at the heart of what economists bring to the causality table. This point was first formalized by Haavelmo (1943) and more recently advanced in Heckman and Pinto (2015).

The mathematics is trivial. The conceptual distinction between (1) (or (5)) and (4) is not trivial and is the source of enormous confusion in statistics and in quarters of economics that follow and implement statistical frameworks that ignore thought experiments. (For a leading example, see Pratt and Schlaifer, 1984.)

Causal model (1) is defined independently of any estimator. In Equation (1), β_1 may or may not be an estimand of a regression of Y on X_1 and X_2 . If the estimand from Equation (4) based on observed data coincides with Equation (5), we are in company with Wigner (1960) in admiring the unreasonable effectiveness of theory (1) in predicting reality. But theory and inference are logically different. See Table 1.

Table 1: Two Distinct Tasks That Arise in the Analysis of Causal Models

Task	Description	Requirements
1	Defining the set of hypotheticals or counterfactuals	A well-specified theory
2	Identifying and estimating causal parameters from data	Mathematical analysis of point or set identification in hypothetical samples along with estimation and testing theory

Appreciating the distinction between functions (like (1) or (5)) and estimands (like (3) and (4)) is key to understanding the contribution of economics to the study of causality. A function g (like (1) or (5)) maps $x \rightarrow y$. Formally,

$$g : x \rightarrow y.$$

Functions are *defined* to be stable maps between X and Y over their entire support, regardless of the variation of its inputs. No special language or additional conditions such as those in “SUTVA” are required (see Pinto and Heckman, 2022). Moreover, systems of interdependent (non-recursive) equations are readily formulated. Mathematics and economic theory are replete with systems of equations (see, e.g., Mas-Colell et al., 1995):

$$F : Z \rightarrow W, \tag{6}$$

where Z is a collection of variables, W is a set of outcomes of any dimension, and F is a vector of functions that may include g from our previous example. State space equations are widely used in engineering, economics, and chemistry to list only a few applications. Simultaneity and interactions are readily characterized by such systems of functions. Many of the approaches current in “causal” analyses in statistics ignore the benefits of abstract theoretical models and thought experiments. They limit the range of causal questions that can be investigated by their practitioners.

This simple example readily generalizes to considering a definition of causality. Abstract theoretical equations need not be linear. Normality is not an essential feature of them. Equation (1) can be replaced by a system of simultaneous equations such as (6). “SUTVA” is irrelevant. Causal effects are by *definition* manipulations of input variables (by a thought

experiment) that generates hypothetical outputs. They are defined independent of any estimation technique or availability of data. Simultaneous causality is readily defined and poses no necessary paradox or contradiction to logic (see Matzkin, 2008 and Heckman and Pinto, 2015).

Human knowledge is produced by constructing counterfactuals and theories. Blind empiricism unguided by theoretical frameworks for interpreting facts leads nowhere. Many statisticians are uncomfortable with counterfactuals. Their discomfort arises in part from the need to specify abstract models to interpret and identify counterfactuals. Many statisticians are not trained in science or social science and adopt as their credo that they “should stick to the facts.” An extreme recent example of this discomfort is expressed by Dawid (2000), who denies the need for, or validity of, counterfactual analysis.

Economists since the time of Haavelmo (1943; 1944) have recognized the need for precise models to construct counterfactuals and to answer causal questions and more general policy evaluation questions, including making out-of-sample forecasts. The econometric framework is explicit about how counterfactuals are generated and how interventions are conducted (i.e., the rules of assigning “treatment”). The sources of unobservables, in both treatment assignment equations and outcome equations, and the relationship between the unobservables and observables are studied. Rather than leaving the rules governing selection of treatment implicit, the econometric approach explicitly models the relationship between the unobservables in outcome equations and the choice of outcome equations to identify causal models from data and to clarify the nature of identifying assumptions. The theory of structural modeling in econometrics is based on these principles. Modeling choice also enables analysts to distinguish objective counterfactuals (e.g., did the drug work?) and subjective counterfactuals (e.g., what is the pain and suffering experienced by users of the drug?). Answers to both questions are valuable in evaluating the impact of any treatment.

Ambiguity in model specification implies ambiguity in the definition of counterfactuals and hence in the notion of causality. The more complete the model of counterfactuals, the more precise the definition of causality. The ambiguity and controversy surrounding discussions of causal models are consequences of analysts wanting something for nothing: a definition of causality without a clearly articulated model of the phenomenon being described (i.e., a model of counterfactuals). They want to describe a phenomenon as being modeled “causally” without producing a clear hypothetical model of how the phenomena being described are generated or what mechanisms select the counterfactuals that are observed in hypothetical or real samples.

In the words of Holland (1986), they want to model the “effects of causes” without modeling the causes of effects. Science is all about constructing models of the causes of effects. Such models are essential in analyzing policy problems, as in our cigarette example. Economic (and scientific) problems dictate the choice of an abstract model and the definitions of causal parameters—not some estimand from one or another procedure.

In summary, causality is a property of a model of hypotheticals. A fully articulated model of the phenomena being studied precisely defines hypothetical or counterfactual

states. A definition of causality drops out of a fully articulated model as an automatic by-product. A model is a set of possible counterfactual worlds constructed under some rules. The rules may be the laws of physics, the consequences of utility maximization, or the rules governing social interactions, to take only three of many possible examples. A model is in the mind. As a consequence, causality is in the mind.

3.2 Policy Evaluation

Heckman: In empirical studies of causality, there are some standard problems: (1) selection bias; and (2) that for any causal question, at most one of many counterfactuals is known empirically (i.e., what is actually observed). These problems have been formalized at least since the time of Cox (1958), if not before. There are three different policy evaluation problems that are fruitfully distinguished but often conflated:

P1 *Evaluating the causal impacts of actual interventions, including their impact in terms of welfare.*

This is the problem of identifying a given treatment effect or a set of treatment effects in a given environment (Campbell and Stanley, 1963). This is the policy question usually addressed in the epidemiological and statistical literatures on causality. A drug trial for a particular patient population is the prototypical problem in that literature where investigators typically focus on objective outcomes, e.g., the effect of a drug treatment on health rather than the subjective wellbeing of the patient.

However interesting that may be, most policy evaluation is designed with an eye toward the future and toward decisions about new policies and application of old policies to new environments. I distinguish a second task of policy analysis:

P2 *Forecasting the impacts (constructing counterfactual states) of interventions implemented in one environment in other environments, including their impacts in terms of welfare (subjective wellbeing).*

Included in these interventions are policies described by generic characteristics (e.g., tax or benefit rates, etc.) that are applied to different groups of people or in different time periods from those studied in previous implementations of policies. This is the problem of *external validity*: transporting a structural parameter or a set of parameters estimated in one environment to another environment. The “environment” includes the characteristics of individuals and their social and economic setting. This is the forecasting problem long studied in economics.

Finally, the most ambitious problem is forecasting the effect of a new policy, never previously experienced:

P3 *Forecasting the impacts of interventions (constructing counterfactual states associated with interventions) never historically experienced to current or different environments, including their impacts in terms of welfare.*

This problem requires that one uses past experience to forecast the consequences of new policies. It is a fundamental problem in knowledge. It requires that one use abstract models to connect the future to ingredients from the past. This is the focus of structural estimation. Marschak (1953) and Domencich and McFadden (1975) are outstanding examples of how economists answer P3. I discuss these different scientific problems in greater detail elsewhere (Heckman, 2008).

P3 has been consistently ignored by statisticians because they cannot or will not deal with abstract models of counterfactuals like Equation (1). Holland (1986) boasts that statistics deals with the “effects of causes”—in the terminology of this paper, outcomes of comparisons of counterfactuals—rather than the causes of the effects (models like (1) that determine the counterfactuals studied).

A prime example of the limits to this approach is the claim that one can never learn the causal effect of race on outcomes because race cannot be randomly assigned. An estimand (the outcome of an RCT) is used to define causal parameters. In contrast, the abstract theory-based approach investigates which factors (X in Equation (1)) are important for producing outcomes. Having determined this, one can conceptually equalize the factors across persons of different race groups. That may not be an easy empirical task, but there is a large scholarly literature examining the factors leading to various outcomes (e.g., productivity in a job). Whatever the quality of the empirical work, the thought experiment remains valid even if a reliable estimate of the causal effect is difficult to obtain.

4. On Dominance of Specific Models

Interviewer: What are the advantages and disadvantages of the recent interest in randomized trials within economics?

Heckman: I have written on this elsewhere (Heckman, 1992, 2020). Experiments are useful for certain problems but are far from being a panacea and are frequently corrupted in practice, producing misleading inference. A good example is the analysis of a recent Head Start experiment (Kline and Walters, 2016). Families randomized out of the program at one location went to other Head Start locations or even better programs. The inconclusive treatment effect reported from the simple mean difference RCT “causal effect” was a result of uncontrolled “contamination bias” (see Heckman et al., 2000). The “control group” participated in programs as good or better than the program being evaluated, so the “gold standard” estimated treatment effects were negative, when in fact estimates that correct for contamination bias show strong positive effects. Often, clean RCTs do not exist, and we always have to judge the causal claims of each study on a case by case basis.

Interviewer: What are your thoughts on the current dominance of propensity scores, diffs-in-diffs, regression discontinuity, and synthetic controls?

Heckman: There are many estimation methods out there, each justified by different assumptions. There is no universal best estimator. The conditional nature of (causal) knowledge alarms some analysts who seek absolute knowledge of “causal effects.” One needs to understand the science behind a phenomena before one can select appropriate estimators. However, this is not often done in many statistical studies. An “effect” is estimated without a clear interpretation of what precisely it captures and how it is generated.

Heckman and Robb (1985, 1986) compare the identifying assumptions underlying a large array of cross-section, repeated cross-section, and panel data estimators, including “diffs-in-diffs.” Many users of the methods you describe do not have a clear problem in mind and instead chant “ATE, TOT,” or whatever sells that day. It’s like a fashion show in Milan with different inferential techniques coming down the runway at different times. Most users have no clear question in hand, just an “estimand” that will get them published. The estimands are well defined; the problems they address often much less so. Whether the estimands are relevant for solving the stated problems is an entirely different issue that is rarely addressed. Many users do not have a question in hand. They just want an estimate of a not so clearly defined something.

A good example of this phenomenon is the literature in economics on estimating the “return” to schooling. It focuses on estimating the coefficient of schooling in an equation relating log earnings (y) to factors determining earnings. A close reading of the literature reveals its focus on one element of β in Equation (4) free of any bias arising from correlations between U and the associated X .

The recent “credibility” revolution in economics focuses on “credible” estimators for β (see, e.g., Angrist and Pischke, 2010). It emphasizes easily computed, easily replicated statistical methods. These are surely desirable features of any inferential procedure. Missing, however, is any clear statement of why β s so obtained answer relevant economic questions. Heckman et al. (2006) demonstrate that the “return” featured in the “credibility revolution” answers interesting economic questions only under very special circumstances and that a deeper analysis is required to estimate the economically interesting return to schooling. The credibility revolutionaries create clean estimators for generally uninterpretable estimands. They conflate tasks 1 and 2 of Table 1.

Interviewer: How do you see the trade-offs between more parametric approaches (SEM) versus less parametric approaches such as DAGs?

Heckman: I don’t agree with the premise of the question. Structural models (SEM) estimating abstract theoretical models can be and often are nonparametric or at least semi-parametric. Heckman and Pinto (2015) and Pinto and Heckman (2022) show that Pearl’s DAGs and the rules of do-calculus do not accommodate instrumental variable methods or selection bias models. They account for simultaneity only by “shutting down equations” without regard for the properties of systems so generated, while structural econometric

models readily accommodate simultaneity, as Matzkin establishes in her many papers (e.g., Matzkin, 2007, 2004, 2008).

5. On Treatment Effects

Interviewer: How would you describe the differences between average causal effects and your “marginal treatment effect” (Heckman and Vytlačil, 1999)?

Heckman: The choice of a parameter should be dictated by the problem one seeks to address, not by some arbitrary convention. The marginal treatment effect (MTE) is useful for determining benefits or costs to people at margins of choice (see, e.g., Eisenhauer et al., 2015). MTE is a building block from which all other conventional treatment effects can be constructed under appropriate support conditions. It does not replace treatment effects—it is a device for unifying them. It also links choice equations to outcome equations. Heckman and Vytlačil (2005) show that a variety of policy relevant causal parameters outside the basic toolkit can be formed from marginal treatment effects.

This approach has its origins in the insights of calculus: the relationship between derivatives and integrals. The marginal treatment effect can be interpreted in some settings as the marginal willingness to pay for a good for a subset of people who are indifferent between buying it or not. The common treatment effects aggregate over all people who buy the good with their different intensity of preferences.

Interviewer: What are the relative advantages and disadvantages of local treatment effects?

Heckman: Local treatment effects are the building blocks from which all treatment effects can be constructed. They can be used to establish the relationships among the various treatment effects using a common conceptual tool. It unifies knowledge. See Heckman and Vytlačil (2007a,b). Local treatment effects characterize choices at margins which are central to economic analysis.

6. On Mediation

Interviewer: Do you find mediation/path-specific inference a useful or promising area of causal research and application?

Heckman: Yes, it is. But note that “mediation” analysis has long been conducted in econometrics and is very useful for many questions. Adelman and Adelman (1959) simulated the dynamic multiperiod Klein-Goldberger (1955) model of the US economy using what is now called “mediation.” They based their analysis on “dynamic impact multipliers,” which chart the impact of policy changes in one period on the output, consumption, and investment of future periods. Long ago, Sewall Wright (Wright, 1934) pioneered the mediation approach, calling it path analysis, and there is a long tradition of using it in

social science. Mediation analysis enables analysts to understand causes of effects rather than stopping at just reporting effects. It is essential for policy analysis that aims to improve outcomes. It tells analysts which levers to pull to make effective policies, and it is valuable for the intellectually curious who seek to know why treatment outcomes occur.

7. On Bounds and Sensitivity Analyses

Interviewer: For informing policy decisions, do you find bounding or sensitivity analyses useful?

Heckman: Yes. But there is usually much more information available than the numerical information on the supports of random variables that is featured in recent work on bounds. This additional information is useful in doing sensitivity analysis (e.g., newspaper accounts, common sense, etc.), but it is rarely used. It often requires too much subject matter knowledge for most “causal analysts.” See Heckman and Singer (2017).

For example in estimating the impact of a new law on social outcomes, it is helpful to get point estimates or bounds on point estimates for an outcome. However, newspaper accounts, related indicators, and witness reports are also informative. For a brilliant example of this approach see Katz and Singer (2007).

8. On Limiting Assumptions

Interviewer: How do you approach the common assumptions in causal inference that appear limiting given real data?

Heckman: I discussed this both thirty years ago and in a recent paper (Heckman and Robb, 1985, 1986; Heckman and Pinto, 2015; Pinto and Heckman, 2022). Many standard methods for analyzing social interactions and general equilibrium effects, simultaneity, and feedback violate the “causal framework” protocols of statistics but answer interesting policy questions outside that straightjacket. The SUTVA assumption is a good example.

One aspect of “SUTVA” is “structural invariance” or “autonomy” developed by Frisch (1938) and Hurwicz (1962). It characterizes functional relationships like Equations (1) or (5). It is a core idea applied more than eighty years ago in econometrics. The other aspect of “SUTVA” is the assumption of no interaction among agents. That rules out a lot of the study of interesting social phenomena. Simultaneity and interaction have been studied for decades. For instance, much research in economics focuses on general equilibrium environments, ruled out by SUTVA. A sizeable literature in economics is devoted to the study of peer-effects (see Blume et al., 2011).

The goal of this literature is precisely the evaluation of the “interference” (really interaction) among randomization units, which is ruled out by SUTVA. Interactions are treated as a problem in statistics but are, in fact, a source of information in economics and social science more broadly. SUTVA is an artifact of the obsession of many statis-

ticians and their followers on single-agent RCTs. I believe statisticians' ignorance of the econometrics literature with respect to simultaneity and social interactions has harmed the advance of knowledge. Information about the mechanisms producing "interference" is useful for analyzing the propagation of disease, economic shocks, network externalities, etc.

Interviewer: Has there been progress on the topic of nonrecursive causal effects?

Heckman: Econometrics offers a clear discussion of non-recursive causal effects. Cowles' Monograph 10 (Koopmans et al., 1950) and Monograph 13 (Hood and Koopmans, 1953) present the basic framework. Rosa Matzkin (2008; 2004; 2007) has substantial work studying systems of equations which are non-recursive models. Fisher (1966) presents a general analysis of identification in both linear and nonlinear simultaneous equations systems. I already mentioned the work of Blume et al. on social interactions. The common analysis of market prices and quantities requires nonrecursive models. Time series econometrics and financial economics abound with nonrecursive models.

9. On Machine Learning

Interviewer: Do you think machine learning will be useful in causality?

Heckman: Machine learning is a useful tool but it's only a computational device for prediction, estimation, and establishing empirical relationships. It offers no insight about causality per se, other than what is learned from careful descriptions of phenomena. It is useful to have good descriptions of phenomena as material on which to build interpretative causal models.

References

- Irma Adelman and Frank L. Adelman. The dynamic properties of the klein-goldberger model. *Econometrica*, 27(4):596–625, 1959. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1909353>.
- Joshua D. Angrist and Jörn-Steffan Pischke. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30, Spring 2010.
- Lawrence E. Blume, William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides. Identification of social interactions. In Jess Benhabib, Alberto Bisin Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1B, chapter 18, pages 1–30. North-Holland, Amsterdam, 2011.
- Donald T. Campbell and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin Company, Boston, MA, 1963.

- Scott Carrell, David Figlio, and Lester Lusher. Clubs and networks in economics reviewing. January 2022. Institute for Policy Research, Northwestern University. WP-22-05.
- David R. Cox. *Planning of Experiments*. Wiley, New York, 1958.
- A.P. Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, June 2000.
- Tom Domencich and Daniel L. McFadden. *Urban Travel Demand: A Behavioral Analysis*. North Holland: Amsterdam, 1975.
- Philipp Eisenhauer, James J. Heckman, and Edward J. Vytlačil. Generalized Roy model and cost-benefit analysis of social programs. *Journal of Political Economy*, 123(2):413–433, 2015.
- Richard Feynman. *The Character of Physical Law*. MIT Press, Cambridge, MA, 1965.
- Richard Feynman. The pleasure of finding things out. BBC Horizon Interview, 1981.
- Franklin M. Fisher. *The Identification Problem in Econometrics*. McGraw-Hill, New York, 1966. doi: 10.2307/1912803.
- Ragnar Frisch. In Olav Bjerkholt and Duo Qin, editors, *A Dynamic Approach to Economic Theory: The Yale Lectures of Ragnar Frisch, 1930*. Routledge, New York, New York, 1930, published 2010.
- Ragnar Frisch. Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics. Paper given at League of Nations. Reprinted in D.F. Hendry and M.S. Morgan (1995), *The Foundations of Econometric Analysis*, Cambridge University Press, 1938.
- Stephen M. Goldfeld and Richard E. Quandt. *Studies in Nonlinear Estimation*. Ballinger Publishing Company, Cambridge, Massachusetts, 1976.
- Trygve Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1–12, January 1943.
- Trygve Haavelmo. The probability approach in econometrics. *Econometrica*, 12 (Supplement):iii–vi and 1–115, 1944.
- James J. Heckman. Randomization and social policy evaluation. In Charles F. Manski and Irwin Garfinkel, editors, *Evaluating Welfare and Training Programs*, chapter 5, pages 201–230. Harvard University Press, Cambridge, MA, 1992.
- James J. Heckman. Econometric causality. *International Statistical Review*, 76(1):1–27, April 2008.

- James J. Heckman. Randomization and social policy evaluation revisited. Discussion Paper 12882, IZA Institute of Labor Economics, January 2020.
- James J. Heckman and Rodrigo Pinto. Causal analysis after Haavelmo. *Econometric Theory*, 31(1):115–151, 2015.
- James J. Heckman and Richard Robb. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1–2):239–267, October–November 1985.
- James J. Heckman and Richard Robb. Alternative identifying assumptions in econometric models of selection bias. In G. Rhodes, editor, *Advances in Econometrics*, volume 5, pages 243–287. JAI Press, Greenwich, CT, 1986.
- James J. Heckman and Burton Singer. Abducting economics. *American Economic Review: Papers and Proceedings*, 107(5):298–302, 2017.
- James J. Heckman and Edward J. Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, 96(8):4730–4734, April 1999.
- James J. Heckman and Edward J. Vytlacil. Structural equations, treatment effects and econometric policy evaluation. *Econometrica*, 73(3):669–738, May 2005.
- James J. Heckman and Edward J. Vytlacil. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume 6B, chapter 70, pages 4779–4874. Elsevier B. V., Amsterdam, 2007a. doi: 10.1016/S1573-4412(07)06070-9.
- James J. Heckman and Edward J. Vytlacil. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs, and to forecast their effects in new environments. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume 6B, chapter 71, pages 4875–5143. Elsevier B. V., Amsterdam, 2007b. doi: 10.1016/S1573-4412(07)06071-0.
- James J. Heckman, Neil Hohmann, Jeffrey Smith, and Michael Khoo. Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics*, 115(2):651–694, May 2000.
- James J. Heckman, Lance J. Lochner, and Petra E. Todd. Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In Eric A. Hanushek and Frank Welch, editors, *Handbook of the Economics of Education*, volume 1, chapter 7, pages 307–458. Elsevier, Amsterdam, 2006.

- Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. ISSN 01621459. URL <http://www.jstor.org/stable/2289064>.
- William C. Hood and Tjalling C. Koopmans. *Studies in Econometric Method*. Wiley, New York, 1953.
- Leonid Hurwicz. On the structural form of interdependent systems. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science*, pages 232–239. Stanford University Press, 1962.
- J. R. Josephson and S. G. Josephson, editors. *Abductive inference: Computation, Philosophy, Technology*. Cambridge University Press, 1996.
- Rebecca Katz and Burton Singer. Can an attribution assessment be made for yellow rain. *Politics and the Life Sciences*, 26(1):24–42, 2007.
- Lawrence Robert Klein and Arthur Stanley Goldberger. *An Econometric Model of the United States, 1929–1952*. North-Holland Publishing Company, Amsterdam, 1955.
- Patrick Kline and Christopher Walters. Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics*, 131(4):1795–1848, 2016.
- Konrad Knopp. *Elements of the Theory of Functions*. Dover Publications, Inc., 2016. English translation of *Elemente der Funktionentheorie*.
- Andreĭ Nikolaevich Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, 1956. Translation Edited by Nathan Morrison. With an Added Bibliography by AT Bharu-cha-reid.
- T. C. Koopmans, H. Rubin, and R. B. Leipnik. Measuring the equation systems of dynamic economics. In T. C. Koopmans, editor, *Statistical Inference in Dynamic Economic Models*, number 10 in Cowles Commission Monograph, chapter 2, pages 53–237. John Wiley & Sons, New York, 1950.
- David Lewis. *Counterfactuals*. Basil Blackwell Ltd, Oxford, UK, 1973.
- Jacob Marschak. Economic measurements for policy and prediction. In William C. Hood and Tjalling C. Koopmans, editors, *Studies in Econometric Method*, pages 1–26. Yale University Press, New Haven, CT, 1953.
- Alfred Marshall. *Principles of Economics*. Macmillan and Company, New York, 1890.
- Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, New York, 1995.

- Rosa L. Matzkin. Unobserved instruments. Unpublished manuscript, Northwestern University, Department of Economics, 2004.
- Rosa L. Matzkin. Nonparametric identification. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume 6B. Elsevier, Amsterdam, 2007.
- Rosa L. Matzkin. Identification in nonparametric simultaneous equations models. *Econometrica*, 76(5):945–978, 2008.
- Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2nd edition, 2009.
- Rodrigo Pinto and James J. Heckman. The Econometric Model for Causal Policy Analysis. 2022. Under review, *Annual Review of Economics*.
- John W. Pratt and Robert Schlaifer. On the nature and discovery of structure. *Journal of the American Statistical Association*, 79(385):9–33, March 1984. doi: 10.1080/01621459.1984.10477054.
- Richard E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):873–880, December 1958.
- Richard E. Quandt and W. Baumol. The demand for abstract transport modes: Theory and measurement,. *Journal of Regional Science*, 6:13–26, 1966.
- Eugene P. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13(1):1–14, 1960.
- Sewall Wright. The method of path coefficients. *Annals of Mathematical Statistics*, 5(3): 161–215, 1934.