LSE CPNSS

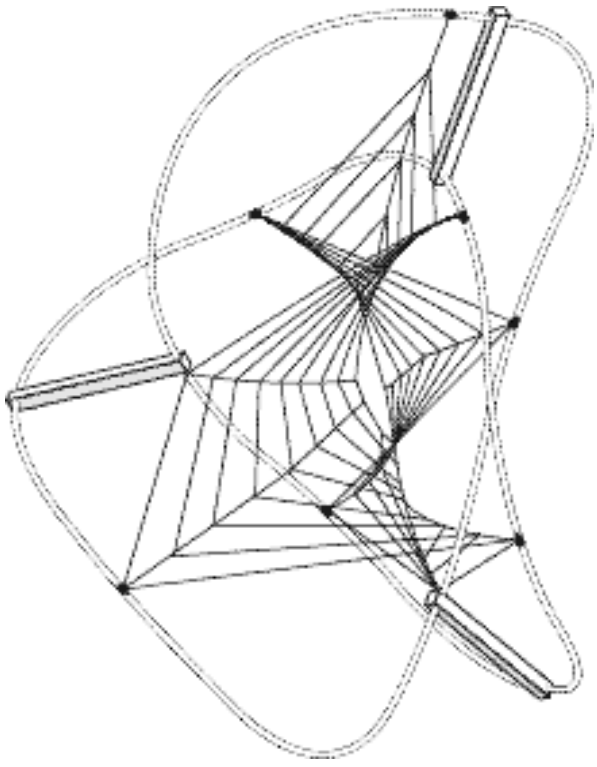
Centre for Philosophy of Natural and Social Science

Causality: Metaphysics and Methods

Technical Report 20/04

Causality in Economics

Stephen F. LeRoy



Editor: Julian Reiss

Causality in Economics*

Stephen F. LeRoy
University of California, Santa Barbara

September 12, 2004

Abstract

Formal analysis of causal relations using graphical methods has become increasingly popular in the natural and social sciences, but has been used much less in economics. The reason may be that heretofore graphical methods have been based on models that are structural in the sense that the equality symbol is taken to denote causation directly. We argue that current economic models are not structural in this sense. This paper proposes a formal analysis of causality that applies to economic models and develops its properties. Among other results, it is shown that Granger causality can be connected to causality only under very strong assumptions. Also, it is shown that graphical methods are not likely to find much application in determining causal orderings as defined in this paper in economic models.

Formal analysis of causal relations using graphical methods has become increasingly popular in the natural and social sciences. Pearl [23] is a good sample of this work. Graphical methods are a development of analytical techniques that originated in economic theory—specifically, they take as their basis structural economic models as defined by the Cowles group half a century ago. The standard reference on causation is Simon [26].

The term "structural" was never defined clearly by the Cowles economists. The term has subsequently been given different meanings by different analysts, as has frequently been observed. At a minimum, the term refers to the distinction between the structural form and the reduced form of a model. In the structural form each internal variable was expressed as a function of some other internal variables and some external variables, whereas the reduced form referred to the solution of a model, in which each internal variable is expressed as a function of the external variables alone. As the term implies, the structural form was viewed as more fundamental than the

*First draft August 25, 1999. I am indebted to Nancy Cartwright for many discussions and much correspondence. I have also received comments from Daniel Hausman, Damien Fennell, Julian Reiss and other participants at a seminar at the London School of Economics.

reduced form. It was seen as containing information not present in the reduced form, such as exclusion restrictions. For example, a structural supply-demand system might specify that one or several of the external variables affect demand but not supply, or vice-versa. These restrictions were used to identify structural coefficients.

In many analyses "structural" has a further meaning: a structural equation is one in which the right-hand side variables are causes of the left-hand side variable. In structural equations so defined the equals sign has the meaning of the assignment operator in computer languages, as Pearl [23] observed. This is distinguished from its meaning in mathematics, which by definition treats the right-hand side and left-hand side of an equation symmetrically. Graphical analyses of causation have heretofore been based on this causal interpretation of structural equations: the graphical representation of the equals sign is an arrow from the right-hand side variables to the left-hand side variable.

With a few exceptions (such as Swanson and Granger [28]), economists have been conspicuous in their absence from these developments in recent years. It is true that the topic of causality has been of passing interest to economists—witness Granger causality—but it is not obvious how the lines of inquiry that economists have pursued are related to the graphical methods used in the other disciplines. It is clear why economists have not adopted the new graphical methods: economic models use the equality symbol with its usual mathematical meaning, not with the meaning of the assignment operator.[1] Therefore economic models are not expressible as graphs, at least insofar as graphs are based on the causal interpretation of the equality symbol. That being so, economists' lack of interest in graphical analyses of causation is not surprising. Further, economic models in the tradition of general equilibrium theory (including modern macroeconomics) make no use of the structural form/reduced form distinction.

In an earlier paper (LeRoy [18]) I presented an alternative treatment of causation that is fully applicable in models of the sort that economists use. However, that presentation did not address many applied questions that must be resolved if the methods outlined there are to be adopted in causal analysis of economic models. Some of these are discussed in the present paper. The paper begins in Section 1 with a review of Simon's treatment of causation, which turns out to be more closely related to the analysis of this paper that to treatments based on the structural interpretation of economic models. Then a self-contained formal treatment of causality, essentially the same as that in LeRoy [18], is presented. In Section 2 econometric considerations are introduced by specifying the observability or lack thereof of external variables and determining observable implications of causal orderings. In Sections 3 and 4 various

[1]Some contemporary studies of causation in economic models carry over the interpretation of right-hand side variables as causes of the left-hand side variable. For example, in a supply-demand model Heckman [12] justified putting quantity on the left-hand side and price on the right-hand side on the grounds that in competitive models individuals are modeled as price-takers. It is not clear to what extent Heckman's formal analysis depends on this interpretation.

definitions of causality are compared. Section 5 points out that graphical methods can be used to test causal orderings under the definition of causality adopted in this paper. Section 6 is the conclusion.

1 Causality

1.1 Notation

One problem that the reader of the literature on causality encounters is that terminology is sometimes not clearly defined and, further, many of the terms are used with different meanings by different authors. To forestall confusion we begin by defining the terminology used here. It seems to me that these definitions are close to standard usage, insofar as standard usage exists, but some analysts appear to disagree (for example, Hoover [15], p. 171, referred to the paper preceding this one, which has substantially the same terminology, as offering a "complex and difficult terminological landscape").

One essential distinction, stressed in logic and mathematics but often blurred in philosophical discourse, is between variables and constants. A *variable* is the argument or value of a function, while a *constant* is a standin for a number. By specifying that a is a constant, the analyst stipulates that alternative values of a will not be considered—doing so would make no more sense than asking what would happen in mathematics if the constant π took on some value other than 3.14159. Variables, in turn, may be classified into *external* and *internal variables*. Internal variables are determined by the model; external variables are taken as model inputs. Therefore a model consists of a map from the space of external variables to that of internal variables (for example, see Debreu [5]). Of course, in some exercises involving formal models, such as data-description exercises, variables are not classified as between external and internal variables, making it impossible to discuss causation. Since causation is the subject here, we will assume that the analyst is willing to make this classification. It is assumed that there are no functional relations linking external variables, but the support of each external variable may be different for different values of other external variables (so that the domain of the map is not necessarily an interval).

The terms "exogenous variable" and "endogenous variable" are often used with the same meaning as "external variable" and "internal variable", and the etymology of the former pair of terms supports this usage. However, econometricians use the term "exogenous variable" with a different meaning: roughly, an exogenous variable is an observable variable that is uncorrelated with some unobservable variable (Engle, Hendry and Richard [6], for example). Therefore we follow Leamer's [17] recommendation that the terms "external variable" and "internal variable" be substituted to avoid confusion.

In multidate models we distinguish between parameters and processes. A *para-*

meter is a variable, while a *process* is a collection of variables, one for each element of some index set representing time. Equivalently, one could define a parameter as a process each element of which is assumed to take on the same value. In many discussions the terms "constant" and "variable" are used where we use "parameter" and "process", but this usage would invite confusion in the present paper because of the different definitions of constant and variable presented above.

Parameters and processes, like variables, may be external or internal. In macroeconomics the terms "deep parameters" and "shallow parameters" are often used with the same meaning we assign to "external parameters" and "internal parameters".

In other discussions the term "parameter" is used with different meanings. For example, Marschak [21] classified a variable as a parameter if it was constant in the past, even if it is subject to change now. A related usage has been adopted in new classical macroeconomics (Lucas [20]). Hoover [15], in contrast, defined a parameter as a variable that is subject to direct control (p. 61). This definition appears to coincide with our definition of an external variable. In some of Hoover's applications parameters are assumed to take on different values at different dates, although Hoover did not time-subscript parameters or identify them as processes, as would seem to be appropriate at least according to the notation adopted here. The merits of Hoover's terminology are not clear; however, our point here is to set out the notation of the present paper, not to argue against other possible choices.

Variables may be either observable to the analyst, or unobservable. External unobservable variables will be assigned probability distributions, and these will induce probability distributions on internal variables, both observable and unobservable (assuming that models have unique solutions, as we do throughout this paper). We will assume that constants are unobservable. The following notation is adopted:

external observable variables or processes	x
internal observable variables or processes	y
unobservable constants	a, b, A, B, C
unobservable external variables or processes	ε
external parameters	θ
internal parameters	ψ

Note that this classification is incomplete; other possibilities, such as internal unobservable processes (latent variables, in some characterizations) are deleted because they are not considered in this paper.

1.2 Simon's Definition

We begin with a (somewhat unconventional) review of Simon's definition of causation. Suppose that a model is representable as a linear operator from R^m, a space of external variables, onto R^n, a space of internal variables, where $m \geq n$. This operator is

assumed to be representable by a rank-n $n \times m$ matrix C of constants:

$$y = Cx. \tag{1}$$

Here (1) is the *reduced form*.

Assume that (1) A is an $n \times n$ matrix of constants with zeros on the main diagonal, (2) B is an $n \times m$ matrix of constants, and (3) A and B satisfy

$$(I - A)^{-1}B = C. \tag{2}$$

Under these assumptions the model (1) can be written in the form

$$y = Ay + Bx, \tag{3}$$

as is readily verified by substituting the left-hand side of (2) for C in (1).

Simon defined causal orderings from (3): for two internal variables y_1 and y_2, y_1 causes y_2—denoted $y_1 \rightarrow y_2$—if y_1 appears in the block of equations that determine y_2, and also in a block of equations of lower order (see Simon [26] for definitions of these terms). For example, in the model

$$y_1 = b_{11}x_1 + b_{12}x_2 \tag{4}$$
$$y_2 = a_{21}y_1 + b_{23}x_3, \tag{5}$$

we have $y_1 \rightarrow y_2$ because y_1 appears in equation (5), which determines y_2, but also in equation (4), which by itself constitutes a lower-order block.[2] Formally, the causal ordering on Y, the set of internal variables, associated with a given structural model is a subset of $Y \times Y$; $y_1 \rightarrow y_2$ means that (y_1, y_2) is in the ordering.

There is a well-known difficulty with this: algebraic operations on the equations of (3) can apparently alter causal orderings. For example, substituting (4) in (5) results in

$$y_2 = a_{21}b_{11}x_1 + a_{21}b_{12}x_2 + b_{23}x_3. \tag{6}$$

In the model (4), (6) neither y_1 nor y_2 causes the other according to Simon's definition. Different algebraic operations result in models in which y_1 and y_2 are simultaneously determined, or obey $y_2 \rightarrow y_1$, even though each of these models represents the same linear operator C. It appears as if apparently innocuous mathematical operations alter causal orderings.

This problem reflects the fact that the matrices A and B associated with a given C by (2) are not unique. Therefore on Simon's definition the causal ordering of the internal variables depends on which of an infinite number of pairs of matrices A and B satisfying (2) is chosen.

[2]This statement, of course, presumes that $a_{21} \neq 0$. Hereafter this qualification will not be repeated.

To avoid the problem posed by the apparent dependence of causal orderings on algebraic operations, we impose a further restriction: we require that *each equation contain at least one external variable not found in any other equation.* Hereafter we refer to this condition as the *exclusion condition.* The exclusion condition rules out algebraic operations that involve more than one equation (because if the original model satisfies the exclusion condition, the modified model will not). For example, the model (4), (6) does not satisfy the exclusion condition: the external variables that appear in (4)—x_1 and x_2—also appear in (6).

If the exclusion condition is satisfied and if in addition the model has the same number of internal as external variables, then causal orderings are unique. To see this, note that (1) can be solved to yield

$$x = C^{-1}y. \tag{7}$$

If F is defined by $F = I - C^{-1}$, where I is the identity matrix, this becomes

$$y = Fy + x. \tag{8}$$

Since F is unique in (8), it is clear that the causal ordering is unique.

If $m \geq n$, with any C there is always associated at least one pair A and B that satisfies the exclusion condition: the fact that C is of rank n implies that one can always find a square nonsingular matrix C_1 and a matrix C_2 such that the external variables x can be partitioned (perhaps after reordering) into (x_1, x_2) and (1) can be written in the form

$$y = C_1x_1 + C_2x_2. \tag{9}$$

Premultiplying by C_1^{-1} results in

$$C_1^{-1}y = x_1 + C_1^{-1}C_2x_2. \tag{10}$$

As before, if F is defined by F as $I - C^{-1}$, (10) can be written as

$$y = Fy + x_1 + C_1^{-1}C_2x_2 \tag{11}$$

Here each x_1 enters one and only one of the equations. The variables in x_2 can enter in any or all of the equations.

Damien J. Fennell [7] pointed out that if $m > n$ causal orderings under Simon's definition are not unique even if the exclusion condition is imposed. This is so because with $m > n$ different subsets of the external variables can be selected to satisfy the exclusion condition, and each choice implies a different causal ordering. To see this, consider the system

$$y_1 = b_{11}x_1 + b_{12}x_2 \tag{12}$$

$$y_2 = b_{22}x_2 + b_{23}x_3, \tag{13}$$

in which y_1 and y_2 are not causally ordered. The exclusion condition is satisfied because of the absence of x_3 from (12) and of x_1 from (13). However, if (13) is solved for x_2 and the result is substituted in (12), there results the system

$$y_1 = b_{11}x_1 + a_{12}y_2 - (b_{12}b_{23}/b_{22})x_3. \tag{14}$$

$$y_2 = b_{22}x_2 + b_{23}x_3, \tag{15}$$

where

$$a_{12} = b_{12}/b_{22}. \tag{16}$$

In the system (14)-(15) the exclusion condition is again satisfied because of the absence of x_2 from (14) and of x_1 from (15). In (14)-(15) we have $y_2 \to y_1$. Since (12)-(13) is mathematically equivalent to (14)-(16), it follows that causal orderings are not unique.

Despite this, causal orderings are unique generically: in (14)-(15) we have $y_2 \to y_1$ for almost all values of a_{12}; it is only in the special case (16) that (14)-(15) is equivalent to the system (12)-(13) in which y_2 does not cause y_1. Since we have already ruled out nongeneric special cases (see note 2), it is seen that Fennell's observation about nonuniqueness of causal orderings when $m > n$ does not involve anything new.

What is noteworthy is that assuming that a model satisfies the exclusion condition is weaker than assuming that it is structural as that term was used by the Cowles economists and is used in the modern literature on causality. Imposition of the exclusion condition allows renormalization of individual equations (i.e., solving them so that a different variable appears on the left-hand side), so it does not matter which variable is located on the left-hand side. It follows that causal orderings as Simon defined them are not altered by such renormalizations. In contrast, under the causality definition based on the causal interpretation of the equality symbol, renormalizations of individual equations result in a different model with a different causal ordering.

The foregoing discussion is very close to Simon's development, at least on a sympathetic reading of Simon. On a superficial comparison of the above discussion with Simon's paper, it appears that the exclusion condition has nothing to do with Simon's Section 6 discussion of when causal structures are "operationally meaningful". In fact, however, Simon's discussion is entirely consistent with the discussion here; the apparent differences are terminological.

Simon's Section 6 marked a change from the discussion that preceded it in his paper. Prior to that section Simon did not explicitly incorporate external variables in his discussion (except in Example 4.2), as that term is used here. His examples contained only variables x and constants a (or α). Simon's x corresponds to our y; his a (or α) corresponds to our x and a. Simon used the terms "exogenous variable" and "endogenous variable", but he assigned them a meaning that is derived from

his definition of causal orderings: on Simon's usage if we have $y_1 \rightarrow y_2$, then y_1 is exogenous in the set of equations that determine y_2, and y_2 is endogenous in that set.

However, in Section 6 in dealing with the fact that algebraic operations can apparently alter causal orderings, Simon considered "interventions" in the a terms, implying that in that section he was viewing the a terms as variables that are external in the sense of this paper, as opposed to constants as in the earlier sections.

Simon did not distinguish between the coefficients and the intercept terms, implying that he was allowing for interventions in either. Here, in contrast, we are simplifying relative to Simon by maintaining the assumption that the coefficient terms are constants (i.e., are not subject to intervention), so that only the intercept terms are treated as external variables. Treating the coefficients as variables would convert what is a linear model into a bilinear model. Following Simon here would complicate the discussion unnecessarily.

For Simon, causal orderings are operationally meaningful only if the equations of a structural model have "individual identities". The equations of a structural model have "individual identities" insofar as interventions can be associated with particular equations or subsets of equations. In the terminology of the present paper, these interventions are associated with external variables. Therefore, translating into the terminology of the present paper, Simon's criterion for operational meaningfulness is that particular external variables be associated with particular equations. This corresponds exactly to our exclusion condition.

Simon stated this explicitly: "The causal relationships have operational meaning, then, to the extent that particular alterations or 'interventions' in the structure can be associated with specific complete subsets of equations" (p. 65). Continuing, "[w]e found that we could provide [a causal] ordering with an operational basis if we could associate with each equation of a structure a specific power of intervention, or 'direct control.' ... Hence, ... structural equations are equations that correspond to specified possibilities of intervention" (p. 66).

Simon's discussion would have been clearer if he had explicitly incorporated this idea in his definition of causal orderings, as we have, rather than implicitly attaching the relevant condition later as a condition for causal orderings to be operationally meaningful. This is, of course, a criticism of exposition, not substance.

As noted, we simplified Simon's discussion by distinguishing intercepts from coefficients: we treated intercepts as external variables and coefficients as constants. In contrast, Simon treated both intercepts and coefficients as external parameters. Our simplification does not alter the substance of Simon's argument: it is easy to see from examination of examples that if $y_1 \rightarrow y_2$ when the coefficients are treated as constants, the same is true when the coefficients are treated as external parameters.[3]

However, a critical assumption that Simon adopted without discussion, that the

[3]However, the converse is not true: changing an external parameter to a constant reduces the set of possible interventions, implying that the causal ordering may become smaller.

coefficients are all external, greatly limits the applicability of his treatment of causation. Macroeconomists in particular frequently deal with linearized versions of models. In such settings the coefficients—the "shallow parameters"—are functions of a few deep parameters representing preferences, production possibilities and the like. With fewer external than internal variables, there is no possibility that the exclusion condition is satisfied. In such models Simon's definition of causality is not applicable or, as he would have put it, not operationally meaningful.

1.3 Causality as Sufficiency[4]

Part of the reason Simon's characterization of causality is not much used currently is that Simon did not provide a clear explanation of what follows if one variable causes another. What interventions are admissible if y_1 causes y_2, but not otherwise? What is the interpretation of these interventions? What does the fact that the cause variable is determined in a lower-order subsystem relative to the effect variable have to do with causation? What is the content of "operationally meaningful" in this context, and what is the connection between this concept and the exclusion condition?

The best way to supply intuitive content to causality is to consider simple examples. We will see that in some such cases it is clear that causal statements are not appropriate, while in other cases it is equally clear that they are. Examination of the difference between these examples will suggest the general principle. This principle is stated more formally in the next subsection.

Consider the supply-demand model:

$$\begin{aligned} q_s &= a_s + a_{sp}p + b_{sw}w && (17)\\ q_d &= a_d + a_{dp}p + b_{di}i && (18)\\ q_s &= q_d = q, && (19) \end{aligned}$$

where q_s is quantity supplied, q_d is quantity demanded, q is equilibrium quantity, i is income, p is price and w is weather. Here weather and income are external and the other variables are internal.

In the system (17)-(19) the question "What is the effect of weather on the equilibrium quantity?" is unambiguous: the effect can be directly calculated from the model. This is so because weather is external. However, if one were to ask "What is the effect of price on equilibrium quantity?" the appropriate response would be that the question is misposed. Price and quantity are both internal; they are simultaneously determined, and neither is causally prior to the other.

The reasoning here is worth elaborating. The assumed intervention results in the price changing from, say, p to $p + \Delta p$. The problem is to infer the effect of this intervention on q. The reason the question is ambiguous is that any of an infinite

[4]The material presented in this and the following subsections is drawn from LeRoy [18]

number of pairs of shifts in the external variables "weather" and "income" could have caused the assumed change in price, and these interventions map onto different values of q. Thus the reason the question is misposed is that it does not give enough information about the intervention being considered to allow a unique answer.

The suggestion is that causal statements involving internal variables as causes are ambiguous *except when all the interventions consistent with a given change in the cause variable map onto the same change in the effect variable.* One is led to define two internal variables as causally ordered when the indicated condition is satisfied, and not otherwise.

Now consider the model

$$\begin{aligned} q_s &= a_s + b_{sw}w + b_{sf}f \quad (20) \\ q_d &= a_d + a_{dp}p + b_{di}i \quad (21) \\ q_s &= q_d = q, \quad (22) \end{aligned}$$

where f is fertilizer. Weather, fertilizer and income are the external variables. Here even though q is internal there is no problem with the assertion that q causes p. This is so because all the interventions in weather and fertilizer consistent with a given change in q map onto the same value of p, as the structure of the model makes obvious.

1.4 A Formal Statement

Let the *external set* $X_j$for a particular internal variable y_j be the minimal set of external variables such that y_j can be written as a function of X_j. Then the model (1) can be written in the form

$$y_j = \beta_j \overline{X}_j \qquad j = 1, ..., n, \quad (23)$$

where $\overline{X}_j$ is a vector of which the elements are the members of X_j, and β_j is a conformable vector. Of course, β_j coincides with the j-th row of C with the zero elements deleted.

Suppose that $X_i \subset\subset X_j$, where $\subset\subset$ means "is a proper subset of". Hereafter we will call this the *subset condition.*[5] Further, assuming that the subset condition is satisfied, define $X_{j,i}$ as $X_j - X_i$ (i.e., as the set consisting of the elements of X_j that are not in X_i). Define $\overline{X}_{j,i}$ as a vector of which the elements are the members of $X_{j,i}$. Suppose in addition that there exists a scalar $\gamma_{j,i}$ and vector $\delta_{j,i}$ such that (23) can be written in the form

[5]The subset condition is essentially the same as Hausman's *independence condition* ([11], p. 64). See also Hoover [14], p. 103 ff.

$$y_j = \beta_j \overline{X}_j = \gamma_{j,i} y_i + \delta_{j,i} \overline{X}_{j,i}. \tag{24}$$

Existence of $\gamma_{j,i}$ and $\delta_{j,i}$ with this property implies that all the interventions in X_i consistent with a given change in y_i have the same effect on y_j. Thus all the information relevant for y_j contained in X_i is summarized in y_i, so that even though the intervention in X_i associated with a given intervention on y_i is ambiguous, there is no ambiguity in the effect of the intervention on y_j. When $\gamma_{j,i}$ and $\delta_{j,i}$ exist that satisfy the above property we will say that y_i is a *simple cause* of y_j, and will write $y_i \Rightarrow y_j$. Thus $y_i \Rightarrow y_j$ means that y_i is sufficient for X_i in the determination of y_j. We will call the condition that there exist $\gamma_{j,i}$ and $\delta_{j,i}$ with the properties just described the *sufficiency condition*. Thus we have $y_i \Rightarrow y_j$ if and only if both the subset condition and the sufficiency condition are satisfied.

In general, $X_i \subset\subset X_j$ does not imply existence of $\gamma_{j,i}$ and $\delta_{j,i}$ satisfying (24). Therefore it will not generally be the case that $X_i \subset\subset X_j$ implies $y_i \Rightarrow y_j$. However, if $X_i \subset\subset X_j$, there may exist some other internal variable y_k such that all the interventions in X_i consistent with a given change in y_i and a given value of y_k map onto the same value of y_j. The sufficiency condition just stated plays the same role here as with simple causation. Then we have *conditional causation*, indicated by $y_i \Rightarrow y_j | y_k$. Still more generally, the set of conditioning variables may include several internal variables rather than just one, and may include external variables also.[6]

Another condition is required for conditional causation: y_i must be *variation-free:* if, on the contrary, the variables held constant completely determine y_i, it makes no sense to talk about the effect of variations in y_i on y_j, ceteris paribus, since variations in y_i are impossible. For example, if the conditioning set includes all the external variables, no variation in y_i is possible.

As long as $X_i \subset\subset X_j$, there will always exist some subset (possibly the null set, if $y_i \Rightarrow y_j$) of the external and internal variables such that y_i causes y_j conditional on that set of variables. (However, the conditioning set may not be unique). We will write $y_i \rightarrow y_j$ if either $y_i \Rightarrow y_j$ or $y_i \Rightarrow y_j | z_k$ for some set of variables z_k. Thus $y_i \rightarrow y_j$ under the definition just given is equivalent to $X_i \subset\subset X_j$.

In some settings conditional causation raises problems of interpretation. The indicated intervention requires a nonzero change in the variables in X_i, with the changes required to satisfy a linear relation so as to hold y_k constant. Existence

[6]Even if the subset condition is satisfied, there may exist no set y_k consisting entirely of internal variables such that the sufficiency condition for a conditional causal ordering is satisfied. For example, consider the system

$$y_1 = b_{11} x_1 + b_{12} x_2 \tag{25}$$

$$y_2 = a_{21} y_1 + b_{21} x_1 + b_{13} x_3. \tag{26}$$

Here $X_1 \subset\subset X_2$, but the sufficiency condition for $y_1 \Rightarrow y_2$ is not satisfied (assuming $b_{21} \neq 0$). Further, there are no other internal variables to condition on.

of such functional relations among external variables appears to conflict with their assumed status as external variables. If there exists a functional relation among the variables in X_i, then assuming that these variables are external is a misspecification. Thus the intervention is inappropriate to the assumed model (or, correspondingly, the model is inappropriate to the assumed intervention). An example will be discussed below.

It is easily verified that the above definition of $y_i \to y_j$ coincides with Simon's definition: assuming that the exclusion condition is satisfied, y_i appears in the block of equations that determines y_j and also in a lower-order block if and only if $X_i \subset\subset X_j$. Thus $y_i \to y_j$ can refer to both conditional causation as defined here and Simon's definition of causation.

The preceding characterization of causal orderings applies to linear models, where the coefficients are constants. The extension to the general case is immediate: we have $y_i \Rightarrow y_j$ if there exists a function expressing y_j as a function of $(y_i, X_{j,i})$. Of course, in the general case there does not exist a constant that expresses the strength of the effect of y_i on y_j for all values of the external variables.

1.5 Causality and Parameter Interventions

Many problems involving causation do not satisfy the linearity restriction. In this subsection we make some observations about the consequences of labeling coefficients as parameters rather than constants.

Neoclassical macroeconomists stress that analysis of macroeconomic policy changes requires identifying and estimating "deep parameters" (labeled here external parameters). This is correct if one is considering changes in policy regimes, and if one is modeling regime change by parameter interventions, as recommended by the Lucas critique (Lucas [20]), at least on some readings. I have argued elsewhere (LeRoy [19]) that policy changes in dynamic models can be appropriately modeled either through interventions on external parameters or external processes, depending on whether the policy intervention involves real-time changes or comparison of different policies each of which obtains over all time.

An important point is that if regime changes are modeled using process rather than parameter interventions, there is no need to model the dependence of internal parameters on external parameters. This is so because if y_{1t} causes y_{2t} when the coefficients are treated as constants, the same is true when the coefficients are treated as parameters, and this is so regardless of the causal ordering among parameters. The only difference is that in the latter case the external parameters, or some subset of them, are included in the exogenous sets, but this change will not cause failure of the subset and sufficiency conditions, assuming that these are satisfied when coefficients are treated as constants. This point underscores the importance of Marschak's [21] observation that analysis of causation does not always require a complete characterization of a model's causal ordering.

Hoover [13] pointed out that a potentially testable implication of causality is that interventions on the (external parameters that determine the) probability distribution of the cause variable should not affect the probability distribution of the effect variable conditional on the cause variable. Cartwright [1], p. 57, took issue with this assertion:

> If x causes y, then in a two-variable model $D(y|x)$ [the distribution of y conditional on x] measures the strength of x's effect on y [Cartwright's notation has been changed]. Clearly the question of the invariance of the strength of this influence across envisaged interventions in x is one of considerable interest in itself. But finding out the answer is not a test for causality, neither in the original situation nor in any of the new situations that might be created by intervention. Even if x causes y in the original situation and continues to do so across all the changes envisaged, there is in general no reason to think that interventions that change the distribution of x will not also affect the mechanism by which x brings about y, and hence also change the strength of x's influence on y.

This observation appears to be incorrect, at least under the implementation of causality analyzed in this paper. Suppose that we write the model as

$$y_1 = \theta_1 + \theta_2 y_2 + \theta_3 \varepsilon \tag{27}$$
$$y_2 = \theta_4 \eta \tag{28}$$

where $\theta_1, ..., \theta_4$ are external parameters (or variables; in a static model there is no basis for distinguishing parameters from variables), and ε and η are unobserved external variables. For concreteness we will take ε and η to have zero mean and unit variance. Taking θ_4 as a parameter rather than a constant allows the analyst to consider interventions on the standard deviation of y_2.

In the model (27)-(28) y_2 causes y_1. To see this, note that we have that the external sets for y_2 and y_1 are

$$e(y_2) = (\theta_4, \eta) \tag{29}$$
$$e(y_1) = (\theta_1, \theta_2, \theta_3, \theta_4, \varepsilon, \eta), \tag{30}$$

which satisfy the subset condition $X_2 \subset\subset X_1$. The sufficiency condition for $y_2 \Rightarrow y_1$ is also satisfied. We have that y_1 conditional on y_2 has mean $\theta_1 + \theta_2 y_2$ and standard deviation θ_3. Therefore an intervention on θ_4 will affect the marginal distribution of y_1, but will not affect $D(y_1|y_2)$.

As Cartwright observed in the quoted passage, it is easy to imagine models in which interventions on θ_4 do affect the parameters of the conditional distribution of y_1. However, in such models by definition the conditions for $y_2 \Rightarrow y_1$ will fail, contrary to the assumption. It appears, then, that Hoover's assertion is correct.

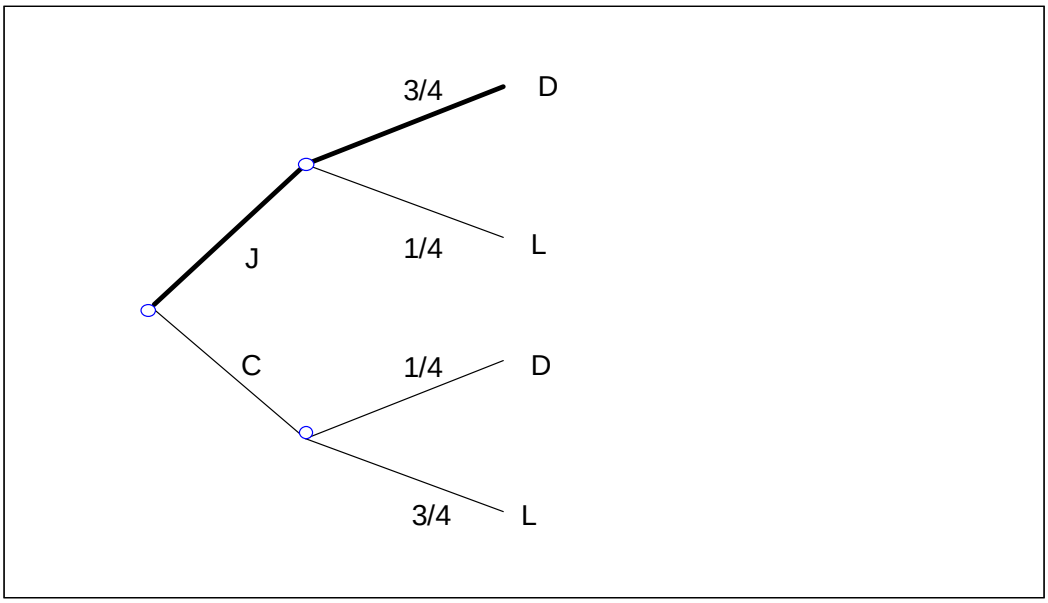


Figure 1:

1.6 Causality and Counterfactuals

Analysis of causality inherently involves counterfactuals (or, as they are sometimes called, hypotheticals). Clear identification of the assumptions that are needed to determine the effect of hypothetical interventions is facilitated by formal analysis. An example, adapted from Dawid [4], makes this point.[7]

A firm has to decide whether to assign employees to the jungle J or the city C. Employees assigned to the jungle will die D with probability 3/4 (and therefore live L with probability 1/4), while those assigned to the city will die with probability 1/4. A given employee is assigned to the jungle, and he dies. The manager asks "What is the probability that he would have died if instead he had been assigned to the city?" Such hypothetical questions are entirely legitimate, and it is worthwhile determining conditions that are required to arrive at specific answers. In Figure 1 the setting just described is presented as a game. The manager decides between assigning the employee to the jungle or the city, and then a chance player determines whether the employee dies. The probabilities assigned to the chance player are as shown, and the heavy line denotes the assumed equilibrium path.

If the manager plays C instead of J, the outcome depends on the draw at the lower chance node, not the upper chance node. Further, the relevant probability is that conditional on D being drawn at the upper chance node. Since we are interested

[7] In the cited paper Dawid's purpose is to argue a very different point: that analysis of counterfactuals can and should somehow be avoided. See Heckman [12] for criticism of this line.

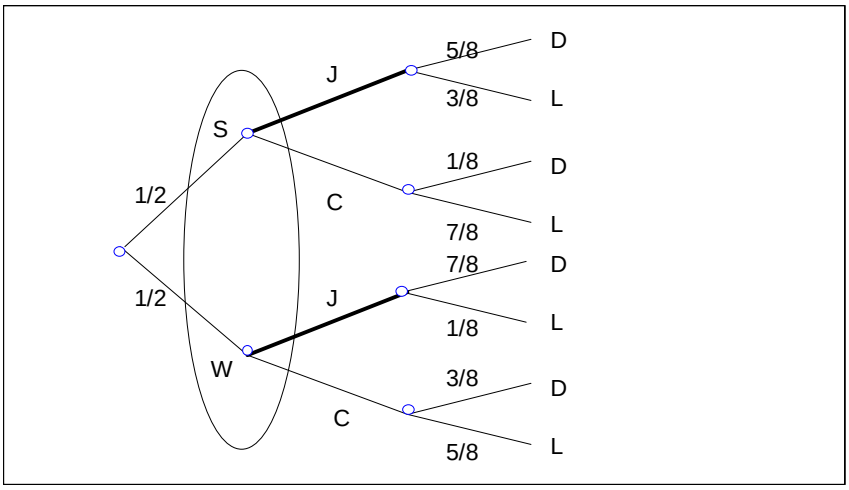


Figure 2:

in a conditional probability, we need to know whether the draws at the two chance nodes are independent. If so, the conditional probability that the employee will die if assigned to the city is 1/4, equal to the corresponding marginal probability. If the analyst is not willing to assume independence, or to make some alternative assumption, then the question cannot be answered.

Specifying whether the draws at the two chance nodes are independent involves a substantive characterization by the analyst of the environment under study. To see this, consider an alternative. Suppose that there are two types of workers, strong S and weak W. Strong workers will die with probability 5/8 if assigned to the jungle, and will die with probability 1/8 if assigned to the city. Weak workers die with probability 7/8 if assigned to the jungle and 3/8 if assigned to the city. The probability that any worker is strong is 1/2. The manager cannot distinguish between strong and weak workers in setting the work assignment. Since 3/4 is the average of 7/8 and 5/8, and 1/4 is the average of 3/8 and 1/8, this specification is identical to that given above to any analyst who cannot distinguish strong from weak workers.

The game is shown in Figure 2. The root node is a chance node determining whether the worker is strong or weak. Then the manager chooses J or C (he cannot tell which node he is at, and therefore must choose J at both nodes or C at both nodes). The rightmost nodes are chance nodes.

Given that the manager chose J and the worker died, an application of Bayes' theorem shows that the posterior probability that the worker is strong is 5/12 rather than 1/2. Consequently, the probability that the worker would have died if assigned to the city is 13/48 rather than 1/4. In this version the analyst no longer treats the event D if C as independent of the event D if J; instead, these events are positively

associated. However, the analysis still depends on the assumption that the chance nodes are mutually independent along alternative paths of play.

The point of this discussion is that in discussing hypotheticals it is necessary to characterize aspects of the environment that can otherwise remain unspecified: in most applications of game theory the question of dependence between the draws at chance nodes on different paths of play does not come up because these paths represent alternatives (in contrast, of course, game theorists routinely assume independence between draws at chance nodes along given paths of play). However, such dependence is obviously relevant when the analysis involves counterfactuals, as here.

2 Causality and Identification

2.1 Identification and Exclusion Restrictions

Exclusion restrictions play a central role in determining causal orderings. As is well known, they also play a central role in determining whether the coefficients in a model are identified (Fisher [8]). Despite the common role of exclusion restrictions in determining causality and identification, the two are very different notions: causality is an ordering on variables, whereas identification has to do with whether the econometrician can make inferences about parameter values from the (population) distribution of observable variables. Causality can be defined and analyzed without even specifying which variables are observable, and in fact this is exactly what we have done up to now. However, estimation of causal parameters (or constants; the distinction is not relevant here) requires assumptions that assure identification, and this goes beyond causality. Whether parameters are identified depends not just on which variables are excluded from which equations, but also on which variables are observable and what distributional assumptions are imposed on those (external) variables that are not observable.

In bringing empirical evidence to bear on assessing the causal relation (or lack thereof) between variables y_i and y_j in a given model, then, two questions must be distinguished. First, are the two variables causally ordered? If not, the question of identification of causal or parameters does not come up because the relevant parameter is undefined. If the two variables are causally ordered, then the associated constant is well defined conceptually, but it is not necessarily identified.

An assumption that plays a central role in assuring identifiability of causal parameters is that *all unobserved external variables are uncorrelated with observed external variables and with each other.* Henceforth we will call this the *uncorrelatedness assumption.*[8] The role of this requirement is to force the model-builder to state explic-

[8]The uncorrelatedness assumption is essentially the same as the principle of the common cause, which says that if two variables are correlated, then either one causes the other or they have a common cause (Reichenbach [25]). Here, however, uncorrelatedness is interpreted as a restriction on formal models rather than as a philosophical proposition. As such, the suitability of the assumption

itly what is assumed about causation, rather than burying causation in uninterpreted correlations.

To understand the role of the uncorrelatedness assumption, consider two observed variables y_1 and y_2. Knowledge of their joint probability distribution obviously does not allow any inference about which variable, if either, causes the other. However, the analyst can certainly define an unobserved variable ε and a parameter β such that y_1 and y_2 satisfy

$$y_2 = \beta y_1 + \varepsilon, \tag{31}$$

where the joint distribution of y_1 and ε and the value of β are chosen so as to generate the assumed joint distribution of y_1 and y_2. If the model-builder is willing to characterize y_1 and ε as external variables, then we have $y_1 \Rightarrow y_2$. Of course, an infinite number of pairs of parameters β and random variables ε are available which satisfy this construction, implying that β is unidentified.

Correpondingly, the model-builder can simply project y_2 onto y_1:

$$y_2 = \delta y_1 + \eta. \tag{32}$$

Here by construction η is uncorrelated with y_1, and the random variable η and parameter δ are unique. However, in the absence of further assumptions this decomposition of y_2 into y_1 and δ has nothing to do with causation; causal variables can be projected onto effect variables as well as vice-versa.

As is obvious, these two decompositions of y_2 into the sum of y_1 multiplied by a constant and a random variable are very different operations: the construction of ε and the assumption that it is external amount to assuming directly that $y_1 \Rightarrow y_2$, but the associated parameter is unidentified. In contrast, the construction of η guarantees its uncorrelatedness with y_1, but has nothing to do with causation. If we are willing to assume further that $\varepsilon = \eta$—or, equivalently, that y_1 and ε are uncorrelated, or that η is external—then we can interpret the coefficient δ of the projection of y_2 on y_1 as coinciding with the causal parameter β.

One implication of the uncorrelatedness assumption is that if one variable causes another, the two variables are necessarily correlated.[9] In some settings this may seem counterintuitive. For example, suppose that y_{1t} is generated according to

$$y_{1t} = a y_{1,t-1} + b y_{2t} + \varepsilon_t, \tag{33}$$

is evaluated according to whether it is analytically fruitful, rather than according to whether it is consistent with, for example, the correlation between British bread prices and the sea level in Venice (see Hoover [16] for discussion).

[9] Of course, the converse is not true: if the external sets for two internal variables have a nonempty intersection, the uncorrelatedness assumption implies that the two variables will be correlated. However, if neither external set is a subset of the other, then the two variables are not causally ordered, either unconditionally or conditionally.

where y_{2t} is a regulator, the behavior of which is generated by

$$y_{2t} = cy_{1,t-1}. \tag{34}$$

Then the behavior of y_{1t} follows

$$y_{1t} = (a + bc)y_{1,t-1} + \varepsilon_t, \tag{35}$$

where the external variables are $\varepsilon_1, \varepsilon_2, \ldots$. and y_{10}. The definitions of causation imply that we have $y_{2t} \Rightarrow y_{1t}$ and $y_{1,t-1} \Rightarrow y_{1t}$ under the (heretofore unstated) assumption that $a + bc \neq 0$.

However, suppose that the regulator is operated by choosing c so as to minimize the unconditional variance of y_t. This results in $c = -a/b$, implying that (35) becomes

$$y_{1t} = \varepsilon_t. \tag{36}$$

In this special case the regulator y_{2t} is no longer correlated with y_{1t}, and we no longer have causation: y_{2t} ; y_{1t} and $y_{1,t-1}$; y_{1t}. This result may seem to run counter to the ordinary-language usage of the term "causality", but it is an unavoidable consequence of the formal definitions.

Under any particular parametrization of a model, imposing the uncorrelatedness requirement is obviously very restrictive. However, imposing uncorrelatedness on some related parametrization of the model amounts only to requiring the modeler to state explicitly what he is or is not willing to assume about causation. This is not an unreasonable requirement if the goal is to arrive at causal conclusions. For example, the modeler who is willing to assume $y_2 \Rightarrow y_1$ would generate y_1 from $y_1 = \gamma y_2 + \xi$, with y_2 and ξ assumed external and uncorrelated. Finally, if the model-builder were unwilling to assume that neither y_1 nor y_2 causes the other, he would specify the parametrization

$$y_1 = \beta y_2 + \varepsilon \tag{37}$$

$$y_2 = \delta y_1 + \eta. \tag{38}$$

Here ε and η are uncorrelated by the uncorrelatedness assumption, but this implies neither $y_1 \Rightarrow y_2$ nor $y_2 \Rightarrow y_1$.

2.2 Causality and the Variation-Free Condition

Assume, as in Subsection 1.2, that vector-valued variables x and y are observed, where x is external and y is internal. Further, define u as a vector of unobserved external variables. It is assumed that the covariance matrix of u is nonsingular, so that the realizations of u are variation-free. Then we can write a linear static model as

$$y = Ay + Bx + u. \tag{39}$$

Barring other restrictions, coefficient a_{ij} is *identified*—its value can be inferred uniquely from x and the distribution of u—if and only if the number of variables excluded from equation i at least equals the number of equations in the system less one (see, for example, Fisher [8]). Coefficient a_{ij} is *underidentified* if there are fewer than this number of exclusion restrictions.

Surprisingly, in such models it turns out that there is a close link between the causal interpretability of coefficients and their identifiability. The basic result is that *coefficient a_{ij} represents the effect of internal variable j on internal variable i, conditional on the other variables in equation i being held constant, if and only if a_{ij} is identified* (Heckman [12]). If, in contrast, a_{ij} is underidentified, so that there are too few exclusions from equation i, then the condition for a conditional causal ordering that y_j be variation-free fails.

Since identifiability guarantees only that coefficients are associated with conditional causation, not simple causation, the envisioned interventions generally involve more than one external variable, and are subject to linear restrictions. We argued above that such interventions are sometimes difficult to interpret. However, in special cases this problem does not occur. For example, consider an equation containing exactly two internal variables, but also all but one of the external variables. This equation is just identified, implying that the variation-free condition is satisfied. Further, the implicit interventions involve only the excluded external variable, so no functional relations among external variables are implicitly assumed. The supply-demand model presented in the next paragraph fits this description.

The supply-demand model analyzed above can be written as

$$q = a_s + a_{sp}p + b_{sw}w \tag{40}$$
$$q = a_d + a_{dp}p + b_{di}i, \tag{41}$$

in which weather w affects supply but is excluded from demand, and vice-versa with income i. Both equations of this model are just identified, so a_{sp} measures the effect of price on quantity holding constant weather, while a_{dp} measures the effect of price on quantity holding constant income. These, of course, are the slopes of the supply and demand functions, respectively.

If i had not been excluded from the supply function, a_{sp} would not measure the effect of variations in p on q, cet. par., since if w and i are held constant no variation in p is possible. Correspondingly, a_{sp} is underidentified.

2.3 Causality and Regression

In LeRoy [18] it was pointed out that, assuming satisfaction of the uncorrelatedness condition, coefficients associated with causation can be estimated by ordinary least

squares; the resulting estimates will be unbiased and consistent.[10] This is so because under the stated restriction variables representing causes are uncorrelated with error terms. It follows that econometric theory can sometimes be used to determine causality: knowledge that ordinary least squares results in bias or inconsistency implies absence of causation. The example given to illustrate this point was the model

$$y_t = \lambda y_{t-1} + \eta_t \tag{42}$$
$$\eta_t = \rho \eta_{t-1} + \varepsilon_t, \tag{43}$$

where the external variables, $\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_1$ and y_0, are assumed to be uncorrelated. When $\rho \neq 0$ we have that y_{t-1} is correlated with η_t, so a least-squares regression of y_t on y_{t-1} will not produce a consistent estimate of λ. From the stated result it follows that y_{t-1} ; y_t.

In response to this, Hoover [14] observed that applying the Koyck transformation results in

$$y_t = (\rho + \lambda) y_{t-1} - \rho\lambda y_{t-2} + \varepsilon_t, \tag{44}$$

in which the parameters $\rho + \lambda$ and $\rho\lambda$ are in fact estimated consistently by ordinary least squares. Further, separate consistent estimates of ρ and λ are easily calculated from the estimated regression coefficients. Hoover appeared to view this result as raising questions about the validity of the inference that failure of ordinary least squares implies nonexistence of causation, although he did not spell out the argument. In fact, the multiple regression of y_t on y_{t-1} and y_{t-2} is different from the univariate regression of y_t on y_{t-1}. The fact that ordinary least squares is valid in (44) suggests[11] that even though we do not have the simple causation $y_{t-1} \Rightarrow y_t$, we might have the conditional causation $y_{t-1} \Rightarrow y_t | y_{t-2}$, and it can be directly verified from the definition of conditional causation that this is the case.

2.4 Granger Causality

A process y_2 *Granger-causes* another process y_1 if lagged values of y_2 predict y_1 conditional on lagged values of y_1. If y_2 fails to Granger-cause y_1 then, according to Granger [10], correlations between the two processes can be taken to represent the causal influence of y_1 on y_2. It shortly was made clear by a number of critics that this conception of causality bore no obvious relation to causality as defined either in

[10] However, least squares estimates will not necessarily be efficient: in the presence of overidentifying restrictions, the uncorrelatedness assumption does not imply that the cause variable is weakly exogenous for the causal coefficient (Engle, Hendry and Richard [6], esp. n. 13).

[11] We use "suggests" rather than "implies" because the converse of the above result—that consistency of ordinary least squares implies causation—is not generally true.

ordinary language or in formal analysis. However, the point was not made as clearly as it might have been because of the lack of a suitable formal definition of causality to compare to Granger causality, or so it appears with hindsight. The definition of causality developed in this paper makes possible a precise comparison with Granger causality.

Suppose that process y_1 and y_2 are generated by

$$y_{1t} = a_{12}y_{2t} + b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + \varepsilon_{1t} \quad (45)$$
$$y_{2t} = a_{21}y_{1t} + b_{21}y_{1,t-1} + b_{22}y_{2,t-1} + \varepsilon_{2t}, \quad (46)$$

where the uncorrelatedness assumption is satisfied (so that the external variables ε_{1t} and ε_{2t} are uncorrelated with each other contemporaneously, with each other's lagged values, and with the initial values y_{10} and y_{20}. This model is underidentified. By inspection the lagged terms are causally prior to the contemporaneous values, conditional on each other: $y_{1,t-1} \Rightarrow y_{1t}|y_{2,t-1}$ and $y_{2,t-1} \Rightarrow y_{1t}|y_{1,t-1}$, but y_{1t} is not causally prior to y_{2t}, either simply or conditionally (because the external sets of these variables are the same, and therefore fail to satisfy the subset condition).[12]

Under the restriction $a_{12} = 0$ the model also generates $y_{1t} \Rightarrow y_{2t}|y_{1,t-1}, y_{2,t-1}$, which gives a precise sense in which the process y_1 is causally prior to the process y_2 and a precise condition under which that causal priority exists. The restriction $a_{12} = 0$, being just-identifying, has no observable implications, and therefore is not testable in the absence of other restrictions.

Granger-causality is defined from the reduced form for the model (45)-(46), written as

$$y_{1t} = c_{11}y_{1,t-1} + c_{12}y_{2,t-1} + \eta_{1t} \quad (47)$$
$$y_{2t} = c_{21}y_{1,t-1} + c_{22}y_{2,t-1} + \eta_{2t}, \quad (48)$$

where, for reference below, we have $c_{12} = (b_{12} + a_{12}b_{22})/(1 - a_{12}a_{21})$. Here y_2 fails to Granger-cause y_1 if $c_{12} = 0$. Plainly $c_{12} = 0$ is neither necessary nor sufficient for $a_{12} = 0$, so Granger-noncausality is neither necessary nor sufficient for $y_{1t} \Rightarrow y_{2t}|y_{1,t-1}, y_{2,t-1}$. Indeed there does not appear to be any sense in which Granger-noncausality implies that y_1 is causally prior to y_2.

It is true that under the maintained restrictions $b_{12} = 0$ and $b_{22} \neq 0$, we have that $c_{12} = 0$ is equivalent to $a_{12} = 0$, so under those maintained restrictions Granger noncausality of y_1 by y_2 is equivalent to $y_{1t} \Rightarrow y_{2t}|y_{1,t-1}, y_{2,t-1}$. However, it is difficult to think of an economic setting in which such restrictions would apply.

[12] Of course, there are similar results with the roles of the variables reversed. Hereafter we do not note these separately.

3 Comparing the Causality Definitions

Assume that we have in hand a model that we are willing to treat as structural (in the sense that it is representable using a causal graph with arrows representing the equals sign). Assume also that the exclusion condition is satisfied (otherwise the Simon definition is inapplicable). Then we have three distinct definitions of causal orderings: (1) that encoded in the graph, (2) that of Simon (which coincides with the definition here of conditional causation, as noted), and (3) the definition of simple causation proposed here. We now show that in general these orderings are different, although they may coincide in special cases.

First we produce an example in which the causal ordering under the definition of Simon is the same as that under the definition offered here, but different from that implied by the associated causal graph. This is the birth-control pills/thrombosis example, which is familiar from the philosophy literature. Taking birth control pills increases the risk of thrombosis, but so does pregnancy. In turn, suppose that the incidence of pregnancy is determined by whether or not a woman takes birth control pills, and also by her rate of sexual activity. In this model pregnancy and thrombosis are internal, while birth control pills and sexual activity are external.

Write the model as

$$t = b_{tb}b + a_{tp}p \tag{49}$$

$$p = a_{pb}b + b_{ps}s. \tag{50}$$

The question is whether pregnancy, an internal variable, causes thrombosis under each of the three definitions. According to the causal graph suggested by the account just given, with the equality symbol directly representing causation, the answer is yes: the structural equation which has the incidence of thrombosis on the left-hand side has pregnancy and birth control pills on the right-hand side, and this is the criterion for causation.

According to Simon's definition, however, the answer is no, and the same is true under the definition of simple causal orderings proposed here (the latter conclusion is immediate because y_1 9 y_2 implies y_1 ; y_2). In terms of Simon's definition, pregnancy is determined in the same subsystem as thrombosis, not a lower-order subsystem. This is so because the exclusion condition fails: each external variable in the exogenous set for p also appears in the exogenous set for t. Therefore pregnancy is not causally prior to thrombosis. In terms of our definition of simple causation, the question "What is the effect of pregnancy on thrombosis?" is inherently ambiguous. The change in pregnancy might be due to a change in either pill use or sexual activity, and these affect the incidence of thrombosis differently. Until we are told why the pregnancy rate changed, there is no way to answer the question.

As this example makes clear, under both Simon's definition of causation and that proposed here, adopting language like "pregnancy is a determinant of thrombosis" to

describe a model informally is not sufficient to justify a formal statement that pregnancy causes thrombosis. One might view this disconnect between an informal usage that is virtually unavoidable and the formal definition as a disadvantage. Perhaps so, but the alternative of using causal language to describe the effect of one variable on another when the envisioned intervention is not adequately characterized does not seem like an improvement.

In other models Simon's definition gives a different causal ordering from the definition of simple causation proposed here. The definition proposed here gives a sparser characterization of causal orderings: whenever y_1 causes y_2 according to our definition, the same is true under Simon's definition (this follows immediately because $X_1 \subset\subset X_2$ immediately implies that y_1 is determined in a lower-order system than y_2). However, the converse is not true. Consider the uninterpreted model

$$\begin{aligned} y_1 &= a_{12}y_2 + b_{11}x_1 && (51)\\ y_2 &= a_{21}y_1 + b_{22}x_2 && (52)\\ y_3 &= a_{31}y_1 + a_{32}y_2 + b_{33}x_3, && (53) \end{aligned}$$

which contains the recursive model

$$\begin{aligned} y_1 &= b_{11}x_1 && (54)\\ y_2 &= a_{21}y_1 + b_{22}x_2 && (55)\\ y_3 &= a_{31}y_1 + a_{32}y_2 + b_{33}x_3 && (56) \end{aligned}$$

as a special case if $a_{12} = 0$. In both these models we have $y_2 \rightarrow y_3$ according to Simon's definition, but not $y_2 \Rightarrow y_3$ according to our definition of simple causation. The fact that $y_2 \rightarrow y_3$ follows from the fact that the subset condition for our definition is satisfied. This, however, does not imply that the sufficiency condition is satisfied, and it is not satisfied in the present case. The variable y_2 is not a sufficient statistic for X_2 for the determination of y_3, since x_1 and x_2 also affect y_3 through y_1.

In these models we have $y_2 \Rightarrow y_3|y_1$. In the recursive model (54)-(56) there is no difficulty determining the intervention associated with this conditional causation: it involves an intervention in x_2. However, in the model (51)-(53) the indicated intervention involves simultaneous changes in x_1 and x_2, with the changes restricted so that y_1 remains constant. If, as this intervention suggests, there exists a functional relation between x_1 and x_2, the model should be altered to incorporate this relation.

4 Comparing with Pearl

As noted in the introduction, Pearl [23] presented an alternative formalization of causality. Like our development, his is based on the Cowles analysis of the 1950s,

particularly that of Simon [26]. Pearl's view is that after a promising beginning during the Cowles years, social scientists lost touch with the idea of structural modeling and failed to develop the original formal analysis of causation. He criticized sharply the tendency of economists—as exemplified in this paper—to interpret the equals sign in (supposedly) structural models as having its usual mathematical meaning, rather than as directly representing causation. He also would reject the assertion in Subsection 1.2 that Simon's analysis of causation is relevant to current practice precisely because it does not depend on the interpretation of structural equations as directly incorporating causation.

Under Pearl's interpretation of a structural model, each structural equation represents a distinct causal law for one of the internal variables. Pearl's method of analyzing interventions is to delete the equation determining the cause variable and setting the value of that variable at a preassigned level. In this paper, in contrast, we have followed the current economics literature in modeling interventions by the straightforward device of simply specifying values for external variables, either directly or indirectly. This is a distinction without a difference when the causal variable is external. When the cause variable is internal, however, Pearl's algorithm can lead to difficulties.

The assumption that it makes sense to delete one or more of the structural equations and replacing the value of the internal variable so determined by a constant without altering the other equations has been termed "modularity".[13] In special cases Pearl's assumption of modularity is satisfied, implying that his algorithm is valid even when the causal variable is internal. For example, modularity for all possible interventions is satisfied if the external sets for the internal variables are disjoint. This property, however, is virtually never satisfied in economic models since each external variable typically affects equilibrium values of more than one internal variable. In fact, it is difficult to think of nontrivial models in any area of research in which the modularity assumption is satisfied (Cartwright [2]). In any case, when modularity is satisfied the resulting causal ordering on internal variables is empty, so causal analysis is rendered trivial.

When modularity fails, Pearl's method of analyzing interventions is valid if the variable Pearl treats as a cause is in fact causally prior to the effect variable in the sense defined in this paper. If not, however, replacing the equation determining the purported cause variable with direct determination of the equilibrium value of that variable amounts to changing the model so as to render an inherently ambiguous question—what is the effect of one variable on another?—unambiguous. There is no reason to expect that causal analysis based on the altered model has any relevance for the original model.

To get a clearer idea of the problems Pearl's algorithm entails when the requisite

[13] This term was used by Cartwright and Reiss [3], whose criticism of Pearl is similar to that presented here.

causal ordering fails, we consider Pearl's application in a supply-demand model like those analyzed above. Pearl (p. 215) wrote the model as follows:

$$q = a_{qp}p + b_{qi}i \tag{57}$$

$$p = a_{pq}q + b_{pw}w, \tag{58}$$

where we have deleted the error terms since they play no role in the analysis.[14] Here (57) is a structural demand equation and (58) is a structural supply equation. As before, i is income and w is weather. These external variables enter the demand equation and the supply equation, respectively. This model conforms to Pearl's interpretation of the equations of structural models as representing distinct causal laws, one for each internal variable; here price causes quantity in the demand equation, whereas quantity causes price in the supply equation. Economists will be puzzled by this asymmetric modeling of supply and demand; however, some such specification is required under Pearl's characterization of structural models.

Pearl noted that three queries can be distinguished:

1. What is the expected value of "the demand q" [quotation marks supplied] if the price is controlled at $p = p_0$?

2. What is the expected value of "the demand q" if the price is reported to be $p = p_0$?

3. Given that the current price is $p = p_0$, what would be the expected value of "the demand q" if we were to control the price at $p = p_1$?

Observe the syntax here: despite Pearl's terminology, the symbol q refers to (equilibrium) quantity, which equals quantity demanded and quantity supplied equivalently, as above. Pearl's use of the phrase "the demand q" reflects his specification that quantity is determined by price in the demand equation (57). In contrast, price determines quantity in the supply equation (58). One wonders whether Pearl would accept the question "What is the expected value of 'the supply q' if the price is controlled at $p = p_0$?" as being equivalent to question 1, on the grounds that quantity demand equals quantity supplied in equilibrium, or whether instead he would regard that question as inapplicable in the system (57)-(58), in which supply quantity is a cause of price, not an effect.

To see that Pearl's method of modeling interventions causes problems when the requisite causal ordering (in the sense of this paper) fails, note that under Pearl's algorithm an intervention in p is necessarily associated with a shift in demand, while an assumed intervention on q would necessarily be identified with a supply shift. But there is nothing in the structure of the model that justifies either of these associations.

[14] Neuberg [22] also discussed this example in his excellent review of Pearl's book.

In a footnote Pearl reported that he has presented this model and these questions to well over one hundred econometrics students and faculty. He found that the respondents had no trouble answering 2, but only one person could solve 1, and none could solve 3. With the exception of the one respondent who could answer question 1 to Pearl's satisfaction, this is exactly the response pattern that one would hope for based on the analysis of this paper: if the unsatisfactory phrase "the demand q" is replaced by simply "q", the correct response to questions 1 and 3 is that they are ambiguous because price does not cause quantity in the system (57)-(58).

Under Pearl's algorithm, however, questions 1 and 3 are not ambiguous. They are answered by deleting (58) and replacing it with the equation $p = p_0$ or $p = p_1$, respectively. Thus the relevant causal parameter is a_{qp}; the supply elasticity a_{pq} by assumption plays no role.

These difficulties arise because, as we have seen, Pearl's representation of an economic model differs in key respects from the representation of a model which most economists would feel comfortable working with. In contrast to Pearl's view, we have argued that defining "structural" models as models that directly encode causal ideas is a dead end.

5 Graphical Methods

We noted in the introduction that graphical analysis of causation has seen few applications in economics. One suspects that the reason is that most economists have difficulty following the logic underlying graphical analyses of causation, at least as set out in most expositions (Pearl [23], Pearl and Verma [24] and Geiger, Verma and Pearl [9] are good examples). In these developments no attempt is made to derive causal orderings from deeper properties of models relating to exogeneity, as Simon did and as we have done here.[15] Even though the notion of causality presumed in graphical analysis is not reduced to any deeper idea, we are still presented with methods for deriving graphs from patterns of conditional independence among internal variables. For example, if one observes $cov(x_1, x_3|x_2) = 0$, we are told that one can infer either $y_1 \Rightarrow y_2 \Rightarrow y_3$ or $y_3 \Rightarrow y_2 \Rightarrow y_1$. The validity of this inference is apparently taken to be self-evident: when x_1 influences x_3 through the mediation of x_2, if we condition on x_2, should not x_1 and x_3 be uncorrelated? However appealing this intuitive argument is, one wonders whether the proposition cannot be demonstrated formally. But taking graphs as primitive seems to rule out this possibility.

In this paper we have derived causal orderings from underlying assumptions about exogeneity. Therefore rather than attempting to construct graphs from statistical independence relations as in the papers just cited, we would derive them from assumed

[15] Despite this, it is clear that the notion of causality underlying graphical methods is not that of Simon: as we have seen, Simon's definition of causality precludes cyclicality in causation, because of the role of the subset condition. However, cyclicality is not ruled out, at least as a logical possibility, in graphical analyses, even if attention usually centers on acyclic graphs.

model specifications, and then inquire whether causal orderings in fact imply the conditional independence relations assumed in graph theory. It turns out that this exercise is straightforward when the model under examination has only simple causality. We show this in the next subsection. Conditional causality raises substantial difficulties, as shown in the following subsection

5.1 Simple Causality

We begin by considering models in which $y_i \Rightarrow y_j$ whenever $y_i \rightarrow y_j$. We construct a graph from a model as follows: first, connect by an arrow each internal variable with each external variable in that variable's external set, except when there exists another internal variable that causes the internal variable under consideration. Second, connect each pair of internal variables in the causal ordering with an arrow, except when there exists another internal variable that is an effect of the cause variable and a cause of the effect variable. This produces a directed acyclic graph.

As observed above, without adopting the uncorrelatedness assumption there is obviously no way to make inferences about causal orderings, either unconditional or conditional, from statistical uncorrelatedness, or vice-versa. However, if we adopt the uncorrelatedness assumption, it turns out that the statistical independences of graph theory are derivable consequences of causal orderings as defined here. For example, suppose we have a 3-variable system parametrized as

$$y = Ay + \varepsilon, \tag{59}$$

where ε is a triple of zero-mean external—and therefore uncorrelated, by the uncorrelatedness assumption—random variables and A is a 3×3 matrix with zeros on the main diagonal. It is easily verified that the internal variables are causally ordered as $y_1 \Rightarrow y_2 \Rightarrow y_3$ if A is nonzero only in the 2,1 and 3,2 places. Assume that this is the case. Then it is easy to compute the covariance matrix of y as a function of the variances of the ε_i and the terms of A. It is immediately verified that $cov(x_1, x_3|x_2) = \sigma_{13} - \sigma_{12}\sigma_{23}/\sigma_{22} = 0$. Thus we have a testable implication of an assumed causal ordering. A similar derivation justifies the other conditional independences of graph theory as consequences of causal orderings.

As an example, return to the price-quantity model (17)-(19)

$$q = a_s + b_{sw}w + b_{sf}f \tag{60}$$
$$q = a_d + a_{dp}p + b_{di}i \tag{61}$$

the graph so constructed is shown as Figure 3. In graph-theoretic terms (Pearl [23], for example), the chain $w \Rightarrow q \Rightarrow p$ (or, equivalently, $f \Rightarrow q \Rightarrow p$) is blocked by q. This means that all the causal influence of weather on price passes through quantity. If the external variables weather, fertilizer and income are uncorrelated, then it may be directly verified that weather and price are uncorrelated, conditional on quantity

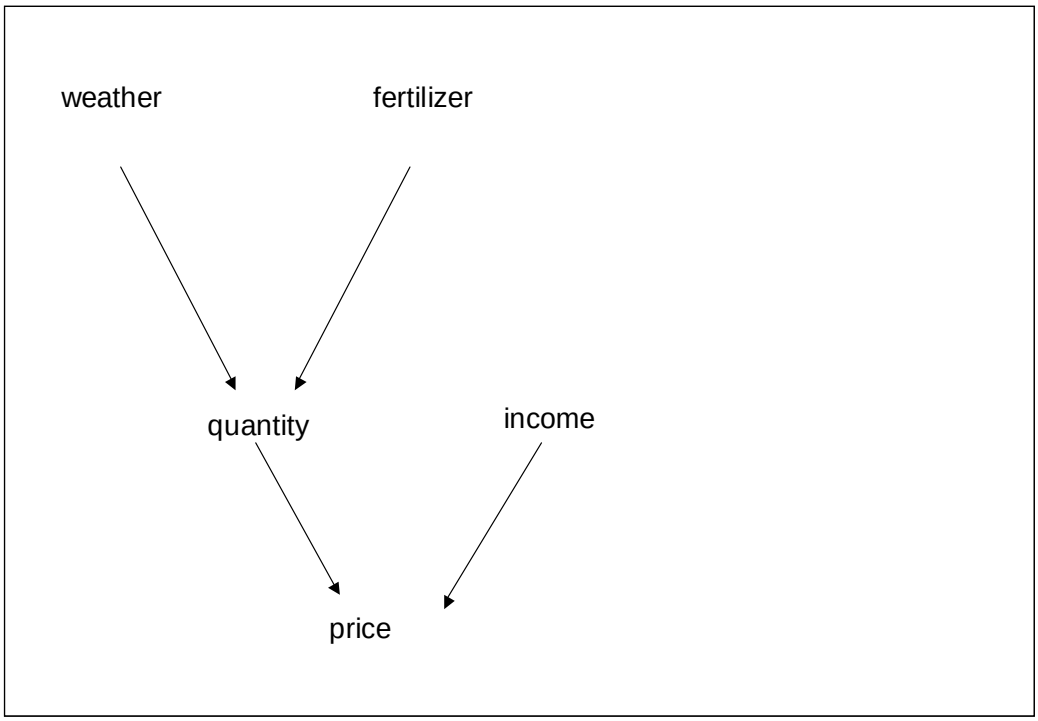


Figure 3:

(even though they are necessarily correlated unconditionally). This is a testable implication of the assumed model.

Causal orderings produce testable implications only when they are overidentifying. The causal ordering $y_1 \Rightarrow y_2 \Rightarrow y_3$ is overidentifying, as we just saw. However, other causal orderings may not be overidentifying, and these will not have testable implications. For example, in 3-variable systems the Wold ordering $y_1 \Rightarrow y_2$ and $y_2 \Rightarrow y_3|y_1$ is produced by a triangular coefficient matrix with the variables ordered as y_1, y_2, y_3. This causal order is just-identifying, as is evident from the fact that any coefficient matrix can be reparametrized in triangular form by applying the Choleski decomposition. Hence the Wold causal ordering cannot be tested.

From the foregoing development, there is no doubt that graph-theoretic methods have a role to play in testing overidentifying restrictions implied by causal models. However, proponents of graph-theoretical methods go farther than this: they assert that computer-based methods can be used to identify causal structures (for example, see Pearl [23] or Spirtes, Glymour and Scheines [27]. The essential assumption that is held to underly such exercises is that of "faithfulness" (Spirtes et al.) or, in Pearl's terminology, "stability". The idea is to rule out parameter values that imply extraneous zero conditional covariances—that is, zero conditional covariances that are not produced by causal orderings—so as to be able to reverse the chain of

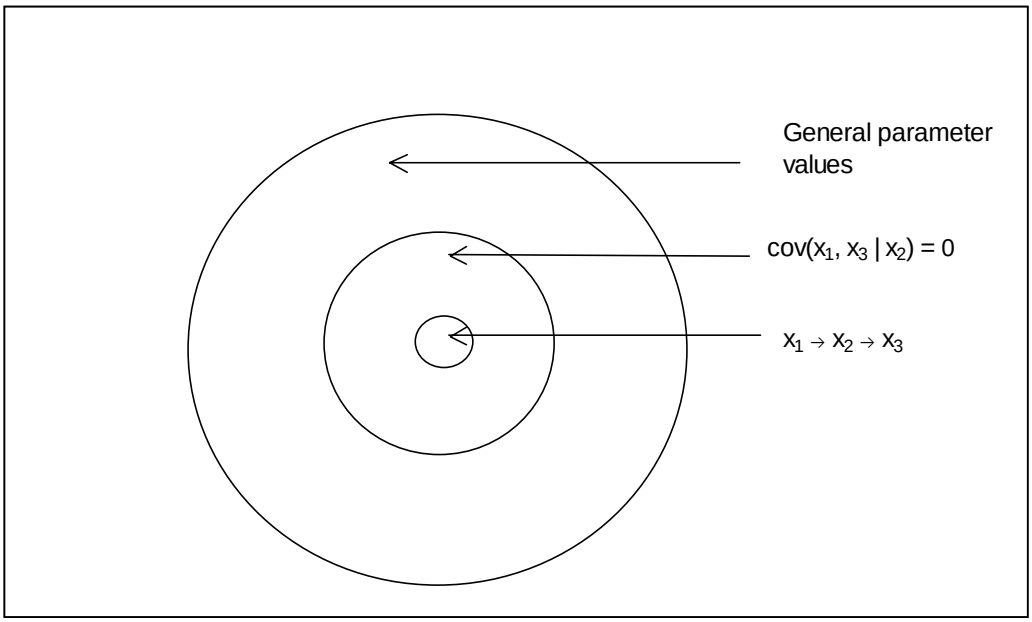


Figure 4:

reasoning and treat zero conditional covariances as evidence in favor of the causal orderings that produce them.

Proponents of graphical methods view the assumption of faithfulness as innocuous. Spirtes et al. assert (p. 41) that all they are doing is ruling out a set of Lebesgue measure zero. To see the basis for this argument, consider the three-variable example (59) analyzed above. The parameter space is 9-dimensional (A has 6 off-diagonal elements, and there are 3 error variances). Generically, all the covariances, both unconditional and conditional, of the variables in this space are nonzero. However, an 8-dimensional manifold of the parameter space will satisfy the single restriction $cov(y_1, y_3|y_2) = 0$; call this region B. Further, a 5-dimensional manifold of the parameter space will satisfy $y_1 \Rightarrow y_2 \Rightarrow y_3$; call this region C. As noted above, C is the set of parameter values such that A is zero except in the $(2,1)$ and $(3,2)$ elements. Since $y_1 \Rightarrow y_2 \Rightarrow y_3$ implies $cov(y_1, y_3|y_2) = 0$, we have that C is a subset of B (see Figure 4).

The assumption of faithfulness amounts to assuming that $C - B$ is the empty set, so that the above reasoning can be reversed, and a finding of $cov(y_1, y_3|y_2) = 0$ can be interpreted as implying $y_1 \Rightarrow y_2 \Rightarrow y_3$. The fact that the 8-dimensional manifold B has zero Lebesgue measure relative to the 9-dimensional parameter space is the basis for Spirtes et al.'s representation of the faithfulness assumption as innocuous.

However, there is another way to view the faithfulness assumption. Essentially, the role of faithfulness is to generate the conclusion that if a parameter set is in the 8-dimensional manifold B, it is necessarily in the 5-dimensional manifold C. Therefore

an empirical finding that $cov(y_1, y_3|y_2) = 0$ allows us to conclude that $y_1 \Rightarrow y_2 \Rightarrow y_3$ (or, of course, $y_1 \Leftarrow y_2 \Leftarrow y_3$; it is not asserted that graph theory can determine the direction of causation in this case). Since C is of zero Lebesgue measure relative to B, from the present vantage the assumption of faithfulness amounts to restricting attention to a set of zero Lebesgue measure, not ruling out a set of zero measure. Doing so is anything but innocuous.

Geiger, Verma and Pearl [9] defended the faithfulness assumption on grounds of Occam's razor: the simplest model that explains the empirical facts is to be chosen. However, it is not clear that models with nonempty causal orderings are preferable on grounds of simplicity to models with empty causal orderings. Thus we are left with the conclusion that algorithms that purport to determine causal orderings depend on the validity of strong and probably unrealistic assumptions. Attributions of causality require strong prior theoretical inputs; whether these are sought from preexisting theory, as the Cowles economists envisioned, or from an assumption of faithfulness, as proponents of graphical methods prefer, may be a matter of taste.

5.2 Conditional Causality

These results, weak as they are, do not extend to the case of conditional causality, and this broader concept is the relevant one in all but the simplest models. To see this, consider the model

$$y_1 = b_{11}x_1 + b_{12}x_2 \tag{62}$$

$$y_2 = a_{21}y_1 + b_{21}x_1 + b_{13}x_3, \tag{63}$$

introduced in note 6. The causal ordering for this model is $y_1 \Rightarrow y_2|x_1$ and $y_1 \Rightarrow y_2|x_2$. (However, we do not have $y_1 \Rightarrow y_2|x_1, x_2$; that would violate the variation-free condition.) Thus in the causal ordering x_1 and x_2 appear symmetrically. However, their roles are different in the model: x_1 affects y_2 through its effect on y_1 and also directly, whereas x_2 affects y_2 only through its effect on y_1. Because x_1 and x_2 play different roles in the model, the conditional correlations involving them will be different. Therefore any graph based on the causal ordering will not uniquely imply the conditional correlations implied by (62) and (63).

It follows also that any graph of the above model based on the causal ordering will coincide with the graph of a model in which the term $b_{21}x_1$ is replaced by $b_{22}x_2$. Plainly we cannot use graphs to determine the observable implications of causal orderings if the same graph can represent substantively different models.

There are ways to avoid these problems, but they involve substantial modifications of the definition of conditional causality, and are not considered here. Pending such modification, the general conclusion must be that graphical methods are not likely to play much of a role in testing or determining causal orderings under the characterization of causality adopted here.

6 Conclusion

We have pointed out that formal analyses of causation fall into two groups: those in which the idea of causation is derivable from exogeneity (that of Simon and that proposed here) and those in which structural equations are taken to connote causation in and of themselves. We have expressed the view that the former characterization of causation is better suited to current economic models. The definition of simple causation proposed here has a particularly straightforward interpretation: it is applicable whenever one wishes to know whether an intervention characterized in terms of a hypothetical variation in an internal variable has an unambiguous effect on some other variable.

Other meanings for causation are less easy to interpret. We noted that when one variable causes another according to Simon's definition but not according to the definition proposed here, the intervention usually involves joint restrictions on the external variables. This, we suggested, appears to conflict with their exogeneity. Graphical analyses along the lines developed by Pearl and others involve still more serious problems: they require a model of a fundamentally different sort than those economists use. However, we showed in the preceding section that graphical methods are applicable under the characterization of causation proposed here, and appear to offer promise in identifying testable implications of causality specifications.

References

[1] Nancy Cartwright. Probabilities and experiments. *Journal of Econometrics*, 67:47–59, 1995.

[2] Nancy Cartwright. Measuring causes: Invariance, modularity and the causal Markov condition. reproduced, London School of Economics, undated.

[3] Nancy Cartwright and Julian Reiss. Uncertainty in economics: Evaluating policy counterfactuals. reproduced, CPNSS, London School of Economics, 2003.

[4] A. P. Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95:407–424, 2000.

[5] Gerard Debreu. *Theory of Value*. Wiley, New York, 1959.

[6] Robert F. Engle, David F. Hendry, and Jean-Francois Richard. Exogeneity. *Econometrica*, 51:277–304, 1983.

[7] Damien J. Fennell. Comments on Stephen LeRoy's 'Causality in Economics'. reproduced, London School of Economics, 2004.

[8] Franklin M. Fisher. *The Identification Problem in Econometrics*. Krieger, New York, 1976.

[9] D. Geiger, T. S. Verma, and J. Pearl. Identifying independence in bayesian networks. *Networks*, 20:507–534, 1990.

[10] Clive W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[11] Daniel M. Hausman. *Causal Asymmetries*. Cambridge U. P., Cambridge, 1998.

[12] James J. Heckman. The scientific model of causality. Reproduced, University of Chicago, 2004.

[13] Kevin D. Hoover. The causal direction between money and prices. *Journal of Monetary Economics*, 27:381–423, 1991.

[14] Kevin D. Hoover. *Causality in Macroeconomics*. Cambridge University Press, Cambridge, 2001.

[15] Kevin D. Hoover. *The Methodology of Empirical Macroeconomics*. Cambridge University Press, Cambridge, 2001.

[16] Kevin D. Hoover. Nonstationary time series, cointegration, and the principle of the common cause. *British Journal of the Philosophy of Science*, 54:527–551, 2003.

[17] Edward E. Leamer. Vector autoregressions for causal inference? volume 22. Carnegie-Rochester Conference Series on Public Policy, 1985.

[18] Stephen F. LeRoy. Causal orderings. In Kevin D. Hoover, editor, *Macroeconometrics: Developments, Tensions and Prospects*. Kluwer Academic Publishers, 1995.

[19] Stephen F. LeRoy. On policy regimes. In Kevin D. Hoover, editor, *Macroeconometrics: Developments, Tensions and Prospects*. Kluwer Academic Publishers, 1995.

[20] Robert E. Lucas. Econometric policy evaluation: A critique. In Karl Brunner and Allan H. Meltzer, editors, *Carnegie-Rochester Conference Series on Public Policy*. North-Holland, 1976.

[21] Jacob Marschak. Economic measurement for policy and prediction. In William C. Hood and Tjalling C. Koopmans, editors, *Studies in Econometric Method*. John Wiley and Sons, Inc., 1953.

[22] Leland Gerson Neuberg. Review of 'Causality: Models, reasoning and inference'. *Econometric Theory*, 19:675–685, 2003.

[23] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, 2000.

[24] Judea Pearl and T. S. Verma. A statistical semantics for causation. reproduced, University of California, Los Angeles, undated.

[25] Hans Reichenbach. *The Direction of Time*. University of California Press, Berkeley, 1956.

[26] Herbert A. Simon. Causal ordering and identifiability. In William C. Hood and Tjalling C. Koopmans, editors, *Studies in Econometric Method*. John Wiley and Sons, Inc., 1953.

[27] Peter Spirtes, Clark Glymour, and R. Schienes. *Causation, Prediction and Search*. Springer-Verlag, New York, 1993.

[28] Norman R. Swanson and Clive W. J. Granger. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association*, 92:357–367, 1997.