Making Progress on Causal Inference in Economics

Chapter · October 2020

CITATIONS
0

READS
48

1 author:

Harold Kincaid
University of Cape Town
108 PUBLICATIONS **1,528** CITATIONS

SEE PROFILE

Forthcoming in **A Modern Guide to Philosophy of Economics (Elgar) Harold Kincaid and Don Ross, editors**

Making Progress on Causal Inference in Economics[1]

Harold Kincaid, School of Economics, University of Cape Town

Abstract

Enormous progress has been made on causal inference and modeling in areas outside of economics. We now have a full semantics for causality in a number of empirically relevant situations. This semantics is provided by causal graphs and allows provable precise formulation of causal relations and testable deductions from them. The semantics also allows provable rules for sufficient and biasing covariate adjustment and algorithms for deducing causal structure from data. I outline these developments, show how they describe three basic kinds of causal inference situations that standard multiple regression practice in econometrics frequently gets wrong, and show how these errors can be remedied. I also show that instrumental variables, despite claims to the contrary, do not solve these potential errors and are subject to the same morals. I argue both from the logic of elemental causal situations and from simulated data with nice statistical properties and known causal models. I apply these general points to a reanalysis of the Sachs and Warner model and data on resource abundance and growth. I finish with open potentially fruitful questions.

Key words: causality, econometrics, structural equation models, directed acyclic graphs, mediation, instrumental variables, economic growth

Biography

Harold Kincaid is Professor of Economics at the University of Cape Town. Early books were Philosophical Foundations of the Social Sciences (Cambridge 1996) and Individualism and the Unity of Science (Rowman and Littlefield 1997). He is the editor of the Oxford Handbook of the Philosophy of Social Science (2013) and coeditor of Scientific Metaphysics (Oxford 2013), What is Addiction? (MIT 2010), Distributed Cognition and the Will (MIT 2007), Toward a Sociological Imagination (University Press, 2002), The Oxford Handbook of the Philosophy of Economics (Oxford 2009), Classifying Psychopathology (MIT 2014), What is Addiction? (MIT 2010), Establishing Medical Reality, (Springer, 2008), Value Free Science (Oxford 2007), the Routledge Companion to the Philosophy of Medicine (2017), the forthcoming Oxford Handbook of the Philosophy of Political Science and numerous journal articles and book chapters in the philosophy of science and social science. In addition to his philosophy of science work, Kincaid is also involved in multiple projects in experimental economics focusing primarily on risk and time attitude elicitation and addiction.

[1] Many thanks to Glenn Harrison, Julian Reiss, and Don Ross for very helpful comments.

Introduction

Enormous progress has been made on causal inference and modeling outside of economics. Starting with work in artificial intelligence and philosophy of science (Pearl 2009, Sprites et. al 2000), the semantics of some causal relations have been clarified to such an extent that it is quite possible to describe the logical rules identifying accurate versus biased causal inference across a range of modeling situations. Structural equation modeling and potential outcome approaches are now known to be logically equivalent with each other; graphical model results show how each can be expressed in graph-based terms and show that they imply the same structures and analyses. Epidemiology, biology, medicine, machine learning, marketing, and other areas have noticed these results and now regularly use them to provide rigorous tools for causal inference, both from experimental and observational data.

Unfortunately, important parts of empirical economics have missed the boat. At its worst, causally uninterpretable or at least causally biased results are produced by multiple regression and other related techniques. At its best, empirical economics using experimental or quasiexperimental methods successfully builds on recent developments in causal inference. These approaches, however, are based on the idea of experimental treatment which sometimes does not translate easily into assessing the causal relations of complex economic models. However, by far the biggest problem is traditional econometric practice on observational data. That is my target here, though some of the points that follow will bleed over into some experimental practices. The difficulties I pinpoint are not inevitable but avoidable; the goal is not just critique but promotion of techniques known to avoid identifiable errors and to make causal reasoning about observational evidence more rigorous. Along the way I describe a number of open issues that should be of interest to both economists and philosophers of science.

The arguments I make about standard observational econometric practice flow from results that are not controversial in some areas outside economics. They rest on provable results and are well known to computer scientists, epidemiologists, and others. My points are mostly illustrations of this work (Elwert and Winship 2014 for example). However, these illustrations have large implications for much current econometric work that often go unnoticed in observational empirical economics, even though some of the basic ideas go back to founders of econometrics such as Haavelmo (Pearl 2014; Heckman and Pinto 2015).

I make my case by two routes, one logical and one empirical. The logical angle describes results in the logic of causal modeling and how they demonstrate when statistical adjustment practices are biased and when they are successful and necessary. The empirical route repeats the logical one by producing simulated data with known ideal properties and then showing that standard econometric practices lead to mistaken inferences about that data.

The discussion that follows also makes and uses some points about philosophy of science. The idea that testing is holistic has played an important role in philosophy of science over the last century (Stanford 2017). More recently, the idea that mechanisms are essential to science has received widespread support (Craver and Tabery 2019). I look at these themes as they surface in debates over causality in economics. The holism of testing does not necessarily entail the drastic skeptical or social constructivist conclusions often drawn; for causal inference,

we can specify sometimes quite clearly how evidential blame and credit is distributed and sometimes can show that a given set of data cannot decide between competing causal models. We can sometimes also show when mechanistic evidence is useful for causal inference, when it is essential, and when it is positively harmful. I will illustrate these points as I analyze causal inference practices.

The chapter proceeds as follows. Section 1 provides some brief relevant philosophy of science background. Section 2 describes advances in causal inferences based on causal graphical models. Three elemental causal relations are discussed in Section 3 using the causal graph approach. Casual logic followed by statistical analysis of simulated data both show that standard econometric practices can lead to systematic error in assessing these fundamental causal structures. Section 4 argues that instrumental variables do not solve the problems of Section 3. Section 5 illustrates the points made about causal inference by looking detail at the Sachs and Warner (1995;1997; 2001) empirical work on the resource curse. I conclude with a sketch of related open questions.

Section 1: Some Brief Philosophy of Science Background

There are some general philosophy of science results that I will use to help frame the discussion that follows. I will start with the more abstract philosophical issues and then go on to more specific issues about causation.

Philosophers have long used the Duhem-Quine thesis that theory testing is holistic to derive various morals. The Duhem-Quine thesis says in one form that every test of a hypothesis requires auxiliary assumptions—hypotheses or knowledge about the experimental setup for example. Another version of the Duhem-Quine thesis makes the stronger assertion that only entire theories are tested. If data contradicts prediction, either the hypotheses or auxiliary assumptions may be guilty and we cannot tell which; if the data fits predictions, that might be due to the truth of the hypothesis or to error in the auxiliaries. From the second, stronger Duhem-Quine thesis many drastic claims have been drawn. If only wholes—complexes of theoretical claims, auxiliary hypotheses about experimental set ups, etc.— are tested, then falsifying evidence can always be handled by making a change somewhere. Thus, theories are underdetermined by their evidence and theory choice is not a fully rational affair. Kuhnian and social constructivist approaches build on this reasoning.

The stronger, second Duhem-Quine thesis does not follow from (and is the not the same as) its weaker relative. That testing is holistic does not entail that only wholes are tested. The very complexity of theory and evidence that holism describes can under the right conditions allow us to distribute credit and blame to parts of the web of assertions that makes up theory. We will see below that graphical causal models can help show us when and where credit and blame can be attributed and when our evidence cannot decide between competing theories—that when and where we can attribute credit depends on local knowledge and context. This general conclusion about the holism of testing instantiates a view about philosophy of science that might be called contextualism (Kincaid 2005, 2006; Williams 1999): assessments of evidence for

particular hypotheses always depends on empirical background knowledge and rarely can be determined by scientific logic alone.[2]

Causal inference in economics also raises another important philosophy of science debate. Over the last 20 years there has been a rising chorus praising the importance of mechanisms in science. On my view (Kincaid 1996, 1997, 2011, 2012), the claims about mechanisms espoused can be in turn important, true but trivial, unclear, inconsistent, and/or unmotivated depending on the notion of mechanisms, the purpose they serve, the empirical background and more. Mechanism claims may be about evidence or explanation, about intervening causes or underlying structures, and many more parameters. At least one relatively clear thesis from this multitude is that causal claims can only be confirmed by providing the intervening mechanism linking cause and effect—showing the causal "channels" from a variable such as resource abundance to the ultimate casual outcome of growth. We will see in our discussion various ways this thesis is relevant to causal inference in economics, both positive and negative.

There are also various points about causation in general that I need to make clear as background to the discussion of graphical causal models. First and quite importantly, causal graphical models and their equivalents are independent of statistical questions in two senses. First, graphical causal models entail relations in the data that logically have nothing to do with traditional statistical estimation and hypothesis testing. The causal logics I describe below are relevant even when we have no samples but instead a fully enumerated population where the values of variables are known with certainty. So, these causal logics can be entirely separate from sampling error, asymptotic properties, and so on. This will be important later on, especially when we discuss instrumental variables. Traditional statistical issues can be combined with graphical causal models. That can be done in in a sample to population estimation framework. We can also bring in statistical inference by thinking of economic processes, for example, as we might think about error probabilities in choice experiments or as random shocks to variables. However, the fundament point remains that causal logic is different and independent of sampling and estimation logic, even if we ultimately want to combine them. Much confusion in the econometrics literature is caused by not keeping the distinction between the two clear.

Graphical causal models are also independent of statistical questions in that they are prior to them. Causal relations seem irreducible to statistical relations—that is the truth in the "correlation is not causation" slogan. Regularity and counterfactual accounts of causation as analyses in purely noncausal terms have well-known fundamental difficulties. However, the correlation is not causation slogan is wrong if it implies that we cannot learn anything about causes from correlations or that we can never have good evidence for causes based on correlations. In practice, appeal to the slogan can be a lazy refusal to do the hard work needed to make causal claims (especially when giving policy advice which has to be causal). The fundamental fact seems to be that, to use Cartwright's (1989) phrase, "no causes in, then no causes out." We can make compelling causal inferences when we have the right background

[2] Examples are the use of time order in observational studies to rule out certain possible causal relations or the estimation of time preferences in experimental setting using in part prior knowledge of risk preferences. Neither causal statistical logic in the first case nor experimental setup in the second by itself allows for justifiable inferences without this prior knowledge.

causal knowledge—for example, time ordering of variables—in addition to statistical information about the relations between variables. But alleged solutions to causal problems such as confounding by common causes that claim to be purely statistical are mistaken and always rely, if they actually work, on implicit causal assumptions.

Furthermore, the causal logics described here apply most conclusively and clearly to specific kinds of causal situations. As philosophers have been fond of pointing out, the term "cause" has diverse connotations. Here the distinction between sufficient causes and what I will call causal complexity is important. Sufficient causes are the garden variety of billiard ball causes where one factor has its own influence on some outcome independently of other causes. Causal complexity refers to more complicated relations such as necessary causes, thresholds, binding constraints, and so on. The latter are not describable as just the combination of a batch of independent separable causes. Casual complexity frequently shows up when we get to concrete details in cases studies, in historical analysis, and so on. There are multiple complexities involved in keeping these different senses of cause clear, and I will return to these problems when I discuss open issues in the last section. But for most of this paper and for the causal logics that follow, at issue is the simplest case where we have multiple, independent sufficient causes.

Finally, we need to make some distinctions when we talk about "identifying causes." In particular, we need to note the difference between identifying the existence of causal effects and identifying effect sizes, between showing there is a causal connection and measuring its magnitude. Though obviously importantly connected, these are not the same. Failing to distinguish them can cause problems.

Just to clarify, the section has set up some philosophy of science ideas used in the rest of the paper:

- Testing is holistic and thus so is testing of causal claims, but both philosophers of science (Kuhn 1962) and economists (e.g. Bardsley et. al 2009) have mistakenly drawn skeptical conclusions about scientific inference that do not follow from the holism of testing
- The different and complex relations between various causal hypotheses and auxiliary assumptions can actually provide the basis for compelling causal inferences in the right circumstances
- Causal inference is never going to be purely a statistical matter but instead relies on background causal assumptions; when those assumptions have their own evidence, in the right circumstances causal inference can be quite reliable
- Causal assertions and evidence logically can be evaluated independently of sampling issues and separating the two is essential avoid confusion
- Understanding causal inference requires separating different senses of cause, e.g. sufficient, additive causes versus necessary, nonaddictive causes and distinguishing the presence of causal effects from casual effect sizes. My focus here is almost entirely on the simplest case of sufficient, additive causes.

Section 2: Graphical Causal Models and the Logic of Causal Inference

While a definition of cause in noncausal terms is unlikely, there has nonetheless been tremendous progress in the last twenty years in producing a semantics for causality. Here by a

'semantics' I mean something akin to the logician's formal semantics: definitions, axioms, and implications therefrom. The key steps seem to have roughly simultaneous routes from two sources: Pearl (2009) and other work in artificial intelligence and computer science and Sprites, Glymour, and Scheines (2000) and others in philosophy of science. The idea, simply put, is that interconnected causal claims have implications for what conditional independencies we should see in the data if those claims are true.

The basic tool to illustrate causal implications is the directed graph, drawing on the graphs of Bayesian networks. The graph has nodes—variables--and potential links between them called edges, with a directed arrow representing a causal relation. Nodes with arrows leading into them are 'descendents' and nodes from which the arrows come are 'parents.' A path is an edge or set of edges linking nodes. A direct path is an edge that goes through no other node and an indirect path does. A graph is acyclic if no node has a path leaving and returning to itself. Nodes which have no edges coming into them represent exogenous causes and those that do have incoming paths are endogenous. Formally, a graph is a set exogenous variables *U*, endogenous variables V, and a set of functions f that defines the edges between them.

A graph with a set of directed edges then entails various claims of dependency (labelled associational, statistical, or probabilistic) and independency, where conditional independencies are essential for causal inference. Conditional independencies are simply the association of two variables given another, e.g. p(A and B)/p(C). For example, causal model 1 graphed in Figure 1 has the following conditional independencies:

$x1 \perp x3 \mid x4$
$x2 \perp x4 \mid x1, x3$

where, $\perp$ signifies probabilistic independence and | signifies conditionality. Holding $x4$ constant, $x1$ and $x3$ are not associated though they may be associated in the data. Obviously, these conditional independencies are very important for testing. If such conditional independencies are found in the data the model is supported and if they are not there, the model is rejected.

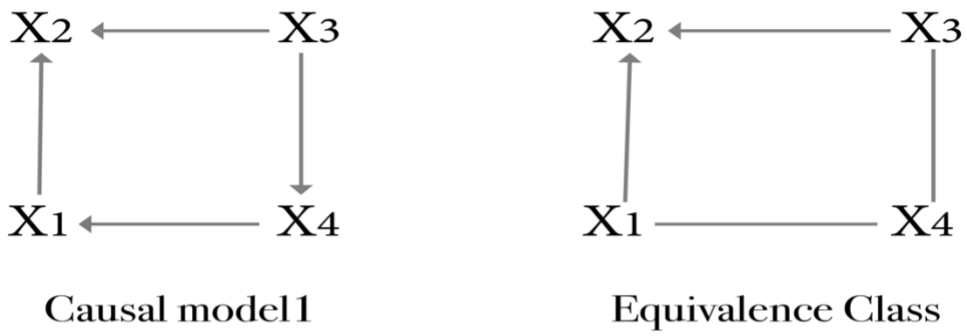


Figure 1. A causal model and its equivalent class of other models consistent with it. Undirected arrows show that causal relations could run in either direction.

Having an explicit model with a list of its implications as in causal model 1 has another virtue: it allows us to see if there are other models with different causal relations that also imply the same statistical conditional independencies. For example, the graph of causal model 1 has the equivalent models illustrated by the equivalence class in Figure1 (an "equivalence class" here refers to all causal models consistent with a set of conditional independencies). The undirected edges in the equivalence class represent possible causal arrows going in either direction. The causal independencies implied by causal model 1 do not tell us which of these possibilities is correct. Thus, there are four other models compatible with the conditional independencies that causal model 1 predicts—causal model 1 is underdetermined by all its possible implications. However, it is incompatible with any model that does not have the causal effects of $x1$ and $x3$ on $x2$.

Thus, stating the model, identifying its implications about conditional independencies, and determining other equivalent models are essential for knowing what the data can and cannot

tell us. The example of Figure 1 shows that while a model as a whole may be underdetermined by the data, not everything in it is.[3] Figure 1 represents a case where the classical thesis of undeterdetermination of theory by evidence is true when we compare models as a whole. However, such underdetermination does not prevent us from knowing about more partial claims, namely, the causal arrows shared by all the models.

Thus, we know that if we start with a proposed model, then we can determine which data in a set are entailed by that model and thus which are incompatible with it. We can also see what other models are compatible with that set of implied relations in the data. This kind of investigation works top down: we have a set of variables and a possible model, and we see what can learn from its entailments about expected relations in the data. In particular, we can derive the effects, positive and negative, of controlling or adjusting for any variable given the possible causal model. It is this point from causal modeling results that I apply in most of the chapter.

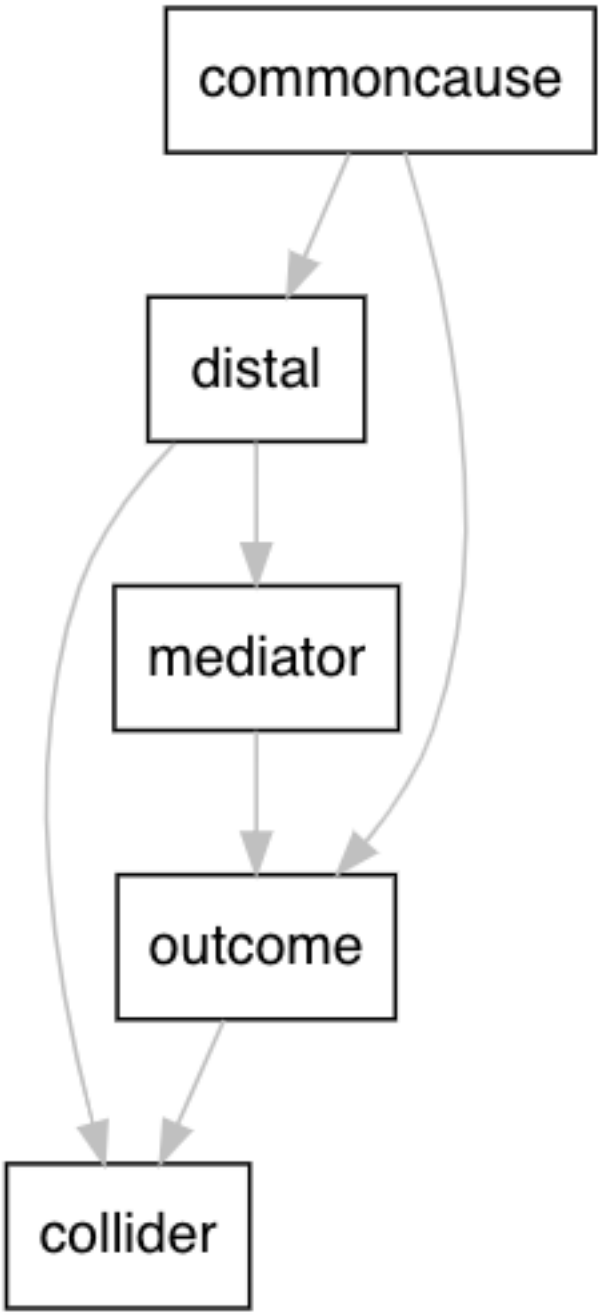
It is, however, possible using graphical models to go a more inductive route that starts with the data and infers to the model or models. Search, however, is constrained by the assumptions of the graphical causal logic--for example, 1.) unexplained correlations, ones which do not go away on conditioning, entail causal relations and 2.) associations that do go away upon conditioning do not represent direct causa links. Various algorithms have been produced that use given data plus causal logic to try to infer possible causal models consistent with that data (Spirtes et. al 2000; Glymour et. al 2019). The outputs are most often equivalence classes like those in Figure 1. They do not estimate parameters but instead tell us whether the data support some causal relation between any pair of variables.

It can be shown, either analytically or by simulation, what specific algorithms, based explicit assumptions, can find about the true causal data generating process. Search procedures start from the bottom—from the data—and generate models consistent with all the data; top down inferences based on possible causal structure tell us what some of the data must be like and what structure is common to any models consistent with it. The top down and bottom up methods can be combined in a pincer movement to narrow down what the evidence shows. However, as noted above, this chapter concentrates almost entirely on using graphical models to determine what they entail about the data, not on searches from the data. The latter are not needed for my main points.

In addition to finding models consistent or inconsistent with the data, causal graphs are also an essential guide in deciding on what variables to "adjust" or control for in any specific statistical analysis of a data set. The independencies and dependencies implied by an explicit causal model are the route to the idea of statistical control or conditioning. When a variable is controlled, it creates conditional independencies or dependencies according to what kind of link is involved. Figure 2 illustrates. When a common cause or a mediator is involved, then *distal* and

[3] The model of Figure 1 is incompatible with the association between x2 and both x1 and x3 disappearing if we hold x4 constant, with the association of x1 and x2 disappearing if we hold x3 constant, and with the association between x3 and x2 if we hold x1 constant. Correlations between x4 and x1 and x4 and x3 are not sufficient for determining which way the causation runs, and there are no conditional independencies that would decide this for us.

outcome are independent of each other given *commoncause* or *mediator*, where given here means controlled or fixed. Additionally, when the link involved is not a common cause or mediator but instead a collider--a node that has two directed arrows coming into it--then controlling or fixing the collider creates a dependency when none existed before (Figure 2).



The implied independencies of Figure 2 are:

Figure 2. A causal model with the three elemental component for causal analysis: mediators, common causes, and colliders.

1. distal ⊥ *outcome* | *commoncause, mediator*
2. common cause ⊥ *mediator* | *distal*
3. common cause ⊥ *collider* | *distal, outcome*
4. *mediator* ⊥ *collider* | *distal, outcome*

These provide a rich set of implications to test against data. Nonetheless, these implications of Figure 2 are still consistent with models that allow *mediator* to cause *distal* and/or *collider* to cause *distal*—thus there are three other models in the equivalence class. Following these implications gives us a natural way to talk about identified and unidentified models that does not mix up causal issues with sample to population estimates. These implications are nonparametric as well. No specific functional forms are required to determine the dependencies and conditional independencies entailed by a graphical causal model and thus those issues are also set aside. Only when we begin to estimate effect sizes with tools like SEMs or propensity scores do we have to make more concrete assumptions.

Pearl's semantics use what he calls a 'do operator.' A do operator analyzes the impact of taking some variable in a graph and setting it to a specific value. This is in fact making a (possibly hypothetical) intervention on some variable X by setting (X=x) and thus "wiping out" all the arrows coming into X. Then the logical implications of causal relations in the graph can show us the effects of adjusting for a variable by asking what the implications are of a do operation. In Figure 1a, $p(X1 \text{ and } X3) = 0/\text{do } (X2 = x)$, i.e. conditional on setting $X2$ to x.

This leads to rules about the appropriate adjustment of variables in a causal graph if we want to find the effect of one variable on another. We use the causal graph to ask what variables Z need to be adjusted to estimate the true causal relation between two others X and Y in the graph. The essential thing is to adjust sufficiently that we ensure that there is what Pearl calls no 'backdoor path' between X and Y. A backdoor path is a biasing path that in a graph is represented by a.) a set of paths, directed or undirected that link X and Y besides their causal link that does not b.) go through a collider. Colliders 'block' paths and prevent them from being biasing; conditioning on colliders opens them up and allows biasing paths between the variables whose arrows point to them. Common causes are open biasing backdoor paths and in need of adjustment in some form.

These are important points for knowing what the data can and cannot tell us. The example of Figure 1 shows that while the model as a whole may be underdetermined by the data, not everything in it is. So, the right side of Figure 1 represents a case where the classical thesis of undeterdetermination of theory by evidence is true when we compare models as a whole. However, such underdetermination does not prevent us from knowing about more partial claims, namely, the causal arrow shared by all the models.

This illustrates our early points about contextualism and holism. Given background knowledge of the set of possible causes, a proposed model of those variables, and the logic of causal inference we can show what we can know and not know. We know that any data consistent with consistent with causal model 1 will be consistent with the set of models

represented by the equivalence class on the right in Figure 1. We also know that given these variables and data on their association consistent with it, *X*1 and *X*3 are causes of X2.

The causal graphical approach and the three elemental relationships—mediators, common causes, and colliders—lend themselves naturally to sophisticated testing methods. Most directly related to the graphical models are structural equation models. These models have an equation for each arrow in the model. The combined set of equations then entails a covariance matrix which can be estimated by OLS regression in the simplest case and by maximum likelihood or other methods in more complex causes. SEMs are relatively unknown in economics despite the long history of simultaneous equation modeling. I will discuss some of the reasons SEMs have not caught on economics in the final section. SEMs will play a role in the next section showing how multiple regression gets things wrong in simulated data.[4]

An alternative to SEMs and graphical causal models is framing causality in potential outcomes or counterfactual terms and using methods such as propensity score (Austin 2011) and G estimation (Naimi 2017). Propensity scores calculate the probabilty of treatment assignment conditional on baseline characteristics; G estimation provides estimation of differences in potential outcomes using less restrictive assumptions than multiple regressions. Much of the work on RCTs, natural experiments and instrumental variables such as found in the mostly harmless econometrics literature takes this approach rather than the causal SEM perspective. These approaches hope to mimic randomized assignment and do not rely directly on causal graphical models. However, they are provably equivalent to graphical models and their methods have limitations that SEMs do not have, though there is still debate in this regard (Imbens 2020; Pearl 2014a, 2014b; Leamer 2010).

I am not going to discuss the potential outcomes approach in any detail for several reasons. In order to keep my main points clear and the scope of discussion manageable, I limit my discussion to graphical models. More importantly, the main problem with the potential outcome approach is that the causal models presupposed are hard to make explicit. As a result, there is a considerable reliance on unformalized intuition; regressions are used in an ad hoc way without clear causal models. The result is that obstacles to inference are hard to make clear and express. The situation is like that with verbal arguments and piecemeal analyses in the face of general equilibrium situations—it is hard to keep clear what is going on until a full formal treatment is given. Causal graphical models do that in the way the potential outcome framework often does not. I will return to this point later.

[4] Structural equation modeling is a form of simultaneous equation modeling, but the SEM approach has a provable causal interpretation and resulting methodological implications and rigor that simultaneous equations used in economics do not. Essential parts of Pearl's (2009) approach do indeed stem in from Haavelmo (1943), but his causal notions were dropped as econometrics progressed after him. Current simultaneous equations methods in econometrics have much the same problems as do single equation multiple regression paradigm. See Heckman and Pinto (2015) and Pearl's (2013) illuminating response.

Section 3 Good and Bad Econometric Practices

Anybody who reads or does empirical economics with purely observational data faces the problem of what variables to include in equations to be tested. The problem arises regardless of whether we are using ordinary least squares, maximum likelihood, or other tools. Various purely statistical criteria can be brought to be bear, but they are usually about estimation from samples, not causality. A high R-squared for example tells us next to nothing about causality, as is obvious from the fact that a high R squared is compatible with a tiny bivariate regression coefficient.

The most common approach in the face of this uncertainty in equations errs on the side of inclusion, thinking that adding more controls cannot hurt and may help. This advice to go for more controls rather than fewer has support from well-known commentators on the econometrics literature (Angrist and Pischke 2015, p. 60).[5] Qualifications are sometimes added, but the problem is the lack of a clear causal framework for guidance. The major worry for inference is taken to be "omitted variable bias" but that is really multiple things and we need a system to sort them out.[7]

This section uses the causal graph framework of the previous section to identify good, bad and neutral inference procedures in certain econometric practices with observational data.[8] We begin with three possible confounding situations: common causes, which is the best-known case, mediating variables, and colliders.[9] In each case I assume we are dealing with only observable or 'manifest' variables to keep things simple and clear. Common causes, mediators, and colliders are each a possible source of error and bias, and properly handled, each a situation where causal inference can go well. These elementary cases keep their properties when embedded in larger models. These three ways of confounding and their extensions can be expressed more formally, more completely, and in ways to allow them to be combined while still being able to trace their interactions by the use of explicit graphical models so that we can for any possible causal relation we can identify the good, bad and neutral statistical judgements (or actual interventions) to estimate the true causal effect. SEMs are one natural way to test these more complex models.

The simplest case is adjusting for common causes. If *X* and *Y* are each caused by a single common cause, then conditioning on the common cause in a multiple regression removes the *spurious correlation and provides an unbiased estimate of their true causal relation. If X and* Y

[5] Also see (Imai, Keele, et al., 2014, pp. 482– 487) "irrelevant covariates may complicate the modeling but does not compromise the identification of causal mediation effects".

[7] Omitted variable bias in the most obvious sense of other independent variables that are also causes of the dependent variable is in general not a big problem. Simulated data with a true regression coefficient of .61 between x1 and x2 leaving out a second cause X3 of X2 produces an estimate in this run of the data of .64. Adding the missing cause improves that to .63. The standard errors are basically unchanged in specifications. Adding further covariates in this case contributes to precision, which is a statistical, not causal, issue.

[9] I have not explored the relation of these kinds of causal errors from including the wrong variables to earlier literature (Leamer 1978) on specification searches. Ideally on my view the causal points made here could/should be embedded in a full Bayesian treatment--there has been significant progress in Bayesian SEMs--but this is a project for the future.

are independent, conditional on the common cause, then there is evidence of no causal relation. This causal logic is obviously clearly understood, and I will not further illustrate it here.

Nonetheless, once common causes are embedded in more complex causal circumstances, inferences about them can be misguided because of other not so obvious complications. One complication, still entirely in the framework of common causes, is worth mentioning, because it illustrates the morals we see below showing that a clear causal model is essential to inference. The example I have in mind shows that an omitted variable in the form of a confounding common cause can be handled without controlling for—conditioning on— the common cause itself and without randomization or instrumental variables. Consider the graph in Figure 3. We can control for *commoncause* without including it in a regression. We can do so by conditioning on *ccmediator*. Doing so blocks the biasing path from *distal* to *outcome*. Valid inferences about causal relations and effect size are then possible. The graphical causal model and the rules of causal inference make this obvious. Yet, the point is hard to see and to describe in a principled way if there is not an explicit causal model with explicit implications. A common inference in regression studies is that when including an additional control variable reduces the correlation between the independent variable and the dependent variable, we have good evidence that the control is a common cause. The example of Figure 3 shows that inference can be a bad one.

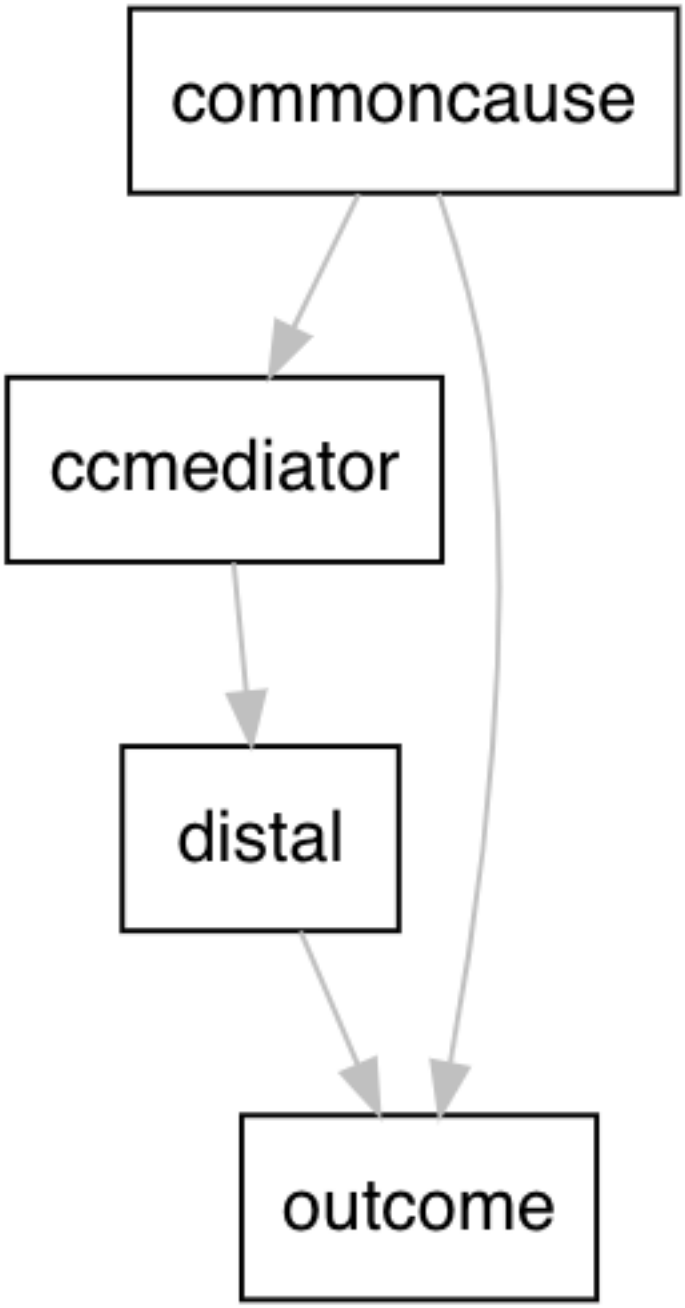


Figure 3. A biasing common cause that works through a mediator.

This example also shows that claims that mechanisms are needed for confirmation are too coarse. What is need for causal inference depends on the causal structure at issue and the data

available. In this case I can increase my confidence without knowing precisely what the confounder is.

move next to mediating variables. Their possible role is also easy to understand but, surprisingly, not clearly studied in the econometrics literature, especially but not exclusively that concerning observational data. Usual talk of "endogeneity" is often ambiguous and most often refers to correlations between a right-hand side variable and the error term or, less often, reverse causation from dependent to independent variables in a regression. Mediation involves neither. It is simply one right hand side variable having some or all of its causal influence through another right-hand side variable. So, graphically, mediation involves either only indirect (mediated) effects (Figure 4a) or both direct and mediated effects (Figure 4b).

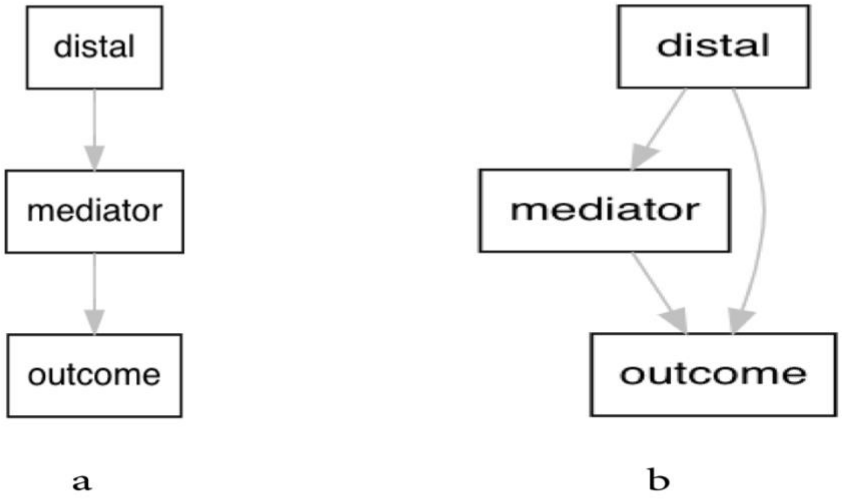


Figure 4. Full mediation (a) and mediation plus direct effect (b).

In mediation model 4a, we know via causal logic that:

distal is independent of *outcome* / *mediator*

This means that conditioning on *mediator* in a multiple regression will remove the correlation between *distal* and *outcome*. If we are basing our causal claims on multiple regression coefficients which regress *outcome* on *distal* and *mediator*, then we will conclude that distal has no causal influence. This results no matter how we calculate the coefficients, i.e. by least squares or maximum likelihood. This inference will be entirely independent of sampling and estimation issues, for it follows from the causal logic even when we know the entire population with certainty.

This causal logic is obvious, but it is useful to see it confirmed by regression results. In this first case and all that follow using simulated data, a sample is generated with 1000 observations randomly selected from normally distributed variables with specified (and standardized) coefficients based on the relevant causal model. So, for the simple bivariate case, values of the independent variable are randomly selected from a normal distribution and then values of the dependent variable are selected consistent with the specified regression coefficient and with random errors with a mean of zero and standard deviation of 1.[11] Thus, the simulated data has all the properties needed for successful inference using OLS and maximum likelihood.

The results of regressing *outcome* on *distal* and *mediator* are unsurprising (Table I).

Table I. Regression results for the model of Figure 4a with no direct effects of distal.

A. *outcome* regressed on *distal* only

	Estimate	Std. Error	t value	Pr(>|t|)
distal	4.791e-01	2.779e-02	17.24	<2e-16 ***

B. *outcome* regressed on *distal* and *mediator*

	Estimate	Std. Error	t value	Pr(>|t|)
distal	7.543e-03	3.565e-02	0.212	0.832

Significance codes:

'***' 0.001 '**' 0.01 '*' 0.05

Coefficients used to generate simulated data:

outcome = .76 * *mediator*

mediator = . 63 * *distal*

In **A** the regression coefficient is .48. This is close to the values used to generate the data (listed at the bottom of the table) because the expected effect of *distal* on *outcome* is simply their product. In B *distal* has lost its statistical significance and is much smaller than the estimated value identified in **A**-it has gone from .48 to .007. Obviously including the mediator in the

[11] TETRAD 6 causal modeling software (Center for Causal Discover, Carnegie Mellon University) was used for simulations and the R package Lavaan 6 (Rosseel 2012) used for the regressions and SEMs. Dagitty (Textor 2016) software for causal graphs and implications was also immensely useful

regression gives us a mistaken causal conclusion, namely, that *distal* has no influence on *outcome*. It does give us evidence that *mediator* is an intervening variable, but does not tell us the relation between *distal* and *outcome*, because Figure 4a entails the same dependencies and independencies as the model of Figure 5, where the causal direction is reversed.

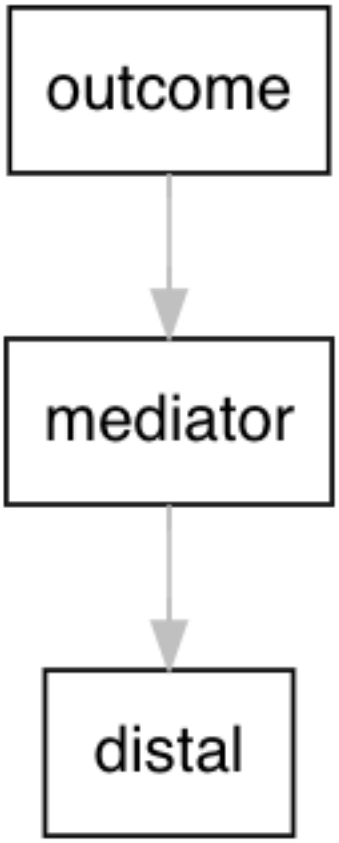


Figure 5. Causal model with same dependency implications as those in Figure 4a.

These results again illustrate our philosophy of science points from the beginning about the holism of testing and the need for mechanisms. The data can bear differentially on parts of the model--they tell us that *mediator* is between *distal* and *outcome* but not the direction of the causal relation.

Turning to the more complex model of Figure 4b where *distal* has both indirect effects through *moderator* and its own direct effects, data were again generated with known causal relations and effect sizes and with the same desirable statistical properties. The regression results when we include either *distal* alone or *distal* and *mediator* as right-hand side variables are in Table II.

Table 2: Regression results using simulated data generated using the model from Figure 4b where *distal* has both direct and indirect effects.

A. *outcome* regressed on *distal* only

	Estimate	Std. Error	t value	Pr(>|t|)
distal	8.478e-01	1.679e-02	50.51	<2e-16 ***

B. *outcome* regressed on *distal* and *mediator*

	Estimate	Std. Error	t value	Pr(>|t|)
distal	5.502e-01	12.534e-02	21.72	<2e-16 ***
mediator	3.723e-01	2.534e-02	14.70	<2e-16 ***

Significance codes:
'***' 0.001 '**' 0.01 '*' 0.05

Coefficients used to generate simulated data:
outcome = .39 * *mediator*
mediator = . 78 * *distal*
outcome = .52 * *distal*

Regression **A** accurately reflects the true total effect in the simulated of *distal* on *outcome*. The total effect in the simulated data is just the direct effect (.52) plus the indirect effect (.78 x .39) and thus has a value of .82. Regression **A** shows a coefficient of .85 of *distal* on *outcome*, thus accurately reflecting the true causal model generating the data. However, including *mediator* in the regression biases the results. Now we might infer that *distal* does cause *outcome*, but our estimate will be biased downward because *mediator* absorbs part of the effect of *distal* on *outcome*. Thus, the coefficient for *distal* in regression **B** of .55 is smaller than the true causal relation which is .82 in the simulated data.

There are straightforward and well-known ways to avoid these wrong conclusions (VanderWeele 2015; Hayes 2013), though it rare to see them explicitly used in economics. For our simple model the correct value can be found by two regressions. Regress *outcome* on *distal* and then on *distal* and *mediator*, though steps have to be taken to correct standard errors. The first regression gives you the total effect of *outcome*. Substracting the value of *distal* in the second regression from its value in the first provides the indirect effect of *outcome*. For complex models, we have to move to full structural equation models and maximum likelihood estimation. A SEM was fitted to the simulated data used for the analyses of Table II. The results are as in Table III.

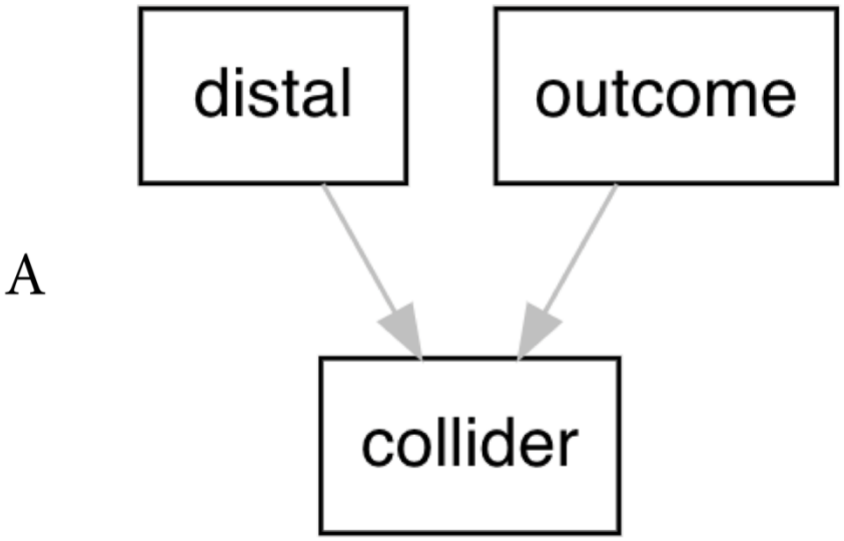
Table III. SEM estimates of simulated data involving *distal* with direct and indirect effects through *mediator*.

	Estimate	Std. Err	z-value	P(>|z|)
outcome ~				
distal	0.550	0.025	21.749	0.000
mediator ~				
distal	0.799	0.019	42.056	0.000
outcome ~				
mediator	0.372	0.025	14.718	0.00
TOTAL EFFECT *outcome* ~ *distal*	.848			

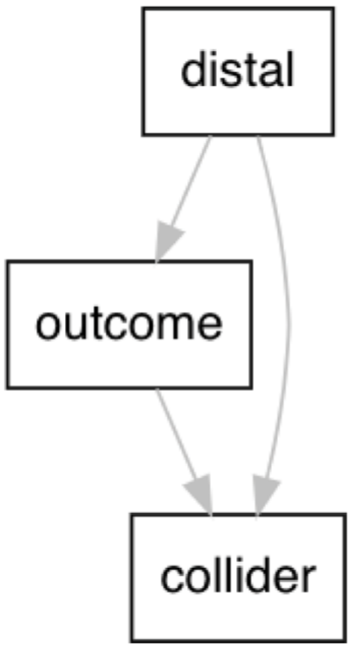
The results of Table III use the standard SEM notation where "~" can be read as "is caused by." The coefficients are standardized estimates. The key point here is that the true values can be found --the SEM for this data set estimated total effects of .85 when the true relation used to simulate data was .82 but with random errors. The standard multiple regression including both the mediator and the independent variable underestimates this value considerably (over 30%).

I have presented the two basic cases involved in mediation. However, as we saw with common causes (and will see with colliders), the basic cases can be expanded into other easily imaginable circumstances. One obvious case for mediators arises when we have data on a mediator's causal effects other than its effect on the outcome. Even if the *distal* cause does not work through these descendants of mediators, conditioning on the effects of mediators will bias the estimated influence of the distal cause on the outcome just as would conditioning on the mediator itself. Intuitively, conditioning on the mediator's effects restricts its variation in the causally relevant range and thus partially holds the mediator constant, thus confounding the relation between the distal cause and the outcome. Anywhere standard econometrics are run using many right-side variables where theory tells us little about their interconnection, this kind of situation will be a risk. Cross-country growth regressions come to mind as a likely example.

Thus far we have covered two of the three basic causal complexities out of which most complex causal accounts have to be built. We turn now to the third elemental causal complexity, namely, colliders. A simple collider is illustrated in Figure 6a and a collider combined with a real direct effect in Figure 6b. Our causal logic tells that conditioning on a collider opens up an otherwise blocked path between variables--conditioning creates an open biasing path and thus correlation between the variables on the path. So, in Figure 6a by conditioning on collider we open a path between *distal* and *outcome* even though there is no causal relation and thus we make a wrong inference. In Figure 6b there is a real causal relation between *distal* and *outcome* and multiple regression will find that relation. However, the causal logic says regression will overestimate the strength of the causal relation due to the bias from conditioning on collider.



A



B

Figure 6. A collider causal relation (A) and collider with direct cause (B).

Regression on simulated data again confirms the problem. Data are produced in the same fashion as before: random samples are taken of the independent variables from a normal distribution with specified parameters and then values of the dependent variables are chosen as a function of the independent variable(s) and an i.i.d. error term. Starting first with the model of Figure 6a, data were simulated to match Figure 6a and regressed producing the results in Table IV.

Table IV. *Outcome* is the dependent variable in a regression on data with no causal relation between it and *distal*.

	Estimate	Std. Error	t value	Pr(>|t|)
distal	-4.529e-01	3.210e-02	-14.11	<2e-16 ***
collider	7.315e-01	3.210e-02	22.79	<2e-16 ***

Significance codes:
'***' 0.001 '**' 0.01 '*' 0.05
Coefficients used to generate simulated data:
collider = .61 * *distal*
outcome = .48 * *distal*

We thus get a statistically significant and large correlation between *distal* and *outcome*, -.45 standardized and significant at the .001 level. But we know there is no such causal relation in our simulated data. They were produced by a causal model--a set of equations--where there is no effect of *distal* on *outcome*.

Look now at the more complex case where there is a *collider* but there is also a direct causal effect of *distal* on *outcome*. By the causal semantics, the effect size should be biased because of the correlation created by conditioning on *collider*. In a regression on simulated data where there is both a collider and direct effect of known values, we get the results of Table V. The actual value of the causal relation between *distal* and *outcome* is .36. Both the sign and size of the coefficients are wrong.

Table V *Outcome* is caused by *distal* but a collider is conditioned on.

	Estimate	Std. Error	t value	Pr(>|t|)
distal	-9.838e-02	2.378e-02	-4.138	3.8e-05 ***
collider	8.310e-01	2.378e-02	34.952	< 2e-16 ***

Significance codes:
'***' 0.001 '**' 0.01 '*' 0.05
Coefficients used to generate simulated data:
collider = .30 * *distal* + .66 * *outcome*
outcome = .36 * *distal*

So, we see once again that the Angrist et. al recommendation that it is generally good to add more covariates is wrong.[10][13] Adding the collider reduces the true coefficient of *distal* on *outcome* of .36 to .09. Similarly, the methodological claim that we need to add mechanisms to have warranted causal claims is wrong. Or, perhaps we should put the point in terms of vagueness: what part of the causal process and with what inferential procedures is a mechanism essential or valuable? Collider processes added to multiple regression inference demonstrably lead to error.

As with common causes and mediators, this kind of bias can be produced without conditioning on a collider if the causal structure produces a collider on conditioning. Consider the causal model in Figure 7. What happens if we condition on—include in a regression—a "potential collider." Though not directly a collider--the two variables whose causal relation we are estimating do not themselves directly causally effect a collider-- conditioning on it produces collider-like bias. The path from *distal* to *potcollider* to *other* to *outcome* is opened up by conditioning. The estimate of the influence of *distal* will be biased upward.

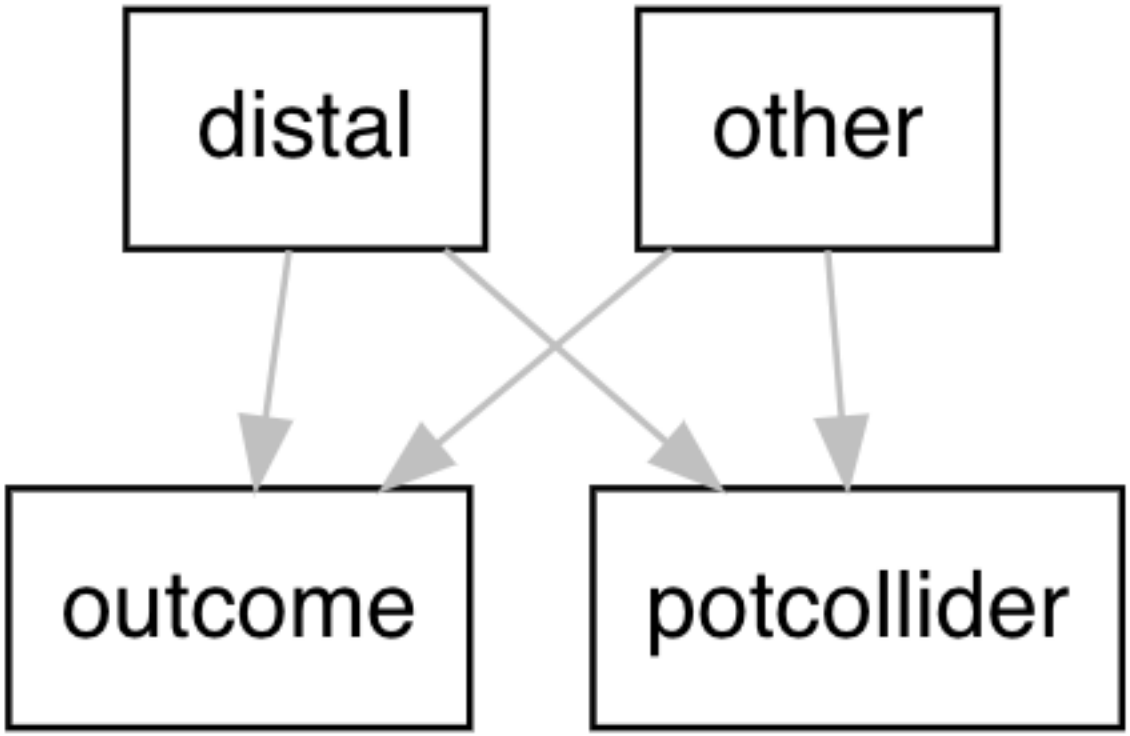


Figure 7. Indirect collider.

The collider case, like the mediator situation, shows that standard econometric practice, when it is unclear about its proposed causal models, can lead at least to uninterpretable results and at worst to clear error. Graphical causal models can help make the causal relations more explicit and can provide the tools, for example in the form of SEMs tested by maximum likelihood, to provide decisive evidence about whether those causal relations are real. These can

make the argumentation economists use considerably more rigorous and disciplined.[14] We would expect this to be mostly likely in circumstances where prior theory cannot simplify what are probably complex relationships with many factors.

We will develop these ideas further in the empirical case studies in Section 4. However, we need next to ask whether the above complexities can be solved by the use of instrumental variables. Many seem to think so.

Section 4: Instrumental Variables to the Rescue?

It is widely claimed that instrumental variables (IVs) can solve 'endogeneity' problems. There has been, the enthusiasm goes, a “credibility revolution” in the practice of econometrics (Angrist and Pischke 2010 for the claim; Leamer 2010 and Deaton 2010 for early critique). We now know, it is claimed, how to avoid all these dreary and unconvincing structural regressions and do proper science. Instrumental variables and associated techniques mimic the logic of experiments. With them, much clearer inferences will surface.

Unfortunately, and unsurprisingly, there is a bit too much hype here. Our old axiom that testing is holistic should make us suspicious of claims for a simple inference strategy that promises to eliminate apparent complexities. It is great if you can get such simple rules, but they are unlikely. Instrumental variables do not solve all problems that fall under the endogeneity label. The complexities discussed in Section 3 are among the ones instrumental variables cannot dissolve.

That IVs are not a panacea should be obvious from the standard criteria for them: the variable in question z should be correlated, preferably strongly, with the independent variable x and not correlated with the error term. These are statistical criteria. We already know that no causes in, no causes out--it takes causal assumptions to get causal conclusions from statistical findinds. So, IVs used purely in terms of associations are unlikely to ensure that we are making the right causal inferences. Reiss (2005) showed some time ago cases where the statistical criteria by themselves cannot ensure the correct causal inference.

The confusion between statistical issues and causal ones has a long history in economics (and elsewhere of course). IVs are often motivated in terms of providing consistent estimates. Providing consistent estimates in the traditional statistical sense of the term is an issue about asymptotic results in estimating from samples to populations--in short, a sampling issue, not a causal one. It is only recently that IVs have gotten a more explicit causal role, and the two roles—providing consistent statistical estimation of population values and identifying causes in situations of potential confounding—are often not clearly separated in the literature.

IVs only give reasonable causal information in the right causal circumstances. We need, as usual, to have specific kinds of causal background knowledge to make them work. Without that aid, they can lead to error. To be more specific, IVs can handle common causes in the right situation. They are no solution for mediators and colliders.

[14] This point was made long ago by the early developers of graphical causal modeling. See Glymour et. al (1994).

As before, I first give the causal logic and then provide empirical illustrations from known simulated data. I start with mediators. Figure 8 illustrates the general situation. The conclusion that instruments do not solve mediator confounding is obvious from our previous discussion of mediators and the fact that instruments are about the true effect of *distal* causes on outcomes. Instruments cannot correct errors from conditioning on mediators.

Once again data were simulated according to the model of Figure 8 with nice statistical properties and a strong instrument ensured. The results from a standard 2SLS are in Table VI.

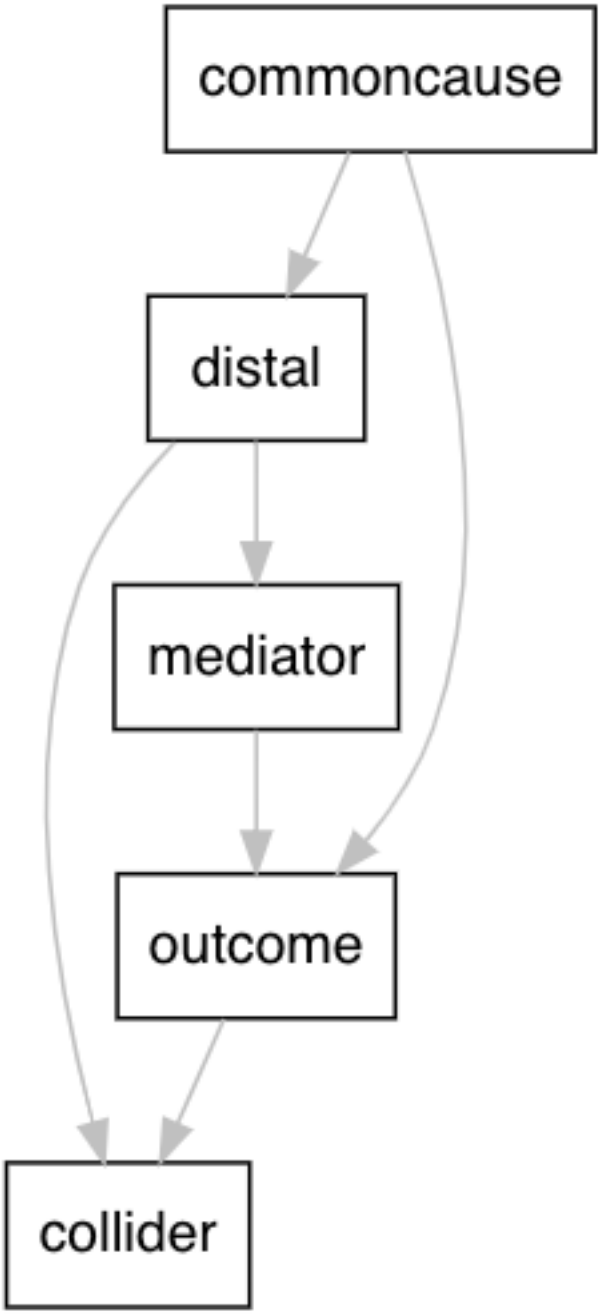


Figure 8. Causal diagram of the use of instrumental variables in situation where mediator is conditioned on.

Table VI: 2SLS on simulated data from causal model with mediator and instrument.

A. *outcome* regressed on *distal* and instrument only

	Estimate	Std. Error	t value	Pr(>|t|)
distal	0.41788	0.03914	10.678	<2e-16 ***

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	998	648.28	< 2e-16 ***
Wu-Hausman	1	997	19.49	1.12e-05 ***

B. *outcome* regressed on *distal*, *mediator* and *instrument*

	Estimate	Std. Error	t value	Pr(>|t|)
distal	-0.03441	0.03583	-0.960	0.337
moderator	0.83043	0.02662	31.190	<2e-16 ***

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	998	648.28	< 2e-16 ***
Wu-Hausman	1	997	19.49	1.12e-05 ***

Significance codes:
'***' 0.001 '**' 0.01 '*' 0.05

Coefficients used to generate simulated data:

outcome = .76 * *mediator* + .53 *commoncause*

mediator = . 56 * *distal*

distal = .64* *instrument*

The diagnostic tests show that the instrument is indeed a strong instrumental variable, which is no surprise since we made it so in the data. The weak instrument test confirms that there is strong correlation with the independent variable; the Wu-Hausman test shows that the instrument is consistent and ordinary OLS is not. Furthermore, our sample size is 1000, so there should not be small sample problems that can arise with instrumental variables. By standard econometric practices our estimate of *distal* on *outcome* should be valid. Part A of Table VI shows that the instrument does indeed work well when the mediating variable is not included in the regression: the estimated value of the coefficient on *distal* is .417 and the value used in the model to simulate the data .426.

However, our instrumental variable gives us the wrong answer when the mediator is included (Table V1 B). Despite being instrumented, *distal* has a statistically insignificant

coefficient of -.034, thus of the wrong sign and far from the size of the true value of .426 in the simulated data. Clearly, in the face of conditioning on mediators, instrumental variables give unreliable results, even when there is no doubt that they cause the independent variable and influence the outcome entirely through it (the instrument diagnostics remain strong when the mediator is included--see Table VI B).

This is confirmed by estimating a SEM model which shows (Table VII) unsurprisingly that the model of Figure 8 fits the data very well and produces accurate estimates of the coefficients used in the simulation of the data, unlike the standard *instrument*al variable regressions with mediators. IVs do not solve the problems caused by including mediators in standard econometric methods.

Table VII. SEM applied to simulated data used in IV regressions of Table VI.

Regressions:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
distal ~						
instrument	0.644	0.024	26.928	0.000	0.644	0.624
commoncause	0.274	0.024	11.296	0.000	0.274	0.262
outcome ~						
commoncause	0.527	0.004	136.832	0.000	0.527	0.526
mediator	0.755	0.004	199.525	0.000	0.755	0.767
mediator ~						
distal	0.541	0.026	21.175	0.000	0.541	0.556

TOTAL EFFECT distal ~ *outcome* .41

Fit statistics:

Comparative Fit Index (CFI)	0.999
Tucker-Lewis Index (TLI)	0.998
RMSEA	0.035
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.066
P-value RMSEA <= 0.05	0.752

A similar analysis applies to colliders. Figure 9 depicts the use of an instrument where we have a confounder, so an instrument is needed. However, there is also a collider. We know from the causal logic that conditioning on colliders creates correlations where there is no causation. Again, a perfect instrument--one that causes the outcome only through the endogenous variable and is strongly correlated with it--does not solve the problem.

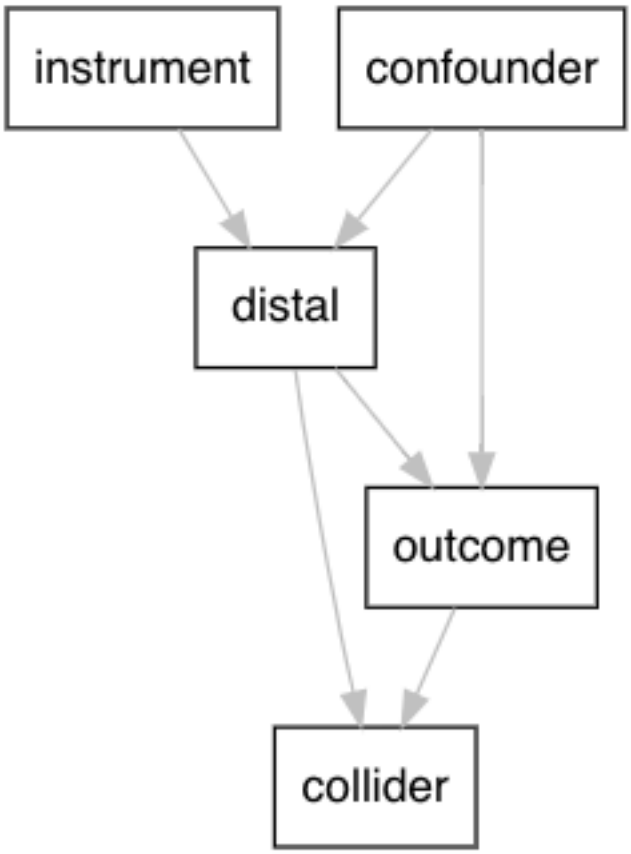


Figure 9. Instrumental variable with confounder and collider.

The statistical analysis again parallels the logic. An instrumental variable analysis of simulated data ensuring a strong instrument nonetheless give us wrong information. *Distal* has a impact on *outcome* in the simulated data of .25 but the instrument analysis gives the results in Table XIII. The instrument used in a regression without the collider gives a slightly upwardly biased estimate. When the collider is included the coefficient doubles: conditioning on *collider*s can correlations where there is no real relation. The instrument in both cases easily passes the standard tests for validity. The SEM results in part C of Table VIII again show that the true values can be recovered from the data if a full causal model is estimated . The model does show strong fit to the data--a good model has CFI and TLI values near 1 and RMSEA shows good fit the closer it is to 0 and that is exactly what we see here.

Table VIII. IV analysis and SEM analysis of data generated with strong instrument and collider.

A. *outcome* regressed on *distal* only with instrument

	Estimate	Std. Error	t value	Pr(>|t|)
distal	0.384	0.042	9.016	<2e-16 ***

Diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	998	518.34	<2e-16 ***
Wu-Hausman	1	997	97.67	<2e-16 ***

B. *outcome* regressed on *distal* and *collider* with instrument

	Estimate	Std. Error	t value	Pr(>|t|)
distal	6.973e-01	2.576e-02	27.07	<2e-16 ***
collider	5.745e-01	1.506e-02	38.13	<2e-16 ***

diagnostic tests:

	df1	df2	statistic	p-value
Weak instruments	1	997	457.09	< 2e-16 ***
Wu-Hausman	1	996	28.27	1.3e-07 ***

C. SEM results from analysis of simulated data

Regressions:

	Estimate	Std.Err	z-value	P(>|z|)
outcome ~				
distal	0.295	0.023	12.607	0.000***

Fit statistics:

Comparative Fit Index (CFI)	1.000
Tucker-Lewis Index (TLI)	1.002
Akaike (AIC)	5068.412
Bayesian (BIC)	5112.582
RMSEA	0.000

Significance codes:
'***' 0.001 '**' 0.01 '*' 0.05

Coefficients used to generate simulated data:

outcome = .25 * *distal* - .6 * *commoncause*

collider = -.86 * *distal* + 1 * *outcome*

distal = .56 * *instrument* + -.61 * *commoncause*

The moral of the section then is that instrumental variables do nothing to prevent some standard kinds of bias in regression. When complex models are used without explicit causal assumptions, regression and related techniques can produce causally unintepretable results, regardless of whether variables have been "instrumented."

Section 5 Some Data Reanalyzed: Making Causal Arguments

In this section I move from making general points using causal logic and simulated data to discussing some of these issues in the context of real data on important topics. My general point across the paper is that empirical work in economics using observational data would be improved by disciplining it with explicit causal models and the appropriate logic and tools.[10] I begin to investigate how that can be done by looking at an important debate over the last 25 years--whether there is a 'resource curse' that partly explains low economic growth. I take publicly available date for key papers about the controversy and use the causal framework from above to reanalyze the conclusions drawn.

Sachs and Warner (1995, 1997, 2001) in several highly cited papers claim to find that an initial large share of natural resources inhibits economic growth over time. Aside from addressing an important issue about growth and being influential on subsequent literature, Sachs and Warner's key papers have been carefully scrutinized for reproducibility (Davis 2013). Their results have almost all been successfully repeated. That does not mean their interpretations of their analyses are correct. However, the successful replications do mean that Sachs and Warner's regression results are computationally dependable. This puts aside worries about which specifications they actually used and whether they are reliable, allowing us to focus on the causal interpretations of their findings.

Sachs and Warner compile an international data set measuring GDP and a set of other national-level covariates over the period from 1970 to 1989. The key covariate in their papers is a measure of natural resource level at the beginning of the panel. Their central conclusion is that resource abundance is the 'main' causal factor explaining slow growth.

Sachs and Warner also make statistical arguments about some of the possible pathways from resource abundance to low growth. They suggest that:

1. higher resources in the initial period contribute to higher corruption which in turn causes slower growth
2. higher resources in the initial period contribute to protectionism which in turn causes slower growth
3. higher resources in the initial period make investment goods more expensive--because higher natural resources means higher prices for nontraded goods-- thus reducing investment and then growth
4. higher resources in the initial period shifts labor away from learning by doing sectors and thus decreases labor productivity and consequently growth

However, Sachs and Warner are explicit that resource abundance on their view also directly causes decreased growth apart from its indirect effects listed in (1) to (4).

The above causal claims are clearer than the presentation in Sach and Warner. Their discussion of causal processes is quite brief and the meaning of claims that resources are the main cause is never clear-- no metric is ever explicitly given.

We do get more clarity from looking at their regressions, illustrating the contextualist idea that 'models' come to different things and have to be understood as embedded in context. The verbal model presented above--a sort of unspecified prior in Bayesian terms-- coexists with the mathematical equations for regression which lead to empirical tests. We have to look at both the verbal and mathematical aspects to see what Sachs and Warner are trying to do.

The key mathematical and empirical components are found in five regression equations (Sachs and Warner 1997, Table 1) that are tested against their data set. The equations listed in their order with regression coefficients in parentheses (all of which are statistically significant) are in Table IX. The full model is equation 1.5.

Table IX. Association between Growth and Other Key Variables

Equation 1.1	Equation 1.2	Equation 1.3	Equation 1.4	Equation 1.5
initial GDP (-.11)	initial GDP (-.96)	initial GDP (-1.34)	initial GDP (-1.76)	initial GDP (-1.79)
initial resources (-9.43)	initial resources (-6.96)	initial resources (-7.29)	initial resources (-10.57)	initial resources (-10.26)
	open to trade (3.06)	open to trade (2.42)	open to trade (1.33)	open to trade (1.34)
		investment (1.25)	investment (1.02)	investment (.81)
			rule of law (.36)	rule of law (.40)
				export prices (.09)

Keeping in mind the results from our analysis of causality, what can be gleaned from these results? I put aside questions about the reliability and validity of the indicators used to measure variables, questions about outliers, and other such statistical issue. My question is what causal conclusion could we draw from these regressions if there were no such measurement issues?

Sachs and Warner (1997) claim that:

1. initial resources "is not simply a proxy for institutional quality or import-substituting industrialization policy" (p. 13)
2. "the adverse effect [of resources] is not operating mainly by lowering investment rates, since the negative correlation is maintained even after controlling for investment rates" (p. 13)
3. "resource intensity is not simply a proxy for adverse trends in global export prices of resource intensive economies" (p. 13)
4. "evidence suggests that the indirect effects are not large; otherwise the additional controls should drive the estimated coefficient on SXP [resources] down as we read from left to right in table 1" (p. 25)
5. that the “estimated direct effect of SXP [resources] on growth is large in comparison with these estimates of the indirect effects.” (p. 26)

Some of these claims are warranted, some are not. I take it that the claim that resource intensity is not a "proxy" for export prices, institution quality, or import substitution policy to means that the correlation between resources and growth is not just the result of other common causes of these two (Sachs and Warner never clarify what they mean by “proxy”). The regressions warrant that claim, since if resource -GDP correlations were just the result of common causes then adding these variables as regressors would remove the correlation between growth and resources. Of course, both could be true--resources could be both a negative cause of growth, and growth and resources jointly caused by other factors. There is no way for these simple regressions in Table IX to support or rule out that possibility. Indeed, also seeing the effect of leaving the resources variable out of the equations in the table would at least be suggestive evidence about their direct effect, though that is not done.

Another warranted claim is that the regressions in Table IX support the conclusion that the negative effect of resources on growth is not entirely mediated by the other variables. If it were, the coefficient on resources would go to zero when other factors are included. It does not. So, Warner and Sachs are right to say that they have some evidence that resources have a direct effect on growth. Of course, the evidence for both of these claims about the role of resources is quite tentative, because other causal stories are not investigated.

Sach and Warner's claims about effect size, however, are unwarranted. They report unstandardized regression coefficients which are nearly uninterpretable as effect sizes between variables because they are on different scales. However, the much bigger problem is that the Sachs and Warner’s regressions do not control for relations between the variables--they do not measure direct and indirect effects and consider neither colliders nor common causes of variables

that do have direct or indirect effects on growth. Without examining these possibilities, claims about effect sizes even of standardized variables are unwarranted.

Sachs and Warner do suggest and look at some evidence for the mechanisms connecting resources and growth. They regress investment on resources and several other variables, they regress the price of investment goods on resources and several other variables, and they regress national savings on resources and other variables. They find statistically significant effects. However, these regressions do not help much. First, they find (Sachs and Warner 1997, Table IX) that there is a positive effect of resources on investment, while their hypothetical mechanism expects the opposite. More importantly, there is no way for these piecemeal regressions to test the presence or effect size of these possible pathways for all the reasons discussed earlier in the chapter.

What Sachs and Warner need is an explicit causal model representing their various suggestions about processes and then to test it against their data. That means they need a structural equation model and statistical means to simultaneously estimate parameters. Maximum likelihood is the natural tool.

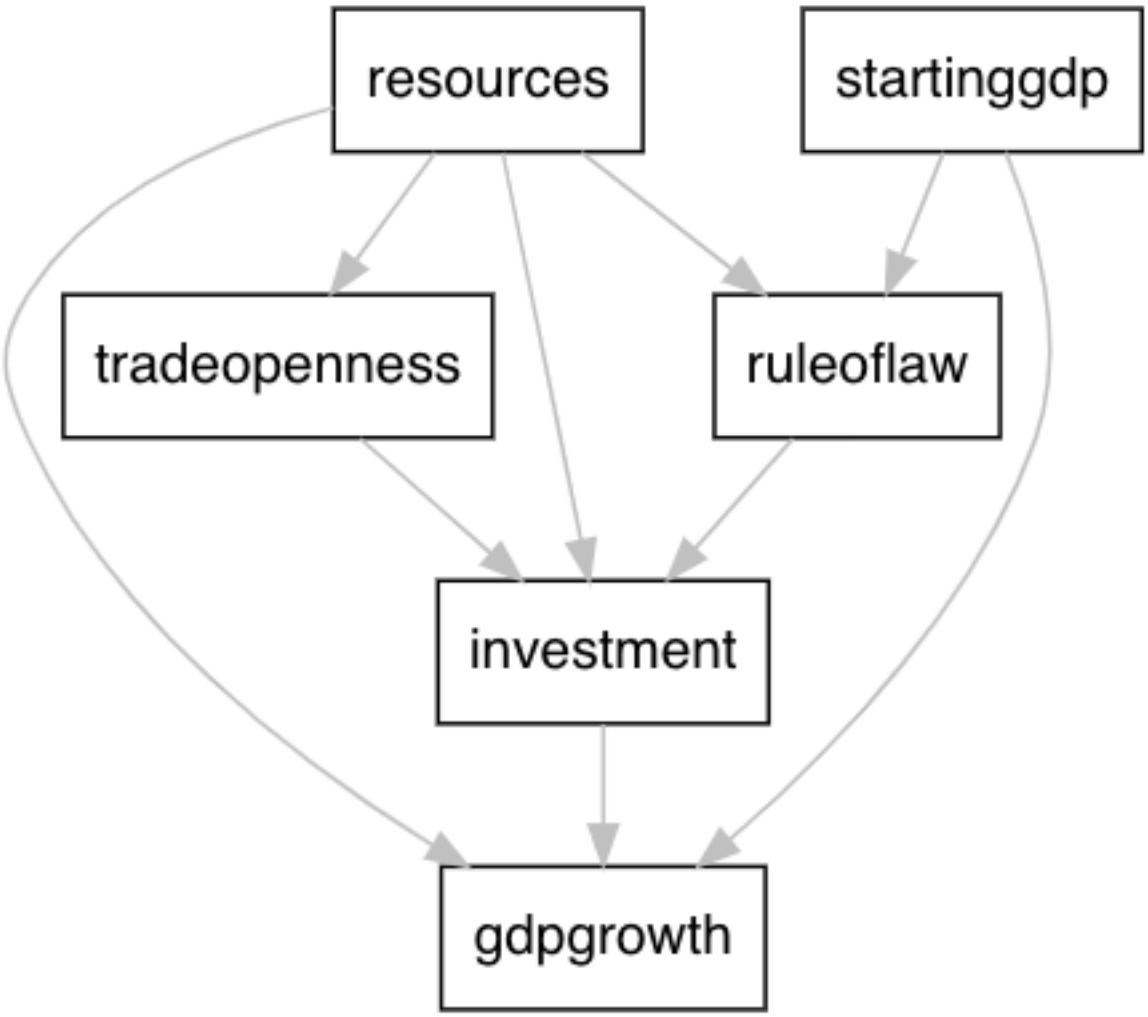


Figure10. A causal model that closely matches the claims of Sachs and Warner

Figure 10 is a graph of the causal model that seems to make the most sense of the various claims that Sachs and Warner make about resource abundance and growth. Initial GDP and resources are strictly exogenous because they measure variables at the beginning of the period studied. The other causal arrows are placed following their proposed mechanisms. So, resource

abundance has a direct effect on growth and an indirect effect mediated by the mechanisms they propose.

This model has the following implications about conditional independencies:

- gdp ⊥ tradeopen | initialgdp, invest, resources, rulelaw
- initialgdp ⊥ resources
- initialgdp ⊥ rulelaw
- initialgdp ⊥ tradeopen
- rulelaw ⊥ tradeopen | resources

These implied independencies in the data can be investigated by maximum likelihood. Because we know a priori that resources and initial GDP are exogenous, and because the model has multiple causal relations, there are no other models that entail the above independencies. In that sense there is not an underdetermination problem. However, that does not mean the full data set--not just the implications listed above--might not support other competing models.

The estimated coefficients from a SEM using the structure in Figure 10 are shown in Figure 11 and the model stats are shown in Table XI.

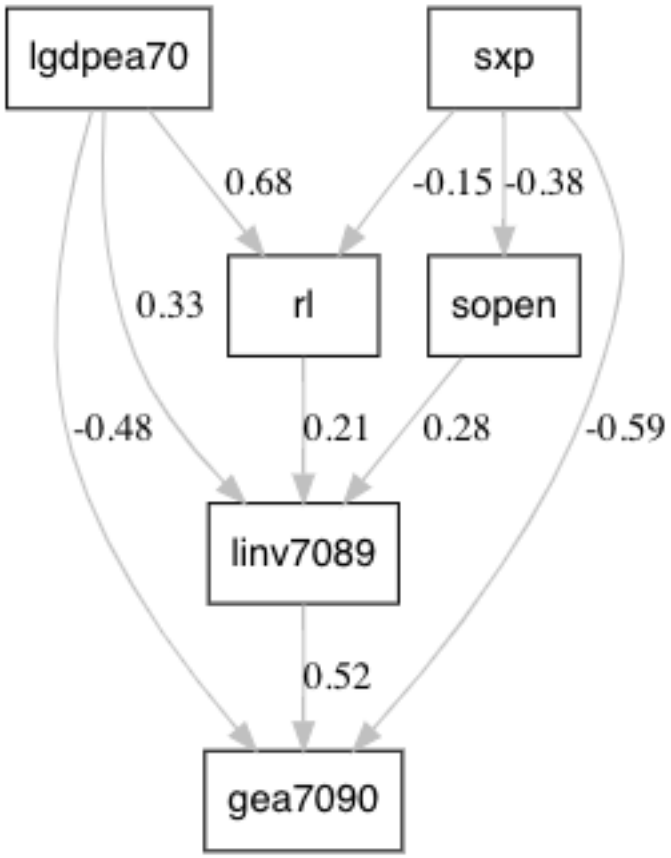


Figure 11. SEM estimates of Sachs and Warner model.

Table XI. Fit statistics for estimated model described in Figure 11.

Comparative Fit Index (CFI)	0.649
Tucker-Lewis Index (TLI)	0.017
Akaike (AIC)	675.594
Bayesian (BIC)	705.547
RMSEA	0.490
P-value RMSEA <= 0.05	0.000

The model makes a sort of intuitive sense, with coefficients consistent with the story that Sachs and Warner tell. However, the model fits the data very poorly according to numerous standard model fit statistics as seen in Table XI.

So, the top down approach testing implied independencies does not support the Sachs and Warner model. If we turn instead to a bottom up causal search, we find the results of Figure 12. These results were produced by running the FCI algorithm (Spirtes et. al 2000) on the Sachs and Warner data. The algorithm is guaranteed to find any causal models consistent with the data under certain assumptions, specifically that correlations that persist when all other variables are conditioned on show causal relationships and when two variables are not correlated, they are not causes of each other (no perfectly offsetting causes in opposite directions). The algorithm does not assume no hidden or unobserved variables. Given that we know the resources ("sxp") and initial GDP (lgdpea70) are from the start of consistent with the data under certain assumptions, specifically that correlations that persist when all other variables are conditioned on show causal relationships and when two variables are not correlated, they are not causes of each other (no perfectly offsetting causes in opposite directions). The algorithm does not assume no hidden or unobserved variables. Given that we know the resources ("sxp") and initial GDP (lgdpea70) are from the start of the time period, we can conclude that the Sachs and Warner data support:

- a negative causal effect of resources on growth
- a causal effect of initial GDP on rule of law
- a causal relation, direction unknown between rule of law and how open a country is to trade
- a causal relation, direction unknown, between how open a country is to trade and GDP growth

Thus, systematically studied for causality, Sachs and Warner's claim that there is a resource curse holds up. Their claims for its effect size is unwarranted. Parts of their proposed mechanisms are not ruled out; other parts—those working through investment, for example, are not supported. These conclusions are derived from the causal logic of models and are not reports about statistical inferences from samples to populations.

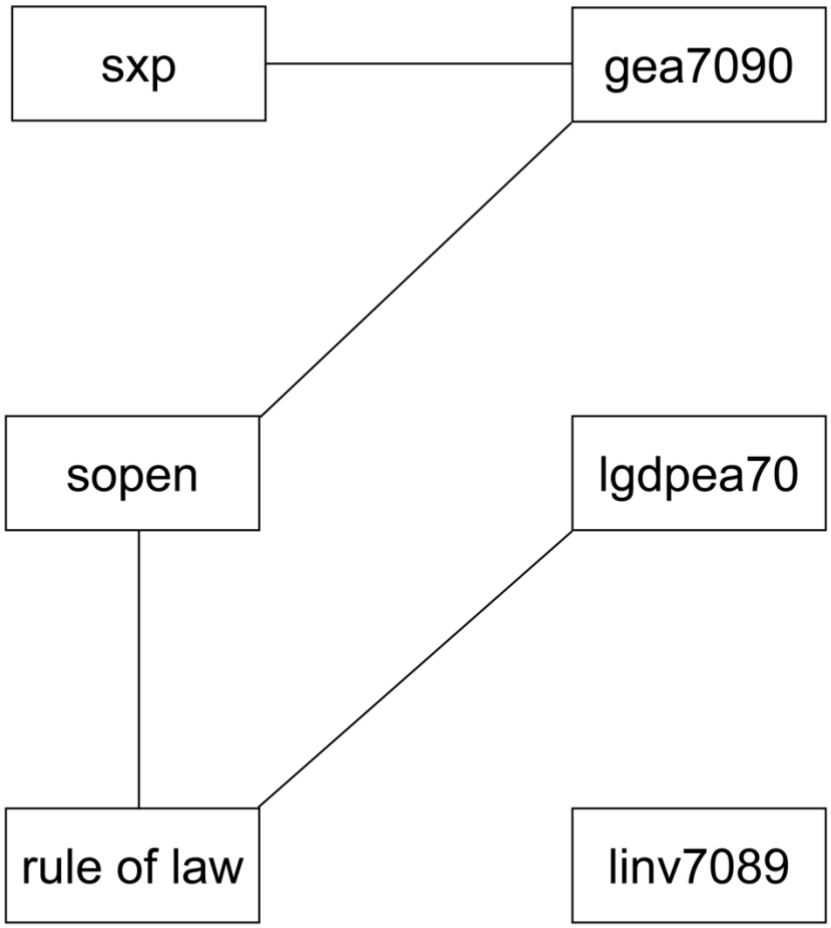


Figure 12. The set of causal models supported by the Sachs and Warner (1997) data according to the FCI search algorithm. Undirected lines suggest causal relations going in either direction.

Conclusion: Open Issues

I have shown that some standard economic practices sometimes may lead to fallacious causal inferences. Put in the affirmative, some economic practice can be improved by more explicit causal reasoning. Economists often make causal arguments and they may often be persuasive. However, they are also often informal. There are now tools to replace informal causal reasoning with much more rigorous logic. I have shown how this is possible for the elemental components of causal analysis and provided one example illustrating how to apply the basic elements to more complex causal situations.

However, I have left many questions hanging or unaddressed. I sketch some of these now in hopes of encouraging others to pursue them. I group the questions into three categories: 1.) where can the points made in this chapter be fruitfully used in thinking about economics, including the trend of randomized field experiments, 2.) history, philosophy and social studies of science questions about why the revolution in causal analysis has deeply influenced epidemiology and other fields but not economics, and 3.) questions about economic causal relations that are more complex than those studied in this chapter.

A key open issue is where and how the kind of analyses I have provided can be applied fruitfully to practices in economics. I think there are many possibilities. The elements of causal inference, causal adjustment, and causal bias and the tools I have described to deal with them should be able to help empirical economists working with observational data produce more rigorous and regemented arguments. Studies looking at the real argumentation from a causal perspective such as that I provided for the resource curse argument should be possible in a wide variety of areas.

At a more principled or theoretical level, the perspective of this chapter should also provide illuminating things to say about the RCT and quasiexperiment trend in empirical economics and about the long tradition of simultaneous equations and empirics in econometric practice. I have said little about either. However, my comments on instrumental variables are certainly relevant to at least assessing so-called natural experiments. Others (Pearl 2018; Leamer 2010) have already made some related points.

The simultaneous equations tradition in economics likewise could be approached with the framework presented here (see Hoover 2001for some efforts). While this tradition has much in common historically with the approach to causality I have outlined, my sense is that it still has trouble separating statistical issues from causal ones and that it still works with piecemeal regressions rather than full causal models that are testable as SEMs. As is hopefully now clear, arguing about causality without fully explicit causal models and sticking only to partial regressions makes progress hard. There should be room for more fruitful interaction between the causal graphical approach and the simultaneous equations approach in economics.

The graphical causal framework and the structural equation modeling that can instantiate it has had little use in economics. An interesting puzzle is why. The roots of these approaches go back in part to Haavelmo (see Pearl 2014). Of course, economists like almost everyone in the

1950s through the 1980s repeated the correlation is not causation mantra (Hoover 2004). Yet, epidemiology, significant parts of biology, and other social sciences have outgrown that and now use explicit causal modeling techniques to draw causal conclusions from correlational data (Shipley 2016 is a very good example). Imbens (2020) claims that the current 'harmless econometrics' movement finds the randomized clinical trial and natural experiments paradigm a more fruitful approach. The use of propensity scores, difference in difference analyses, and so on follows from that. However, it is provable that these potential outcomes frameworks along with related counterfactual approaches are logically equivalent to SEMs and graphical models. Or, perhaps better put, RCTs, natural experiments, propensity scores, SEMs, and so on can, as tools for causal inference, all be brought under the umbrella of causal graphical models. Thus, presumably, a more nuanced account is needed to explain the lack of take up.

The puzzle about take up leads naturally to the next question about the current enthusiasm for "harmless econometrics" and RCTs. Much has already been said about these approaches as used in areas outside economics. The discussion above about instruments could be fruitfully extended to the methods of harmless econometrics. That discussion showed that instruments without theory--without explicit motivated causal models--lead to error. This conclusion fits and amplifies doubts about the current excitement over trials and quasi-experiments from the structural econometric point of view. However, filling in the details with the insights of graphical causal models applied to these trends remains a topic worth further exploration.

Yet a further set of large open questions concerns concepts of causation that are more complex or different from those presupposed by regression and structural equation modeling. This chapter has discussed 'sufficient' causation. By that I mean causes that have an influence on their own, independent of other causal influences. This does not preclude other causes of the same effect; sufficient causes also need not be operative at all times and places.

However, the idea of sufficient causes as just defined does not fit easily with other causal notions that economists (and everybody else) often use. Anscombe (1957) long ago noted that many of our causal concepts are what might be called "thick"--they presuppose a variety of connotations specific to the causal concept in question as embodied in terms such as 'prevent,' 'constrain,' 'enable,' and so on. So, in development economics it is common to talk about factors that are preconditions for growth--they do not ensure growth but you cannot have growth without them. Education is probably such a factor--Cuba has very high levels of education by international standards, but low growth; education is necessary but not sufficient for development. Or, we identify thresholds for factors, binding or constraining factors, forcing factors, contingent causes, and so on (Rodrik 2009 on causes of growth is a good illustration). In all these cases it is not easy to think of these causes as independent elements that have a specific effect on their own.

At this point it is not clear, at least to me, how these complex types of causality can be analyzed in the sufficient cause framework exemplified by regression and SEMs (see Vander Weele 2015 for some efforts here). DAGs as developed by Pearl are nonparametric--they only describe dependencies, not their functional forms. In that sense they allow more complex causality. How that works out when turning DAGs into models for the estimation of effect sizes

is unclear. Interaction terms are sometimes included in SEMs. However, that work is undeveloped and interaction terms are hard to interpret causally in general. The sufficient cause framework wants to see every cause as having its own part of the effect that is constant regardless of the level of other causes. That makes things statistically much easier. Yet that is not the metaphysics of thick causal concepts like 'threshold' or 'binding constraint'.

Sociologists and political scientists have developed over the last two decades innovative ways to think about complex causality. Sometimes called qualitative comparative analysis (QCA, Regin 1987; Rihouz 2013), boolean analysis of necessary and sufficient conditions is used to analyze complex causality. There is ongoing debate about its evidential status for causal claims (Hug 2013). However, it clearly at least adds some discipline to complex causal claims. QCA is basically unknown in economics. But its possible applications in areas such as development economics or economic history seem obvious.

Thus, as I have hopefully shown, the moral is that there is still much to learn about causality in economics and we have some good tools to do so.

References

Anscombe, G.E.M. 1957. Intention. Oxford: Blackwell.

Angrist, J. and Pischke, J.-S. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. Journal of Economic Perspectives 24, Number 2, pp. 3–30.

Angrist, J. and Pischke, J.-S. 2015. Mastering 'Metrics. Princeton: Princeton University Press.

Austin, P. 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research, 46:399–424.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C. and Sugden, R. 2009. Experimental Economics: Rethinking the Rules. Princeton: Princeton University Press.

Cartwright, N. 1989. Nature Capacities and Their Measurement. Oxford: Oxford University Press.

Craver, C. and Tabery, T. 2019. "Mechanisms in Science", The Stanford Encyclopedia of Philosophy (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>.

Davis, G. 2013. "Replicating Sachs and Warner's Working Papers on the Resource Curse." Journal of Development Studies, 49:12, 1615-1630.

Deaton, A. 2010. Instruments, Randomization, and Learning about Development. Journal of Economic Literature 48 (June 2010): 424–455.

Elwert, F. and Winship, C. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." Annual Review Sociology 40:31–53.

Glymour, C., Spirtes, P. and Scheines, R. 1994. In Place of Regression. In Patrick Suppes: Scientific Philosopher. Volume 1. Probability and Probabilistic Causality. Paul Humphreys, Ed. Dordrecht, Springer: 339-366.

Glymour, C., Zhang, K. and Spirtes, P. 2019. "Review of Causal Discovery Methods Based on Graphical Models." Frontiers of Genetics 10 June 4 https://doi.org/10.3389/fgene.2019.00524

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. Econometrica 11 1Reprinted in D.F. Hendry and M.S. Morgan (Eds.), The Foundations of Econometric Analysis, Cambridge University Press, 477-490.

Hayes, A. 2013. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (Methodology in the Social Sciences). New York: Guilford Press.Harrison, G. 2011. Randomization and Its Discontents. Journal of African Economics 20 (4): 626-652.

Heckman, J. and Pinto, R. 2015. Causal Analysis After Haavelmo. Economic Theory 31(1): 115.

Hoover, K. 2001. Causality in Macroeconomics. Cambridge: Cambridge University Press.

Hoover, K. 2004. Lost Causes. Journal of the History of Economic Thought. 26 (2): 149-164.

Hug, S. 2013. "Qualitative Comparative Analysis: How Inductive Use and Measurement Error Lead to Problematic Inference". Political Analysis. 21 (2): 252–265.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2014). Commentary on Pearl (2014): Practical implications of theoretical results for causal mediation analysis. Psychological Methods, 19, 482–487. doi:10.1037/met0000021

Imbens, G. 2020. "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics," arXiv:1907.07271

Kincaid, H. 1996. Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research (Cambridge: Cambridge University Press, 1996).

Kincaid, H. 1997. Individualism and the Unity of Science: Essays on Reduction, Explanation, and the Special Sciences (Lanham, MD: Rowman and Littlefield, 1997).
Kincaid, H. 2005. "Contextualism, Explanation and the Social Sciences," Philosophical Explorations 7, pp. 201-219.
Kincaid, H. 2006. "Contextualist Morals and Science," in Value Free Science: Ideals and Illusions, ed. H. Kincaid, J. Dupre, and A. Wylie. Oxford: Oxford University Press, pp. 218-238.
Kincaid, H. 2011. "Causal Modeling, mechanism, and probability in epidemiology," Causation in the Sciences, ed. P. Illari, F. Russo, and J. Williamson. Oxford: Oxford University Press, pp. 70-91.
Kincaid, H. 2012. "Mechanisms, Causal Modeling, and the Limitations of Traditional Multiple Regression," in Kincaid, H. (ed.), Oxford Handbook of the Philosophy of the Social Sciences (Oxford University Press 2012), pp. 46-64.
Kuhn, T. 1962. Structure of Scientific Revolutions. Chicago: University of Chicago Press.
Leamer, E. 1978. Specification Searches. New York: Wiley.
Leamer, E. 2010. Tantalus on the Road to Asymptopia. Journal of Economic Perspectives 24 (2): 31–46.
Lin, W. 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. Annals of Applied Statistics Vol. 7, No. 1, 295-318.
Naimi, A., Cole, S. Kennedy, E. 2017. An introduction to g methods. International Journal of Epidemiology, 2017, 756–762
Pearl, J. 2009. Causality. Oxford: Oxford University Press.
Pearl, J 2014. "Trygve Haavelmo and the Emergence of Causal Calculus." Economic Theory.
Pearl, J. 2014a. October 27 "Are economists smarter than epidemiologists? (Comments on Imbens's recent paper)." Causal Analysis Theory and Practice blog.
Pearl, J. 2014b. Reply to Commentary by Imai, Keele, Tingley, and Yamamoto Concerning Causal Mediation Analysis. Psychological Methods. Vol. 19, No. 4, 488–492.
Pearl, J. 2018. Challenging the hegemony of randomized controlled trials: A commentary on Deaton and Cartwright Social Science & Medicine 210 (2018) 60–62
Ragin, Charles C. (1987). The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. University of California Press.
Reiss, J. 2005. Causal instrumental variables and interventions. Philosophy of Science 72964-976.
Rihoux, Benoît (2013), "QCA, 25 Years after "The Comparative Method": Mapping, Challenges, and Innovations--Mini-Symposium", Political Research Quarterly, 66: 167–235, doi:10.1177/1065912912468269.
Rodrik, D. 2009. One Economy, Many Recipes. Princeton: Princeton University Press.
Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling." Journal of Statistical Software, 48(2), 1–36. http://www.jstatsoft.org/v48/i02/.
Sachs, J., and A.Warner (1995), "Natural Resource Abundance and Economic Growth", NBER Working Paper 5398, Cambridge, MA.
Sachs, J., and A.Warner (1997), "Natural Resource Abundance and Economic Growth", Revised version, Unpublished manuscript, Harvard Institute for International Development, Cambridge, MA.
Sachs, J. and Warner, A. 2001. "Natural Resources and Economic Development: The curse of natural resources." European Economic Review 45 (2001) 827-838.
Shipley, B. 2016. Cause and Correlation in Biology. Cambridge: Cambridge University Press.
Sala-I-Martin, X. 1997. I Just Ran Two Million Regressions. The American Economic Review 87 (2): 178-183.

Spirtes, P. Glamour, C. and Scheines, R. 2000. Causation, Prediction, and Search. Cambridge: MIT Press.

Spirtes, P. 2005. “Graphical models, causal inference, and econometric models.” Journal of Economic Methodology 12:1, 1–33.

Stanford, Kyle, "Underdetermination of Scientific Theory", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination/>

TETRAD 6, Carnegie Mellon, Center for Causal Discovery

Textor, J., van Zander, B., Gilthorpe, M., Liskiewicz, M., and Ellison, G. 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. International Journal of Epidemiology 45(6):1887-1894.

VanderWeele, T. 2015. Explanation in Causal Inference. Oxford: Oxford University Press.

Williams, M. 1999. Groundless Belief. Princeton: Princeton University Press.