Implementation-Neutral Causation

Stephen F. LeRoy
University of California, Santa Barbara
August 14, 2014

Abstract

The most basic question one can ask of a model is “According to the model, what is the effect of variable $y_1$ on variable $y_2$?”, where $y_1$ and $y_2$ are two variables determined by the model. Causation is “implementation neutral” when all interventions on external variables that lead to a given change in $y_1$ have the same effect on $y_2$. We derive conditions for implementation neutrality in simple formal models.

Imposition of implementation neutrality implies that many of the most familiar ideas of causal analysis, based on approaches that ignore implementation neutrality, do not apply. For example, in our setting causality is transitive, implying that a cause variable cannot be linked to an effect variable by both a direct path and a distinct indirect path. This result contradicts the presumption to the contrary in many causality papers, according to which direct and indirect causal paths can be distinguished.

From there the discussion turns to empirical aspects of implementation neutrality. Again the results are unexpected. For example, it follows from the fact that causality here is defined from the solution form of a model (rather than its structural form) that if one variable causes another the coefficient representing that causal link is always identified. Finally, as an application of the analysis, the conditions for implementation-neutral causation are compared with those required for Granger noncausation. The required condition is stronger than strict exogeneity of the cause variable, once more contrary to received doctrine.

The most basic question one can ask of a model is “According to the model, what is the effect of an intervention on variable $y_1$ on variable $y_2$?”,
where $y_1$ and $y_2$ are two variables determined by the model.\footnote{Abbreviated versions of this material were presented in LeRoy [13] and [14] (these papers are discussed at some length in Cartwright [1]). Here more detail is supplied, and the packaging is different.} Two answers are possible. The first involves observing that many possible interventions on the model’s external variables could have led to the assumed change in $y_1$, and in general the effects of these interventions on $y_2$ are different. Therefore the question “What is the effect of $y_1$ on $y_2$?” does not have an unambiguous answer: the information given about the intervention—its effect on $y_1$—is insufficient to characterize its effect on $y_2$.

The second answer is that even though the intervention is not completely characterized for the reason just noted, all interventions consistent with the assumed change in $y_1$ may map onto the same change in $y_2$. In that case the question “What is the effect of $y_1$ on $y_2$?” has a well-defined answer. In linear systems, to which our attention will be restricted in this paper, the effect is captured by a single constant, here labeled $a_{21}$. This coefficient gives the effect on $y_2$ of a unit change of $y_1$, regardless of what intervention on the external variables caused the change in $y_1$.

If, as in the second case above, the effect of a change in $y_1$ on $y_2$ is independent of how the change in $y_1$ is implemented—in other words, independent of the specific interventions on the external variables that determine the assumed change in $y_1$—we will say that the effect of $y_1$ on $y_2$ is implementation-neutral, and will write $y_1 \rightarrow y_2$. Thus, as the notation indicates, we incorporate implementation neutrality in the definition of causation. We do this on the grounds that implementation neutrality is implicit in the idea of causation: if implementation neutrality fails one cannot meaningfully characterize the effect of the intervention on $y_2$, that not being uniquely implied by the assumed change in $y_1$.

If implementation neutrality fails one can only analyze the effect on $y_2$ of various specific interventions on the determinants of $y_1$.\footnote{Note that throughout this paper no notational distinction is made between the name of a variable and the values that variable takes on.} Doing so amounts to abandoning the attempt to characterize $y_1$ as a cause of $y_2$, instead shifting the discussion to treating the external variables which determine $y_1$ as the causes of $y_2$.

Restricting the meaning of causation to cases where interventions are implementation neutral is, of course, purely a matter of definition: if one prefers to say instead that $y_1$ causes $y_2$ but only an implementation-specific
account of the magnitude of the effect is available, no harm is done. However, doing so seems misdirected in view of the fact that the causal links are between the external determinants of \( y_1 \) and \( y_2 \), not between \( y_1 \) itself and \( y_2 \).

The question of implementation neutrality is of central importance in applied scientific work: analysts want either to measure the impact of the treatment on the patient or to know that doing so is impossible because the effect of that treatment depends on how it is applied. Restricting the term “causation” to settings where causation is implementation neutral amounts only to rejecting application of the term in settings in which it has no clear meaning.\(^3\)

Use of diagrammatical methods in causal analysis has become widespread in recent years, due to work by Pearl [15], Spirtes, Glymour and Scheines [17], Woodward [19], Hausman [9], Cartwright [1] and others. These authors do not include implementation neutrality in their definition of causation. As we will see, implementation-neutral causation is antisymmetric, so it can be used to define directed acyclic diagrams of the type in common use. Therefore one has the option of imposing implementation neutrality in the derivation of directed acyclic diagrams and comparing the causal diagrams so derived with diagrams obtained under characterizations of causation that do not impose implementation neutrality. We conduct this analysis in the context of simple examples.

1 Characterization of Implementation-Neutral Causation

A distinction that is central in any model that deals with issues of causation is that between internal and external variables. Internal variables are those determined by the model, while external variables are those taken as given; i.e., determined outside the model.\(^4\) We will use \( y \) to denote internal vari-

\(^3\)The opposite has been argued. Pearl [15], p. 136, took the view that linking causation with implementation neutrality “denies any causal reading to most of the structural parameters that economists and social scientists labor to estimate.” On the contrary, if the argument here is accepted attempts by economists and social scientists to estimate parameters associated with causation are hopeless when implementation neutrality fails because in that case these parameters are not well defined.

\(^4\)In the earlier literature the terms “endogenous” and “exogenous” were often used in place of “internal” and “external”. The earlier usage is consistent with the etymology of the
ables and \( x \) to denote external variables. All changes in solution values of internal variables are assumed to be attributable to interventions on external variables, as opposed to alterations of equations. Implementing this attribution requires the analyst to be explicit about which hypothetical alterations in the model are permitted and which are ruled out. Of course, the analyst can always implement a shift on any of the equations of the model simply by specifying that the relevant equation has an external shift variable. In that case the shift variable is a cause of any internal variable that depends on it. Doing so, of course, is not the same as converting one of the internal variables to an external variable, which constitutes an alteration of the model.

External variables are assumed to be variation-free: that is, the analyst is free to alter them independently. Independent variation corresponds to the assumption that by definition external variables are not linked by functional relations; otherwise they would be classified as internal.

The solution form of a model expresses each internal variable as a function of the set of external variables that determine it.\(^5\) We will refer to the set of external variables that determine any internal variable as its external set, and will denote the external set for \( y_i \) as \( \mathcal{E}(y_i) \). We will assume that the external set for any internal variable consists of at least two external variables. Otherwise the internal variable is a rescaling of the external variable (assuming linearity); a model containing an internal variable the external set of which consists of one external variable can be simplified by deleting the internal variable.\(^6\)

There is no difficulty in defining causation when the cause variable is external: \( x_i \) causes \( y_j \) whenever \( x_i \) is in the external set for \( y_j \). In that case, by virtue of linearity, a unique constant \( b_{ji} \) gives the effect of a unit change.

\(^5\)We thus distinguish between the solution form and the reduced form, in which current-date internal variables are expressed as functions of lagged internal variables and external variables. In static models the solution form and reduced form coincide.

\(^6\)Also, we will assume below that internal variables are observable and external variables are not. Allowing equations in which the external set of some internal variable consists of a single variable would raise the question whether that variable is observable or unobservable.
in \(x_i\) on \(y_j\) for any values of the external variables. If \(x_i\) is not in the external set for \(y_j\) the former does not cause the latter.

The ambiguity comes up when the cause variable is internal, since then an assumed change in the cause variable could come from interventions on any or all of the variables in its external set, and in general the effect of the interventions on \(y_j\) is different for each possible set of interventions. This is so even if all the contemplated interventions on external variables are restricted to have the same combined effect on the cause variable. Given this ambiguity, we cannot attribute causation in this case: the intervention is not described with sufficient detail to generate a clear characterization of the effect.

However, consider a special case in which two conditions are satisfied. These conditions involve two internal variables, \(y_i\) and \(y_j\), and their external sets \(E(y_i)\) and \(E(y_j)\). The first is the subset condition, which requires that the external set for \(y_i\) be a proper subset of that of \(y_j\). The subset condition guarantees that any external variable that affects \(y_i\) also affects \(y_j\), but not vice-versa. Hoover [10] in particular emphasized this condition, which assures the antisymmetry of causation.\(^7\)

The second is the sufficiency condition, which states that the map from \(E(y_j)\) to \(y_j\) can be expressed as the sum of two functions. The first function is the composition of a function from \(E(y_i)\) to \(y_i\) and a function from \(y_i\) to \(y_j\), while the second is a function from \(E(y_j)\) to \(y_j\). If such functions exist \(y_i\) is a sufficient statistic for \(E(y_i)\) for the purpose of determining \(y_j\), meaning that for the purpose of determining \(y_j\) an intervention on any or all of the variables in \(E(y_i)\) is adequately characterized by the resulting induced change in \(y_i\).

If the subset and sufficiency conditions are satisfied we can write \(y_i \rightarrow y_j\).

\(^7\)In this paper the subset condition is a condition we impose on models to assure that causation is antisymmetric. Hausman ([9], Ch. 4) had a different take on what we call the subset condition. Hausman’s independence condition states that “if \(a\) causes \(b\) ... then \(b\) has a cause that is distinct from \(a\) and not causally connected to \(a\).” Hausman appears to view the independence condition, not as an assumption in a model, but as a proposition about the world that may or may not be true: “As a metaphysical claim about patterns of lawlike connections found in nature, [the independence condition] seems incredible, and its truth miraculous.” (p. 64).

However, he went on to consider another possible interpretation, that the failure of the independence condition implies only that there may exist lawlike relations in the world that are not specifically causal relations. This is so because causality inherently involves antisymmetry, and antisymmetry may not occur if the independence condition fails. This latter interpretation is closer to the position taken here.
In that case causation is implementation neutral, since any intervention on the elements of $\mathcal{E}(y_i)$ that results in a fixed $\Delta y_i$ will result in the same $\Delta y_j$.

In the *causal form* of a model the equations are written in such a way as to reflect the model’s causal structure. Starting from the solution form of the model, one can readily derive its causal form. First one derives the causal ordering, which consists of determining for each $i$ and $j$ whether or not we have that $y_i$ is a parent of $y_j$. In the causal form of the model the equation for each internal variable $y_j$ that has no internal variables as causal parents coincides with the corresponding equation in the solution form of the model (i.e., consists of a map from $\mathcal{E}(y_i)$ to $y_i$). The causal form for internal variables $y_j$ that have one or more internal variables as causal parents consists of a map from the parent, or from each of the parents, to $y_j$, plus a map to $y_j$ from the elements of $\mathcal{E}(y_j)$ that are not in the external sets of any of the parents of $y_j$. In the linear setting assumed here the equations of the causal form of a model can be written in the form

$$y_j \leftrightarrow a_{ji}y_i + b_{jk}x_k. \tag{1}$$

Here $y_i$ is the (single, in this case) internal variable that causes $y_j$, and $x_k$ is an external variable (again assumed to be a single variable). The cases in which $y_j$ has more than one parent, or $\mathcal{E}(y_j) - \mathcal{E}(y_i)$ contains more than one external variable, are handled by expanding (1) appropriately. Note our substitution of $\leftrightarrow$ for $=$; since causation is irreflexive and antisymmetric it is inappropriate to use the equality relation in writing the causal form of a model.

Observe that the conditions for causation differ from those defining recursive models. Any model can be put in recursive form by using algebraic operations on the model’s equations, but doing so does not imply that internal variables appearing on the right-hand side of the equation cause the variable on the left-hand side, as is required for the causal form. As we have seen, the causal form depends on assumed parameter restrictions on the solution form of the model. It is unique, unlike the recursive form, which depends on the ordering of the internal variables. Below we will expand on the point that causality cannot necessarily be associated with recursivity.

One can represent the causal form of the model by a causal diagram. For variables $y_i$ without internal variables as parents this consists of arrows

---

8External (internal) variable $x_1$ ($y_1$) is an *ancestor* of internal variable $y_2$ if $x_1 \to y_2$ ($y_1 \to y_2$). It is a *parent* of $y_2$ if it is an ancestor and in addition there is no internal variable $y_3$ such that $x_1 \to y_3 \to y_2$ ($y_1 \to y_3 \to y_2$).
drawn to \( y_i \) from each element of \( \mathcal{E}(y_i) \), as in a diagram of the solution form. For variables with internal parents the arrows run to \( y_j \) from the parent(s) of \( y_j \), and also to \( y_j \) from each variable that is an element of the external set of \( y_j \) but is not in the external sets of any of its parents.

Note that under our convention the causal form does not include as arguments internal variables that are ancestors when these are not also parents. To include both a parent and its ancestors in a single causal equation would involve double counting. The corresponding convention applies to causal diagrams: no arrow directly connects variables with their ancestors when these are not parents.

An example will make this clear.

### 1.1 Example

Consider the following model, written in solution form:

\[
\begin{align*}
y_1 &= b_{11}x_1 + b_{12}x_2 \\
y_2 &= b_{21}x_1 + b_{22}x_2 + b_{23}x_3.
\end{align*}
\]

(2) 

(3)

The external sets for \( y_1 \) and \( y_2 \) are \( \mathcal{E}(y_1) = \{x_1, x_2\} \) and \( \mathcal{E}(y_2) = \{x_1, x_2, x_3\} \). The former is a strict subset of the latter, so the subset condition for \( y_1 \rightarrow y_2 \) is satisfied. Without parameter restrictions the sufficiency condition for \( y_1 \rightarrow y_2 \) is not satisfied. However, if the condition

\[
b_{21}/b_{11} = b_{22}/b_{12}
\]

(4)

obtains the sufficiency condition is satisfied. In that case we can define \( a_{21} \) by

\[
a_{21} \equiv b_{21}/b_{11} = b_{22}/b_{12},
\]

(5)

allowing replacement of (3) with

\[
y_2 = a_{21}y_1 + b_{23}x_3.
\]

(6)

We have \( y_1 \rightarrow y_2 \).

To repeat, \( y_1 \) causes \( y_2 \) in this case because all interventions in \( x_1 \) and \( x_2 \) consistent with a given change in \( y_1 \) have the same effect on \( y_2 \). If, on the other hand, the condition (4) fails then different interventions in \( x_1 \) and \( x_2 \) consistent with a given change in \( y_1 \) lead to different effects on \( y_2 \), so the
outcome of the assumed change in $y_1$ on $y_2$ is ambiguous. In that case we will say that $y_1$ does not cause $y_2$.

If the condition (4) is satisfied the causal form of the model consists of equations (2) and (6), while if it is not the causal form coincides with the solution form, which is (2) and (3). The upper panel of Figure 1 shows the causal diagram if the restriction (4) is satisfied; the lower panel shows the causal diagram if the restriction is not satisfied.

The subset condition guarantees that if $y_i \rightarrow y_j$ then there exists at least one external variable that causes $y_j$ but does not cause $y_i$. This condition implies that causation is antisymmetric ($y_i \rightarrow y_j$ implies $y_j \rightarrow y_i$). This means that in diagrams of models containing blocks of simultaneous equations there will not exist causal arrows connecting the internal variables in these blocks. In the extreme case of a completely simultaneous model the causal form will be the same as the solution form. This makes sense: in simultaneous models without parameter restrictions the effect of one internal variable on another is ambiguous.

An important aspect of the definition of causation just presented is that causation is transitive: if $y_1 \rightarrow y_2$ with coefficient $a_{21}$ and $y_2 \rightarrow y_3$ with coefficient $a_{32}$, then $y_1 \rightarrow y_3$ with coefficient $a_{31} = a_{21}a_{32}$. The transitivity of causation has the implication that an internal variable never has both an indirect effect on another variable via a causal chain involving one or more third variables, and also a distinct direct effect; rather, the direct effect is
always the composition of the indirect effects. This outcome, although highly
counterintuitive, is an implication of implementation neutrality: any model
which purports to represent a setting with both direct and indirect causation
(with the former distinct from the composition of the latter) must involve the
ambiguity in causal attributions that accompanies failure of implementation
neutrality.

2 Causation and Projection

As noted above, in many sources causation is directly inferred from a model
written in recursive form. Here, in contrast, (implementation-neutral) cau-
sation is inferred from restrictions on the solution-form coefficients. It may
not be clear that invoking coefficient restrictions is really necessary for the
characterization of causation. To show that doing so is in fact necessary,
we follow the Cowles economists and subsequent analysts of causation and
write down a model, recast it so as to put it in recursive form, and then
determine whether the variable on the right-hand side causes the variable on
the left-hand side.

On the face of it, general considerations indicate that the exercise of
solving a model for its recursive form and then attributing causation in the
resulting model is likely to run into problems. This is so because the indicated
construction is too easy: under the received definition of causation, for any
pair of internal variables it would appear that either variable can be made
a cause of the other by replacing one of the equations of the model with
an equation representing the effect variable as the sum of its projection on
the cause variable plus a projection error. Inasmuch as any purported causal
ordering can be generated in this way, it follows that the construction implies
that the idea of causation is empty—any variable can cause or be caused
by any other. We argue that identifying causation with recursivity, with
the latter generated by representing equations determining effect variables
via projections, is nothing more than a sophisticated—or, perhaps, not so
sophisticated—version of the correlation-implies-causation fallacy.

We begin with a demonstration that even if one of the equations of the
model consists of a projection of $y_2$ on $y_1$, it does not follow that $y_1$ causes $y_2$.
Consider the solution-form model (2)-(3), where the external variables $x_1$, $x_2$
and $x_3$ are pairwise independent random variables. Replace eq. (3) with an
equation expressing $y_2$ as the sum of its projection on $y_1$ and a projection
error. This can be written as

\[ y_2 = \beta y_1 + y_3. \]  

(7)

Here \( \beta \) is the projection coefficient, given by

\[ \beta = \frac{\text{cov}(y_1, y_2)}{\text{var}(y_1)} = \frac{b_{11}b_{21}\sigma_1^2 + b_{12}b_{22}\sigma_2^2}{b_{11}^2\sigma_1^2 + b_{12}^2\sigma_2^2}. \]  

(8)

In these equations \( \sigma_1^2 \) is the variance of \( x_1 \), and similarly for \( \sigma_2^2 \). The projection error \( y_3 \) is defined to equal \( y_2 \) minus the projection of \( y_2 \) onto \( y_1 \):

\[
y_3 \equiv y_2 - \beta y_1 = b_{21}x_1 + b_{22}x_2 + b_{23}x_3 - \beta y_1 \\
= (b_{21} - \beta b_{11})x_1 + (b_{22} - \beta b_{12})x_2 + b_{23}x_3.
\]  

(9)

(10)

Note that the projection error, being a function of all three external variables, is an internal variable, and this is indicated by its label \( y_3 \).

We see that the model (2)-(7) is recursive. The question now is whether this recursivity implies that \( y_1 \) can be taken to cause \( y_2 \) in this model, so that an intervention consisting of replacing a given value of \( y_1 \) with \( y_1 + \Delta y_1 \), holding constant \( y_3 \), can be considered to be causal. Determining this involves ascertaining whether the conditions for implementation neutrality are satisfied. The subset condition is satisfied, since the external set for \( y_1 \), consisting of \( \{x_1, x_2\} \), is a proper subset of the external set for \( y_2 \), which is \( \{x_1, x_2, x_3\} \). However, the sufficiency condition is not satisfied: there are an infinite number of \( \{x_1, x_2, x_3\} \) triples that generate a given \( y_1 \) and \( y_1 + \Delta y_1 \) with a fixed \( y_3 \) (because the \( \{x_1, x_2, x_3\} \) satisfy two equations, the set of \( \{x_1, x_2, x_3\} \) associated with the intervention lie on a line). These generally map onto different values of \( y_2 \). Thus the fact that the model is recursive with \( y_3 \) orthogonal to \( y_1 \) is consistent with failure of the sufficiency condition, and therefore does not imply existence of a causal relation.

As would be expected, the restriction (4) on the solution-form coefficients shown above to imply \( y_1 \to y_2 \) plays exactly the same role here. The restriction implies that \( \beta \) reduces to \( b_{21}/b_{11} (= b_{22}/b_{12}) \), so that the two terms in parentheses in (10) equal zero. Consequently, \( x_1 \) and \( x_2 \) drop out of the equation for the projection error, which becomes \( y_3 = b_{23}x_3 \). Under the coefficient restriction the causal links from \( x_1 \) and \( x_2 \) to \( y_2 \) pass through \( y_1 \), with the projection error unaffected by the intervention. Accordingly, the
The point here is that it is the parameter restriction (4), not the recursive structure of the model, that generates $y_1 \rightarrow y_2$.

The fact that the projection error $y_3$ is orthogonal to the $y_1$ could be invoked to justify the assumption that $y_3$ is an external variable.$^9$ Suppose one makes this assumption. Correspondingly, one would also convert $x_3$ into an internal variable (otherwise the model is overdetermined). The model so transformed looks similar to (2) – (3). We have $y_1 \rightarrow y_2$ in both cases (however, note that the causal coefficient in the latter case, $\beta$, is restricted by (8), whereas $a_{21}$ in (3) is unrestricted. However, the point is that in relabeling variables between external and internal one has altered the model in a fundamental way. Even though the two models under discussion have the same equations linking the external and internal variables, the fact that the specifications of which variables are external and which are internal are different implies that the two models will not necessarily have the same causal ordering.

3 Comparison with Simon

It is instructive to compare the representation of causation just presented with that of Simon’s classic [16] paper.

Simon observed that any linear structural model contains an ordered set of self-contained sub-models, with some (or all) of the internal variables determined in each sub-model. These self-contained sub-models are ordered: each sub-model contains the internal variables determined in that sub-model and, except for the lowest-ordered sub-models, also some or all of those internal variables determined in lower-ordered sets. Fully recursive models, in which each internal variable is determined by a single equation, are the most extreme special case. Block-recursive models constitute the general case. Under Simon’s definition we have $y_1 \rightarrow y_2$ if $y_1$ appears in the sub-model that determines $y_2$, but is determined in a lower-order sub-model.

There exist two major differences between the treatment of causation here and that of Simon. The first is that causation here is defined from the

---

$^9$We momentarily depart from the convention of labeling external variables $x$ and internal variables $y$. A better notation would be to redefine $\{x_1, x_2, x_3\}$ and $\{y_1, y_2, y_3\}$ as $\{z_1, \ldots, z_6\}$, where $z_i$ can be either external or internal, and consider the implications of changing which variables are external and which internal.
solution form of the model, as opposed to its structural form as with Simon. The fact that we base the definition of causation on the solution form allows us to sidestep debates about the meaning of the term “structural model”, a question that occupied the early writers on simultaneous equation models.¹⁰

Pending discussion below, we provisionally define the structural form of a model as consisting of a model in which each equation has one of the internal variables on its left-hand side. Its right-hand side consists of some of the other internal variables and one or more of the external variables.¹¹ It is easily verified that Simon’s criterion for causation coincides with our subset condition: if \( y_1 \) is determined in a lower-order sub-model than \( y_2 \), then the external set for \( y_1 \) is a proper subset of the external set for \( y_2 \). However, and this is the second difference between our treatment and Simon’s, there is no analogue in Simon’s discussion for our sufficiency condition: under Simon’s definition the subset condition by itself is necessary and sufficient for causation. Thus in comparing the definition here with that of Simon the question becomes to determine the consequences of (1) working with the solution form of a model rather than its structural form, and of (2) imposing the sufficiency condition.

Our first result is that in the simplest setting Simon’s characterization of causation leads to the same causal ordering as ours. This occurs because in the simplest case Simon’s specification of a structural-form rather than solution-form model turns out to be equivalent to imposing the parameter

¹⁰ Despite their emphasis on the importance of the concept of structural models, the Cowles analysts never provided an adequate motivation. Friedman [5] is one of the few who have raised questions explicitly: “[S]ome economists, particularly a group connected with the Cowles Commission ..., have placed great emphasis on a division of this step of selecting a hypothesis consistent with known evidence into two substeps: first, the selection of a class of admissible hypotheses from all possible hypotheses; second, the selection of one hypothesis from this class (the choice of a “structure”). This subdivision may be heuristically valuable in some kinds of work.... From a methodological point of view, however, it is an entirely arbitrary subdivision of the process of deciding on a particular hypothesis that is on par with many other subdivisions that may be convenient for one purpose or another or that may suit the psychological needs of particular investigators.” It is not easy to improve on this formulation.

I am indebted to Tom Cooley and Boyan Jovanovic for this reference.

¹¹ We allow structural equations to contain more than one external variable so as to ensure that the external sets associated with each internal variable have at least two variables. As discussed in Section 1, this specification renders unambiguous the characterization of variables as external vs. internal.
restriction that implies satisfaction of the sufficiency condition.

To see this, consider the model defined above. In the model consisting of (2) and (6) $y_1$ is determined in a lower-order block than $y_2$, implying that $y_1$ causes $y_2$ by Simon's criterion. But formulating the model in recursive form (rather than deriving the recursive form from a projection, as in the preceding section), as these equations do, implies that when one calculates the solution form of the model, the solution coefficients (2)-(3) satisfy (4). Therefore specifying the structural model guarantees that the sufficiency condition for a causal ordering is satisfied. Thus we have $y_1 \rightarrow y_2$ under either definition of causation.

Our major result is that in larger models the equivalence between Simon's characterization of causation and that proposed here breaks down. It follows that causation as Simon defined it is not generally implementation neutral. To see this, consider the triangular structural model

\begin{align*}
y_1 &= b_{11}x_1 + b_{12}x_2 \\
y_2 &= a_{21}y_1 + b_{23}x_3 \\
y_3 &= a_{31}y_1 + a_{32}y_2 + b_{34}x_4.
\end{align*}

By Simon's criterion the causal ordering here is $y_1 \rightarrow y_2$, $y_1 \rightarrow y_3$ and $y_2 \rightarrow y_3$. It is immediately clear that this causal ordering cannot occur in this model under our definition of causation. Under our characterization of causation, $y_1 \rightarrow y_2$ and $y_2 \rightarrow y_3$ imply $y_1 \rightarrow y_3$ as the composition of the two causal functions, as observed above. In the causal ordering under Simon's definition we see that, in contrast, $y_1 \rightarrow y_3$ appears separately, with a causal coefficient that is not functionally related to the causal coefficients associated with $y_1 \rightarrow y_2$ and $y_2 \rightarrow y_3$. It follows that in the model just formulated the causal ordering that occurs under Simon's definition cannot be characterized as implementation-neutral under any specification of coefficient restrictions on the solution form.

The solution form of the structural model just presented can be written as
In the absence of restrictions on the solution-form coefficients the causal ordering by our definition is empty. If we have $b_{21}/b_{11} = b_{22}/b_{12}$ we can define $a_{21}$ by $a_{21} \equiv b_{21}/b_{11} = b_{22}/b_{12}$ and can replace (15) by $y_2 = a_{21} y_1 + b_{23} x_3$, corresponding to $y_1 \rightarrow y_2$. In that case the causal form of the model consists of (11), (12) and (16). Under only this restriction $y_1$ and $y_2$ are not causally connected to $y_3$.

Under further restrictions one may have in addition $y_1 \rightarrow y_3$ or $y_2 \rightarrow y_3$, or both. If we have both, then $y_1 \rightarrow y_3$ is the composition of $y_1 \rightarrow y_2$ and $y_2 \rightarrow y_3$, contrary to the implication of (11)-(13) (for generic parameter values). Thus there exists no set of restrictions on (14)-(16) that results in (11)-(13) as representing the causal form of the model. We conclude that in general causation under Simon’s definition differs from causation as defined here.

Connecting the definition of causation here with that of the Cowles researchers requires determining further consequences of the fact that causation here is defined from the solution form of the model, whereas the Cowles economists defined causation from the structural form. Up to now we have identified structural models as consisting of equations written in recursive form. For the Cowles economists the significance of structural equations had to do with the role they assigned to the coefficients (see, for example, Hurwicz [11]). They modeled behavioral changes in structural models—interventions—as consisting of shifts in coefficients. This involves interpreting the coefficients of purportedly linear models as external variables, rather than constants as in our treatment. Thus in the Cowles usage structural models effectively have two types of external variables: (1) the variables explicitly labeled as external and (2) the model coefficients. The former represent the routine functioning of the model, and the latter are used in modeling interventions. The treatment here, in contrast, does not distinguish these two types of external variables.

As a result of this reinterpretation of coefficients as variables rather than constants, models that are formally represented as linear in variables are actually bilinear. This bilinearity property would have greatly complicated...
the analysis of causation if the Cowles economists had elected to acknowledge it explicitly. We avoided the problem of defining causation in bilinear models by adopting a linear setting, which involves treating coefficients as constants rather than external variables.

Bilinear models can be linearized, and doing so allows a comparison of our analysis with Simon’s that is accurate up to the accuracy of the linear approximation. For example, consider the model

\[ y_1 = b_{11}x_1 + b_{12}x_2 \quad (17) \]
\[ y_2 = a_{21}y_1 + b_{23}x_3. \quad (18) \]

As already noted, interpreted as a causal model these equations have \( y_1 \rightarrow y_2 \). If this model is considered to be structural, \( a_{21} \) is interpretable as an external variable rather than a constant. Under our notation, following general practice, variables are indicated by letters from the end of the alphabet (\( x \) and \( y \)) and constants are indicated by letters from the beginning of the alphabet, so we relabel \( a_{21} \) as \( x_4 \). Making the indicated change, there results the bilinear model

\[ y_1 = b_{11}x_1 + b_{12}x_2 \quad (19) \]
\[ y_2 = x_4y_1 + b_{23}x_3. \quad (20) \]

The linearized version of this model (with constant terms deleted, as throughout this paper) is

\[ y_1 = b_{11}x_1 + b_{12}x_2 \quad (21) \]
\[ y_2 = a_{21}y_1 + b_{23}x_3 + b_{24}x_4, \quad (22) \]

where the \( a \) an \( b \) coefficients are redefined. This model has \( y_1 \rightarrow y_2 \). Again we see that in the simplest case Simon’s definition of a causal ordering coincides with ours. In larger models one can implement the same exercise, consisting of relabeling some or all of the coefficients as external variables, linearizing the resulting model, solving for the solution form and determining the causal ordering. We have already seen that in larger models the causal orderings derived under Simon’s definition will generally differ from those derived here.
4 Empirical Aspects of Causation

Up to this point we have considered models in which variables are specified as to their status as internal or external. We have not specified which variables are observable or what we are assuming about the probability distributions of those that are not. That we could postpone discussion of observability to this point reflects the fact that, for any pair of internal variables, the existence or nonexistence of causation depends only on whether the conditions for implementation neutrality are satisfied. It does not depend on which variables are observable or what is assumed about those that are not. However, without specifying which variables are observable and characterizing the probability distribution of unobserved external variables there is no way to estimate causal coefficients empirically: in the absence of identifying information the correlations among internal variables implied by the model’s causal structure cannot be disentangled from those induced by correlations among the external variables.

The most direct way to launch an investigation of the empirical aspects of causation is to specify, first, that external variables are unobservable and internal variables are observable. This specification covers many of the cases of interest. However, it does not cover all: assuming that external variables are not observable rules out at least some kinds of nonstationary models. Further, assuming that internal variables are observable rules out latent variables, a topic of major interest in applied work. The motivation for adopting this restriction is to simplify the analysis.

Second, it is assumed that the external variables are statistically independent random variables. This assumption implies that whatever correlations exist among the model’s internal variables are generated by the equations of the model, not by q correlations among external variables. Without this assumption, or something similar, there is no way to separate causation in the model from uninterpreted correlations among external variables.

The assumptions just listed imply that if we have $y_1 \rightarrow y_2$ the coefficient measuring the effect of $y_1$ on $y_2$ is identified (apart from special cases, as discussed below), and can be estimated consistently and without bias using a least-squares regression of $y_2$ on $y_1$. This is so because the external variable(s) in $\mathcal{E}(y_2) - \mathcal{E}(y_1)$—the constituents of the error term in the regression—is (are) independent of $\mathcal{E}(y_1)$, and therefore of $y_1$ itself.

The finding that causal coefficients are always identified differs from the conclusion of the Cowles economists. The reason for the difference is that, as
noted, the Cowles economists treated causal coefficients as synonymous with the coefficients of structural equations, which may or may not be identified. Here, in contrast, we avoid any reference to structural equations.

The result that causal coefficients are always identified should not be taken to imply that identification is not a major problem in the analysis of causation. Obviously, there exist coefficients associated with causation only when the associated variables are causally ordered, and whether two variables are causally ordered depends on the coefficients that link observed internal variables to unobserved external variables. These coefficients are generally not identified, implying that there is generally no way to check the conditions for causation directly. Thus the fact that causal coefficients, when they exist, are always identified does nothing to mitigate the difficulty of ascertaining whether two variables are causally related in the first place.

Under causation as characterized here, as with other definitions of causation, the restrictions justifying an assumed causal ordering can in principle be tested indirectly by identifying pairs of variables that are or are not statistically independent according to the model, and then determining whether these independence implications are satisfied empirically. We now consider whether powerful empirical tests of causal models along these lines are likely to be available. It appears that they are not: only in special cases is it possible to characterize independence or the lack thereof among internal variables as testable implications of causal models.

Among the few results that are available is the obvious fact that any two internal variables for which the external sets are disjoint are statistically independent. As an implication, if an internal variable has two ancestors, then either one ancestor causes the other or the two are statistically independent. To see this, suppose that $y_1 \rightarrow y_3$ and $y_2 \rightarrow y_3$, so that $y_3$ has ancestors $y_1$ and $y_2$. If $\mathcal{E}(y_1)$ and $\mathcal{E}(y_2)$ are disjoint, then $y_1$ and $y_2$ are statistically independent. Suppose instead that $\mathcal{E}(y_1)$ and $\mathcal{E}(y_2)$ have a nonempty intersection that contains external variable $x$. Then because (1) $x \in \mathcal{E}(y_1)$, and (2) $\mathcal{E}(y_1)$ is a proper subset of $\mathcal{E}(y_3)$, there exists a path from $x$ to $y_3$ that includes $y_1$. Similarly, there exists a path from $x$ to $y_3$ that includes $y_2$. These must be the same path, since if the path included $y_1$ but not $y_2$ then $y_2$ could not be a sufficient statistic for $\mathcal{E}(y_2)$, contradicting $y_2 \rightarrow y_3$. Thus there is a single path connecting $x$ and $y_3$, and that path includes both $y_1$ and $y_2$. This can occur only if $y_1 \rightarrow y_2$ or $y_2 \rightarrow y_1$.\(^{12}\)

\(^{12}\)In the passage discussed in note 5, Hausman [9] observed that common parents of a
Past this there are not many results available about correlation of variables in causal models. Assume that $y_1$ and $y_2$ have $y_3$ as a common ancestor. If also $y_1 \rightarrow y_2$, then we have $y_3 \rightarrow y_1 \rightarrow y_2$. In that case we have that all three variables are correlated since their external sets have a nonempty intersection (consisting of the external set for $y_3$). If, on the other hand, $y_1 \not\rightarrow y_2$ the causal coefficient associated with $y_1 \rightarrow y_2$ is not defined. In the absence of causality, no inference about the correlation among variables is possible.

A related point is that even under the assumption that external variables are independently distributed, causation as characterized here does not satisfy the causal Markov condition (which says that any variable is statistically independent of any other variable that is not among the causal descendants of the first variable, conditional on the first variable’s parents). This is obviously true, for example, in fully simultaneous systems. In that case the causal ordering is empty, so the causal Markov condition would imply that all internal variables are statistically independent. This is generally not the case.

Implementation neutrality implies that the causal Markov condition may fail even in systems that are not fully simultaneous. For instance, in the example of Section 3 with $y_1 \rightarrow y_2$ but not $y_2 \rightarrow y_3$, we have that $y_2$ and $y_3$ are not statistically independent, contrary to the implication of the causal Markov condition.

The fact that imposing implementation neutrality invalidates the causal Markov condition raises questions about the practice in the causation literature of taking the causal Markov condition as an axiom (as in Spirtes, Glymour and Scheines [17], for example). The observation that the causal Markov condition fails under a definition of causation—that proposed here—that captures an important aspect of the ordinary-language meaning of causation suggests that the causal Markov condition depends on considerations not inherent in the idea of causation. We have not been informed as to what these conditions are. Accordingly, we are led to question the association of causation with the causal Markov condition.

Despite the foregoing discussion, it happens that some of the techniques of diagrammatical analysis developed in the causation literature do carry over variable are “probabilistically independent”. However, we noted that by this he meant that they are not causally connected. Whether probabilistic independence in this sense is equivalent to statistical independence depends on whether the causal Markov condition is satisfied. As noted below, it is not satisfied under implementation-neutral causality. Thus the result here is different from Hausman’s.
in the present setting. For example, it is shown in the received literature that if two internal variables are connected only by paths that are “blocked” because each contains a “collider” (a variable with incoming arrows from both directions), those variables are independent. That result appears to carry over here. An example will demonstrate this.

4.1 Example

Consider the following model:

\[ y_1 = x_1 + x_2 \]  \hspace{1cm} (23)
\[ y_2 = x_3 + x_4 \]  \hspace{1cm} (24)
\[ y_3 = x_1 + x_2 + x_3 + x_4 \]  \hspace{1cm} (25)
\[ y_4 = x_1 + x_2 - x_3 - x_4 \]  \hspace{1cm} (26)

(note that here we have supplied specific coefficient values as well as external sets). The causal form of this model is

\[ y_1 \leftarrow x_1 + x_2 \]  \hspace{1cm} (27)
\[ y_2 \leftarrow x_3 + x_4 \]  \hspace{1cm} (28)
\[ y_3 \leftarrow y_1 + y_2 \]  \hspace{1cm} (29)
\[ y_4 \leftarrow y_1 - y_2, \]  \hspace{1cm} (30)

with Figure 2 as its causal diagram. Here \( y_1 \) and \( y_2 \) are statistically independent due to the fact that their external sets are disjoint. We have that \( y_1 \) and \( y_2 \) are parents of \( y_3 \) (and also of \( y_4 \)), so the result illustrates the general fact noted above that if any internal variable has more than one ancestor, either these are independent or one ancestor is a cause of the other.

This independence result can be generated using the diagrammatical techniques developed by Pearl and others for analysis of causation in settings where implementation neutrality is not imposed. In the example there exist two paths from \( y_1 \) to \( y_2 \), but both are blocked by existence of the colliders \( y_3 \) and \( y_4 \). Therefore these paths do not transmit association. Independence of \( y_1 \) and \( y_2 \) results. Note that here the diagrammatical analysis applies by virtue of the assumption that the external variables are independently distributed, not because of the causal Markov condition, which as we have seen does not generally apply under implementation neutrality even if external
variables are assumed independent. The result suggests that even though the conditions for causation analyzed here are different from those in the received literature, at least some of the diagrammatical techniques for analysis of causation carry over.

The independence result does not extend to the children $y_3$ and $y_4$ except in special situations. For example, if the $x_i$ are normally distributed and all have the same variance, $y_3$ and $y_4$ are independent. However, if $x_1$ and $x_2$ have higher (lower) variance than $x_3$ and $x_4$, then $y_3$ and $y_4$ will be positively (negatively) correlated.

5 Conditioning on Internal Variables

The result in the preceding section that the coefficient associated with any causal relation is identified and can be estimated consistently using least squares depends critically on the underlying assumption that external variables are independently distributed and internal variables are fully observable. If some internal variable $y_i$ is observed only when it lies in a certain region, the distribution for the external variables that is relevant for determining the identifiability of causal coefficients is that conditional on this restriction, not the unconditional distribution.

The joint distribution of the external variables conditional on $y_i$ will generally display statistical dependence even if the unconditional distribution of the external variables incorporates independence. This situation will not affect the causal ordering of the variables, but it does invalidate the result that the coefficients associated with the causal ordering can be estimated consistently by least squares. This is so because failure of independence in the
external variables implies that the error term covaries with the explanatory variable in the relevant regression, inducing bias and inconsistency.

As an extreme case, suppose that the analyst only has data in which \( y_i \) takes on a single value, for some \( i \). Obviously the coefficient associated with \( y_i \rightarrow y_j \) or \( y_j \rightarrow y_i \) for some \( y_j \) is not identified, there being no variation in the cause variable in one case or the effect variable in the other. A more common situation occurs when the data for \( y_i \) are truncated, as by \( y_i \geq 0 \). In that case the sample regression coefficient associated with \( y_j \rightarrow y_k \) is not a consistent estimate of the associated causal coefficient if either \( y_j \) or \( y_k \) has external sets that overlap with that of \( y_i \). This is so because if \( y_i \) is subject to a restriction like \( y_i \geq 0 \) the relevant joint distribution of the external variables in \( \mathcal{E}(y_i) \) is that conditional on \( y_i \geq 0 \), and this does not generally have any independence property.

A simple example (adopted from Elwert [3]) illustrates this. Suppose that movie actors become famous if they are good looking or can act well, or both. Assume, probably realistically, that being good looking and being a good actor are independently distributed. If the analyst has a data set consisting only of actors who are famous, then any actor in that set who is not good looking must be a good actor, since otherwise he would not be famous. Thus in the data set of famous actors there will be a negative correlation between being good looking and being a good actor, even though by assumption there is no such correlation in the general population. Any statistical exercise that makes no allowance for this effect will be biased.

We will not discuss statistical procedures to deal with this problem since the problem does not directly involve causal issues. The point here is only to demonstrate that the attractive statistical properties of least squares in estimating causal coefficients do not apply universally when data on internal variables are not fully observed.

6 Comparison with “Fixing”

The analysis of causation outlined in this paper differs in major respects from what is found in the literature. Most important, interventions here consist of hypothetical alterations in the assumed values of external variables. In contrast, as noted above the usual treatment in the literature (based on Haavelmo [8] and Strotz and Wold [18]) involves modeling policy interventions on, say, \( y_1 \) by deleting from the model the equation determining \( y_1 \) and
replacing it with the specification that $y_1$ is external.

This practice of “fixing” internal variables and deleting equations when analyzing interventions seems misdirected. It violates the autonomy assumption (which consists of the assertion that the model equations are assumed invariant to assumed interventions). It does not make sense to claim to analyze interventions using a model if doing so involves changing the model to accommodate the intervention.

Fixing internal variables involves a troubling inconsistency between how model solutions are generated in the routine operation of the model—via realizations of external variables—and how they are modeled under a policy intervention—via relabeling internal variables as external and suppressing equations. What is it about policy interventions that motivates this difference in treatment? We are not told. As suggested above, it seems simpler and more satisfactory to be consistent about carrying over the attribution of assumed interventions on internal variables to underlying changes in the external variables that determine them, and thereby to avoid altering the equations of the model.

Besides this, there are several major problems with modeling interventions by fixing internal variables. Most obviously, it applies only in recursive systems, since in the presence of simultaneity $y_1$ is determined jointly with other variables in a group of several or many equations. In that case there does not exist any obvious way to identify which equation is to be deleted.

The Strotz-Wold procedure assumes that causal models are modular, meaning that causal relations can be modified individually without invalidating the other equations of the model (modularity has been discussed widely in the philosophical literature on causation; see, for example, Cartwright [1] and the works cited there). Under our treatment, in contrast, the existence or nonexistence of causal effects is calculated from the solution form, not the structural equations that purportedly express the modularity assumption. Since our analysis does not involve replacing equations with exogeneity specifications, the question of modularity does not come up.

Modeling interventions by respecifying internal variables as external—the Strotz-Wold procedure—implies that causation will be treated as if it were implementation neutral whether or not this treatment is justified when causality is modeled as proposed here. If implementation neutrality fails coefficients will be interpreted as causal when they do not support that interpretation under the present causal definition. It is far from clear why one would want to take this route. In general the question “What is the effect of
y_1 on y_2?” is properly viewed as possibly depending on what brings about the change in y_1. The model encodes exactly this information in the equations determining y_1. Therefore the analyst can determine whether the question of causation has an unambiguous answer, but only if he avoids deleting the equations that determine y_1 from the model. Strotz and Wold and other proponents of analyzing interventions by reclassifying internal variables as external do not explain why it makes sense to adopt a framework that implies suppression of the question of whether causation is unambiguous.

7 Application: Granger Causation

Granger [6] proposed a definition of causation that can be implemented empirically without relying on theoretical restrictions: a stochastic process (i.e., sequence of random variables) y_1 = \{y_{1t}\} Granger-causes another process y_2 if the optimal prediction of future values of y_2 based on past values of y_2 alone can be improved by including current and lagged values of y_1 as explanatory variables. It is asserted that if y_1 does not Granger-cause y_2, then y_{2t} can be treated as strictly exogenous with respect to y_{1t}, so that correlations between the two can be interpreted as reflecting the causal effect of y_2 on y_1.

The problem here is to determine the relation between Granger-causation and causation as defined in this paper.

Analysts recognized immediately that Granger-causation is not the same as causation as that term is used in ordinary discussion. For example, Granger pointed out that under the definition just stated cattle stamping their hooves before an earthquake implies that the cattle Granger-cause the earthquake. Granger termed such cases “spurious causation”, implying that the question of how to define non-spurious causation remained open.

We need to understand the relation between Granger causation and causation as modeled here. We formulate a two-variable vector autoregression generating the values of the money stock m = \{m_t\} and gross domestic product y = \{y_t\} (note that in this section we are using y to denote GDP, not as representing a general internal variable):
\[ m_t = a_{my}y_t + b_{mm}m_{t-1} + b_{my}y_{t-1} + x_{1t} \]  
\[ y_t = a_{ym}m_t + b_{ym}m_{t-1} + b_{yy}y_{t-1} + x_{2t} \]  

(31)

Here the external variables \( x_{1t} \) and \( x_{2t} \) are independent, and are independent over time. The reduced form corresponding to this system is

\[ m_t = c_{mm}m_{t-1} + c_{my}y_{t-1} + u_{1t} \]  
\[ y_t = c_{ym}m_{t-1} + c_{yy}y_{t-1} + u_{2t} \]  

(33)

(34)

GDP fails to Granger-cause the money stock if

\[ c_{my} = \frac{a_{my}b_{yy} + b_{my}}{1 - a_{my}a_{ym}} = 0. \]  

(35)

The money stock is strictly exogenous with respect to GDP if \( a_{my} = b_{my} = 0 \). Strict exogeneity implies that GDP shocks do not feed back into the equation determining money, either currently or with a lag. Granger non-causation is a necessary condition for strict exogeneity, but not a sufficient condition.

We are interested in determining what parameter restrictions are necessary for \( m_t \rightarrow y_t \) under the definition of causation proposed here. To do so we first write the solution form of the model under the assumption that \( m_t \) is strictly exogenous:

\[ m_t = x_{1t} + b_{mm}x_{1,t-1} + ... \]  
\[ y_t = a_{ym}x_{1t} + (a_{ym}b_{mm} + b_{ym})x_{1,t-1} + x_{2t} + b_{yy}x_{2,t-1} + ... \]  

(36)

(37)

Implementation-neutral causation requires that the ratio of the coefficients of \( x_{1t} \) in determining \( m_t \) and \( y_t \) equal the corresponding ratio for \( x_{1,t-1} \):

\[ \frac{1}{a_{ym}} = \frac{b_{mm}}{a_{ym}b_{mm} + b_{ym}}. \]  

(38)

Here the reasoning is exactly the same as in Subsection 1.1. This equality is satisfied if and only if \( b_{ym} = 0 \).\(^{14}\)

\(^{14}\)This condition is unchanged if more lagged terms of \( x_1 \) and \( x_2 \) are included explicitly in the solution form of the model.
Thus even strict exogeneity of \( m \) is not a sufficient condition for interpreting the coefficient of \( m_t \) in equation (32) for \( y_t \) as the causal coefficient associated with \( m_t \rightarrow y_t \). This is so because if \( b_{ym} \neq 0 \) the lagged values of \( x_1 \)—the external variables that determine \( y_t \) through their effect on \( m_t \)—also affect \( y_t \) via \( m_{t-1} \). Thus we have a failure of implementation neutrality: if \( b_{ym} \neq 0 \) characterizing an intervention as a hypothesized change in \( m_t \) does not give enough information about the intervention to determine the resulting change in \( y_t \). Avoiding this outcome requires imposing the implementation-neutrality condition \( b_{ym} = 0 \) in addition to the strict exogeneity of \( m \), so as to shut down \( m_{t-1} \) as a determinant of \( y_t \).

We see that to make the transition from Granger-noncausation to causation in the sense of this paper, one has to make two further restrictions on the model (31)-(32), beyond \( c_{my} = 0 \). The first is that \( c_{my} = 0 \) must be strengthened to \( a_{my} = b_{my} = 0 \). Analysts aware of the distinction between strict exogeneity and Granger non-causality frequently state that \( c_{my} = 0 \) is consistent with \( a_{my} = b_{my} = 0 \), but then incorrectly go on to treat “is consistent with” as having the same meaning as “implies”. Second, as we have just seen implementation neutrality requires that one rule out \( m_{t-1} \) as an argument in the equation for \( y_t \).

The conclusion is that Granger causation is a specialized—and, to be sure, a very useful—form of forecastability, but it cannot be directly interpreted as having anything to do with implementation-neutral causation without strong restrictions on model coefficients.

It may be that we are being too narrow in trying to relate Granger-causation to causation between current values of \( m \) and \( y \) as defined here. The definition of causation here relates a single cause variable and a single effect variable at the same date, whereas Granger causation involves the stochastic processes \( m \) and \( y \). The suggestion is that a more general notion of causation is required. If so, the task at hand for proponents of Granger causation would seem to be to propose a more general characterization of (true) causation and then relate Granger causation to that.

8 Conclusion

In this paper we propose a specific meaning for causation: one variable causes another if the effect of all interventions that produce a given alteration on the cause variable induce the same alteration on the effect variable (of course, the
alterations in the cause variable and the effect variable are generally unequal). If this condition is satisfied the answer to the question “What is the effect of a change in $y_1$ on $y_2$?” does not depend on what caused the assumed change in $y_1$. This, it seems to us, captures what scientists want to know when they investigate questions dealing with causation. If the condition is not satisfied the effect of $y_1$ on $y_2$ is undefined, implying that one can only discuss the effects of the determinants of $y_1$ on $y_2$.

Philosophers sometimes reject limiting use of the term “causation” to settings in which interventions are implementation-neutral. For example, Cartwright [1] states that “[w]e must be careful ... not to be misled by [LeRoy’s] own use of the language of ‘causal order’ to suppose it tells us whether and how much one quantity causally contributes to another” (p. 246). Why are we misled by this supposition? How much one quantity causally contributes to another is exactly what causation as defined here tells us, and is exactly what we want to know. And what meaning can we attach to a purported measure of the effects of an intervention on an internal variable if the model is such that the intervention is not implementation-neutral? In that case there is no alternative to redirecting the analysis to implementation-specific interventions on the external variables.

We saw that the algorithm proposed in this paper can be implemented to produce a directed acyclic diagram in any equation system. Under the construction here, systems with large simultaneous-equations blocks will have few or no causal arrows, reflecting the fact that in such systems interventions usually cannot be adequately characterized by their effects on any one internal variable. The fact that causal orderings in completely simultaneous models are completely empty is not a flaw of the conception of causation developed here, as one is led to conclude by some discussions. Rather, an empty causal ordering is a simple acknowledgement of the fact—familiar to economists from elementary analysis of supply-demand systems—that in simultaneous systems internal variable cannot intelligibly be interpreted as causing each other. However, we noted in several places above that causal diagrams can sometimes be used in the same way under our definition of causation as under other definitions. It would be of some interest to conduct a systematic comparison of inference from causal diagrams in the two

---

15 All of this is in sharp contrast to the case under the received construction, in which simultaneity produces graphs that are cyclic. Thus our algorithm avoids the implication of the received construction that in nonrecursive models internal variables cause themselves, albeit indirectly, which seems to be an unfortunate implication.
cases, although there would remain the question of how to attach meaning to causation when implementation neutrality fails.

The question of how to implement the definition of causation proposed here is a difficult one. At a minimum, the analysis here can play the role of raising questions about discussions of causation when there is no attempt to justify the implicit assumption of intervention neutrality. The idea is to encourage clear communication about what exactly is involved in causal assertions.

References


