Written for Econometric Theory, special issue on Haavelmo Centennial.

TECHNICAL REPORT
R-391
February 2012

Trygve Haavelmo and the Emergence of Causal Calculus

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
(310) 825-3243
judea@cs.ucla.edu

Abstract

Haavelmo was the first to recognize the capacity of economic models to guide policies. This paper describes some of the barriers that Haavelmo's ideas have had (and still have) to overcome, and lays out a logical framework for capturing the relationships between theory, data and policy questions. The mathematical tools that emerge from this framework now enable investigators to answer complex policy and counterfactual questions using embarrassingly simple routines, some by mere inspection of the model's structure. Several such problems are illustrated by examples, including misspecification tests, identification, mediation and introspection. Finally, we observe that modern economists are largely unaware of the benefits that Haavelmo's ideas bestow upon them and, as a result, econometric research has not fully utilized modern advances in causal analysis.

1 Introduction

To students of causation, Haavelmo's paper "The statistical implications of a system of simultaneous equations," (Haavelmo, 1943) marks a pivotal turning point, not in the statistical implications of econometric models but in their causal counterparts. Causal implications, which prior to Haavelmo's paper were cast to the mercy of speculation and intuitive judgment have thus begun their quest for full membership in the good company of scientific discourse.

Haavelmo introduced three revolutionary insights in 1943.

First, when an economist sits down to write a structural equation he/she envisions, not statistical relationships but a set of hypothetical experiments, qualitative aspects of which are then encoded in the system of equations. Second, an economic model thus constructed is capable of answering policy intervention questions, with no further assistance from the modeller. Finally, to demonstrate the feature above, a mathematical method was proposed that takes an arbitrary model, combines it with data, and derives answers to policy questions in a way that is consistent with both the data and the model (unless incompatibility is discovered).

1.1 What is an economic model?

This first idea, that an economic model depicts a series of hypothetical experiments was expressed more forcefully in Haavelmo's 1944 paper (The probabilistic Approach to Econometrics) where he states:

> "What makes a piece of mathematical economics not only mathematics but also economics is, I believe, this: When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved, we have in mind some actual experiment, or some *design* of an experiment, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes." (1944, p. 5)

But the seed of this idea was planted already in 1943, when Haavelmo tried to explain what a modeller must have in mind in putting together two or more simultaneous equations, say

$$y = ax + \epsilon_1 \tag{1}$$
$$x = by + \epsilon_2 \tag{2}$$

Haavelmo first showed that, contrary to naive expectation, the term ax is not equal to $E(Y|x)$ and, so, asked Haavelmo, what information did the modeller intend a to carry in Eq. (1), and what information would a provide if we are able to estimate its value.

In posing this question, Haavelmo addressed the dilemma of incremental model construction. Given that the statistical content of a can only be discerned (if at all) by considering the entire system of equations, how can a modeller write down one equation at a time, without knowing what the meaning of the coefficients is in each equation. "What *is* then the significance of the theoretical equations..." Haavelmo asked (1943, p. 11) and answered it immediately: "To see that, let us consider, not a problem of passive predictions, but a problem of government planning."

In modern terms, Haavelmo rejected the then-ruling paradigm that parameters are conveyors of statistical information and prepared the ground for the causal definition of a

$$a = \frac{\partial}{\partial x} E(Y|do(x)) \tag{3}$$

which refers to a controlled experiment in which an agent (e.g., Government) is controlling x and observing y.[1] In such experiment, the average slope of Y on X (i.e.,

[1]More precisely, the definition of a is $a = \frac{\partial}{\partial x}[Y_x(u)]$ when $Y_x(u)$ is the counterfactual "Y if x" for unit u (Pearl, 2009a, Ch. 7). However, counterfactuals were rather late in penetrating structural modeling (Simon and Rescher, 1966; Balke and Pearl, 1995; Heckman, 2000).

a) bears no relationship to the regression slope (i.e., $\frac{\partial}{\partial x}E(Y|X=x)$) in the population prior to intervention. Whereas the statistical content of a (if identified) may come from many equations, its causal content is local – to the great relief of most economists who think causally, not statistically.

This simple truth, which today is taken (almost) for granted, took a long time to take roots. To illustrate, the fierce debate between prominent statisticians and economists that flared up in 1992, fifty years after Haavelmo's paper, revolved precisely around this issue of interpreting the meaning of a. The economist in the debate, Arthur Goldberger (1992), claimed that ax in Eq. (1) may be interpreted as the expected value of Y "if x were fixed," so that the a parameter "has natural meaning for the economist." The statistician, Nanny Wermuth (1992), argued that, since $ax \neq E(Y|X=x)$, "the parameters in (1) cannot have the meaning Arthur Goldberger claims they have." Summarizing their arguments, Wermuth concluded that structural coefficients have dubious meaning, and Goldberger retorted that statistics has dubious substance. Remarkably, each side quoted Haavelmo to prove the other wrong, and both sides were in fact correct; structural coefficients have no meaning in terms of properties of joint distribution functions, the only meaning that statisticians were willing to accept in the 1990's. And statistics has no substance, if it excludes from its province all aspects of the data generating mechanism that do not show up in the joint distribution, for example, a, or $E(Y|do(x))$.

The confusion did not end in 1992. The idea that an economic model must contain extra-statistical information, that is, information that can not be derived from joint densities, and that the gap between the two can never be bridged, seems to be very slow in penetrating the mind set of mainstream economists. Hendry, for example, wrote: "The joint density is the basis: SEMs are merely an interpretation of that" (Hendry, 1998, personal communication). Spanos (2011), expressing similar sentiments, hopes to "bridge the gap between theory and data" through the teachings of Fisher, Neyman and Pearson. Youth of our generation continue to be enticed by such promises, and have not internalized the hard fact that statistics, however revered, cannot provide the causal information that economic models must encode to be of use to policy making.

A highly popular econometrics textbook writes: "A state implements tough new penalties on drunk drivers: What is the effect on highway fatalities?... [This effect] is an unknown characteristic of the population joint distribution of X and Y" (Stock and Watson, 2007, Chapter 4, p. 111). The fact that "effects" are *not* characteristics of population joint distributions, so compellingly demonstrated by Haavelmo (Eq. (3)), would probably come as a surprise to modern authors of econometric texts. To witness, almost seventy years after Haavelmo defined a model as a set of hypothetical experiments, the Wikipedia's definition of "Econometric Models" reads (February 18, 2012): "An econometric model specifies the statistical relationship that is believed to hold between the various economic quantities pertaining to particular economic phenomenal under study."[2]

[2] I was tempted to correct this sentence in the Wikipedia, but decided to keep it as a witness to prevailing views, and as an incentive for the editors of *Econometric Theory* to bring the issue to

1.2 An oracle for policies or an aid to forecasters?

Haavelmo's second and third insights also took time to be fully appreciated. Even today, the idea that an economic model should serve as an oracle for interventional questions tends to evoke immediate doubts and resistance: "How can one predict outcomes of experiments that where never performed?" Ask the skeptics. And if the modeller's assumptions possess such clairvoyant powers, why not answer policy questions directly, rather than engage in modeling and analysis? How can a set of ordinary equations encapsulate the information needed for predicting the vast variety of interventions that a policy maker may wish to evaluate? How is this vast amount of information encoded, and what means do we have to extract it from its encoding?

To a large extent, this typical resistance stems from the absence of distinct mathematical notation for marking the causal assumptions that enter into an economic model; the syntax of the equations appears deceptively algebraic, similar to that of regression models, hence void of causal content. Some economists, lured by this surface similarity, were led to conclude: "We must first emphasize that, disturbance terms being unobservable, the usual zero covariances "assumptions" generally reduce to mere definitions and have no necessary causality and exogeneity implications." (Richard, 1980, p. 3)

The absence of distinct notation for causal assumptions further compelled economists to assume that, to qualify for policy analysis, an economic model must be hardened by some extra ingredients; the equations themselves were deemed too simplistic or "fragile" to convey interventional information.

The literature on "exogeneity" (e.g. Richard, 1980; Engle, Hendry, and Richard, 1983; Hendry, 1995) for example, sought such extra power in the notion of "parameter invariance." Cartwright (2007) sought it in the notion of "modularity" (see Pearl 2010a). And, in general, one would be hard pressed to find an economic textbook that encourages readers to answer policy questions from the equations themselves, without resorting to meta-mathematical disclaimers or preconditions that reside outside the model.

This lack of confidence in the ability of economic models to guide policies has threatened the utility of the entire enterprise of economic modeling for, taken to extreme, it commits economic analysis to statistical extrapolation of time series data. I doubt Haavelmo would agree to such restriction. Indeed, what is the point of parameter estimation if at the end of such exercise one must appeal to judgment to decide which parameter is invariant and which is not, or, lacking such judgment, to physically trying out the policy and observing its effect on various parameters.

A more reasonable alternative, one that I have advocated in (Pearl, 2000, 2009a) and that is gaining support among economists (e.g., Heckman, 2000, 2003, 2008; Leamer, 2010; Keane, 2010) is to treat an economic model as an oracle for *all* causally related queries, including questions of prospective and introspective counterfactuals and, simultaneously, insist on encoding the assumptions needed for answering such queries within the model itself, not external to it. In other words, the model can be totally wrong and still capable of issuing logically sound and practically useful

public discussion and collective revision.

conclusions, as long as each conclusion is understood to be contingent on the validity of a distinct set of assumptions, and as long as we articulate each assumption explicitly and transparently.

"And what if an intervention changes the very equation that purports to predict its effect?" ask the critics (e.g., Lucas Jr. (1976)). The answer is that, since the model provides the facility for encoding side-effects associated with particular implementations of the intervention evaluated, failure to encode them in the model constitutes a case of *query misspecification*, no less damaging than model misspecification. The burden of properly specifying queries rests with the query provider not with the oracle.

1.3 The algorithmization of interventions

Modern days interest in causal models and their tentative conclusions, owes its renaissance to Haavelmo's third insight – a concrete procedure for eliciting answers to policy questions from the model equations. This he devised at the end of his 1943 paper:

> "Assume that the Government decides, through public spending, taxation, etc., to keep income, r_t, at a given level, and that consumption u_i and private investment v_i continue to be given by (2.5) and (2.6), the only change in the system being that, instead of (2.7), we now have
>
> $$r_i = u_i + v_i + g_i \tag{2.7$'$}$$
>
> where g_i is Government expenditure, so adjusted as to keep r constant, whatever be u and v,..." (1943, p. 12)

This idea of simulating an intervention on a variable by modifying the equation that determines that variable while keeping all other equations in tact is the basis of all modern approaches to causal analysis. Haavelmo's proposal of adding an adjustable term to the equation was later transformed by Strotz and Wold (1960) into the operation of "wiping out" the equation altogether, and was further translated into graphical models as "wiping out" incoming arrows (Spirtes, Glymour, and Scheines, 1993; Pearl, 1993). This operation has subsequently led to *do*-calculus (Pearl, 1994, 2009a) and to the structural theory of counterfactuals (Balke and Pearl, 1995; Pearl, 2009a, Ch. 7), which unifies structural equation modeling with the potential outcome paradigm of Neyman (1923) and Rubin (1974) and the possible-world semantics of Lewis (1973).

Key to this unifying framework has been a symbolic procedure for reading counterfactual information in a system of economic equations, as articulated in the following Definition:

Definition 1 (unit-level counterfactuals**)** (Pearl, 2000, p. 98)
Let M be a fully specified structural model and X and Y two arbitrary sets of variables in M. Let M_x be a modified version of M, with the equation(s) of X replaced by

$X = x$. Denote the solution for Y in the modified model by the symbol $Y_{M_x}(u)$. The counterfactual $Y_x(u)$ (Read: "The value of Y in unit u, had X been x") is define by

$$Y_x(u) \triangleq Y_{M_x}(u). \tag{4}$$

In words: The counterfactual $Y_x(u)$ in model M is defined as the solution for Y in the modified submodel M_x, with the exogenous variables held at $U = u$.

We see that every structural equation, say $y = ax + \epsilon_1$ (Eq. (1)), carries counterfactual information, $Y_{xz}(u) = ax + \epsilon_1$, where Z is any set of variables that do not appear on the right hand side of the equation. Naturally, when the exogenous variables U in a model are random variables, the counterfactual Y_x will be a random variable as well, the distribution of which is dictated by the distribution $P(u)$ and the structure of the model M_x. This interpretation permits us to define joint distributions of counterfactual variables and to detect conditional independencies of counterfactuals directly from the structure of the model (Pearl, 2009a, Ch. 7).

Equation (4) constitutes the bridge between the structural interpretation of counterfactuals and the potential outcome framework advanced by Neyman (1923) and Rubin (1974), which takes the controlled randomized experiment as its guiding paradigm. The essential difference between the two frameworks is that counterfactuals, as well as critical assumptions such as "ignorability," "sequential ignorability," or "instrumentality" can actually be *derived* from the economic model, rather than be at the mercy of guesswork and convenience. Another difference is that the antecedent x in the structural interpretation of $Y_x(u)$ need not be a manipulable treatment but may consist of any exogenous or endogenous variable (e.g., sex, genetic traits, race, earning) that affects Y as part of a social or biological process (Heckman, 2008). This interpretation has extended Haavelmo's theory of interventions from linear to nonparametric analysis and permitted questions of identification, estimation, and generalization to be handled with mathematical precision and algorithmic simplicity (see Section 3).

Haavelmo did not deem his intervention theory to be revolutionary, but natural. In his words:

> "That is, to predict consumption ... under the Government policy,... we may use the 'theoretical' equations obtained by omitting the error terms..."
> "this is only natural, because now the Government is, in fact, performing 'experiments' of the type we had in mind when constructing each of the two equations." (1943, p. 12)

I do consider it revolutionary in that it encodes the effect of interventions not in terms of the model's parameters but in the form of a *procedure* (or "surgery") that modifies the structure of the model. It thus liberates economic analysis from its dependence on parametric representations and permits a totally nonparametric calculus of causes and counterfactuals that makes the connection between assumptions and conclusions explicit and transparent.

In the next section I will give a brief summary of nonparametric structural models and the wealth of mathematical tools that they now offer to economists and other policy-minded data analysts.

2 The Logic of Structural Causal Models (SCM)

This section describes a coherent theory of causal inference that I propose to call Structural Causal Model (SCM). It takes seriously the original insights of Haavelmo and the subsequent philosophy of the Cowles Commission program and, enriched with a few ideas from logic and graph theory, provides a unifying framework for all known approaches to causation.

A simple way to view SCM is to imagine a logical machine, or an inference engine, that takes three inputs and produces three outputs. The inputs are:

I-1. A set A of qualitative causal *assumptions* that the investigator is prepared to defend on scientific grounds, and a model M_A that encodes these assumptions. (Traditionally, M_A takes the form of a set of structural equations with undetermined parameters. A typical assumption is that certain omitted factors, represented by error terms, are uncorrelated, or that no direct effect exists between a pair of variables (i.e., an "exclusion restriction").

I-2. A set Q of *queries* concerning causal and counterfactual relationships among variables of interest. Traditionally, Q concerned the magnitudes of structural parameters but, in general, Q may address causal relations more directly, e.g.,

Q_1 : What is the effect of treatment X on outcome Y?

Q_2 : Is this employer guilty of gender discrimination?

Formally, each query $Q_i \in Q$ should be computable from a fully specified theoretical model M in which all functional relationships are given, together with the joint distribution of all omitted factors. Non-computable queries are inadmissible.

I-3. A set D of experimental or non-experimental *data*, presumably generated by a process consistent with A.

The outputs are

O-1. A set A^* of statements which are the logical implications of A, prior to obtaining any data. For example, that X has no effect on Y if we hold Z constant, or that Z is an instrument relative to a pair $\{X, Y\}$.

O-2. A set C of data-dependent *claims* (or conclusions) concerning the magnitudes or likelihoods of the target queries in Q, each conditional of A. C may contain, in the simple case, the estimated mean and variance of a given structural parameter, or the expected effect of a given intervention or, to illustrate a counterfactual query, the probability that a student trained in a given program who now earns 50K per year would not have reached a salary level greater than 30K had he/she not been trained (Pearl, 2009a, Ch. 9).

Auxiliary to C, SCM also generates an estimand $Q_i(P)$ for each query in Q, or a determination that Q_i is not identifiable from P, the joint density of observed variables.

O-3. A list T of testable statistical implications of A, and the degree $g(T_i), T_i \in T$, to which the data agrees with each of those implications. A typical implication would be the vanishing of a specific regression coefficient, or the invariance of such coefficient to the addition or removal of a given regressor; such constraints can be read from the model M_A and confirmed quantitatively by the data.

The structure of this inferential exercise is shown schematically in Fig. 1.

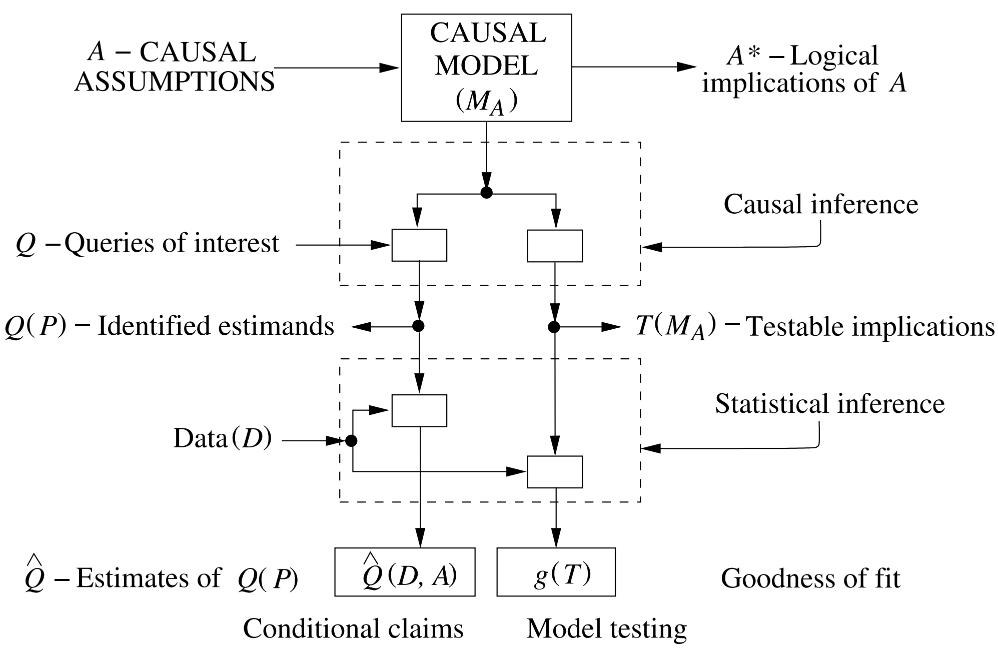

Figure 1: Should go SCM methodology depicted as the an inference engine converting assumptions (A), queries (Q), and data (D) into logical implications (A^*) Conditional claims (C) and data-fitness indices ($g(T)$).

Several observations are worth noting before illustrating these inferences by examples. First, SCM is not a traditional statistical methodology, typified by hypothesis testing or estimation, because neither claims nor assumptions are expressed in terms of probability functions of realizable variables (Pearl, 2009a).

Second, all claims produced by SCM are conditional on the validity of A, and should be reported in conditional format: "If A then C_i" for any claim $C_i \in C$. Such claims assert that anyone willing to accept A, must also accept C_i out of logical necessity. Moreover, no other method can do better, that is, if SCM analysis finds that a subset A' of assumptions is necessary for inferring a claim C_i, no other methodology can infer C_i with a weaker set of assumptions. This follows from casting the relationship between A and C in a formal mathematical system, coupled with the completeness theorems of Halpern (1998) and Shpitser and Pearl (2008).[3]

Thirdly, passing a goodness-of-fit test is not a prerequisite for the validity of the conditional claim "If A then C_i," nor for the validity of C_i. While it is important to know if any assumptions in A are inconsistent with the data, M_A may not have any testable implications whatsoever. In such a case, the assertion "If A then C_i" may still be extremely informative in a decision making context, since each C_i conveys quantitative information extracted from the data compared with the qualitative assumptions A with which the study commences. Moreover, even if A turns out inconsistent with D, the inconsistencies may be entirely due to portions of the model which have nothing to do with the derivation of C_i (Marschak, 1953). It is therefore important to identify which statistical implication of A is responsible for the inconsistency; global tests for goodness-of-fit hide this information (Pearl 2004; 2009a, pp. 144-45).

Finally, and this has been realized by many researchers in the 1980's, there is nothing in SCM's methodology to protect C from the inevitability of contradictory equivalent models, namely, models that satisfy all the testable implications of M_A and still advertise claims that contradict C. Modern developments in graphical modeling have devised visual and algorithmic tools for detecting, displaying, and enumerating these equivalent models (Verma and Pearl, 1990; Kyono, 2010; Ali, Richardson, and Spirtes, 2009). Researchers should keep in mind therefore that only a tiny portion of the assumptions behind each SCM lends itself to scrutiny by the data; the bulk of it must remain untestable, at the mercy of scientific judgment.

3 Causal Calculus, Tools, and Frills

By "causal calculus" I mean mathematical machinery for performing the computational tasks described in the inference engine of Fig. 1.

These include:

1. Tools of reading and explicating the causal assumptions embodied in structural models as well as the set of assumptions that support each individual causal claim.

2. Methods of identifying the testable implications (if any) of the assumptions in (1), and ways of testing, not the model in its entirety, but the testable implications of the assumptions behind each causal claim.

3. Methods of deciding, prior to taking any data, what measurements ought to be taken, whether one set of measurements is as good as to another, and which ad-

[3]This is important to emphasize in view of often heard critics that, in SCM, one must start with a model in which all causal relations are presumed known, at least qualitatively. Other methods must rest on the same knowledge, though some tend to hide the assumptions under catch-all terms such as "ignorability," "nonconfoundedness," "exchangeability," "quasi-experiment," "exogeneity," and the like.

justments need to be made so as to render our estimates of the target quantities unbiased.

4. Methods for devising critical statistical tests by which two competing theories can be distinguished.

5. Methods of deciding mathematically if the causal relationships of interest are estimable from non-experimental data and, if not, what additional assumptions, measurements or experiments would render them estimable.

6. Methods of recognizing and generating equivalent models.

7. Methods of locating instrumental variables for any relationship in a model, or turning variables into instruments when none exists.

8. Methods of evaluating "causes of effects" and predicting effects of choices that differ from the ones actually made.

9. A solution to the so called "Mediation Problem," which estimates the degree to which specific mechanisms contribute to the transmission of a given effect, in models containing both continuous and categorical variables, linear as well as nonlinear interactions (Pearl, 2001, 2012a).

10. A solution to the problem of "external validity" (Campbell and Stanley, 1963), namely, deciding if a causal relation estimated in one population can be transported to another population, in which experimental conditions are different (Pearl and Bareinboim, 2011).

A full description of these techniques is given in (Pearl, 2009a) as well as in recent survey papers (Pearl, 2010b,c). Here I will demonstrate by examples how some of the simple tasks listed above are handled in the nonparametric framework of SCM.

3.1 Two models for discussion

Consider a nonparametric structural model defined over a set of endogenous variables $\{Y, X, Z_1,$
$Z_2, Z_3, W_1, W_2, W_3\}$, and unobserved exogenous variables $\{U, U_1, U_2, U_1', U_2', U', U_1'', U_2'', U_3''\}$. The equations are assumed to be structured as follows:

Model 1

$$\begin{array}{rclcrcl}
Y &=& f(W_3, Z_3, W_2, u) & \quad & X &=& g(W_1, Z_3, u'') \\
W_3 &=& g_3(X, u_3'') & & W_1 &=& g_1(Z_1, u_1') \\
Z_3 &=& f_3(Z_1, Z_2, u_3') & & Z_1 &=& f_1(u_1) \\
W_2 &=& g_2(Z_2, u_2') & & Z_2 &=& f_2(u_2)
\end{array}$$

$f, g, f_1, f_2, f_3, g_1, g_2, g_3$ *are arbitrary, unknown functions, and all exogenous variables are mutually independent but otherwise arbitrarily distributed.*

For the purpose of our illustration, we will avoid assigning any economic meaning to the variables and functions involved, thus focusing on the formal aspects of such models rather than their substance. The model conveys two types of theoretical (or causal) assumptions:

1. Exclusion restrictions, depicted by the absence of certain variables from the arguments of certain functions, and

2. Causal Markov conditions, depicted by the absence of common U-terms in any two functions, and the assumption of mutual independence among the U's.

Given the qualitative nature of these assumptions, the algebraic representation is superfluous and can be replaced, without loss of information, with the diagram depicted in Fig. 2.[4] To anchor the discussion in familiar grounds, we also present the

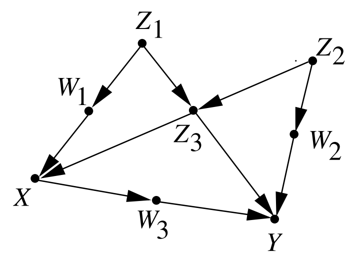

Figure 2: A graphical representation of Model 1. Error terms are assumed mutually independent and not shown explicitly.

linear version of Model 1

Model 2 *(Linear version of Model 1)*

$$
\begin{array}{rclcrcl}
Y & = & aW_3 + bZ_3 + cW_2 + u & \qquad & X & = & t_1W_1 + t_2Z_3 + u'' \\
W_3 & = & c_3X + u_3'' & & W_1 & = & a_1'Z_1 + u_1' \\
Z_3 & = & a_3Z_1 + b_3Z_2 + u_3' & & Z_1 & = & u_1 \\
W_2 & = & c_2Z_2 + u_2' & & Z_2 & = & u_2
\end{array}
$$

All u's are assumed to be uncorrelated.

While the orthogonality assumption renders these equations regressional, we can easily illustrate non-regressional models by assuming that some of the endogenous variables are not measurable.

3.2 Illustrating typical question-answering tasks

Given the model defined above, the following are typical questions that an economist may wish to ask.

[4]This is entirely optional; readers comfortable with algebraic representations are invited to stay in their comfort zone.

3.2.1 Testable implications (misspecification tests)

a. What are the testable implications of the assumptions embedded in Model 1?

b. Assume that only variables X, Y, Z_3, and W_3 are measured, are there any testable implications?

c. The same, but assuming only variables X, Y, and Z_3 are measured,

d. The same, assuming all but Z_3 are measured.

e. Assume that an alternative model, competing with Model 1, has the same structure, with the $Z_3 \rightarrow X$ arrow reversed. What statistical test would distinguish between the two models?

f. What regression coefficient in Model 2 would reflect the test devised in (e)?

3.2.2 Equivalent models

a. Which arrows in Fig. 2 can be reversed without being detected by any statistical test?

b. Is there an equivalent model (statistically indistinguishable) in which Z_3 is a mediator between X and Y (i.e., the arrow $X \leftarrow Z_3$ is reversed)?

3.2.3 Identification

a. Suppose we wish to estimate the average causal effect of X on Y

$$ACE = P(Y = y|do(X = 1)) - P(Y = y|do(X = 0)).$$

Which subsets of variables need to be adjusted to obtain an unbiased estimate of ACE?

b. Is there a single variable that, if measured, would allow an unbiased estimate of ACE?

c. Assume we have a choice between measuring $\{Z_3, Z_1\}$ or $\{Z_3, Z_2\}$, which would be preferred?

3.2.4 Instrumental variables

a. Is there an instrumental variable for the $Z_3 \rightarrow Y$ relationship?
If so, what would be the IV estimand for parameter b in Model 2?

b. Is there an instrument for the $X \rightarrow Y$ relationship?
If so, what would be the IV estimand for the product $c_3 c$ in Model 2?

3.2.5 Mediation

a. What variables must be measured if we wish to estimate the direct effect of Z_3 on Y?

b. What variables must be measured if we wish to estimate the indirect effect of Z_3 on Y, mediated by X?

c. What is the estimand of the indirect effect in (b), assuming that all variables are binary?

3.2.6 Regressional digressions

Consider the linear version of our model (Model 2)

Question 1: Name three testable implications of this model

Question 2: Suppose X, Y, and W_3 are the only variables that can be observed. Which parameters can be identified from the data?

Question 3: If we regress Z_1 on all other variables in the model, which regression coefficient will be zero?

Question 4: If we regress Z_1 on all the other variables in the model and then remove Z_3 from the regressor set, which coefficient will not change?

Question 5: ("robustness") (A more general version of Question 4.)

Model 2 implies that certain regression coefficients will remain invariant when an additional variable is added as a regressor. Identify five such coefficients with their added regressors.[5]

3.2.7 Counterfactual reasoning

a. Determine if X is independent of the counterfactual Y_x conditioned on all the other endogenous variables.

b. Determine if X is independent of the counterfactual $W_{3,x}$ conditioned on all the other endogenous variables.

c. Determine if the counterfactual $P(Y_x|X = x')$ is identifiable, assuming that only X, Y, and W_3 are observed.

[5] According to White and Lu (2010) "A common exercise in empirical studies is a 'robustness check,' where the researcher examines how certain 'core' regression coefficient estimates behave when the regression specification is modified by adding or removing regressors." "of the 98 papers published in The American Economic Review during 2009, 76 involve some data analysis. Of these, 23 perform a robustness check along the lines just described, using a variety of estimators." Since this practice is conducted to help diagnose misspecification, the answer to Question 5 is essential for discerning whether an altered coefficient indicates misspecification or not.

3.3 Solutions

The problems posed in Section 3.2 read like homework problems in Economics 101 class. They should be! Because they are fundamental, easily solvable, and absolutely necessary for even the most elementary exercises in nonparametric analysis. Readers should be pleased to know that with the graphical techniques available today, these questions can generally be answered by a quick glance at the graph of Fig. 2 (see, for example, Pearl (2010b,c, 2012b), Greenland and Pearl (2011), or Kyono (2010)).

More elaborate problems, like those involving transportability or counterfactual queries may require the use of *do*-calculus or counterfactual logic. Still, such problems have been mathematized, and are no longer at the mercy of metaphysical thinking.

It should also be noted that, with the exception of our regressional digression (3.2.6) into Model 2, all queries were addressed to a purely nonparametric model and, despite the fact that the form of our equations and the distribution of the U's are totally arbitrary, we were able to extract answers to policy-relevant questions in a form that is estimable from the data available.

For example, the answer to the first identification question (a) is: The set $\{W_1, Z_3\}$ is sufficient for adjustment and the resulting estimand is:

$$P(Y = y|do(X = x)) = \sum_{w_1, z_3} P(Y = y|X = x, Z_3 = z_3, W_1 = w_1)P(Z_3 = z_3, W_1 = w_1).$$

This can be derived algebraically using the rules of *do*-calculus or seen directly from the graph, using the back-door criterion (Pearl, 1993). When a policy question is not identifiable, the *do*-calculus is guaranteed to discover it and exit with failure.

The nonparametric nature of these exercises represents the ultimate realization of what Heckman calls the Marschak's Maxim (Heckman, 2010), referring to an observation made by Jacob Marschak (1953) that many policy questions do not require the estimation of each and every parameter in the system – a combination of parameters is all that is necessary and, moreover, it is often possible to identify the desired combination without identifying the individual components. The exercises presented above show that Marschak Maxim goes even further – the desired quantity can often be identified without ever specifying the functional or distributional forms of these economic models.

4 Remarks on the "Structuralists" vs. "Experimentalists" Debate

The Spring 2010 issue of the *Journal of Econometric Perspectives* (Vol. 24, No. 2) presented an interesting discussion on causal inference between two camps of economists: the "structuralists" and the "experimentalists;" the former acknowledge their reliance on modelling assumptions, the latter argue that they don't, or claim to minimize such dependence. Angrist and Pischke (2010) represented the "experimentalist" position and Leamer (2010), Nevo and Whinston (2010), Keane (2010), and Sims (2010) defending the structural approach.

Viewed from the SCM perspective, the debate is rhetorical. It is an axiomatic wisdom and, by now, also a logical theorem, that any causal conclusion drawn from observational studies must rest on untested causal assumptions. Therefore, whatever relation an instrumental design bears to an ideal controlled experiment is just one such assumption and, to the extent that the "experimental" approach is valid, it is a routine exercise in structural economics. However, the philosophical basis of the "experimentalist" approach, as it is currently marketed, is misguided and potentially dangerous, for it takes surface similarity to the randomized controlled trial ideal to be its guiding principle, as opposed to explicitly examining the validity of the assumptions. The fallibility of this paradigm has surfaced in a number of examples (e.g., Pearl 2009b, 2011a,b) and has given birth to a school of research that in the name of mimicking controlled experiments avoids making modelling assumptions transparent.

Another take on the "experimental - structural" debate is provided by Heckman (2010) who reiterates the superiority of the structural over the Neyman-Rubin model, but stops short of identifying the key element for that superiority. This is important because, after all, the structural and potential-outcome approaches are logically equivalent, differing only in the languages used to encode assumptions; the former using equations, the latter using counterfactual independencies (see Pearl 2009a, pp. 230–234). So why did the "experimentalists" end up with the primitive, single-equation exercises reported in Angrist and Pischke (2010)? Why did they not import the rich knowledge that structural modellers encode in their equations, to make their assumptions compelling, explicit and transparent?

I believe the answer lies in an observation made by Sims (2010): "using instrumental variable formulas while simply listing the instruments, with little or no discussion of what kind of larger multivariate system would justify isolating the single equation or small system to which the formulas are applied, was, and to some extent still is, a common practice."

By rejecting structural equations as a language for expressing substantive economic knowledge, and confining themselves exclusively to the language of potential outcomes[6] "experimentalists" have in effect cut themselves off from the one language in which large number of relationships can be expressed meaningfully and reasoned about.

At the very least, "experimentalists" could have acquired the basic tools of identifying instrumental variables in a system of equations and could have learned to solve elementary problems such as those posed in 3.2.4. This they refuse to do citing theology (Rubin, 2010) and lack of evidence (Imbens, 2010; Pearl, 2009b); at the price of

[6] The potential outcome language has been shown to be logically equivalent to the language of nonparametric structural equations (Galles and Pearl, 1998; Halpern, 1998); a theorem in one is a theorem in another, and an assumption in one has a corresponding assumption in the other. The two differ only in how substantive information is encoded. The potential outcome language insists on encoding such information in the form of conditional independence statements about counterfactual variables, a cognitively formidable task, while the structural equation model permits modelers to encode this information in the form of causal-effect relationships representing economic processes. A simple translation between the two is given in (Pearl, 2009a, pp. 231–234) which "experimentalists" have so far ignored.

having to confine their analysis to "using instrumental variables formulas while simply listing the instrument." It is not lack of good intention, but lack of modern mathematical tools that prevents the "experimentalists" from conducting a "discussion of what kind of larger multivariate system would justify" their formulas.[7]

5 Conclusions

This paper traces the logic and mathematical machinery needed for causal analysis from the original insights advanced by Haavelmo to the nonparametric analysis of Structural Causal Models (SCM). We have demonstrated by examples the type of queries the SCM framework can answer, the assumptions required, the language used for encoding those assumptions and the mathematical operations needed for deriving causal and counterfactual conclusions.

Not surprisingly, graphical formalism was found to be the most succinct, natural and effective language for representing nonparametric structural equations; it highlights the assumptions and abstracts away unnecessary algebraic details. It is for these reasons that graphical representations have become an indispensable second language in the health sciences (Greenland, Pearl, and Robins, 1999; Greenland and Pearl, 2011 or Kyono, 2010) and are making their way towards the social and behavioral sciences (Morgan and Winship, 2007). I am convinced therefore that, once the power of graphical tools is recognized through simple examples, economists too will add them to their arsenal of formal methods and be able to reap the benefits of causal analysis, parametric as well as nonparametric.[8] Acquiring these tools enables researchers to recognize the testable implications of a system of equations, locate instruments in such systems, decide if two such systems are equivalent, if causal effects are identifiable, if two counterfactuals are independent given another, whether a set of measurements will reduce bias, and, most importantly, reading the causal and counterfactual information that such systems convey.

The development of powerful mathematical tools for deriving or predicting the logical ramifications of untested theoretical assumptions will enable us to reverse-engineer our inferences and learn to minimize sensitivity to those assumptions.

[7]The potential outcome language, is rather inept for capturing substantive knowledge of the kind carried by structural equation models. The restricted vocabulary of "ignorability," "treatment assignment" and "missing data" that has ruled (and still rules) the potential-outcome paradigm is not flexible enough to specify transparently even the most elementary models (say a three-variable Markov chain) that researchers wish to hypothesize (Pearl, 2011c).

[8]Frankly, a recent survey of econometric textbooks have made me more pessimistic on whether economists can lift themselves to the age of modernity. Research leadership is zealously guarding the tradition from imported new tools (Pearl, 2009a, pp. 374-80) and textbooks continue to conflate regressional and structural vocabulary with increasing frequency (Chen and Pearl, 2012).

Acknowledgment

This paper has benefited from discussions with Ed Leamer, James Heckman, and Hal White.

This research was supported in parts by grants from NIH #1R01 LM009961-01, NSF #IIS-0914211 and #IIS-1018922, and ONR #N000-14-09-1-0665 and #N00014-10-1-0933.

References

Ali, R., Richardson, T. and Spirtes, P. (2009). Markov equivalence for ancestral graphs. *The Annals of Statistics* **37** 2808–2837.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* **24** 3–30.

Balke, A. and Pearl, J. (1995). Counterfactuals and policy analysis in structural models. In *Uncertainty in Artificial Intelligence 11* (P. Besnard and S. Hanks, eds.). Morgan Kaufmann, San Francisco, 11–18.

Campbell, D. and Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research*. Wadsworth Publishing, Chicago.

Cartwright, N. (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press, New York, NY.

Chen, B. and Pearl, J. (2012). Regression and causation: A critical examination of econometric textbooks. Tech. Rep. R-395, Department of Computer Science, University of California, Los Angeles, CA. In process.

Engle, R., Hendry, D. and Richard, J. (1983). Exogeneity. *Econometrica* **51** 277–304.

Galles, D. and Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundation of Science* **3** 151–182.

Goldberger, A. (1992). Models of substance; comment on N. Wermuth, 'On block-recursive linear regression equations'. *Brazilian Journal of Probability and Statistics* **6** 1–56.

Greenland, S. and Pearl, J. (2011). Adjustments and their consequences – collapsibility analysis using graphical models. *International Statistical Review* **79** 401–426.

Greenland, S., Pearl, J. and Robins, J. (1999). Causal diagrams for epidemiologic research. *Epidemiology* **10** 37–48.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11** 1–12. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 477–490, 1995.

Haavelmo, T. (1944). The probability approach in econometrics (1944)*. Supplement to *Econometrica* **12** 12–17, 26–31, 33–39. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, New York, 440–453, 1995.

Halpern, J. (1998). Axiomatizing causal reasoning. In *Uncertainty in Artificial Intelligence* (G. Cooper and S. Moral, eds.). Morgan Kaufmann, San Francisco, CA, 202–210. Also, *Journal of Artificial Intelligence Research* 12:3, 17–37, 2000.

Heckman, J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* **115** 45–97.

Heckman, J. (2003). Conditioning causality and policy analysis. *Journal of Econometrics* **112** 73–78.

Heckman, J. (2008). Econometric causality. *International Statistical Review* **76** 1–27.

Heckman, J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* **48** 356–398.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press, New York.

Imbens, G. W. (2010). An economist's perspective on shadish (2010) and west and thoemmes (2010). *Psychological Methods* **15** 47–55.

Keane, M. P. (2010). A structural perspective on the experimentalist school. *Journal of Economic Perspectives* **24** 47–58.

Kyono, T. (2010). Commentator: A front-end user-interface module for graphical and structural equation modeling. Tech. Rep. R-364, <http://ftp.cs.ucla.edu/pub/stat_ser/r364.pdf>, Master Thesis, Department of Computer Science, University of California, Los Angeles, CA.

Leamer, E. E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives* **24** 31–46.

Lewis, D. (1973). *Counterfactuals*. Harvard University Press, Cambridge, MA.

Lucas Jr., R. (1976). Econometric policy evaluation: A critique. In *The Phillips Curve and Labor Markets* (K. Brunner and A. Meltzer, eds.), vol. CRCS, Vol. 1. North-Holland, Amsterdam, 19–46.

Marschak, J. (1953). Economic measurements for policy and prediction. In *Studies in Econometric Method* (W. C. Hood and T. Koopmans, eds.). Cowles Commission Monograph 10, Wiley and Sons, Inc., 1–26.

Morgan, S. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. Cambridge University Press, New York, NY.

Nevo, A. and Whinston, M. D. (2010). Taking the dogma out of econometrics: Structural modeling and credible inference. *Journal of Economic Perspectives* **24** 69–82.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science* **5** 465–480.

Pearl, J. (1993). Comment: Graphical models, causality, and intervention. *Statistical Science* **8** 266–269.

Pearl, J. (1994). A probabilistic calculus of actions. In *Uncertainty in Artificial Intelligence 10* (R. L. de Mantaras and D. Poole, eds.). Morgan Kaufmann, San Mateo, CA, 454–462.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. 2nd edition, 2009.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, San Francisco, CA, 411–420.

Pearl, J. (2004). Robustness of causal claims. In *Proceedings of the Twentieth Conference Uncertainty in Artificial Intelligence* (M. Chickering and J. Halpern, eds.). AUAI Press, Arlington, VA, 446–453.

Pearl, J. (2009a). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Pearl, J. (2009b). Myth, confusion, and science in causal analysis. Tech. Rep. R-348, University of California, Los Angeles, CA. <http://ftp.cs.ucla.edu/pub/stat_ser/r348.pdf>.

Pearl, J. (2010a). Review of N. Cartwright 'Hunting causes and using them'. *Economics and Philosophy* **26** 69–77.

Pearl, J. (2010b). The foundations of causal inference. *Sociological Methodology* **40** 75–149.

Pearl, J. (2010c). An introduction to causal inference. *The International Journal of Biostatistics* **6** DOI: 10.2202/1557–4679.1203. Available at: <http://ftp.cs.ucla.edu/pub/stat_ser/r354–corrected–reprint.pdf>.

Pearl, J. (2011a). Invited commentary: Understanding bias amplification. *American Journal of Epidemiology* Published online, DOI: 10.1093/aje/kwr352.

Pearl, J. (2011b). Principal stratification a goal or a tool? *The International Journal of Biostatistics* **7**. Article 20, DOI: 10.2202/1557-4679.1322. Available at: <http://ftp.cs.ucla.edu/pub/stat_ser/r382.pdf>.

Pearl, J. (2011c). Graphical models, potential outcomes and causal inference: Comment on Lindquist and Sobel. *Statistics in Medicine* **58** 770–771.

Pearl, J. (2012a). The causal mediation formula – a guide to the assessment of pathways and mechanisms. *Prevention Science* In Press. Available at: <http://ftp.cs.ucla.edu/pub/stat_ser/r379.pdf>.

Pearl, J. (2012b). The causal foundations of structural equation modeling. In *Handbook of Structural Equation Modeling* (R. H. Hoyle, ed.). Guilford Press, New York. In press.

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI-11).* Menlo Park, CA. Available at: <http://ftp.cs.ucla.edu/pub/stat_ser/r372a.pdf>.

Richard, J. (1980). Models with several regimes and changes in exogeneity. *Review of Economic Studies* **47** 1–20.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701.

Rubin, D. (2010). Reflections stimulated by the comments of shadish (2010) and west and thoemmes (2010). *Psychological Methods* **15** 39–46.

Shpitser, I. and Pearl, J. (2008). Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research* **9** 1941–1979.

Simon, H. and Rescher, N. (1966). Cause and counterfactual. *Philosophy and Science* **33** 323–340.

Sims, C. A. (2010). But economics is not an experimental science. *Journal of Economic Perspectives* **24** 59–68.

Spanos, A. (2011). Revisiting haavelmo's structural econometrics: Bridging the gap between theory and data. Tech. rep., Department of Economics, Virginia Tech, Blacksburg, VA. Presented at the Haavelmo's Centennial Symposium, Oslo. Dec. 13 2011.

Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction, and Search.* Springer-Verlag, New York.

Stock, J. and Watson, M. (2007). *Introduction to Econometrics*. 2nd ed. Addison-Wesley, New York.

Strotz, R. and Wold, H. (1960). Recursive versus nonrecursive systems: An attempt at synthesis. *Econometrica* **28** 417–427.

Verma, T. and Pearl, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*. Cambridge, MA. Also in P. Bonissone, M. Henrion, L.N. Kanal and J.F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 6*, Elsevier Science Publishers, B.V., 255–268, 1991.

Wermuth, N. (1992). On block-recursive regression equations. *Brazilian Journal of Probability and Statistics* (with discussion) **6** 1–56.

White, H. and Lu, X. (2010). Robustness checks and robustness tests in applied economics. Tech. rep., Department of Economics University of California, San Diego, CA. Discussion paper.