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.\(^1\) 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 \( y_1 \) implementation neutral causes \( y_2 \), and will write \( y_1 \rightarrow y_2 \). Hereafter this term is abbreviated IN-causes, so that \( y_1 \) IN-causes \( y_2 \). If the implementation neutrality condition fails, or may fail, we will say that \( y_1 \) causes \( y_2 \). In that case different interventions on the determinants of \( y_1 \) have different effects on \( y_2 \), implying that we cannot characterize the effect of \( y_1 \) on \( y_2 \) without knowing more about the intervention. “Knowing more about the intervention” amounts to redirecting the discussion from the causal relation between \( y_1 \) and \( y_2 \) to the causal relation between the determinants of \( y_1 \) and \( y_2 \).

If we know only that \( y_1 \) causes \( y_2 \)—that is, if we do not have implementation neutrality—we know that interventions that affect \( y_1 \) also affect \( y_2 \), but we cannot identify a unique coefficient that gives the effect of \( y_1 \) on \( y_2 \). For many—arguably, most—scientific purposes it is useful to have implementation neutrality, so as to know that the effect of \( y_1 \) on \( y_2 \) does not depend on what caused the change in \( y_1 \).

Use of diagrammatical methods in causal analysis has become widespread.

\(^1\)Abbreviated versions of this material were presented in LeRoy [11] and [12] (these papers are discussed at some length in Cartwright [1]). Here more detail is supplied, and the packaging is different.
in recent years, due to work by Pearl [13], Spirtes, Glymour and Scheines [15], Woodward [17], Hausman [8], Cartwright [1] and others. These authors do not include implementation neutrality in their definition of causation (at least not explicitly; see discussion below). 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 IN-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. We will use $y$ to denote internal variables 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, a specification that is essential in inquiries dealing with causation. Of course, the analyst can always implement a shift on any of the equations of the model simply by specifying that the relevant equation includes 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, and which, as discussed below, we will

\[2\] 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 terms, but econometricians have implemented a change in their meaning (see Granger [6], wherein I am reprimanded for using the earlier terms in LeRoy [11]). To avoid ambiguity, economists now use “internal” and “external” when the earlier meaning is intended, as here.

For discussion of various definitions of exogeneity and endogeneity see Leamer [10]. For a statement of the definition of exogeneity currently favored by econometricians see Engle, Hendry and Richard [4].
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.\(^3\) 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 \(E(y_i)\). In examples we will adopt the convention that the external set for any internal variable consists of at least two external variables.\(^4\)

There is no difficulty in defining causation when the cause variable is external: \(x_1\) causes \(y_1\) whenever \(x_1\) is in the external set for \(y_1\). In that case, by virtue of linearity, a unique constant \(b_{11}\) gives the effect of a unit change in \(x_1\) on \(y_1\) for any values of the external variables. If \(x_1\) is not in the external set for \(y_1\) the former does not cause the latter.

The ambiguity comes up when the cause variable is internal, because 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 of the external variables of \(y_1\) on \(y_2\) 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 effect on \(y_1\). Given this ambiguity, we cannot associate causation with a single number giving the effect of \(y_1\) on \(y_2\): 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, \(y_1\) and \(y_2\), their external sets \(E(y_1)\) and \(E(y_2)\), and the functions relating the former to the latter. The first is the *subset condition*, which requires that the external set for \(y_1\) be

---

\(^3\)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.

\(^4\)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. 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.
a proper subset of that of \( y_2 \). The subset condition guarantees that any external variable that affects \( y_1 \) also affects \( y_2 \), but not vice-versa. Hoover [9] in particular emphasized this condition, which assures the antisymmetry of causation.\(^5\) If the subset condition is satisfied we will say that \( y_1 \) causes \( y_2 \), and will write \( y_1 \rightarrow y_2 \).

The second is the \textit{sufficiency condition} (the definition of which presumes satisfaction of the subset condition). The sufficiency condition states that the map from \( \mathcal{E}(y_1) \) to \( y_2 \) can be expressed as two functions. The first is the composition of a function from \( \mathcal{E}(y_1) \) to \( y_1 \) and a function from \( y_1 \) to \( y_2 \), while the second is a function from \( \mathcal{E}(y_2) - \mathcal{E}(y_1) \) to \( y_2 \). If such functions exist then \( y_1 \) is a sufficient statistic for \( \mathcal{E}(y_1) \) for the purpose of determining \( y_2 \), meaning that for the purpose of determining \( y_2 \) an intervention on any or all of the variables in \( \mathcal{E}(y_1) \) is adequately characterized by the resulting induced change in \( y_1 \). If \( y_1 \rightarrow y_2 \) and in addition the sufficiency condition is satisfied, we will say that \( y_1 \) IN-causes \( y_2 \), and will write \( y_1 \Rightarrow y_2 \).

The theme of this paper is that for evaluation of the magnitude of causal effects one is interested primarily in IN-causation (requiring satisfaction of both the subset and the sufficiency conditions), not just causation (requiring satisfaction of just the subset condition).

If \( y_1 \rightarrow y_2 \) we can calculate the effects \( \Delta y_1 \) and \( \Delta y_2 \) of an intervention \( \Delta x_k \) on \( x_k \) for any \( x_k \) in the external set of \( y_1 \). The effect of \( y_1 \) on \( y_2 \) induced by the intervention on \( x_k \) then equals \( (\Delta y_2/\Delta x_k)/(\Delta y_1/\Delta x_k) \). If the sufficiency condition is satisfied this expression will take on the same value for all \( k \), so the lack of specificity about how the change in \( y_1 \) is implemented is immaterial. Accordingly, we have \( y_1 \Rightarrow y_2 \). If, on the other hand, the sufficiency condition fails then the expression will take different values for

\(^5\) In this paper the subset condition is a condition we impose on models to assure that causation is antisymmetric. Hausman ([8], Ch. 4) had a different take on what we call the subset condition. Hausman’s \textit{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.
different $k$, reflecting the fact that different implementations of $\Delta y_1$ imply different effects on $y_2$. If we have only $y_1 \rightarrow y_2$ we know that any external variable that affects $y_1$ also affects $y_2$, but no single constant connects $y_1$ and $y_2$ causally.

In the causal form of a model the equations are written so as to reflect the model’s IN-causal structure. Starting from the solution form of the model and having in hand a set of restrictions on the parameters of that 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 $y_j$.\footnote{External (internal) variable $x_1$ ($y_1$) is an ancestor of internal variable $y_2$ if $x_1 \Rightarrow y_2$ ($y_1 \Rightarrow y_2$). It is a parent of $y_2$ if it is an ancestor and in addition there is no internal variable $y_3$ such that $x_1 \Rightarrow y_3 \Rightarrow y_2$ ($y_1 \Rightarrow y_3 \Rightarrow y_2$).} In the causal form of the model, like the solution form, each equation has one of the internal variables on the left-hand side. The equation for each internal variable $y_i$ that has no internal variables as causal parents coincides with the corresponding equation in the solution form of the model (that is, 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 can be written in the form

$$y_j \leftarrow a_{ji}y_i + b_{jk}x_k.$$ \hfill (1)

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

Rather than deriving the IN-causal form from assumed restrictions on the coefficients of the solution form, one can start from a specification of the IN-causal form and derive the associated parameter restrictions by calculating the solution form. In doing this, observe that even though IN-causal forms are structural (in the sense that every internal variable is written as a function
of the other internal variables and one or more external variables), not every structural model qualifies as a causal form. We will discuss examples below. If one starts from a structural model that is not interpretable as a causal model, one determines the causal ordering by calculating the solution form and determining the causal ordering from whatever parameter restrictions are available from the specified structural model.

Any model can be written in structural form by applying algebraic operations to the model’s solution equations. In fact, any model can be written in recursive form by imposing an ordering on the internal variables and replacing each equation by the projection of one of the internal variables on the external variables and the internal variables lower in the ordering. Doing so does not change the causal ordering. In particular, doing so does not imply that the internal variables appearing on the right-hand side of an equation IN-cause the variable on the left-hand side, as is required for the causal form.

Beginning with a structural model, we have an algorithm to determine whether its coefficients incorporate IN-causation. The analyst (1) computes the solution form of the model, (2) determines its causal ordering by checking whether the conditions for \( y_i \Rightarrow y_j \) are satisfied for all \( i, j \), and (3) computes the causal form. If one ends with the same model that one began with, causation is implementation-neutral. If not, implementation neutrality fails for at least one of the pairs \( y_i \rightarrow y_j \).

One can represent the causal form of the model by a causal diagram. For variables \( y_i \) without internal variables as IN-parents this consists of arrows drawn to \( y_i \) from each element of \( \mathcal{E}(y_i) \), as in a diagram of the solution form. For variables with internal IN-parents the arrows run to \( y_j \) from the IN-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 IN-parents.

Observe that under our characterization the IN-causal form does not include as arguments internal variables that are ancestors of some internal variable when these are not also parents. The corresponding convention applies to causal diagrams: no arrow directly connects variables with their ancestors when these are not direct parents.

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 IN-causation just presented is that it 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 IN-causation has the implication that an internal variable never has both an indirect effect on another variable via an IN-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 IN-causation (with the former distinct from the composition of the latter) must involve the ambiguity in causal attributions that accompanies failure of implementation neutrality.

Examples will make these results clear.

### 1.1 Examples

Consider the following model written in solution form:

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

The external sets for \( y_1 \) and \( y_2 \) are \( E(y_1) = \{x_1, x_2\} \) and \( E(y_2) = \{x_1, x_2, x_3\} \). The former is a strict subset of the latter, so the subset condition is satisfied, and we have \( y_1 \Rightarrow y_2 \).

Without parameter restrictions the sufficiency condition for \( y_1 \Rightarrow y_2 \) is not satisfied. However, if the condition

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

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

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

allowing replacement of (3) with

\[
y_2 = a_{21}y_1 + b_{23}x_3.
\]
We have $y_1 \Rightarrow y_2$. The IN-causal form of the model is

\begin{align*}
y_1 &\leftarrow b_{11}x_1 + b_{12}x_2 \\
y_2 &\leftarrow a_{21}y_1 + b_{23}x_3.
\end{align*}

To repeat, $y_1$ IN-causes $y_2$ in this case because all interventions on $x_1$ 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 on $x_1$ 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. If the condition (4) is not satisfied the IN-causal form of the model consists of the following equations:

\begin{align*}
y_1 &\leftarrow b_{11}x_1 + b_{12}x_2 \\
y_2 &\leftarrow b_{21}x_1 + b_{22}x_2 + b_{23}x_3,
\end{align*}

coinciding with the solution equations.

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.

As observed above, one can equally well begin by specifying a model in IN-causal form, as in (7)-(8). Using (2) to eliminate $y_1$ in (6) results in
Comparing this equation with the solution equation (3) for $y_2$ results in $a_{21}b_{11} = b_{21}$ and $a_{21}b_{12} = b_{22}$, agreeing with (5). Thus writing a model in IN-causal form is equivalent to assuming the parameter restrictions on the solution form associated with the assumed causal ordering.

We noted above that some models that are acceptable as structural models are inadmissible as IN-causal models. For example, consider the model

$$y_1 = a_{12}y_2 + b_{11}x_1$$  \hspace{1cm} \text{(12)}

$$y_2 = a_{21}y_1 + b_{22}x_2$$  \hspace{1cm} \text{(13)}

$$y_3 = a_{31}y_1 + a_{32}y_2 + b_{33}x_3.$$  \hspace{1cm} \text{(14)}

This is an acceptable block-recursive structural model, but not an acceptable IN-causal model. Rewritten as an IN-causal model (by replacing $=$ with $\Leftarrow$), the model (12)-(14) would take the form

$$y_1 \Leftarrow a_{12}y_2 + b_{11}x_1$$  \hspace{1cm} \text{(15)}

$$y_2 \Leftarrow a_{21}y_1 + b_{22}x_2$$  \hspace{1cm} \text{(16)}

$$y_3 \Leftarrow a_{31}y_1 + a_{32}y_2 + b_{33}x_3.$$  \hspace{1cm} \text{(17)}

But this is inadmissible because it states that $y_1 \Rightarrow y_2$ and also $y_2 \Rightarrow y_1$, violating the antisymmetry of IN-causation.

Barring coefficient restrictions, the IN-causal form of the model (12)-(14) is

$$y_1 \Leftarrow b_{11}x_1 + b_{12}x_2$$  \hspace{1cm} \text{(18)}

$$y_2 \Leftarrow b_{21}x_1 + b_{22}x_2$$  \hspace{1cm} \text{(19)}

$$y_3 \Leftarrow b_{31}x_1 + b_{32}x_2 + b_{34}x_4 + b_{33}x_3,$$  \hspace{1cm} \text{(20)}

coinciding with the solution form. Therefore if one begins with a model like (12)-(14) that is not interpretable as an IN-causal model one cannot generally view the coefficients $a_{ij}$ of that model as measuring IN-causal effects due to failure of the implementation neutrality condition.
It might be thought that all structural models that are recursive, which the model (12)-(14) is not, would qualify as causal models. This is not so. Triangular models, which are recursive, provide a counterexample. Triangular models may specify causation between variables and ancestors that are not parents; in our definition of causal models such links are deleted. Consider the model

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

(21) (22) (23)

Here we have $y_1 \Rightarrow y_2$ and $y_2 \Rightarrow y_3$, implying also $y_1 \Rightarrow y_3$ by transitivity. Note that $y_1$ is excluded from the equation for $y_3$, reflecting the fact that although $y_1$ is an ancestor of $y_3$, it is not a parent as required for the causal form of a model as defined above.

If we had included $y_1$ in the equation for $y_3$, presumably with a coefficient $a_{31}$ not necessarily equal to $a_{21}a_{32}$, we would have a valid structural model, but not a model written in IN-causal form. The causal form for the respecified 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*}

(24) (25) (26)

is

\begin{align*}
y_1 & \leftarrow b_{11}x_1 + b_{12}x_2 \\
y_2 & \leftarrow a_{21}y_1 + b_{23}x_3 \\
y_3 & \leftarrow b_{31}x_1 + b_{32}x_2 + b_{33}x_3 + b_{34}x_4,
\end{align*}

(27) (28) (29)

so that the only internal variable pair in the ordering is $y_1 \Rightarrow y_2$. This reflects $y_1 \not\Rightarrow y_3$ and $y_2 \not\Rightarrow y_3$ due to failure of implementation neutrality.

2 Comparison with Simon

It is instructive to compare the representation of causation just presented to that of Simon’s classic [14] paper.
Simon characterized a structural model as a partially ordered set of self-contained sub-models, with some (or all) of the internal variables determined in each sub-model. 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 sub-models. Triangular models, in which each sub-model consists of a single equation, are the most extreme special case. In triangular models a complete ordering is defined on the internal variables, with the explanatory internal variables for each internal variable consisting of internal variables that are lower in the ordering. Under Simon’s definition of causation $y_1$ causes $y_2$ if $y_1$ enters the sub-model that determines $y_2$, and is determined in a lower-order sub-model.

There exist two major differences between the treatment of IN-causation here and that of Simon. The first is that IN-causation is defined from the solution form of the model, as opposed to its structural form as with Simon. It is easily verified that Simon’s criterion for causation coincides with our subset condition: if $y_1$ enters the equation for $y_2$ and 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 $y_2$. Thus we can use the same symbol, $\rightarrow$, to denote causation for Simon and to denote causation that is not necessarily implementation neutral, as defined above.

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 satisfaction of the subset condition by itself is necessary and sufficient for causation. Consequently causation under Simon’s definition may or may not be implementation neutral.

Whether causation as defined by Simon is implementation neutral depends on whether the assumed structural model is a model in IN-causal form. If it is, as with the model (7)-(8), then causation is implementation neutral. If it is not, as with the model of (12)-(14), then at least one of the causal pairs $y_i \rightarrow y_j$ is not implementation neutral. In that case, as noted above, one determines the IN-causal ordering by calculating the solution form of the model and applying whatever coefficient restrictions are available.
3 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 unobserved external variables. 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 IN-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 IN-causal coefficients empirically: the correlations among internal variables implied by the model’s causal structure cannot be disentangled from those induced by correlations among unobserved 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 and provide some preliminary results.

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 uninterpreted correlations among external variables.

An analyst who is uncomfortable with the assumption that the external variables $x_1$ and $x_2$ in two equations are independent can replace $x_2$ with $x_2 + \lambda x_1$, which allows for correlation even if $x_1$ and $x_2$ are independent. Of course, adopting such flexible specifications results in sparse causal orderings. As always, the analyst must deal with a tradeoff between how restrictive a model’s specification is and how rich its empirical implications are.

The assumptions just listed imply that if we have $y_1 \Rightarrow y_2$ the IN-causal coefficient measuring the effect of $y_1$ on $y_2$ is identified (apart from special cases in which observability is limited, as discussed below), and can be estimated consistently 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. Therefore the conditions for the Gauss-Markov theorem of linear regression are satisfied and least-squares regression coefficients provide optimal estimators.

The contrapositive of this statement is that the existence of econometric problems in the estimation of a parameter implies that the parameter is not one associated with IN-causation. For example, consider the system

\begin{align}
y_1 &= b_{11}x_1 + b_{12}x_2 \\
y_2 &= a_{21}y_1 + b_{21}x_1 \\
y_3 &= b_{32}x_2 + b_{33}x_3.
\end{align}

Here the external variables $x_1$, $x_2$ and $x_3$ are assumed to be independently distributed. Analysis of the solution form of this model reveals that the population parameter $a_{21}$ does not equal $\text{cov}(y_2, y_1)/\text{var}(y_1)$, the population regression coefficient of $y_2$ on $y_1$. This is so because $y_1$ and $x_1$ are correlated due to the presence of $x_1$ in the external set for $y_1$. Therefore $a_{21}$ is not estimated consistently by least squares on (31). Further, if $y_3$ and (32) are dropped from the model, then $a_{21}$ is not even identified. This can be seen by inspection of the solution form of the model (30)-(31).

However, in the presence of $y_3$ and (32) we have $a_{21} = \text{cov}(y_2, y_3)/\text{cov}(y_1, y_3)$, implying that $a_{21}$ is identified and can be estimated consistently by taking $y_3$ as an instrument. Here we make use of the fact that $y_3$ is correlated with $y_1$, due to the common presence of $x_2$ in their external sets, but not with $x_1$.

The result that the least-squares estimate of $a_{21}$ is not estimated consistently by least squares reflects the fact that $y_1$ does not IN-cause $y_2$, a fact that is also easily verified directly from the definition of IN-causation. Thus the inconsistency of the least-squares estimate of $a_{21}$ via a regression of $y_2$ on $y_1$ does not contradict our assertion that coefficients associated with IN-causal orderings are identified and easy to estimate by least squares.

The finding that IN-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 used a broader conception of causation—one that did not include implementation neutrality—than we focus on here. Parameters that are causal in the Cowles sense may or may not be identified and may or may not correspond to coefficients associated with IN-causation.
Here our attention is restricted to the smaller set of coefficients that are IN-causal.

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 IN-causation only when the associated variables are in fact IN-causally ordered, and whether two variables are IN-causally ordered depends on the coefficients that link observed internal variables to unobserved external variables. These coefficients, in contrast to those linking observed internal variables that are known (or assumed) to be causally ordered, are generally not identified. Therefore there may be no way to directly test models that make particular specifications of causation.

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 IN-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 the two are statistically independent or one ancestor causes the other. 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 \( y_2 \Rightarrow y_1 \).

\[ \text{In the passage discussed in note 5, Hausman [8] observed that common parents of a variable are "probabilistically independent". However, we noted that by this he meant} \]
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 pairs of these 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 IN-causation, no inference about the correlation among variables is possible.

A related point is that even under the assumption that external variables are independently distributed, IN-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). It is obviously true that the causal Markov condition fails in fully simultaneous systems. In that case no pair of internal variables is in the causal ordering, so the causal Markov condition would imply that all internal variables are statistically independent. This is generally not the case.

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 [15], 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 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.

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.
3.1 Example

Consider the following model:

\[ y_1 = x_1 + x_2 \]  
\[ y_2 = x_3 + x_4 \]  
\[ y_3 = x_1 + x_2 + x_3 + x_4 \]  
\[ y_4 = x_1 + x_2 - x_3 - x_4 \]

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

\[ y_1 \leftarrow x_1 + x_2 \]  
\[ y_2 \leftarrow x_3 + x_4 \]  
\[ y_3 \leftarrow y_1 + y_2 \]  
\[ y_4 \leftarrow y_1 - y_2 \]

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 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 gen-
erally 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.

4 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 vari-
ables are independently distributed and internal variables are fully observ-
able. 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 deter-
mining 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 gen-
erally display statistical dependence even if the unconditional distribution of
the external variables incorporates independence. This situation will not af-
fect the causal ordering of the variables, but it does invalidate the result that
the coefficients associated with the causal ordering can be estimated consist-
tently 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
observed values of 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 an external set that overlaps 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 $E(y_i)$ is that conditional on $y_i \geq 0$, and this does not generally have any independence property.

Berkson’s Paradox illustrates this. Suppose, following Elwert [3], 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.

## 5 Comparison with “Fixing”

The analysis of IN-causation outlined in this paper differs in major respects from what is found in the literature. Most important, interventions here consist exclusively of hypothetical alterations in the assumed values of external variables. In contrast, as noted above the usual treatment in the literature (based on Haavelmo [7] and Strotz and Wold [16]) 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 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. It corresponds to measuring a person’s height using a yard-
stick that expands or shrinks according to the height being measured.

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, doing so 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 equations are to be deleted. In contrast, our analysis of IN-causation applies in non-recursive models, although of course IN-causal relations among internal variables are likely to be sparse in models with large simultaneous blocks.

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 question of modularity does not come up because we are not modifying the model.

Modeling interventions by respecifying internal variables as external implies that causation is treated as if it were implementation neutral whether or not this treatment is justified. If implementation neutrality fails coefficients will be interpreted as IN-causal when they do not support that interpretation. It is far from clear why one would want to take this route. In general the answer to the question “What is the effect of \( y_1 \) on \( y_2 \)?” is properly viewed as possibly, but not necessarily, 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.

6 Application: Granger Causation

8
Granger [5] 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 IN-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 causation that is not spurious remained open.

To determine the relation between Granger causation and IN-causation, 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 to represent a general internal variable as above):

\[
\begin{align*}
m_t &= a_{my} y_t + b_{mm} m_{t-1} + b_{my} y_{t-1} + x_{1t} \quad (41) \\
y_t &= a_{ym} m_t + b_{ym} m_{t-1} + b_{yy} y_{t-1} + x_{2t}. \quad (42)
\end{align*}
\]

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

\[
\begin{align*}
m_t &= c_{mm} m_{t-1} + c_{my} y_{t-1} + u_{1t} \quad (43) \\
y_t &= c_{ym} m_{t-1} + c_{yy} y_{t-1} + u_{2t}. \quad (44)
\end{align*}
\]

GDP fails to Granger-cause the money stock if

\[8\] This section draws heavily on Cooley and LeRoy [2], although some of the discussion there is altered to accommodate the treatment here of causality.
The money stock is \textit{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. From (45) Granger non-causation is a necessary condition for strict exogeneity, but not a sufficient condition.

To determine what parameter restrictions are necessary for \( m_t \Rightarrow y_t \) we first write the solution form of the model under the assumption that \( m_t \) is strictly exogenous:

\[
m_t = x_{1t} + b_{mm} x_{1,t-1} + ... \tag{46}
\]

\[
y_t = a_{ym} x_{1t} + (a_{ym} b_{mm} + b_{ym}) x_{1,t-1} + x_{2t} + b_{yy} x_{2,t-1} + ... \tag{47}
\]

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 \( x_{1,t-1} \):

\[
\frac{1}{a_{ym}} = \frac{b_{mm}}{a_{ym} b_{mm} + b_{ym}}. \tag{48}
\]

Here the reasoning is exactly the same as in Subsection 1.1. This equality is satisfied if and only if \( b_{ym} = 0 \).\(^9\)

Thus even strict exogeneity of \( m \) is not a sufficient condition for interpreting the coefficient of \( m_t \) in equation (42) 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 IN-causation, one has to make two further restrictions on the model (41)-(42),

\(^9\)This condition is unchanged if more lagged terms of \( x_1 \) and \( x_2 \) are included explicitly in the solution form of the model.
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 IN-causation.

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.

7 Conclusion

In this paper we distinguish between two conceptions of causation, one a restricted version of the other. As is conventional, we use the term “causation” if any intervention that produces a change in the cause variable also produces a change in the effect variable. We direct attention to a restricted meaning for causation: IN-causation. One variable IN-causes another if the effect of all interventions that produce a given change in the cause variable induce the same change in the effect variable. 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 changes in the determinants of $y_1$ on $y_2$, which is unambiguous.

Philosophers sometimes reject this focus on 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 lan-
guage 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 IN-causation 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, so that that measure is not well defined? In that case there is no alternative to redirecting the analysis to implementation-specific interventions on the external variables, avoiding reference to the intermediate variable—the purported cause—which in fact plays no role in the causation.

It is not difficult to find passages in the philosophy literature where the idea of implementation neutrality is implicitly introduced. Further, it is not unusual to find use of the term “causation” reserved to settings in which implementation neutrality is satisfied. For example, Woodward [18] listed “invariance” among the requirements for causation: the effect of the cause variable on the effect variable should be invariant to interventions on other variables. He observed that “[o]ne condition for a successful intervention is that the intervention $I$ on $X$ [the cause variable] with respect to $Y$ should not cause $Y$ via a route that does not go through $X$, and that $I$ should be independent of any variable $Z$ that causes $Y$ but not through a route that goes through $I$ and $X$”. The language here is unclear, but if one reads $I$ as consisting of a variable in the external set of $X$, then Woodward’s criterion for a “successful intervention” corresponds to that for our implementation-neutral causation.\footnote{Woodward gave an example. Suppose that patients are treated or not treated for a medical condition based on a randomized assignment mechanism such as a coin toss. So stated, the assignment mechanism is an IN-cause (assuming that the treatment is effective) of remission of the condition. But suppose that another doctor influences the outcome of the coin toss using a magnet, and does so to ensure that patients with a strong immune system get the treatment. This alteration invalidates implementation neutrality. In our terminology the state of the patient’s immune system is an external variable for the use of the magnet, and the external set for the use of the magnet is a proper subset of the external set of the variable representing the assignment mechanism. The sufficiency condition for causation of the remission variable by the assignment variable is not satisfied. This is so because the variable representing the strength of the immune condition also affects the remission variable via a direct path.}

Critics of the analysis of causation presented here express the view that we
are departing from the ordinary-language usage of “causation”. The opposite is the case. Under the ordinary-language usage of “causation” the question “What is the effect of $y_1$ on $y_2$?” would be “It depends on what causes the variation in $y_1$” in settings where the conditions for implementation-neutral causation fail. It would make sense to avoid the term “causation” when the conditions for implementation-neutrality fail, although we have not done so.

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 IN-causal-form models coincide with the solution form in completely simultaneous models is not a flaw of the conception of causation developed here, as one is led to conclude by some discussions. Rather, the absence of pairs of internal variables in that case is a simple acknowledgement of the fact—familiar to economists from elementary analysis of supply-demand systems—that in simultaneous systems internal variables cannot necessarily be interpreted as causing each other.\(^{11}\)

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 interest to conduct a systematic comparison of inference from causal diagrams in the two 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 that use purported measures of causal magnitudes 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; a great deal remains to be done.

\(^{11}\)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.
References


