C

Causality in Economics and Econometrics

Kevin D. Hoover

Abstract

Economics was conceived as early as the classical period as a science of causes. The philosopher–economists David Hume and J. S. Mill developed the conceptions of causality that remain implicit in economics today. This article traces the history of causality in economics and econometrics, showing that different approaches can be classified on two dimensions: process versus structural approaches, and a priori versus inferential approaches. The variety of modern approaches to causal inference is explained and related to this classification. Causality is also examined in relationship to exogeneity and identification.

Keywords

Aristotle; Causal inference; Causality in economics and econometrics; Correlation; Cowles Commission; Econometrics; Edgeworth, F. Y.; Endogeneity and exogeneity; Granger–Sims causality; Graph theory; Hume, D.; Identification; Index numbers; Induction; Instrumental variables; Jevons, W. S.; Microfoundations; Mill, J. S.; Mises, L. von; Natural experiments; Observational equivalence; Process analysis; Quetelet, A.; Rational expectations; Regression; Robbins, L. C.; Simon, H.; Smith, A.; Statistical inference; Structural vector autoregressions; Tinbergen, J.; Vector autoregressions; Wold, H. O. A

JEL Classifications

B4

Philosophers of Economics and Causality

The full title of Adam Smith's great foundational work, *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), illustrates the centrality of causality to economics. The connection between causality and economics predates Smith. Starting with Aristotle, the great economists are frequently also the great philosophers of causality. Aristotle's contributions to economics are found principally in the *Topics*, the *Politics*, and the *Nicomachean Ethics*, while he lays out his famous four causes (material, formal, final and efficient) in the *Physics*. Material and formal causes are among the concerns of economic ontology, a subject addressed by philosophers of economics (see, for example, Mäki 2001) albeit rarely by

This chapter was originally published in *The New Palgrave Dictionary of Economics*, 2nd edition, 2008. Edited by Steven N. Durlauf and Lawrence E. Blume

Palgrave Macmillan (ed.), *The New Palgrave Dictionary of Economics*,
DOI 10.1057/978-1-349-95121-5_2227-1

practicing economists. Sometimes, as for example in Karl Marx's grand theory of capitalist development, economists have appealed to final causes or teleological explanation (for a defence, see Cohen 1978; for a general discussion, see Kincaid 1996). But, for the most part, taking physical sciences as a model, economics deals with efficient causes. What is it that makes things happen? What explains change? (See Bunge 1963, for a broad account of the history and philosophy of causal analysis.)

The greatest of the philosopher/economists, David Hume, set the tone for much of the later development of causality in economics. On the one hand, economists inherited from Hume the sense that practical economics was essentially a causal science. In 'On Interest', Hume (1742, p. 304) writes:

> it is of consequence to know the principle whence any phenomenon arises, and to distinguish between a cause and a concomitant effect. Besides that the speculation is curious, it may frequently be of use in the conduct of public affairs. At least, it must be owned, that nothing can be of more use than to improve, by practice, the method of reasoning on these subjects, which of all others are the most important; though they are commonly treated in the loosest and most careless manner.

On the other hand, Hume doubted whether we could ever know the essential nature of causation 'in the objects' (Hume 1739, p. 165). Coupled with a formidable critique of inductive inference more generally, Hume's scepticism has contributed to a wariness about causal analysis in many sciences, including economics (1739, 1777). The tension between the epistemological status of causal relations and their role in practical policy runs through the history of economic analysis since Hume.

History

Hume's Foundational Analysis

Although Hume's dominant concerns are moral, historical, political, and social (including economic), physical illustrations serve as his paradigm causal relationships. *A* (say, a billiard ball) strikes *B* (another ball) and causes it to move. Any analysis must address two key features of causality: first, causes are asymmetrical (in general, if *A* causes *B*, *B* does not cause *A*). Hume sees temporal succession (the movement of *A* precedes the movement of *B*) as accounting for asymmetry. Second, causes are effective. A cause must be distinguished from an accidental correlation and must bring about its effect. Hume sees spatial contiguity (the balls touch) and necessary connection (the movement of *B* follows of necessity from the movement of *A*) as distinguishing causes from accidents and establishing their effectiveness.

Hume was famously sceptical of any idea that could not be traced either to logical or mathematical deduction or to direct sense experience. He asks, whence comes the idea of the necessary connection of cause and effect? It cannot be deduced from first principles. So, he argues that our idea of necessary connection, which he concedes is the most characteristic element of causality, can arise only from our experience of the constant conjunction of particular temporal sequences. But this then implies that causality stands on a very weak foundation. For one corollary of Hume's belief that all ideas are based either in logic or sense experience was that we do not have any secure warrant for inductive inference. Neither logic nor experience (unless we beg the question by implicitly assuming the truth of induction) gives us secure grounds from observing instances to inferring a general rule. Therefore, what we regard as necessary connection in causal inference is really more of habit of mind without clear warrant. Causes may be necessarily connected to effects; but, for Hume, we shall never know in what that necessary connection consists.

While later philosophers have differed with Hume on the analysis of causality, his views were instrumental in setting the agenda, not only for philosophical discussions, but for practical causal analysis as well.

The 19th Century: Logic and Statistics

Even more influential than Hume in shaping economics, John Stuart Mill, another philosopher/economist, was less sceptical about causal inference in general, but more sceptical about its application to economics. In his *System of Logic* (1851), Mill advanced his famous canons of induction: the methods of (*a*) agreement, (*b*) difference, (*c*) joint (or double) agreement and difference, (*d*) residues, and (*e*) concomitant variations. For example, according to the method of difference, if we have two sets of circumstances, one in which a phenomenon occurs and one in which it does not, and the circumstances agree in all but one respect, that respect is the cause of the phenomenon. Mill's canons are essentially abstractions from the manner in which causes are inferred in controlled experiments. As such, Mill doubted that the canons could be easily applied to social or economic situations, in which a wide variety of uncontrolled factors are obviously relevant. Mill argued that economics was what Daniel M. Hausman (1992) has called an 'inexact and separate science', whose general principles were essentially known a priori and which held only subject to *ceteris paribus* clauses. Mill's apriorism proved to be hugely influential in later economics. Lionel Robbins (1935) expressed considerable scepticism about the place of empirical studies within economic science. Some Austrian economists, such as Ludwig von Mises (1966), went so far as to deny that economics could be an empirical discipline at all. Mill's apriorism also influenced those economists who see economic theory as similar to physical theory as a domain of universal laws.

Other 19th-century economists were less sceptical about the application of causal reasoning to economic data. For instance, W. Stanley Jevons (1863) pioneered the construction of index numbers as the core element of an attempt to prove the causal connection between inflation and the increase in worldwide gold stocks after 1849. Jevons's investigation can be interpreted as an application of Mill's method of residues (see Hoover and Dowell 2001). He saw the various idiosyncratic relative price movements, owing to supply and demand for particular commodities, as cancelling out to leave the common factor that could only be the effect of changes in the money stock.

The 19th century witnessed extensive development in the theory and practice of statistics (Stigler 1986). Inference based on statistical distributions and correlation measures was closely connected to causality. Adolphe Quetelet envisaged the inferential problem in statistics as one of distinguishing among constant, variable, and accidental causes (Stigler 1999, p. 52). The economist Francis Ysidro Edgeworth pioneered tests of statistical significance (in fact Edgeworth may have been the first to use this phrase). He glossed the finding of a statistically significant result as one that 'comes by cause' (Edgeworth 1885, pp. 187–8).

The 20th Century: Causality and Identification

Further developments of statistical techniques, such as multiple correlation and regression, in the 20th century were frequently associated with causal inference. It was fairly quickly understood that, unlike correlation, regression has a natural direction: the regression of Y on X does not produce coefficient estimates that are the algebraic inverse of those from the regression of X on Y. The direction of regression should respect the direction of causation.

By the early 20th century, however, the dominant vision of economics was one in which prices and quantities are determined simultaneously. This is as much true for Alfred Marshall (1930), who is often described (not perfectly accurately) as an advocate of partial equilibrium analysis, as it is for Léon Walras (1954), the principal font of modern general equilibrium analysis. Simultaneity does not necessarily rule out causal order, though it does complicate causal inference. Although regressions may have a natural causal direction, there is nothing in the data on their own that reveal which direction is the correct one – each is an equally eligible rescaling of a symmetrical and non-causal correlation. This is a problem of observational equivalence. And it is the obverse side of the now familiar problem of econometric identification: in this case, how can we distinguish a supply curve from a demand curve? The problem of identification was pursued throughout most of the first half of the 20th century until the fairly complete treatment by the

Cowles Commission at mid-century (Koopmans 1950; Hood and Koopmans 1953; see Morgan 1990, for a thorough treatment of the history of the identification problem).

The standard solution to the identification problem is to look for additional causal determinants that discriminate between otherwise simultaneous relationships. Both the supply of milk and demand for milk depend on the price of milk. If, however, the supply also depends on the price of alfalfa used to feed the cows and the demand also on the daily high temperature (which affects the demand for milk to make ice cream), then supply and demand curves can be identified separately. Identification can be viewed through the glasses of simultaneous equations, pushing causality into the background, or it can be viewed as a problem in causal articulation. In the first case, economists frequently use the language of exogenous variables (the price of alfalfa, the temperature) and endogenous variables (the price and quantity of milk). Exogenous variables can also be regarded as the causes of the endogenous variables. From the 1920s to the 1950s, different economists placed different emphasis on the causal aspects of identification (Morgan 1990) and the various papers reprinted in Hendry and Morgan (1995).

Modern econometrics can be dated from the development of structural econometric models following the pioneering work in the 1930s of Jan Tinbergen, the conceptual foundations of probabilistic econometrics in Trgyve Haavelmo's (1944) 'Probability approach to econometrics', and the technical elaboration of the identification problem in the two Cowles Commission volumes. Structural models did not in themselves necessarily favor the language of identification over the language of causality. Indeed, in Tinbergen's (1951) textbook, dynamic, structural models are explicated with a diagram that uses arrows to indicate causal connections among time-dated variables. Nevertheless, after the econometric work of the Cowles Commission, two approaches can be clearly distinguished.

One approach, associated with Hermann Wold and known as *process analysis,* emphasized the asymmetry of causality, typically grounded it in Hume's criterion of temporal precedence (Morgan 1991). Wold's process analysis belongs to the time-series tradition that ultimately produced Granger causality and the vector autoregression (see section "Alternative Approaches to Causality in Economics").

The other approach, associated with the Cowles Commission, related causality to the invariance properties of the structural econometric model. This approach emphasized the distinction between endogenous and exogenous variables and the identification and estimation of structural parameters. Implicitly, structural modellers accepted Mill's a priori approach to economics. While they differed from Mill in their willingness to conduct empirical investigations, the selection of exogenous (or *instrumental*) variables was seen to be the province of a priori economic theory – a maintained assumption rather than something to be learned from data itself.

In his contribution to one of the Cowles Commission volumes, Herbert Simon (1953) showed that causality could be defined in a structural econometric model, not only between exogenous and endogenous variables, but also among the endogenous variables themselves. And he showed that the conditions for a well-defined causal order are equivalent to the well-known conditions for identification. Despite the equivalence, with the demise of process analysis and the ascendancy of structural econometrics – aided indirectly perhaps by a revival of Humean causal scepticism among the logical-positivist philosophers of science – causal language in economics virtually collapsed between 1950 and about 1990 (Hoover 2004).

Alternative Approaches to Causality in Economics

Different approaches to causality can be classified along two lines as shown in Fig. 1. One the one hand, approaches may emphasize structure or process. On the other hand, approaches may rely on a priori identifying assumptions or they may seek to infer causes from data. The upper left cell, the a priori structural approach, represented by the

Causality in Economics and Econometrics, Fig. 1 Classification of approaches to causality in economics

	Structural	*Process*
A Priori	Cowles Commission: Koopmans (1953); Hood and Koopmans (1953)	Zellner (1979)
Inferential	Simon (1953) Hoover (1990; 2001) Favero and Hendry (1992) Natural experiments: Angrist and Krueger (1999; 2001)	Granger (1969) Vector autoregressions: Sims (1980)

Cowles Commission, dominated economics for most of the postwar period. But since we already discussed it at some length in section "History", and since it was largely responsible for turning the economics profession away from *explicit* causal analysis, we add nothing more about it here and instead turn to the other cells in Fig. 1.

The Inferential Structural Approach

The most important of the inferential structural approaches is due to Simon (1953). Simon eschews temporal order as a basis for causal asymmetry and, instead, looks to recursive structure. As we observed in section "History", Simon's account is closely related to the Cowles Commission's structural approach. Consider the bivariate system:

$$Y_t = \theta X_t + \varepsilon_{1t}, \tag{1}$$

$$X_t = \varepsilon_{2t}, \tag{2}$$

where the random error terms ε_{it} are independent, identically distributed and θ is a parameter. Simon says that X_t causes Y_t, because X_t is recursively ordered ahead of Y_t. One knows all about X_t without knowing about Y_t, but one must know the value of X_t to determine the value of Y_t. Equations (1) and (2) also appear to show that any intervention in (2), say a change in the variance of ε_{2t}, would transmit to (1); while any intervention in (1), say a change in θ or the variance of ε_{1t}, would not transmit to (2). Apparently, X_t could then be used to control Y_t.

Unfortunately, merely being able to write an accurate description of the two variables in the form of (1) and (2) does not guarantee either the apparent asymmetry of information or control. The same data can be repackaged into a statistically identical form with an apparently different causal order. For example, consider the following related system:

$$Y_t = \omega_{1t}, \tag{3}$$

$$X_t = \delta Y_t + \omega_{2t}, \tag{4}$$

where $\delta = \dfrac{\theta \mathrm{var}(\varepsilon_2)}{\theta^2 \mathrm{var}(\varepsilon_2) + \mathrm{var}(\varepsilon_1)}$, $\omega_{1t} = \varepsilon_{1t} + \theta\varepsilon_{2t}$, and $\omega_{2t} = (1 - \delta\theta)\,\varepsilon_{2t} - \delta\varepsilon_{1t}$.

Equations (3) and (4) are derived from Eqs. (1) and (2). The details of the algebra are not important. Essentially, (3) and (4) are linear combinations of (1) and (2) with multiplicative factors carefully chosen, so that the error terms ω_{1t} and ω_{2t} are uncorrelated. Such linear combinations preserve the values of X_t and Y_t and their statistical likelihood (that is, the two systems of equations have the same reduced form) and, so, describe the data equally well. Equations (3) and (4) have a form analogous to (1) and (2); but, on Simon's criterion, it appears that Y_t causes X_t on Simon's criterion. While it looks like the key parameters for (3) and (4) are derived from those of (1) and (2), we could have taken (3) and (4) as the starting point and derived (1) and (2) symmetrically. What we would like to do is to replace the equal signs with arrows that show that the causal direction runs from the right-hand to the left-hand sides in the regression equations in one of the systems, but not in the other. Unfortunately, there is no way to do this, no choosing between the systems, on the basis of a single set of data by itself. This is the problem of observational equivalence again.

The a priori approach of the Cowles Commission relies on economic theory to provide appropriate identifying assumptions to resolve the observational equivalence. Christopher Sims (1980) attacked the typical application of the Cowles Commission's approach to structural macroeconometric models as relying on 'incredible' identifying assumptions: economic theory was simply not informative enough to do the job. But Simon, who was otherwise supportive of the conception of causality in the Cowles Commission, took a different tack.

Simon sees the problem as choosing between two alternative sets of parameters: which set contains the structural parameters, $\{\theta$ and the variances of the $\varepsilon_{it}\}$ or $\{\delta$ and the variances of the $\omega_{it}\}$? Simon suggested that experiments – either controlled or natural – could help to decide. If, for example, an experiment could alter the conditional distribution of X_t without altering the marginal distribution of Y_t, then it must be that Y_t causes X_t, because this would be possible only if a structure like (3) and (4) characterized the data. If it did, a change in the conditional distribution would involve either δ or the variance of ω_{2t}, neither of which would affect the variance of ω_{1t}. In contrast, if (1) and (2) truly characterized the causal structure of the data, a change to the conditional distribution of X_t would, in fact, involve a change to the variance of ε_{2t}, which, according to the equivalences above, would alter either δ or the variance of ω_{2t}. Similar relationships of stability and instability in the face of changes to the marginal distribution can also be demonstrated (Hoover 2001, ch. 7). The appeal to experimental evidence is what marks Simon's approach out as inferential rather than a priori.

Hoover (1990, 2001) generalizes Simon's approach to the type of nonlinear systems of equations found in modern rational-expectations models. He shows that Simon's idea of natural experiments can be operationalized by coordinating historical, institutional, or other non-statistical information with information from structural break tests on what, in effect, amounts to the four regressions corresponding to (1), (2), (3), and (4) above generalized to include lagged dynamics. With allowances for complications introduced by rational expectations, the key idea is that, in the true causal order, interventions that alter the parameters governing the true marginal distribution do not transmit forward to the conditional distribution (characterized by (1) or (4)) nor do interventions in the true conditional distribution transmit backward to the marginal distribution (characterized by (2) or (3)). Since the true structural parameters are not known a priori, non-statistical information is important in identifying an intervention as belonging to the process governing one variable or another.

Although avoiding the term 'causality', Favero and Hendry's (1992) analysis of the Lucas critique in terms of 'super-exogeneity' is also a variant on Simon's causal analysis (Ericsson and Irons 1995; Hoover 2001, ch. 7). Super-exogeneity is essentially an invariance concept (Engle et al. 1983). Favero and Hendry find evidence against the Lucas critique (non-invariance in the face of changes in policy regime) in the super-exogeneity of conditional probability distributions in the face of structural breaks in marginal distributions – the same sort of evidence that Hoover cites as helping to identify causal direction.

The recent revival of causal analysis in microeconomics in the guise of 'natural experiments', although apparently developed independently of Simon, nonetheless proceeds in much the same spirit as Hoover's version of Simon's approach (Angrist and Krueger 1999, 2001). This literature typically employs the language of instrumental variables. A natural experiment is a change in a policy or a relevant environmental factor that can be identified non-statistically. Packaged as an econometric instrument, the experiment can be used – in much the same way that variations in alfalfa prices and temperature were used in the example in section "History" – to identify the underlying relationships and to measure the causally relevant parameters.

While the development of structural approaches in econometrics has largely been independent, there is some cross-fertilization between economists and philosophers (for example, Simon and Rescher 1966); and recently philosophers of causality have looked to economics for inspiration and examples (for example, Cartwright 1989; Woodward 2003).

The Inferential Process Approach

Perhaps the most influential explicit approach to causality in economics is due to Clive W. J. Granger (1969). Granger causality is an inferential approach, in that it is data-based without direct reference to background economic theory; and it is a process approach, in that it was developed to apply to dynamic time-series models (see Granger–Sims causality in this dictionary for technical details). Granger–Sims causality is an example of the modern probabilistic approach to causality, which is a natural successor to Hume (for example, Suppes 1970). Where Hume required constant conjunction of cause and effect, probabilistic approaches are content to identify cause with a factor that raises the probability of the effect: A causes B if $\mathrm{P}(B|A) > \mathrm{P}(B)$, where the vertical '|' indicates 'conditional on'. The asymmetry of causality is secured by requiring the cause (A) to occur before the effect (B) (but the probability criterion is not enough on its own to produce asymmetry since $P(B|A) > P(B)$ implies $\mathrm{P}(A|B) > \mathrm{P}(A)$).

Granger's (1980) definition is more explicit about temporal dynamics than is the generic probabilistic account, and it is cast in terms of the incremental predictability of one variable conditional on another:

X_t Granger-causes Y_{t+1} if $\mathrm{P}(Y_{t+1}|$ all information dated t and earlier$) \neq \mathrm{P}(Y_{t+1}|$ all information dated t and earlier omitting information about X).

This definition is conceptual, as it is impracticable to condition on *all* past information.

In practice, Granger causality tests are typically implemented through bivariate regressions. As an illustration, consider the regression equations:

$$Y_t = \Pi_{11} Y_{t-1} + \Pi_{12} X_{t-1} + \upsilon_{1t}, \qquad (5)$$

$$X_t = \Pi_{21} Y_{t-1} + \Pi_{22} X_{t-1} + \upsilon_{2t}, \qquad (6)$$

where the Π_{ij} are parameters, and the υ_{it} are random error terms. In practice, lag lengths may be larger than one, but far less than the infinity implicit in the general definition. X_t Granger-causes Y_{t+1} if $\Pi_{12} \neq 0$, and Y_t Granger-causes X_{t+1} if $\Pi_{21} \neq 0$.

Sims (1972) famously used Granger causality to demonstrate the causal priority of money over nominal income. Later, as part of a generalized critique of structural econometric models, Sims (1980) advocated vector autoregressions (VARs) – atheoretical time-series regressions analogous to Eqs. (1) and (2), but generally including more variables with lagged values of each appearing in each equation. In the VAR context, Granger causality generalizes to the multivariate case.

While Granger causality has something useful to say about incremental predictability, there is no close mapping between Granger causality and structural notions of causality on either the Cowles Commission's or Simon's accounts (Jacobs et al. 1979). Consider a structural model:

$$Y_t = \theta X_t + \beta_{11} Y_{t-1} + \beta_{12} X_{t-1} + \varepsilon_{1t}, \qquad (7)$$

$$X_t = \gamma Y_t + \beta_{21} Y_{t-1} + \beta_{22} X_{t-1} + \varepsilon_{2t}, \qquad (8)$$

where ε_{1t} and ε_{2t} are identically distributed, independent random errors and θ, γ, and the β_{ij}s are structural parameters. The independence of the parameters and the error terms implies that causality runs from the right-hand to the left-hand sides of each equation. Equations (5) and (6) can be seen as the reduced forms of (7) and (8).

We focus on X causing Y. X structurally causes Y if either θ or $\beta_{12} \neq 0$. And X Granger causes Y if $\Pi_{12} = \frac{\beta_{12} + \theta\beta_{22}}{1 - \theta\gamma} \neq 0$. Thus, if X Granger causes Y, then X structurally causes Y. Note, however, that this result is particular to the case in which (7) and (8) represents the universe, so that (5) and (6) represent the complete conditioning on past histories of relevant variables. If the universe is more complex and the estimated VAR does not capture the true reduced forms of the structural system, which in practice they may not, then the strong connection suggested here does not follow.

More interestingly, even if (5), (6), (7), and (8) are complete, structural causality does not necessarily imply Granger causality. Suppose that $\beta_{12} = \beta_{22} = 0$, but $\theta \neq 0$, then X structurally causes Y, but since $\Pi_{12} = 0$, X does not Granger cause Y.

Now suppose that X does not Granger cause Y. It does not necessarily follow that X does not structurally cause Y, since if θ, β_{12}, and $\beta_{22} \neq 0$, and $-\beta_{12}/\beta_{22} = \theta$, then it will still be true that $\Pi_{12} = 0$. This may appear to be an odd special case, but in fact conditions such as $-\beta_{12}/\beta_{22} = \theta$ arise commonly in optimal control problems in economics.

A simple physical example makes it clear what is happening. Suppose that X measures the direction of the rudder on a ship and Y the direction of the ship. The ship is pummeled by heavy seas. If the helmsman is able to steer on a straight course, effectively moving the rudder to exactly cancel the shocks from the waves, the direction of the rudder (in ignorance of the true values of the shocks) will not predict the course of the ship. The rudder would be structurally effective in causing the ship to turn, but it would not Granger-cause the ship's course.

The a Priori Process Approach

The upper right-hand cell of Fig. 1 is represented by Arnold Zellner's (1979) account of causality (cf. Keuzenkamp 2000, ch. 4, s. 4). Zellner's notion of causality is borrowed from the philosopher Herbert Feigl (1953, p. 408), who defines causation '. . . in terms of predictability according to law (or more adequately, according to a set of laws)'. On the one hand, Zellner opposes Simon and sides with Granger: predictability is a central feature of causal attribution, which is why his is a process account. On the other hand, he opposes Granger and sides with Simon: an underlying structure (a set of laws) is a crucial presupposition of causal analysis, which is why his is an a priori account.

Much obviously depends on what a law is. Zellner's own view is that a law is a (probabilistic) description of a succession of states of the world that holds for many possible boundary conditions and covers many possible circumstances. He couches his position in an explicitly Bayesian theory of inference. Feigl identifies causality with lawlikeness or predictability. It is the fact that formulae fit previously unexamined cases, as well as examined ones, which constitutes their lawlikeness. This is close to Simon's invariance criterion (the true causal order is the one that is invariant under the right sort of intervention).

The central problem, then, is how to distinguish laws from false generalizations or accidental regularities – that is, how to distinguish conditional relations invariant to interventions from regularities that are either not invariant or are altogether adventitious. Zellner believes that a theory serves as the basis for discriminating between laws and casual generalizations. Although Zellner's approach permits us to learn some things from the data, in keeping with the spirit of Bayesian inference, it does so within a narrowly defined framework (cf. Savage's 1954, pp. 82–91, 'small world' assumption). Economic theory in Zellner's account restricts the scope of an investigation a priori.

Zellner objects to Granger causality for two reasons. First, it is not satisfactory to identify cause with temporal ordering, as temporal ordering is not the ordinary, scientific or philosophical foundation of the causal relationship. Second, Granger's approach is atheoretical. In order to implement it practically, an investigator must impose restrictions – limit the information set to a manageable number of variables, consider only a few moments of the probability distribution (in our exposition, just the mean), and so forth. For Zellner, if these restrictions cannot be explained theoretically, Granger's methods will discover only accidental regularities.

Zellner explicitly criticizes Granger for ignoring the need for theoretical basis for empirical investigation – implicitly focusing on only one side of a process in which theory informs empirics and empirics inform theory. He criticizes Simon for defining cause to be a formal property of a model (recursive order) without making essential reference to empirical reality. Zellner's criticism is, however, more aptly directed at the Cowles

Commission's approach, since (as we saw in section "The Inferential Structural Approach") Simon distinguishes himself through tying causal order to empirical inference.

Structural Vector Autoregressions

Not all approaches to causality fall quite neatly into the cells of Fig. 1; or, more to the point, an approach that falls into one cell may morph into one that falls into another cell. The history of Sims's VAR program is an important case.

Sims (1980) advocated VARs as a reaction to the manner in which the Cowles Commission programme, which identified structural models through a priori theory, had been implemented (see section "The Inferential Process Approach"). From a causal perspective, it was closely related to Granger's analysis. Starting with VAR such as Eqs (5) and (6), Sims wished to work out how various 'shocks' would affect the variables of the system. This is complicated by the fact that the error terms in (5) and (6), which might be taken to represent the shocks, are not in general independent, so that a shock to one is a shock to both, depending on how correlated they are. Sims's initial solution was to impose an arbitrary orthogonalization of the shocks (a Choleski decomposition). In effect, this meant transforming (5) and (6) into a system like (6) and (7) and setting either θ or γ to zero. This amounts to imposing a recursive order on X_t and Y_t, such that the covariance matrix of the error terms is diagonal (that is, ε_{1t} and ε_{2t} are uncorrelated). A shock to X can then be represented by a realization of ε_{1t} and a shock to Y by a realization of ε_{2t}.

Initially, Sims treated the choice of recursive order as a matter of indifference. Criticizing the VAR program from the point of view of structural models, Leamer (1985) and Cooley and LeRoy (1985) pointed out that the substantive results (for instance, impulse-response functions and innovation accounts) depend on which recursive order is chosen. Sims (1982, 1986) accepted the point and henceforth advocated Structural vector autoregressions (SVARs). SVARs can be identified through the contemporaneous causal order only. So, for example, to identify (5) and (6), it is enough to assume that either θ or γ in (7) or (8) is zero; one need not make any assumptions about the β_{ij}s. Ironically, since the initial impulse behind the VAR programme was to avoid theoretically tenuous identifying assumptions, the choice of restrictions on contemporaneous variables used to transform the VAR into the SVAR are typically only weakly supported by economic theory.

Nevertheless, the move from the VAR to the SVAR is a move from an inferential to an a priori approach. It is also a move from a fully non-structural, process approach to a partially structural approach, since the structure of the contemporaneous variables, though not of the lagged variables, is fully specified. The SVAR approach can, therefore, be seen as straddling the cells on the first line of Fig. 1.

The Graph-Theoretic Approach to Causal Inference

A final approach to causality in economics sometimes provides another example of an inferential structural approach, and sometimes straddles the cells on the second line of Table 1. Graph-theoretic approaches to causality were first developed outside of economics by computer scientists (for example, Pearl 2000) and philosophers (for example, Spirtes et al. 2000), but have recently been applied within economics (Swanson and Granger 1997; Akleman et al. 1999; Bessler and Lee 2002; Demiralp and Hoover 2003).

The key ideas of the graph-theoretic approach are simple (see Demiralp and Hoover 2003 or Hoover 2005 for a detailed discussion). Any structural model can be represented by a graph in which arrows indicate the causal order. Equations (1) and (2) are represented by $X \rightarrow Y$ and Eqs. (3) and (4) by $Y \rightarrow X$. More complicated structures can be represented by more complicated graphs. Simultaneity, for instance, can be represented by double-headed arrows. The graphs allow us easily to see the dependence or independence among variables. Pearl (2000) and Sprites et al. (2000) demonstrate the isomorphism between causal graphs and the independence relationships encoded in probability distributions. This isomorphism allows conclusions about probability distributions to be derived from theorems proven using the mathematical techniques of graph theory.

Many of the results of graph-theoretic analysis are straightforward. Suppose that $A \rightarrow B \rightarrow C$ (that is, A causes B causes C). A and C would be probabilistically dependent; but, conditional on B, they would be independent. Similarly for $A \leftarrow B \leftarrow C$. In each case, B is said to *screen* A from C. Suppose that $A \leftarrow B \leftarrow C$. Then, once again A and C would be dependent, but conditional on B, they would be independent. B is said to be the *common cause* of A and C. Now suppose that A and B are independent conditional on sets of variables that exclude C or its descendants, and $A \rightarrow C \leftarrow B$, and none of the variables that cause A or B directly causes C. Then, conditional on C, A and B are dependent. C is called an *unshielded collider* on the path ACB. (A *shielded* collider would have a direct link between A and B.) These are the simplest relationships of probabilistic dependence and independence. More complex ones may also obtain in which A is independent of B only conditional on more than one other variable (say, C and D).

A number of causal search algorithms have been developed (Sprites et al. 2000). These start with information about correlations (or other tests of unconditional and conditional statistical independence) among variables. The most common of these, the PC algorithm, assumes that graphs are strictly recursive (known in the literature as *acyclical*) and starts with a graph in which all variables are causally connected with an unknown causal direction (represented by the headless arrow, '—'). It then tests for independence among pairs of variables, conditioning on sets of zero variables, then one, then two, and so forth until the set of variables is exhausted. Whenever it finds independence, it removes the causal connection between the variables in the graph. Once the graph is pared down as far as it can be, it considers triples of variables in which two are conditionally independent but are connected through a third. If conditioning on that third variable renders the variables conditionally dependent, then that variable is an unshielded collider and it is connected to the other two variables with causal arrows running toward it. After all the unshielded colliders have been identified, further logical analysis can be used to orient additional causal arrows. For example, we might reason as follows: suppose we have a triple $A \rightarrow C \rightarrow B$; unless the causal arrow runs away from C toward B, C would be identified as an unshielded collider; but C was not identified as an unshielded collider earlier in the search; therefore, the causal arrow must run away from C towards B, so that the graph becomes $A \rightarrow C \rightarrow B$.

Sometimes the data allow the complete orientation of a causal graph, but sometimes some causal connections are left undirected. In this case, the graph marks out an equivalence class, and the algorithm has identified 2! causal graphs consistent with the empirical probability distribution, where $n =$ the number of undirected causal connections.

While most applications of graph-theoretic methods assume that the true causal structures are recursive (that is, strictly acylical), economics frequently treats variables that are cyclical or simultaneously determined. Although the recursiveness assumption is restrictive, it is an assumption that is also frequently made in the SVAR literature. Some progress has been made in developing graph-theoretic search algorithms for cyclical or simultaneous causal systems (Pearl 2000, pp. 95–6, 142–3; Richardson 1996; Richardson and Spirtes 1999).

Swanson and Granger (1997) showed that estimates of the error terms of the VAR (the υ_{it} in Eqs (5) and (6)) can be treated as the original time-series variables purged of their dynamics. A causal order identified on such variables corresponds to the causal order necessary to convert a VAR into an SVAR. Demiralp and Hoover (2003) present Monte Carlo evidence that the PC algorithm is effective at selecting the true causal connections among variables and, when signal strengths are high enough, moderately effective at directing them correctly. Search algorithms can, therefore, reduce or even eliminate the need to appeal to a priori theory when identifying the causal order of an SVAR.

Where Simon's approach looked for relatively important interventions as a basis for causal inference to a structure, the graph-theoretic approach uses relatively routine random variations to identify patterns of conditional independence that map out causal structures. The two approaches are complementary: Simon's approach may be used to resolve the observational equivalence reflected in causal connections that remain undirected after the application of a causal search algorithm.

From Metaphysics to Econometric Practice

The analysis of causation was originally a branch of metaphysics. In moving from the scholastic to the practical, two deep divisions appeared among economists.

The first is the divide between those who believed that causality in economics could be characterized by relatively simple uniformities (the process approaches) and those who believed that it must be characterized by a rich understanding of the underlying mechanisms (the structural approaches). Economists debate the appropriate level at which to characterize either the uniformities or the mechanisms – individual or aggregate. But this debate over the microfoundations of macroeconomics is another story. The second divide is between those who believe that economic logic itself gives privileged insight into economic behaviour (a priori approaches) and those who believe that we must learn about economic behaviour principally through observation and induction (the inferential approaches).

These are old debates – unlikely to be resolved decisively to the satisfaction of all economists in the near future. How one aligns oneself in them largely determines which particular approaches to causality appear to be compelling in practical economic research.

See Also

► Endogeneity and Exogeneity
► Granger–Sims Causality
► Graph Theory
► Hume, David
► Identification
► Mill, John Stuart
► Simon, Herbert A.
► Structural Vector Autoregressions
► Vector Autoregressions

Bibliography

Akleman, D.G., D.A. Bessler, and D.M. Burton. 1999. Modeling corn exports and exchange rates with directed graphs and statistical loss functions. In *Computation, causation, and discovery*, ed. C. Glymour and G.F. Cooper. Menlo Park/Cambridge, MA: MIT Press and American Association for Artificial Intelligence.

Angrist, J.D., and A.B. Krueger. 1999. Empirical strategies in labor economics. In *Handbook of labor economics*, ed. O. Ashenfelter and D. Card, Vol. 3-A. Amsterdam: North-Holland.

Angrist, J.D., and A.B. Krueger. 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* 15(4): 69–85.

Bessler, D.A., and S. Lee. 2002. Money and prices: U.S. data 1869–1914 (a study with directed graphs). *Empirical Economics* 27: 427–446.

Bunge, M. 1963. *The place of the causal principle in modern science*. Cleveland: Meridian Books.

Cartwright, N. 1989. *Nature's capacities and their measurement*. Oxford: Clarendon Press.

Cohen, G.A. 1978. *Karl Marx's theory of history: A defense*. Princeton: Princeton University Press.

Cooley, T.F., and S.F. LeRoy. 1985. Atheoretical macroeconomics: A critique. *Journal of Monetary Economics* 16: 283–308.

Demiralp, S., and K.D. Hoover. 2003. Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and Statistics* 65(Supplement): 745–767.

Edgeworth, F.Y. 1885. Methods of statistics. In Royal Statistical Society of Britain, *Jubilee volume of the Statistical Society.* London: E. Stanford.

Engle, R.F., D.F. Hendry, and J.-F. Richard. 1983. Exogeneity. *Econometrica* 51: 277–304.

Ericsson, N., and J. Irons. 1995. The Lucas critique in practice: Theory without measurement. In *Macroeconometrics: Developments, tensions and prospects*, ed. K.D. Hoover. Boston: Kluwer.

Favero, C., and D.F. Hendry. 1992. Testing the Lucas critique: A review. *Econometric Reviews* 11: 265–306.

Feigl, H. 1953. Notes on causality. In *Readings in the philosophy of science*, ed. H. Feigl and M. Brodbeck. New York: Appleton-Century-Crofts.

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

Granger, C.W.J. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2: 329–352.

Haavelmo, T. 1944. The probability approach in econometrics. *Econometrica* 12(Supplement), iii–vi, 1–115.

Hausman, D.M. 1992. *The inexact and separate science of economics*. Cambridge: Cambridge University Press.

Hendry, D.F., and M.S. Morgan, ed. 1995. *The foundations of econometric analysis*. Cambridge: Cambridge University Press.

Hood, W., and T. Koopmans, ed. 1953. *Studies in econometric method*, Cowles commission monograph no. Vol. 14. New Haven: Yale University Press.

Hoover, K.D. 1990. The logic of causal inference: Econometrics and the conditional analysis of causality. *Economics and Philosophy* 6: 207–234.

Hoover, K.D. 2001. *Causality in macroeconomics*. Cambridge: Cambridge University Press.

Hoover, K.D. 2004. Lost causes. *Journal of the History of Economic Thought* 26: 149–164.

Hoover, K.D. 2005. Automatic inference of the contemporaneous causal order of a system of equations. *Econometric Theory* 21: 69–77.

Hoover, K.D., and M.E. Dowell. 2001. Measuring causes: Episodes in the quantitative assessment of the value of money. *History of Political Economy* 33: 137–161.

Hume, D. 1739. In *A treatise of human nature,* ed. L.-A. Selby-Bigge. Oxford: Clarendon Press, 1888n.d.-a.

Hume, D. 1742. Of interest. In *Essays: Moral, political, and literary,* ed. E.F. Miller. Indianapolis: Liberty Classics, 1985n.d.-c.

Hume, D. 1777. An enquiry concerning human understanding. In *Enquiries concerning human understanding and concerning the principles of morals*, 2nd ed., ed. L.A. Selby-Bigge. Oxford: Clarendon Press, 1902n.d.-b.

Jacobs, R.L., E.E. Leamer, and M.P. Ward. 1979. Difficulties in testing for causation. *Economic Inquiry* 17: 401–413.

Jevons, W.S. 1863. A serious fall in the value of gold ascertained, and its social effects set forth. In *Investigations in currency and finance,* 1884, Reprinted. New York: Augustus M. Kelley, 1964.

Keuzenkamp, H.A. 2000. *Probability, econometrics, and truth*. Cambridge: Cambridge University Press.

Kincaid, H. 1996. *Philosophical foundations of the social sciences: Analyzing controversies in social research*. Cambridge: Cambridge University Press.

Koopmans, T., ed. 1950. *Statistical inference in dynamic economic models,* Cowles Commission Monograph No. 10. New York: Wiley.

Leamer, E.E. 1985. Vector autoregressions for causal inference. In *Understanding monetary regimes*, Carnegie-Rochester conference series on public policy, ed. K. Brunner and A.H. Meltzer, Vol. 22. Amsterdam: North-Holland.

Mäki, U. 2001. *The economic world view: Studies in the ontology of economics*. Cambridge: Cambridge University Press.

Marshall, A. 1930. *Principles of economics: An introductory volume*. 8th ed. London: Macmillan.

Mill, J.S. 1848. In *Principles of political economy with some of their applications to social philosophy*, ed. W.-J. Ashley, 1909. London: Longman, Green.

Mill, J.S. 1851. *A system of logic, ratiocinative and deductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. Vol. 1. 3rd ed. London: John W. Parker.

von Mises, L. 1966. *Human action: A treatise on economics*. 3rd ed. Chicago: Henry Regnery.

Morgan, M.S. 1991. The stamping out of process analysis in econometrics. In *Appraising economic theories: Studies in the methodology of research programs*, ed. N. De Marchi and M. Blaug. Aldershot: Edward Elgar.

Morgan, M.S. 1990. *The history of econometric ideas*. Cambridge: Cambridge University Press.

Pearl, J. 2000. *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Richardson, T. 1996. A discovery algorithm for directed cyclical graphs. In *Uncertainty in artificial intelligence: Proceedings of the twelfth congress*, ed. F. Jensen and E. Horwitz. San Francisco: Mogan Kaufman.

Richardson, T., and P. Spirtes. 1999. Automated discovery of linear feedback models. In *Computation, causation and discovery*, ed. C. Glymour and G.F. Cooper. Menlo Park: AAAI Press.

Robbins, L. 1935. *An essay on the nature and significance of economic science*. London: Macmillan.

Savage, L.J. 1954. *The foundations of statistics*, 1972. New York: Dover.

Simon, H.A. 1953. Causal order and identifiability. In Hood and Koopmans.

Simon, H.A., and N. Rescher. 1966. Causes and counterfactuals. *Philosophy of science* 33: 323–340.

Sims, C.A. 1972. Money, income and causality. *American Economic Review* 62: 540–552.

Sims, C.A. 1980. Macroeconomics and reality. *Econometrica* 48: 1–48.

Sims, C.A. 1982. Policy analysis with econometric models. *Brookings Papers on Economic Activity* 1982(1): 107–152.

Sims, C.A. 1986. Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review* 10(1): 2–15.

Spirtes, P., C. Glymour, and R. Scheines. 2000. *Causation, prediction, and search*. 2nd ed. Cambridge, MA: MIT Press.

Stigler, S.M. 1986. *The history of statistics: Measurement of uncertainty before 1900*. Cambridge, MA: Belknap Press.

Stigler, S.M. 1999. *Statistics on the table*. Cambridge, MA: Harvard University Press.

Suppes, P. 1970. *A probabilistic theory of causality. Acta Philosophica Fennica 24*. Amsterdam: North Holland.

Swanson, N.R., and C.W.J. Granger. 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions.

Journal of the American Statistical Association 92: 357–367.
Tinbergen, J. 1951. *Econometrics*. New York: Blakiston Company.
Walras, L. 1954. *Elements of pure economics*. London: Allen and Unwin.
Woodward, J. 2003. *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Zellner, A.A. 1979. Causality and econometrics. In *Three aspects of policy making: Knowledge, data and institutions*, Carnegie-Rochester Conference Series on Public Policy, ed. K. Brunner and A.H. Meltzer, Vol. 10. Amsterdam: North-Holland.