Causality in Economics*

Stephen F. LeRoy
University of California, Santa Barbara

October 3, 2006

The initiation of quantitative analysis of formal structural models in the social sciences is generally attributed to the Cowles Commission economists in the 1950s (see Hood and Koopmans [12] for a collection of some of the most important papers). Since then the new methods have been extended and refined and applied in the other social sciences. Graphical methods based on the Cowles developments are increasingly used in the social and biological sciences. The theme of this paper is that recent purported refinements of the Cowles analysis, particularly as it relates to representations of causal relations in formal models, departed from the Cowles analysis to a greater extent than is commonly realized. The suggestion is that the Cowles development as interpreted here compares favorably to more recent developments.

1 Causation

The term "structural" was given several distinct, though related, meanings by the Cowles economists, 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 is 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 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.

A second meaning of "structural", one more basic than that discussed above, has to do with invariance under intervention. The Cowles economists implicitly distinguished two classes of hypothetical experiments: (1) determining the effects of

*First draft August 25, 1999. I have received helpful comments from Nancy Cartwright, Daniel Hausman, Damien Fennell, Julian Reiss and two referees..

routine realizations of external variables, and (2) analyzing changes in model structure. There is no formal justification for this distinction since hypothetical model shifts can always be parametrized, allowing the analyst to represent an intervention on the structure of a model via a hypothetical shift in an external variable. Thus in a model that is structural in this sense coefficients can be treated as if they were external variables: an intervention on one or several model coefficients by definition leaves other coefficients unchanged. The Cowles economists were often unclear as to when they were treating coefficients as external variables and when they were treating them as constants.

This conception of structural models has left its tracks in current practice in macroeconomics: the Lucas critique, with its assertion that policy change is properly modeled as an intervention on parameters rather than variables, is one example. The distinction between deep and shallow parameters is another.

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 [22] observed (Pearl's work is discussed below). This is distinguished from its meaning in mathematics, under which equality 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, economists have been conspicuous in their absence from these developments in recent years. It is true that the topic of causation has been of passing interest to economists—witness Granger causality—but there is essentially no connection between the lines of inquiry that economists have pursued and 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.

These issues require more careful examination. We begin with Simon, whose analysis of causation is in fact completely different from that implicit in the graphical

[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. This is clearly so with Pearl [22]. In a supply-demand model Heckman [11] 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. Pearl and Heckman's work is discussed below.

treatments. Contrary to the implication in the preceding paragraph and the presumption in many expositions of graphical analysis of causation, Simon's definition of causal orderings does not require a model that is structural in the sense that the equals sign is interpreted as causal.

1.1 Notation

One problem that the reader of the literature on causation 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.[2]

One essential distinction, stressed in logic and mathematics but often blurred in philosophical discourse and economic analysis, is that between variables and constants.[3] A *variable* is the argument or value of a function, while a *constant* is a standin for a number. By specifying that θ is a constant, the analyst stipulates that it makes no sense to consider alternative values of θ—doing so makes no more sense than asking what would happen in mathematics if π took on a 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. 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.[4] Since causation is the subject here, we will assume that the analyst is willing to make this categorization.

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 sometimes use the term "exogenous variable" with a different meaning: roughly, an exogenous variable is an observed variable of which an unobserved equation error is independent or mean-independent. Leamer [17] itemized the various meanings of "exogenous" and "endogenous" in the economics and econometrics literature. We follow his recommendation that the terms "external variable" and "internal variable" be substituted to avoid confusion.

It is assumed that all interventions on a model—that is, all hypothetical experiments involving the model—can be characterized, either explicitly or implicitly, as interventions on the external variables. Direct interventions on the internal variables

[2] For example, Hoover [15], p. 171, referred to the paper preceding this one (LeRoy [18]), which has substantially the same terminology, as offering a "complex and difficult terminological landscape".

[3] As observed above, the Cowles economists were sometimes unclear as to whether model coefficients were to be interpreted as constants or external parameters.

[4] See Shafer [24] for an extended formal discussion of causation that dispenses almost completely with the distinction between external and internal variables.

are inadmissible precisely because these variables are determined by the external variables. Thus hypothesized interventions on internal variables must be implicitly attributed to interventions on the external variables that determine them. Our analysis of causal orderings is predicated on this characterization of interventions.

It is assumed that external variables satisfy the "variation-free" condition: the external variables can be intervened on independently. If this condition fails, the interpretation is that there exists a functional relation linking the external variables. If so, that relation should be included in the model. The variation-free condition asserts that the model is "invariant under intervention", to use a phrase favored by the Cowles economists: a hypothesized change in one external variable leaves the other external variables unchanged, and does not otherwise alter the structure of the model.

In multidate models we distinguish between parameters and processes. A *parameter* is a variable which one wants to distinguish from a constant, 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.

Observe that in this usage the term "variable" is used in static but not dynamic models, while the opposite is the case for the terms "parameter" and "process". In contrast, in much applied work involving static models the term "parameter" is used with the same meaning as "variable" under the above definition. Also, in multidate models the term "variable" is often used with the meaning of "process" as defined above. Generally there is no harm in either of these usages, but in the present context they would cause confusion: in static models variables as defined here are the same as parameters, so one of these terms should be deleted, while in multidate models we have already defined processes as collections of variables.

In some discussions the term "parameter" is used with a different meaning. For example, Hoover [15] 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 (or an external process). 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 same practice is followed in many other sources (Engle, Hendry and Richard [4] is an example). 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 observed by the analyst, or unobserved. External unobserved variables will be assigned probability distributions, and these will induce probability distributions on internal variables, both observed and unobserved (assuming that models have unique solutions, as we do throughout this paper). We will assume that constants are unobserved. In writing down uninterpreted models, the following notation is adopted:

external observed variables or processes	x
internal observed variables or processes	y
unobserved constants	$a, b, A, B, C, D, \alpha, \beta$
unobserved external variables or processes	u
external parameters	θ
internal parameters	ψ

(with interpreted models it is sometimes easier to depart from this notation so as to use variable names that evoke the meaning of the variable). Note that this classification is incomplete; other possibilities, such as internal unobserved processes (latent variables, in some characterizations) are deleted because they are not considered in this paper.

Below we will be considering the consequences of classifying terms as constants or coefficients. Now, as a logical matter we should be specifying a name for an entity that could be either a variable or a constant ("term" and "entity" are unsatisfactory), and we should also add a new symbol so as to avoid the need to write a or x. However, expanding the notation in this way is obviously unweildy. In dealing with models that are linear in variables we can use the term "coefficient", so as to be able to state that a linear model becomes bilinear if coeffients are treated as variables rather than constants. In some contexts below we will use the term "coefficient" in this sense even when the context does not require the limitation to linear/bilinear settings. No confusion should result.

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, into R^n, a space of internal variables. This operator is assumed to be representable by an $n \times m$ 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 [25] for definitions of these terms). For example, in the model

$$\begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 && (4) \\ y_2 &= a_{21}y_1 + b_{23}x_3, && (5) \end{aligned}$$

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. 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.

In identifying the model (4)-(5) with the causal ordering $y_1 \rightarrow y_2$ we are implicitly assigning generic values to the coefficients. In special cases (for example, in the case $a_{21} = 0$ above) y_1 does not cause y_2. For reference below, note that the assumption that the coefficients are nonzero is not sufficient to assure the uniqueness of causal orderings; we will see that if coefficients obey certain restrictions, but restrictions that do not involve zero values, causal orderings are altered. Since these restrictions are nongeneric, assuming genericity assures uniqueness. This is demonstrated below. This qualification is not repeated below, but it is assumed.

There is a well-known difficulty with the above account of causation: 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.

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*.[5] 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).

Satisfaction of the exclusion condition requires that C have rank n. 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 under the assumed condition C is nonsingular, so (1) can be solved to yield

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

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

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

Since D 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 D is defined by $D = I - C_1^{-1}$, (10) can be written as

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

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 [5] 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

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

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

[5]The exclusion condition is essentially the same as Hausman's *independence condition* ([9], p. 64). See also Hoover [14], p. 103 ff.

in which y_1 and y_2 are not causally ordered. The exclusion condition is satisfied by the presence of x_1 in (12) and x_3 in (13). However, if (13) is solved for x_2 and the result is substituted in (12), we obtain

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

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

where

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

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

Despite this, causal orderings are unique generically: in (14)-(15) we have $y_2 \rightarrow y_1$, but that version of the model is nongeneric because of the restriction (16). Since we have already ruled out nongeneric special cases, it is seen that Fennell's observation about nonuniqueness of causal orderings when $m > n$ does not involve anything new.

It is noteworthy that assuming that a model satisfies the exclusion condition is weaker than assuming that it is structural in the sense that the equality symbol is asymmetric: imposition of the exclusion condition allows renormalization of individual equations (i.e., expressing 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. Causal orderings as Simon defined them are not altered by such renormalizations. In contrast, under the causality definition based on the asymmetric interpretation of the equality symbol, renormalizations of individual equations result in a different model with a different causal ordering. That Simon's definition of causation does not rely on an unconventional interpretation of the equality symbol is an attractive feature of his treatment.

The foregoing discussion is very close to Simon's development. 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 marks 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, so that they are not subject to intervention, implying 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 (a bilinear model is considered briefly in Sec. 1.5).

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.

Our simplification (relative to Simon) of treating coefficients as constants rather than external variables 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 variables.[6] Assuming that the model is linear in variables (a consequence of treating the coefficients in Simon's model as constants) limits the direct applicability of the analysis; contemporary models are likely to be nonlinear. Again, however, the analysis can

[6] However, the converse is not true: changing an external variable to a constant reduces the set of possible interventions, implying that y_1 may no longer cause y_2.

be extended to the general case. The principal difference between the analysis of causation in linear vs. nonlinear models is that in the latter case causal orderings are no longer associated with constants measuring the strength of causal effects: in general the magnitude of a given change in the cause variable on the effect variable depends on the values of all external variables.

The theme of this paper is that Simon's analysis of causation differs in major respects from more recent treatments. At this stage we point out some of the distinguishing features of Simon's treatment. First, Simon made clear that he was analyzing causation in the context of a formal model, not causation as it applies directly to reality or perceived reality. In contrast, in virtually all discussions in the philosophy literature, and in some in the economics literature, causation is discussed as a direct feature of reality. Second, under Simon's definition causality is not a matter of how a model is interpreted or applied: rather, the causal ordering implied by a model can be inferred unambiguously from its formal structure. Third, under Simon's treatment models are written in the usual form as maps from external to internal variables. As noted above, under some alternative treatments of causation the equals sign is interpreted as asymmetric, with cause variables on the right-hand side and effect variables on the left-hand side. In all three respects we follow Simon's lead in this paper.

These features of Simon's treatment of causality have the implication that some sentences that are customarily interpreted as causal do not satisfy the formal requirements for causation. For example, consider the statement "I drank too much yesterday (D); as a result I fell asleep while smoking (S), and my doing so caused my home to burn down (B)". Here D is clearly external, and it would be natural to diagram this sentence as "$D \rightarrow S \rightarrow B$". The problem is that the exclusion condition for "$S \rightarrow B$" fails: there does not exist an external variable that affects B but not S. Equivalently, S is determined in the same block of equations as B, implying that S and B are appropriately treated as simultaneous, not causally ordered. It follows that the inference that S caused B is a feature of the interpretation of the model, not its formal structure.

The fact that sometimes we are willing to infer causation in settings where Simon's formal analysis does not justify this inference does not reflect any shortcoming in Simon's treatment. Informal statements of causation, such as that just given, generally presume unstated background conditions ("the fire extinguisher did not work (E), so I could not put out the fire"). Explicit incorporation of variables representing background conditions generally allows satisfaction of the exclusion condition for a causal ordering. In this case E would be included as an external variable that appears in the external set of B but not that of S. Under this modification we would have $S \rightarrow B$[7].

[7] In this example causality is expressed as a relation among events rather than variables. The reader can supply the indicated modification of the formal structure set out above so as to deal with this case.

1.3 Causality as Sufficiency[8]

Part of the reason Simon's characterization of causation is not much used currently is that Simon did not provide a clear explanation of what follows if one variable causes another. 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? What interventions are admissible if y_1 causes y_2, but not otherwise? What is the interpretation of these interventions?

The best way to supply intuitive content to causation is to consider simple examples. We will see that in some 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

$$q_s = a_{sp}p + b_{sw}w \tag{17}$$
$$q_d = a_{dp}p + b_{di}i \tag{18}$$
$$q_s = q_d = q, \tag{19}$$

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 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, and therefore inadmissible, *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.

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

Now consider the model

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

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

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

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 constants. 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*. Further, 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 constant $\gamma_{j,i}$ and a vector of constants $\delta_{j,i}$ such that (23) can be written in the form

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

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 many possible interventions in X_i could have caused the variation in y_i, each possible intervention has the same effect 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 that satisfies the subset condition 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. Then we have *conditional causation*, indicated by $y_i \Rightarrow y_j|y_k$. Still more generally, the conditioning set may include several internal variables rather than just one, and may include one or more of the external variables.

Two conditions are required for conditional causation. First, y_i must be *variation-free:* if 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. For example, if the conditioning set includes all the external variables, no variation in y_i is possible. More precisely, the variation-free condition is satisfied for $y_i \Rightarrow y_j|y_k$ if the model permits independent variation in y_i and y_k without restricting y_j

Second, if we are to have $y_i \Rightarrow y_j|y_k$, the conditioning set must be such as to ensure that all values of the external variables X_i consistent with a given change in y_i and a given level of y_k produce the same change in y_j. This requirement, which corresponds to the sufficiency requirement for simple causation, is needed to avoid ambiguity in the effect of variations in the external set for y_i on y_j.

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. 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 scalar or vector z_k. Thus $y_i \rightarrow y_j$ under the definition just given is equivalent to $X_i \subset\subset X_j$.

If $y_i \rightarrow y_j$, there exists a set of conditioning variables z_k such that $y_i \Rightarrow y_j|z_k$, but that set is not necessarily unique. For example, in the model

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

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

where x_1, x_2 and x_3 are external variables, we have $y_1 \Rightarrow y_2|x_1$, but also $y_1 \Rightarrow y_2|x_2$. In the first case the intervention is on x_2, while in the second case it is on x_1. The coefficient associated with $y_1 \Rightarrow y_2|x_1$ is clearly a_{21}. However, in the case of $y_1 \Rightarrow y_2|x_2$ matters are more complicated: we must distinguish between the direct effect of x_1 on y_2, which has coefficient b_{21}, and its indirect effect. The indirect effect has coefficient a_{21} if the cause variable is identified with y_1, and $a_{21}b_{11}$ if it is identified with x_1.

Conditional causation may raise 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 of such functional relations among external variables appears to conflict with the assumption that the external variables are variation-free. 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.

In other cases this problem does not arise. For example, in the important case when the matrix A in the model (3) is triangular (not just block-triangular), each y_i causes y_j $(i < j)$ conditional on $y_1, y_2, ...y_{i-1}, y_{i+1}, ...y_{j-1}$. However, the indicated intervention involves only one external variable, so there is no violation of the variation-free condition.

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.

1.5 Causality and Parameter Interventions

Up to now we have taken coefficients to be constants rather than variables or parameters. This assumption was for convenience only. Many problems involving causation do not satisfy this restriction. For example, as soon as coefficients are treated as external variables rather than constants, models become bilinear rather than linear. In this subsection we make some observations about the consequences of treating coefficients as parameters in multidate models.

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 whether policy changes in dynamic models are appropriately modeled through interventions on external parameters or on external processes depends on the question being asked: if the intervention is intended to apply to the past as well as the present and future, parameter interventions are indicated. If only the present and/or future is to be affected, process interventions are indicated.

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. To see this, suppose that we have a model that is linear in variables if the coefficients are treated as constants. As observed above, if y_{1t} causes y_{2t} in this setting, then the same remains true when the coefficients are treated as parameters. This is so regardless of the causal ordering among parameters. The only difference between the two cases 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 causation 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 causation, 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.

The last sentence of this passage appears to be incorrect, at least under the implementation of causation analyzed in this paper. Cartwright is certainly correct that determining that a conditional distribution appears to be invariant over time does not constitute a test for causation. However, it is easy to show that if y_1 does cause y_2, then the distribution of y_2 conditional on y_1 will be invariant to interventions on X_1, as Hoover asserted. To see this, consider the model

$$y_1 = x_1 + x_2 u_1 \tag{27}$$

$$y_2 = x_3 + x_4 y_1 + x_5 u_2, \tag{28}$$

where $x_1, ..., x_5$ are external variables, and u_1 and u_2 are independently distributed unobserved external variables. For concreteness we will take u_1 and u_2 to have zero mean and unit variance. Taking x_2 as a parameter rather than a constant allows the analyst to consider interventions on the standard deviation of y_1.

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

$$X_1 = (x_1, x_2, u_1) \tag{29}$$

$$X_2 = (x_1, x_2, x_3, x_4, x_5, u_1, u_2), \tag{30}$$

which satisfy the subset condition $X_1 \subset\subset X_2$. The sufficiency condition for $y_1 \Rightarrow y_2$ is also satisfied. An intervention on x_2 will affect the marginal distribution of y_1, but will not affect $D(y_2|y_1)$.

As Cartwright observed in the passage just quoted, it is easy to imagine models in which interventions on x_2 do affect the parameters of the conditional distribution of y_1. However, in such models by definition the conditions for $y_1 \Rightarrow y_2$ will fail, contrary to the assumption. It appears, then, that Hoover's assertion is correct.

2 Causality and Identification

2.1 Identification and Exclusion Restrictions

We have seen that exclusion restrictions play a central role in determining causal orderings. As is well known, they also play a central role in determining whether the parameters in a model are identified (Fisher [7]). Despite the common role of exclusion restrictions in determining causality and identification, the two are very different notions. Causation is an ordering on internal variables, whereas identification has to do with whether the econometrician can make inferences about parameter values from the (population) distribution of observed variables. Causation can be defined and analyzed without even specifying which variables are observed, and in fact this is exactly what we have done up to now. However, empirical estimation of causal parameters requires assumptions that assure identification, and this is different from causation. Whether parameters are identified depends not just on which variables are excludable from which equations, but also on which variables are observed and what distributional assumptions are imposed on those (external) variables that are not observed.

As just noted, discussion of identifiability requires specifying which variables are observed by the econometrician and which are not. Both internal and external variables can be either observed or unobserved. Probability distributions are assigned to unobserved external variables as part of model specification, and these induce distributions on both observed and unobserved internal variables. Henceforth we will use u to denote unobserved variables, and all other letters to denote observed variables.

To illustrate the difference between causality and identifiability, consider the supply-demand model

$$q = a_1 p + b_1 x_1 + b_2 x_2 + u_1 \tag{31}$$

$$q = a_2 p + b_3 x_1 + b_4 x_2 + u_2, \tag{32}$$

where x_1 and x_2 are external observed shift variables (weather, income and fertilizer played this role above) and a_1, a_2, b_1 and b_2 are constants. As it stands, none of the coefficients in this model are identified: observations on p, q, x_1 and x_2 cannot be used to estimate the coefficients. Under the restriction $b_2 = b_3 = 0$, all coefficients are identified, including a_1 and a_2, the price elasticities of demand and supply. However, these coefficients do not measure the strength of causal relations because p is not causally prior to q. We see that in general identification is an issue in estimating not just constants associated with causation, but also those associated with simultaneous determination of internal variables.

In bringing empirical evidence to bear on evaluating 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 coefficients does not come up because there is no causal coefficient. If the two variables are causally ordered, then the associated coefficient is well defined conceptually, but it is not necessarily identified.

An assumption that plays a central role in assuring identifiability of causal coefficients is that *all unobserved external variables are (statistically) independent of observed external variables and of each other*. Henceforth we will call this the *independence assumption*.[9] The role of this requirement is to force the model-builder to state explicitly what is assumed about causation, rather than burying causation in uninterpreted correlations among external variables.

To understand the role of the independence 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 u_1 and a constant a_{21} such that y_1 and y_2 satisfy

$$y_2 = a_{21} y_1 + u_1, \tag{33}$$

where the specified distribution of u_1 and the chosen value of a_{21} generate the joint probability distribution of y_1 and y_2. If the model-builder is willing to interpret y_1 and u_1 as external variables, then we have $y_1 \Rightarrow y_2$. Here, of course, the assumption that y_1 and u_1 are external variables is not implied by equation (33); it is a separate assumption. An infinite number of pairs of parameters a_{21} and random variables u_1 are available which generate a given distribution of y_1 and y_2, implying that a_{21} is unidentified.

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

$$y_2 = c_{21} y_1 + u_2. \tag{34}$$

Here by construction u_2 is uncorrelated with y_1, and the random variable u_2 and parameter c_{21} are unique. However, in the absence of further assumptions this decomposition of y_2 into y_1 and c_{21} has nothing to do with causation; cause variables can be projected onto effect variables as well as vice-versa.

As is obvious, these two decompositions of y_2 into y_1 and an unobserved random variable are very different operations: the construction of u_1 and the assumption that it is external amount to assuming that $y_1 \Rightarrow y_2$, but the associated coefficient is

[9] The independence assumption is essentially the same as the principle of the common cause, which says that if two variables are correlated, then either one variable causes the other or the two variables have a common cause (Reichenbach [23]). Here, however, independence is interpreted as a formal restriction on models rather than as a philosophical proposition. As such, the suitability of the assumption 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).

unidentified. In contrast, the construction of u_2 guarantees its uncorrelatedness with y_1, but has nothing to do with causation. If we are willing to assume further that $u_1 = u_2$—or, equivalently, that y_1 and u_1 are uncorrelated, or that u_2 is external—we can interpret the coefficient c_{21} of the projection of y_2 on y_1 as a causal coefficient.

One implication of the independence assumption is that if one variable causes another, the two variables are necessarily correlated.[10] In some settings this may seem counterintuitive. Suppose that y_{1t} is generated according to

$$y_{1t} = \theta_{11} y_{1,t-1} + \theta_{12} y_{2t} + u_t, \tag{35}$$

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

$$y_{2t} = \theta_{21} y_{1,t-1}. \tag{36}$$

Then the behavior of y_{1t} follows

$$y_{1t} = (\theta_{11} + \theta_{12}\theta_{21}) y_{1,t-1} + u_t, \tag{37}$$

where the external variables are $u_t, u_{t-1}, ...$ and y_0. The definitions of causation imply that we have $y_{1,t-1} \Rightarrow y_{1t}$ under the (heretofore unstated) assumption that $\theta_{11} + \theta_{12}\theta_{21} \neq 0$.

However, suppose that the regulator is operated by choosing c so as to minimize the unconditional variance of y_t. This results in $\theta_{11} + \theta_{12}\theta_{21} = 0$ or, equivalently, $\theta_{21} = -\theta_{11}/\theta_{12}$, implying that (37) becomes

$$y_{1t} = u_t. \tag{38}$$

In this special case the regulator y_{2t} is no longer correlated with y_{1t}, and we no longer have causation: $y_{1,t-1} \not\Rightarrow y_{1t}$. This result may seem to run counter to the ordinary-language usage of the term "causality", but it is an unavoidable consequence of the definitions.

Under any particular parametrization of a model, imposing the independence requirement is obviously very restrictive. However, imposing independence 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 insofar as the goal is to arrive at causal conclusions. For example, the modeler who is willing to assume $y_2 \Rightarrow y_1$ instead of $y_1 \Rightarrow y_2$ would generate y_1 from $y_1 = a_{12} y_2 + u_3$, with y_2 and u_3 assumed external and independent. Finally, if the model-builder were unwilling to assume that either y_1 or y_2 causes the other, he could specify the parametrization

[10] Of course, the converse is not true: if the external sets for two internal variables have a nonempty intersection, the independence assumption implies that 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.

$$
\begin{aligned}
y_1 &= a_{12} y_2 + u_1 && (39) \\
y_2 &= a_{21} y_1 + u_2. && (40)
\end{aligned}
$$

Here u_1 and u_2 are independent by the independence assumption, but this implies neither $y_1 \Rightarrow y_2$ nor $y_2 \Rightarrow y_1$.

2.2 Causation and Regression

In LeRoy [18] it was pointed out that, under the condition that all unobserved external variables are independently distributed, coefficients associated with causation can be estimated consistently by ordinary least squares. This is so because under the stated restriction variables representing causes are statistically independent of error terms. It follows that econometric theory can sometimes be used to determine causation: knowledge that ordinary least squares results in inconsistency implies absence of (simple) causation. The example given to illustrate this point was the model

$$
\begin{aligned}
y_t &= \theta_1 y_{t-1} + u_{1t} && (41) \\
u_{1t} &= \theta_2 u_{1,t-1} + u_{2t}, && (42)
\end{aligned}
$$

where the unobserved external variables $u_{2t}, u_{2,t-1}, ..., u_{21}, u_{20}$ and y_0 are assumed to be independent. When $\theta_2 \neq 0$ we have that y_{t-1} is correlated with u_{1t}, so a least-squares regression of y_t on y_{t-1} will not produce a consistent estimate of θ_1. From the stated result it follows that $y_{t-1} \not\Rightarrow y_t$. Checking, the subset condition for $y_{t-1} \Rightarrow y_t$ is satisfied, but the sufficiency condition fails: the elements of the external set for y_{t-1} (these are $u_{2,t-1}, ..., u_{20}, y_0$) affect y_t through u_{1t} as well as through y_{t-1}. Therefore we have $y_{t-1} \not\Rightarrow y_t$.

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

$$
y_t = (\theta_1 + \theta_2) y_{t-1} - \theta_1 \theta_2 y_{t-2} + u_{2t}, \tag{43}
$$

in which the parameters $\rho + \lambda$ and $\rho\lambda$ are in fact estimated consistently by ordinary least squares. Further, separate consistent estimates of θ_1 and θ_2 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 point out an error in the reasoning or 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 (43) suggests[11] that even though we do not have the

[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.

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.

3 Alternative Treatments of Causality

As noted in the introduction, many expositions of contemporary macroeconometric practice as it relates to causation give the impression that it is essentially a refinement and extension of the Cowles insights, particularly those of Simon. The essentials, we are told, were taken over as a whole; subsequent developments added precision and amplified details, but did not affect the substance. On the contrary, contemporary discussions are much sketchier than the Cowles treatment (although the Cowles economists are not beyond criticism in this respect), and they generally differ in essential respects from the Cowles treatment. The reader who is persuaded by this assessment will be motivated to understand the differences in the various treatments so as to determine which line offers the best prospect for improving analytical practice. It will shortly be clear that the view here is that to the (considerable) extent that modern practice differs from the Cowles analysis, the differences do not represent clear improvements.

3.1 Sims-Granger

The Cowles Commission economists emphasized that in the absence of other identifying information, causal orderings are not empirically testable: two models with different causal orderings may be observationally equivalent (it remains true that, in conjunction with other restrictions, causal orderings may be overidentifying, hence testable). Subsequently various economists have challenged this dictum, apparently asserting that causation is directly testable. For example, Sims [26], p. 24, wrote:

> When, as econometricians estimating models ordinarily do, someone asserts that a particular variable or group of variables is strictly exogenous in a certain regression, that assertion is, in time series models, testable. "Exogeneity" here is given its standard econometrics textbook definition. Exogeneity tests are thus an easily applied test for specification error, powerful against the alternative that simultaneous-equations bias is present. The usefulness of these specification tests ought not to be controversial....

Despite the claim here that the term "exogeneity" has a standard meaning and the presumption in this literature that this meaning is closely related to that of causation, definitional issues come to the forefront. 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 [8], 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 ordinary language or in formal analysis.[12] However, the analysis was not as sharp as it might have been because of the lack of a suitable formal definition of causation to compare to Granger causality, or so it appears with hindsight. The definition of causation developed in this paper makes possible a precise comparison with Granger causality.

Suppose that processes y_1 and y_2 are generated by

$$y_{1t} = a_{12}y_{2t} + b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + u_{1t} \quad (44)$$

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

where the independence assumption is satisfied (so that the unobserved external variables u_{1t} and u_{2t} are uncorrelated with each other contemporaneously, with their own and each other's lagged values, and with the initial values y_{10} and y_{20}). This model is underidentified.

The reduced form of the model is

$$y_{1t} = c_{11}y_{1,t-1} + c_{12}y_{2,t-1} + u_{3t} \quad (46)$$

$$y_{2t} = c_{21}y_{1,t-1} + c_{22}y_{2,t-1} + u_{4t}. \quad (47)$$

From this it is clear that we have $y_{1,t-1} \Rightarrow y_{1,t}|y_{2,t-1}$, with associated parameter c_{11}. The other parameters c_{12}, c_{21} and c_{22} are associated with similar elements of the causal ordering. These parameters, of course, are identified. Note that these elements of the causal ordering obtain despite the lack of identifiability of the "structural form" (44)-(45), and whether or not either y_1 or y_2 Granger-causes the other.

Granger-causality is defined from the reduced form (46)-(47). We have that y_2 fails to Granger-cause y_1 if $c_{12} = 0$, where $c_{12} = (b_{12} + a_{12}b_{22})/(1 - a_{12}a_{21})$. If $c_{12} = 0$, $y_{1,t-1} \Rightarrow y_{1,t}|y_{2,t-1}$ simplifies to $y_{1,t-1} \Rightarrow y_{1,t}$, so Granger-causality is necessary and sufficient for $y_{1,t-1} \Rightarrow y_{1,t}$. However, we are interested in causation involving y_1 as a cause and y_2 as an effect, so the relation $y_{1,t-1} \Rightarrow y_{1,t}$ is not of much interest.

Under the restriction $a_{12} = 0$ the model 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. The restriction $a_{12} = 0$, being just-identifying, by itself has no observed implications, and therefore is not testable in the absence of other restrictions. In particular, $c_{12} = 0$ is neither necessary nor sufficient for $a_{12} = 0$ ($b_{12} + a_{12}b_{22} = 0$ is neither necessary nor sufficient for $a_{12} = b_{12} = 0$), so Granger-noncausality is neither necessary nor sufficient for $y_{1t} \Rightarrow y_{2t}|y_{1,t-1}, y_{2,t-1}$.

[12] Recently, for example, the point was made in passing by Heckman [10] in reviewing the Cowles contributions to econometric theory. Heckman's paper is discussed below.

Under the restriction $a_{12} = b_{12} = 0$ we have $y_{1t} \Rightarrow y_{2,t+j}|y_{1,t-1},\ j = 1, 2, ...$. Granger-noncausality is necessary for $a_{12} = b_{12} = 0$, so $c_{12} \neq 0$ is evidence against $y_{1t} \Rightarrow y_{2,t+j}|y_{1,t-1}$, subject to the usual caveats about sampling error and the like, under the maintained assumptions of the model. However, Granger noncausality is not sufficient for $a_{12} = b_{12} = 0$, and it is easy to construct theoretical examples with "spurious exogeneity" (Granger-noncausality without $a_{12} = b_{12} = 0$). Sims [26] expressed the view that spurious exogeneity is "unlikely". Since this opinion was rendered in the context of a model written in abstract form (like (44)-(45)), it is clear that he intended this judgment to apply to economic models in general. In conclusion, there are connections between Granger-noncausality and causality, but they are somewhat remote and not easily interpreted.

There remains the point that the causal elements discussed above involve conditional rather than simple causation. As such they involve interventions on subsets of the external variables subject to linear restrictions (see Section 1.4). We observed above that such interventions are not easily reconciled with the assumed variation-free status of the external variables. Thus the question of what economic meaning can be attached to the indicated interventions remains open. This point weakens still further the link between causality and Granger-causality.

3.2 Pearl

Pearl [22] presented an alternative formalization of causality. He viewed his development as based on the Cowles analysis of the 1950s, particularly that of Simon [25], as here. 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. For Pearl the definition of structural models implies that the equals sign denotes causation. He also would reject the assertion in Section 1.2 that Simon's analysis of causation is relevant to current practice precisely because it does not depend on the interpretation of the equality sign 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. For Pearl interventions are analyzed by deleting the equation determining a particular internal variable and setting the value of that internal 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 causal variables.

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 replace 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 on a given equation 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 jettisoning the model in which the question at issue is inherently ambiguous—what is the effect of one variable on another?—in favor of a different model in which that question is unambiguous. There is no reason to presume 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 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:

$$q = a_{qp}p + b_{qi}i \tag{48}$$

$$p = a_{pq}q + b_{pw}w, \tag{49}$$

where we have deleted the error terms since they play no role in the analysis. Here (48) is a structural demand equation and (49) 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 $p = p_0$?

[13] This term was used by Cartwright and Reiss [3], whose criticism of Pearl is similar to that presented here. Fennell [6] argued against Pearl's sweeping assertion that structural models are inherently modular, and in particular against Pearl's attribution of modularity to Simon.

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 (48). In contrast, price determines quantity in the supply equation (49). 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 (48)-(49), in which supply quantity is a cause of price, not an effect.

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 (48)-(49).

Under Pearl's algorithm, however, questions 1 and 3 are not ambiguous. They are answered by deleting (49) 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.

3.3 Heckman

Heckman [10] analyzed causality in the context of the standard supply-demand model, as here. The example Heckman used to analyze causation can be written as follows:

$$q_s = a_{sp}p_s + b_{sw}w \tag{50}$$
$$q_d = a_{dp}p_d + b_{di}i. \tag{51}$$

This structure is similar to (17)-(18) above except that the supply price p_s is distinguished from the demand price p_d. Because the two equations of this model have no common variables, they can be analyzed separately. Heckman did so: he interpreted a_{sp} as measuring the causal effect of p_s on q_s, just as b_{sw} measures the causal effect of w on q_s. The interpretation of the demand function is similar.

In characterizing equilibrium the equations for supply and demand are combined by appending the identities $p_s = p_d = p$ and $q_s = q_d = q$. Note the contrast between this treatment and that proposed here. We have analyzed causation from the system as a whole, whereas Heckman analyzed causation from each equation taken separately. Heckman was explicit about this:

> If prices are fixed outside of the market, say by a government pricing program, we can *hypothetically* vary p_d and p_s to obtain causal effects for (50) and (51) as partial derivatives or as finite differences of prices holding other factors constant (p. 10; emphasis in original).

Under the definitions proposed here, the parameter a_{sp} is not associated with causation because, with $p_s = p_d = p$ and $q_s = q_d = q$ added to the model, p does not cause q. Rather, these variables are determined simultaneously.

In this example Heckman treated p_s and p_d as external variables for the purpose of analyzing causation, even though p is internal when the equilibrium conditions are imposed. This treatment leads to puzzles. For example, suppose the equations are renormalized on prices rather than quantities (as observed in note 1, it is not clear that Heckman would accept the renormalized version of this equation as equivalent to the original version).

. Then would the parameter $1/a_{sp}$ be interpretable as measuring the effect of q_s on p_s? Can a_{sp} measure the effect of p_s on q_s at the same time as $1/a_{sp}$ measures the effect of q_s on p_s, or do we have to choose? Either way, under Heckman's treatment there appears to be no asymmetry involved with causation. Causal orderings are no longer orderings in the mathematical sense. In contrast, the treatment here is fully in the spirit of simultaneous-equations modeling; a supply-demand model with both price and quantity as internal variables is a different animal from a demand (or supply) equation with price taken as external, and one cannot substitute one for the other in analyzing causation.

4 Conclusion

In this paper we have presented a relatively detailed exposition of the received Cowles account of causation in social science models, together with a rationale for the Cowles treatment of causal orderings as stating conditions under which interventions are or are not unambiguous. Also, we have compared the Cowles analysis of causality with more recent discussions, concluding that the more recent discussions differ in essential respects both from the Cowles treatment and from each other. Considering that it is exactly the purpose of social science models to provide a framework for the disciplined analysis of causation, this is not a very satisfactory situation. One hopes that analysts with an interest in philosophical inquiry will try to pull together these various lines of thought in the analysis of causal structure.

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

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

[6] Damien J. Fennell. Causality, mechanisms and modularity: Structural models in econometrics. reproduced, London School of Economics, 2006.

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

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

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

[10] James J. Heckman. Causal parameters and policy analysis in economics: A twentieth century retrospective. National Bureau of Economic Research, 1999.

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

[12] William C. Hood and Tjalling C. Koopmans, editors. *Studies in Econometric Method*. John Wiley and Sons, Inc., 1953.

[13] Kevin D. Hoover. The causal direction between money and prices. *Journal of Monetary Economics*, 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] Judea Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, Cambridge, 2000.

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

[24] Glenn Shafer. *The Art of Causal Conjecture.* Cambridge University Press, Cambridge, 1996.

[25] 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.

[26] Christopher A. Sims. Exogeneity and causal ordering in macroeconomic models. In Christopher A. Sims, editor, *New Methods of Business Cycle Research: Proceedings from a Conference.* Federal Reserve Bank of Minneapolis, 1977.