Economic Thought 6.1: 56-82, 2017

Graphs as a Tool for the Close Reading of Econometrics (Settler Mortality is not a Valid Instrument for Institutions)

Michael Margolis,* Universidad de Guanajuato, Mexico
mmargolis6@gmail.com

Abstract

Recently developed theory using directed graphs permits simple and precise statements about the validity of causal inferences in most cases. Applying this while reading econometric papers can make it easy to understand assumptions that are vague in prose, and to isolate those assumptions that are crucial to support the main causal claims. The method is illustrated here alongside a close reading of the paper that introduced the use of settler mortality to instrument the impact of institutions on economic development. Two causal pathways that invalidate the instrument are found not to be blocked by satisfactory strategies. The estimates in the original paper, and in many that have used the instrument since, should be considered highly suspect.

JEL codes: C18, O11, B52

Keywords: causation, development economics, econometrics, graphs

1. Introduction

The need to measure causal effects without experiments arises often for economists, and econometric theory may justly be said to contain some of the clearest statements of when and how this can be done.[1] In practice, however, we are rather forgiving as to whether applied work quite fulfils the theoretical requirements. Rightly so, perhaps: hyperfastidiousness would leave much potentially valuable work unpublished. But the dissonance between theory and practice makes our rhetorical tradition less clear than it could be. I believe a body of theory developed outside of economics can help; I also believe this theory provides an easier way to teach much econometrics, but the latter point is relevant to this discussion only in that the two beliefs share a common source. This lies in the use of a mathematical language that is unambiguously causal (as algebraic equations are not) alongside the algebra required for parametric statements. Causal assertions are most naturally encoded in directed graphs, as represented by diagrams with variable names linked by arrows. Many econometricians have drawn such diagrams to give an idea of what they have in mind, without being aware that the drawing often contains in full the information needed to answer important questions about the causal interpretation of statistics. The theory governing this interpretation has been under development by computer scientists, philosophers and statisticians beginning in the late

[1] I refer in particular to the work associated with the Cowles Commission efforts just after the World War II, as exemplified by Tinbergen (1940); Haavelmo (1943, 1944); Marschak (1950); Koopmans et al. (1950).

1980s (Glymour et al., 1987; Pearl, 1988; 1995; Spirtes et al., 2000, *inter alia*) and is now, in several important ways, quite mature.

Economists have not entirely ignored this work, but those who have applied it have chiefly been drawn to its most exotic branch, known as 'inferred causality'.[2] It is this branch that promises a truly new type of conclusion, ideally taking the form of a probability for the existence of each arrow possible in the causal graph linking observed variables. To infer causality is to let the data answer questions that generally must be answered theoretically, and it is remarkable that this is ever possible. It is thus no surprise that this branch has attracted most attention.

My argument here is for graphs in more conventional analysis, where statistics are given causal interpretation conditional on assumptions dictated by theory. The results I will discuss are part of a non-parametric generalisation of the structural equations approach tracing back at least to Haavelmo (1943; 1944) and they have close analogues in established econometrics. The graphical requirements for causal interpretation of linear models are equivalent to the orthogonality, rank and order conditions from the Generalised Method of Moments. And without linearity, results of a graphical analysis can be translated into the 'conditional ignorability' conditions invoked when matching estimators are viewed in the 'Potential Outcomes' framework associated with Jerzy Neyman (1923) and Rubin (1990, 2011).[3]

So why bother? Why learn new ways to derive known results, a new mathematical language in which to conduct conversations we are already having?

What I aim to show is that graphical language inhibits a sort of ambiguity now common in econometric writing. The conventional mix of algebra and prose, with no clear algebraic representation of causality, makes that ambiguity easy; and the prestige gained from strong causal claims based on sophisticated methods makes it tempting. This is but one of the dangers of which the readers of econometric papers ought to beware of. Deirdre McCloskey has catalogued many such dangers in her study of economic rhetoric, and to this end imported the art of 'close reading' developed in the humanities (McCloskey, 1998). In brief, this means close inspection of the language, taking note of the devices by which authors seek to persuade. For the student of McCloskey's rhetoric, my point can be concisely expressed as follows: graphs are a good tool for the close reading of econometric papers. They help readers form a thorough, concise and organised set of observations to structure their judgment of the paper's claims.

It is not just that drawing a directed graph is a concise way of recording causal assertions, although that simple observation might be enough to justify the effort made below. The graphical theory brings to the fore some simple and important points that are easily lost when causal assumptions are mixed with parametric and statistical assumptions in conventional presentations of structural equation models. Perhaps the most important example – simple enough that I will state the result in full in the brief overview below – is the question of when adding a variable to a regression (or controlling on it non-parametrically) can increase bias rather than decrease it. (I do not mean where omission of one variable compensates some other source of bias by happenstance.)

---

[2] Perhaps most prominently Swanson and Granger (1997) built on early work by Glymour et al. (1987), Pearl and Verma (1991) and Pearl et al. (1991) to devise tests for the contemporaneous causal ordering of shocks in a vector autoregression. Kevin Hoover has developed this further with several co-authors (Hoover, 1991; Demiralp and Hoover, 2003; Hoover, 2005), and David Bessler and co-authors have applied related methods to ask such questions as whether some regional markets are causally prior to others and whether credit booms cause recessions (Babula et al., 2004; Haigh and Bessler, 2004; Zhang et al., 2006) Other examples include Wyatt (2004); Bryant et al. (2009); Kima and Besslerb (2007); Queen and Albers (2008); Tan (2006); White (2006); Wilson (1927); Eichler (2006).

[3] The translation is illustrated in Pearl (2009) Ch. 3.

Although the example I develop here deals with a linear model, the larger advantage comes when non-parametric estimation is feasible. I noted above that the *results* of a graphical analysis can often be *translated into* the conditional ignorability assumptions typically invoked to introduce some sort of matching analysis (say, with propensity scores). This is quite different from saying that the two analyses proceed by the same assumptions. In the graphical case, the assumptions are structural equations – statements about the mechanisms that determine endogenous variables. In a potential outcomes analysis, one states assumptions in terms of the joint distribution of hypothetical variables. This means, for example, defining $Y_i(0)$ as the health of subject i if she does not take medication at a particular moment and $Y_i(1)$ as the same person's health if she does take medication at the same moment; and then proclaiming that the probability – conditional on some set of covariates – of getting medicated at that moment is independent of, say, $Y_i(1) - Y_i(0)$.

This is a fairly natural way to generalise from the intuition of a truly randomised experiment, for the distinguishing feature of such an experiment is that one can say with great confidence that the probability of treatment is independent of everything except the experimenter's coin flip. But for the kind of applications most of interest to economists, this calls on us to think in rather awkward terms. My own enthusiasm upon discovering the graphical approach was due largely to the fact that equally rigorous results could be obtained using the more intuitive idea of causality familiar from structural equations. I had become quite sensitive to the issue after both producing and consuming analyses using propensity score matching when it was still fairly exotic. No very clear idea of why the conditional ignorability assumptions should be believed – why this particular set of covariates was chosen – was being required of us. And when we did dedicate time to that matter, it always came down to intuition about the mechanisms causing treatment and outcome. A directed graph is the clearest expression of such intuition.

The research method advocated by Pearl (2009) is to begin with a graph encoding assertions about what causes what and derive as results that some statistical expressions measure the strength of causal impacts (or that no such measure can be made). In turning this into a method of close reading, I am in a sense reversing the order.

The idea is quite simple. While reading an econometric paper, graph the assumption set crucial to support each causal claim of interest. As you examine the prose being used to defend the approach, refer back to your graph and ask whether the right questions are being addressed. I have found it most useful to sketch what I call the 'fatal graph' representing the causal links which, if present, would invalidate the paper's claims. It is a way of doing more systematically (and perhaps alone) what economists have traditionally done haphazardly in seminars.

In general it is wrong to think of close reading as a way of undermining the papers examined. Not all rhetoric is cheap tricks; much of the analysis in McCloskey (1998) leaves us feeling that classic arguments are persuasive for good reasons we had not previously appreciated. But here I do aim to cast doubt on an influential paper: of the several works to which I have applied this method, I have chosen for illustration one that seems, upon close reading, to have attained unwarranted and potentially dangerous influence.

The paper is Daron Acemoglu, Simon Johnson, and James A. Robinson's (2001) 'Colonial Origins of Comparative Development', hereafter referred to as 'AJR'. Their fundamental argument is that the mortality of settlers centuries ago can be used as an instrument to measure the extent to which the income of ex-colonies around the year 2000 has been determined by such enduring features as how well the legal system protects private property. They find the influence of these features to be large. The paper is singled out for praise by Angrist and Pischke (2010) as an example of the 'credibility revolution' in

macroeconomics and had been cited more than 9,200 times according to Google Scholar by the end of 2016. Prominent among those papers are several that make use of the purported instrument to draw conclusions that may guide policy aimed at expanding opportunity for the world's poorest people. I will argue that those conclusions are profoundly unreliable.

Beyond demonstrating a close reading method and undermining AJR, I aim to support the criticism Chen and Pearl (2012) aimed at econometric textbooks for their equivocal treatment of the causal interpretation of regression models. This case is dispersed throughout the text; the reader will find several quotations, each roughly expressing a valid point about causal inference, but through imprecision buttressing a claim of dubious validity. The imprecision is conventional econometric language, in which we speak of variables being correlated with vaguely defined 'error terms', where the graph theorist makes plainer statements about how variables are linked in causal networks. The econometric tradition is not wrong – the graphical methods are entirely generalisations of Haavelmo (1944), Koopmans et al. (1950) and Marschak (1950), not corrections, and this intellectual debt is much honored (Pearl, 2009). What Chen and Pearl (2012) denounce is the hedged language in which these methods are now taught. They argue that this results from a sort of taboo on causal language that came to dominate statistical science in the generation after those great econometricians. Whether or not that is the reason, much of the language of econometric theory equivocates about causality. The examples below will substantiate that this remains a source of obfuscation even at a very high level, even in the hands of authors whose causal intuition is deep.

Finally, I hope to provide a gentle introduction to the whole field of causal graphs. I thus give the minimum necessary theory in the next two sections. I then review the crucial background in development economics in Section 4. All that can be skipped by a reader with the needed expertise. The original part, the close reading itself, is contained entirely in Section 5, which is followed by a few final remarks.

2. The Necessary Background

The philosophical treatment of causality – from what it is to how we might learn of it – is quite vast, and I am going to make no attempt at a thorough review. This is a paper for economists: we have practical questions in front of us, and would rather not be distracted by epistemology at all, except that we keep accidentally coming to conclusions that aren't true. I will describe just one branch of this work, Judea Pearl's non-parametric generalisation of the econometric structural equation tradition. This is known, for reasons that will become evident, as the 'do-calculus'.

To the subtle question of what 'cause' means Pearl gives a simple answer: it is about manipulation. To say that X causes Y means that anyone who can get control of X, who can set it to a chosen value, can thereby alter the value of Y. There may be other causes of Y, and some of them may be unobservable, so in most cases such a manipulator could never know what Y value would result. But perhaps she could learn how her manipulation affects the probabilities of each Y value. The act of manipulating the variable X to a specific value x is denoted $do(X - x)$. By the *causal effect* of X on Y, we mean the probability distribution resulting from such manipulation $P(Y = y|do(X = x), W = w)$ – i.e., a function giving the probability Y takes on each value y, given the manipulation $do(X = x)$, in an environment where the set of variables W takes on the (vector) value w. Often, the more compact form $P(y|do(x), w)$ is unambiguous, in which case we use it. The do-calculus is all about the relation between this causal effect and the observable distributions familiar from statistics,

such as $P(y|x,w)$. The latter is a function giving the probability Y takes on any value y given we observe $X = x$ rather than manipulating it. Operators which have meaning in terms of a conventional probability distribution have the obvious meanings in terms of the manipulation distribution, e.g.

$$E(y|do(x),w) = \sum_y yP(y|do(x),w) \text{ (if } Y \text{ is discrete).}$$

This is perhaps the place to proclaim my eternal neutrality in a plethora of philosophical discussions, chief among them whether this manipulation metaphor is adequate to cover the whole of what we mean by 'cause'. (See Spohn (2000) and Cartwright (2007) if interested.) Certainly, the metaphor does not exhaust the set of useful observations we have made about causality. Most notably, a set of equilibrium conditions each of which is in itself symmetrical (and thus not causal) can obtain a causal ordering when combined; and the causal ordering of variables present in any given equation can change depending on what other equations are included in the system. This subtle point has been given consequential econometric treatment by Simon (1977), which is by now quite well known.[4] Outside of economics, Dawid (2000) has argued that all useful causal questions can be answered without reference to anything so 'metaphysical' (which means roughly unobservable even in theory) and Robins (1986; 2003) argues for a narrower concept which would disallow manipulating simultaneously two variables that in reality cannot be decoupled.[5]

For present purposes it is really not important whether these subtle points can be treated within the confines of a manipulation-based concept of causality; what is important is that a great number of causal points can be, and quite naturally. Although Pearl has engaged in the philosopher's debate with great vigour, he is again very much in the economist's tradition here. This was how Marschak (1950) understood his structural equations. It is also how Wold (1954) defined causality: 'The relationship is then defined as causal if it is theoretically permissible to regard the variables as involved in a fictive controlled experiment,' i.e. if the cause can be hypothetically manipulated.[6]

The idea of a hypothetical intervention is almost the same as the potential outcomes at the foundation of the Rubin causal model, but the two frameworks encourage different ways of thinking. In Rubin's model, each of the potential outcomes is a distinct variable – for example, $Y_i(1)$ is person i's health had he served in the military, and $Y_i(0)$ is the same person's health had he not served (Angrist et al., 1996). In the graphical approach these are thought of as two values of the same variable, but translation is straightforward: a hypothetical intervention that put someone in the military, but left him otherwise quite the same person,

[4] Wyatt (2004) gives it an interesting graphical treatment, in which the constituent relations are not symmetric but there are rules for reversing causality, which suggests that this too may be folded into the do-based theory.

[5] The word has also been used to refer to entirely different concepts, most notably by Granger (1969): 'Y_t is causing X_t if we are better able to predict X_t using all available information than if the information apart from Y_t had been used', where 'better' prediction meant lower variance, end of story. I am not alone in wishing he had used some other word to describe that useful concept, but by now we are used to saying 'Granger-cause' and knowing it does not refer to our usual idea of causality.

[6] More recently, it is Ed Leamer's definition; or, at least, the definition Sherlock Holmes, as written by Leamer, proclaimed to Dr Watson: 'The word "cause" is a reference to some hypothetical intervention, like putting a gun to the head of the weather forecaster and making her say "sunny". If we actually carried out this experiment, we could get some direct evidence whether or not weather forecasts cause the weather' (Leamer, 2008, p. 176). But Holmes constricts things rather more than Pearl when he insists that the intervention must actually be possible. For Holmes (and Leamer?) the claim that reduced spending on homes caused a recession requires we have in mind that the spending reduction itself is caused by something controlled by actual people, something like taxes or Presidential jawboning. For Pearl, it is only required that we be able to make coherent hypothetical statements: 'Suppose everyone suddenly chose to spend less on homes, for reasons quite unrelated to all their other choices...'

would shift his future health from $Y_i(0)$ to $Y_i(1)$. And both frameworks allow for randomness in the determination of the outcomes: Rubin works with distributions over the two (or more) variables, say $Pr(Y_i(0))$ and $Pr(Y_i(1))$; Pearl with $Pr(Y|do(x))$.

Again, I want to evade the question of whether one framework is superior, embracing only the lesser burden that do, which is nearly unknown, deserves a place. But I will say I find the graphical approach more in line with my causal intuition, which is fundamentally about mechanisms. The statement, 'If my name were Michelle, I would have been born a girl' fits the potential outcomes formalities just as well as does 'Had I been born a girl, my parents would have named me Michelle.' This generality can sometimes be convenient. But it seems rather dangerous, and more so as we move into the subtleties of economic theor, where intuition is so often far from obvious. The second statement conveys clearly the causal mechanism that renders the first statement coherent: we know parents choose names considering a baby's gender. In every context I have ever thought about, if you can't tell the difference between these two orderings, you are missing something crucial about the mechanism under study; and if you can tell the difference, you can state your assumptions in a graph, and thus convey more about them to your readers.

More generally, I noted above that a graphical statement of identifying assumptions can be translated into a conditional ignorability statement; but the reverse is not always the case. The graph may have many variables, with every present and every missing link stating something about the mechanism causing each. The conditional ignorability statement drops most of this information, and just proclaims one feature of the several potential outcomes. This is of course, in some cases, a strength: the identifying assumptions are stated in their most general form. What I am arguing is that sometimes – actually, often – it is valuable to link these statements to precise claims about causal beliefs.

Directed graphs provide the most natural language in which to express such claims. Formally, a directed graph is defined as a set of nodes and a set of directed edges, each of which consists of an ordered pair of nodes. For our purposes, the nodes are variables and the edges denote causation. A graph is most conveniently represented in a diagram, with variable names for the nodes and arrows connecting two nodes for the edges.

Several conventions are useful to note.

First, as already used (and as in Bayesian custom) if a capital letter denotes a variable, then the corresponding lower-case letter denotes a particular value. Thus if we might write $P(Y = 3|do(X = 1))$ to specify a particular aspect of the causal effect, to denote the whole thing we write $P(Y = y|do(X = x))$ or simply, $P(y|do(x))$. This can also represent continuous variables, where a specific value would be something like $P(Y \in (2.9, 3.1)|do(X = 1))$. The convention extends straightforwardly to allow a capital letter to represent a set of several variables, with the lower case then a vector.

Second, in a diagram a two-headed, dashed, curved arrow, known as 'confounding arc,' represents some set of variables that cause both those connected which the analyst has chosen not to represent explicitly. One can always replace such an arc with, say, U for an unobserved set of variables. The use of the arc simply announces that U will not be used in any computations.

Finally, placing a symbol (or perhaps a more complex algebraic expression) next to an arrow indicates the corresponding causal effect is assumed linear. The symbol is then the coefficient in the algebraic representation of the structural equations. (If all the edges have symbols and graph has no cycles, then we can apply Sewall Wright's methods of 'Path Analysis' (Wright, 1921).)

A 'path' in a graph consists of a sequence of edges (some of which may be confounding arcs) such that each has a node in common with the edge just before and after it in

the sequence. The ordering of edges that defines the path has nothing to do with the direction of the arrows. There are thus three types of edge adjacency possible in a path. Figure 1 contains one of each: a *fork* at X (edges pointing away from the shared node); a *chain* at Y (one points in, one away); and a *collider* at Z (both point in).

Figure 1

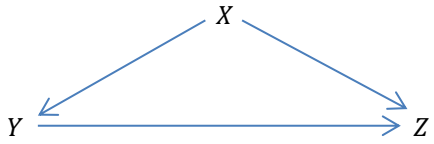


Figure 1 shows a simple directed graph representing a system of three structural equations $x = u_x, y = f_y(x, u_y)$ and $z = f_z(x, y, u_z)$ where the u_i indicate unrepresented ancestors assumed to be unconnected in the full causal network. This unconnectedness assumption is represented by the absence of confounding arcs.

Here is the first key definition. We say that a set of nodes Z 'blocks' (or 'd-separates') X from Y if there is no path between X and Y that does not contain either a member of Z or a collider; and if there is no member of Z, or ancestor of a member of Z, that is itself a collider. This is the central concept needed to validate the use of covariates to calculate causal effects from observational data.

The intuition is this. A path between two variables in a causal network will give rise to a correlation between those variables unless the path contains a collider; and if it does contain a collider, than it gives rise to *conditional* correlation (i.e., conditional on the collider variable). Thus, suppose Y and Z are only connected by a fork or chain through X. Then there will be a statistical association between Y and Z – i.e, knowledge of a unit's Y value (given knowledge of the system) will strengthen prediction of Z and *vice versa*. But in a sample chosen to have only a single X value, this association disappears. In the contrary case, if Y and Z are connected only by a collider, there is no unconditional association; but if we chose a sample to have limited range of the collider variable then values of Y causing an increase in the collider variable will tend to be associated with values of Z causing a decrease.

In the case of a linear system with Z and Y linked only by a non-collider X, regression of Z on Y or *vice versa* will reveal an association corresponding to the product of the correlations of each variable with X, while a regression including X as covariate will indicate (for an infinite sample) zero conditional correlation between Z and Y. If, as in Figure 1 there is also an arrow from Y to Z, the regression with X included will measure the causal effect of Y, while one without X will measure instead the sum of the correlations induced by the two paths – i.e., it will be confounded.

That much is a graphical representation of matter well understood by most econometricians. One of the things that graph theory makes clearer than conventional treatments is the circumstance in which including a regressor can create bias. This is the other half of the definition of 'blocking'. In Figure 1 regression of Y on X measures the causal effect of X, while inclusion of Z confounds causal interpretation. In this case, the error would be avoided by the simple advice not to include a variable influenced by the outcome, but subtler cases exist (one is shown in Figure 7(b) below).

I have reminded the reader that one *must* control on a latent common cause in order to give a measure of association causal meaning; and one must *not* control on a common causal descendent. It is natural to ask what set of conditions are necessary and sufficient to insure that a set of potential controls validates causal interpretation.

In the case where the graph has no cycle (i.e., no sequence of edges that can be followed forward to where it began) a set of three fairly simple rules has been shown to be 'complete': applying them can always prove whether a conditioning set is acceptable (Huang

and Valtorta, 2006). Although we will use only a fraction, I include the whole set of three rules here as inspiration: this is all you need to know to determine what causal effects are non-parametrically identifiable once you have expressed causal assumptions in a graph.

Let $G_{\underline{T}}$ denote the graph constructed from G by removing arrows out of T; let $G_{\overline{Z}}$ denote that constructed by removing arrows into Z; and G the combination. The notation $(T \perp Y|W)_G$ is read 'W blocks all paths between T and Y on graph G'. A causal effect can be non-parametrically identified if and only if it can be transformed to a do-free expression by applying the following three rules:

Rule 1 (removal or insertion of conditioning nodes): if $(Y \perp W|T,Z)_{G_{\underline{T}}}$

$$P(y|do(t),w,z) = P(y|do(t),z)$$

Rule 2 (replacement of action with conditioning): if $(Y \perp T|W,Z)_{G_{\overline{Z}\underline{T}}}$

$$P(y|do(t),do(z),w) = P(y|t,do(z),w)$$

Rule 3 (removal or insertion of action): if $(Y \perp T|W,Z)_{G_{\overline{X}\overline{Z(W)}}}$

$$P(y|do(t),do(z),w) = P(y|do(t),w),$$

where $\overline{Z(W)}$ is the set of all nodes in Z that are not ancestors of any W node in $G_{\overline{X}}$.

Each of these rules allows you to consult your causal graph and determine whether a particular step can be taken. If in a sequence of such steps you are able to derive an equation with a do expression on one side and no do expression on the other, than you have shown that causal effect is measured by an observable statistical distribution. What statistical treatment to give – a series of conditional means, linear regression, Probit, or what have you – is a subsequent question, subject to exactly the considerations governing a purely descriptive discussion of associations, but with its causal interpretation now assured (conditional on the graph, of course). If there is not a set of steps that result in such an equation, the causal effect cannot be identified from the assumptions about causal ordering alone (Huang and Valtorta, 2006). In that case, turn to Matzkin (2008) for the more general (and much more difficult) theory which can allow identification conditional on parameterisations, for example, linearity, sign restrictions, etc.

The aspect of the above rules most used, and perhaps the most intuitive aspect, is known as the 'back-door' criterion, which is implicit in Rule 2. A back-door path between a treatment and outcome is one containing an arrow into the treatment. The idea behind the terminology is that such a path allows an association to enter through the 'back door' – i.e., through the causes of the treatment. What we want is to be able to pretend that the treatment has no causes, as though it were an experiment, so the key thing we need is to eliminate is any association between treatment and outcome that arises from this back door. This is done by including as controls a set of variables that blocks all back-door paths. In the linear case this means using them as covariates in the regression, in other cases – matching or binning. The back-door criterion has been shown valid beyond the acyclic case.

This brief discussion is, of course, intended to give the gist, not to convince anyone or to convey the whole result; see Pearl (2009) for a thorough treatment. For present purposes the most important aspect of the graphical approach is that, as the following discussion will illustrate, graphs capture our usual causal intuition at its simplest; the development of rigorous methods with graphical foundations spares us from dangerous imprecisions that easily arise in the language of traditional econometrics, in which causal concepts often seem awkwardly added to statistical theory.

3. Instrumental Variables in Graphs and Econometrics

The method of instrumental variables is a core tool of modern econometrics. It lies at the heart of all claims to measure causal relations among variables that need be seen as simultaneous outcomes. The classic example is the relation between the amount of some good sold in a competitive market and its price. These two variables are taken to be jointly determined in the observed system – the 'market' – but price is supposed to cause quantity in each of the subsystems ('supply' and 'demand') composed of the decisions of particular agents. An instrumental variable is one that enters these subsystems in such a way as to allow these causal impacts to be measured.

The method was first developed by economist Phillip Wright, influenced, if not assisted, by his son Sewall (Stock and Trebbi, 2003). Sewall Wright was the biologist who developed path analysis, which is one of the sources from which the modern graphical theory of causality descends. One might, then, expect the two methods to have evolved together and econometricians to have benefited from the clarity of Sewall's graphical language. That is not what happened.

Instead, Phillip Wrights's work was largely ignored, and after the World War II the method of instrumental variables was rediscovered independently by Cowles Commission researchers. Although Haavelmo (1943), Marschak (1950) and Koopmans et al. (1950) were quite clear about causality in prose, they eschewed graphs in favour of an algebraic language in which causal equations looked just like equilibrium conditions. In subsequent decades statisticians came to shun causal concepts entirely, and econometric theorists (if not practitioners) uneasily followed suit.

Thus, for example, the subject of instrumental variables was introduced in my graduate school econometrics text with the following language:

> 'Suppose that in the classical model $y_i = \mathbf{x}_i'\beta + \epsilon_i$, the K variables $\mathbf{x}_i$ may be correlated with ϵ_i. Suppose as well that there exists a set of L variables $\mathbf{z}_i$, where L is at least as large as K, such that $\mathbf{z}_i$ is correlated with $\mathbf{x}_i$ but not with ϵ_i. We cannot estimate β consistently by using the familiar least squares estimator. But we can construct a consistent estimator of β by using the assumed relationships among $\mathbf{z}_i$, $\mathbf{x}_i$, and ϵ_i' (Greene, 2003, p 75).

Few of us knew just what it meant that a variable was 'correlated with ϵ_i' except that one was not then allowed to use $E(\mathbf{x}\epsilon) = 0$ to simplify subsequent expressions as was permitted in the contrary case. Sometimes it seemed that ϵ was standing for the causal influence of all the unobserved variables, but even then it was unclear just what was being assumed by asserting that this summary of the unobserved was uncorrelated with some explanatory variable. And in many cases ϵ seemed just to be a technical convenience required by the fact that the regression line cannot be a perfect fit. Indeed, it is exactly that in the many non-causal applications of regression, as in exploration and forecasting.

Some of our teachers, and some applied papers, spoke more plainly: Z, they might say, can be an instrument for the effect of X on Y if Z causes X, but is not for any other reason correlated with Y. But correlation and causality coexisted uneasily in the formal treatments of the textbooks and theoretical papers. So we learned to write in hedged and ambiguous language which, as will become clear below, still misleads some of our best practitioners.

In the plain causal language of directed graphs, Z can serve as an instrument if all of the following hold in any graph G of the true causal network (Pearl, 2009, p. 248):

- There is no unblocked path from Z to Y that does not contain an arrow into X.
- Z and X are not causally independent – i.e., there exists an unblocked path between them.
- The relations among Z, Y and X are all linear.

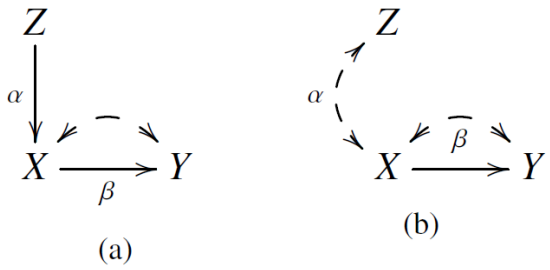
Figure 2 illustrates the two alternative assumption sets that can justify the use of Z as an instrument. Recall that the algebraic symbols beside each path denote causal linearity. Thus, for example, $dE(y|do(x))/dx = \beta$. (To emphasise, the do means we are talking about the derivative of the structural equation, here $y = \beta x$. There are other valid meanings of dy/dx, such as the whole association including that induced by the confounding arc.) By a convention going back to Wright (1921) which sacrifices no generality, the variables are measured in standardised form (difference from mean over standard deviation) so the structural equations have no constant terms.

Consider panel (a). The covariance between Z and Y is due only to the path through X, and is the product of two causal effects, that of Z on X and that of X on Y, labelled α and β. That is, $cov(ZY) = \alpha\beta$. (The linearity assumption means that these are scalar constants.) Since α is the covariance between Z and X, the causal impact of X on Y can be measured as

$$\beta = \frac{cov(ZY)}{cov(ZX)}$$

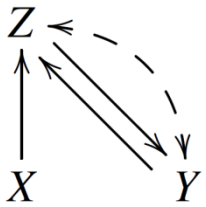
The alternative assumption set (b) is using essentially the same logic, but with Z standing in as a proxy for the unnamed common ancestor represented by the confounding arc.

Figure 2 Alternative conditions for Z to be a valid instrument to measure the effect of X on Y



As is generally true, some of the most important assumptions are represented by the arrows that are *not* present. For the purpose of close reading, I find it useful to represent these explicitly as in Figure 3. I call this the 'fatal graph' for the proposed instrument, since the presence of any of the represented links kills the claim that the instrument is valid. The links in the fatal graph can be crossed out as we read, whenever we are convinced by an argument that the link is absent or blocked. If any remains when we are done, we ought not to believe the paper's claims.

Figure 3 Fatal graph for the use of Z as instrument to measure the effect of X on Y. If any of these links is present, Z is not a valid instrument for measuring the impact of X on Y.



This is perhaps the place to note that such analysis might be taken much further. If instead of crossing out links, one were to mark each with a number representing the subjective probability of its absence, opinions from very diverse parts of one's intuition, and indeed from the intuitions of diverse people, could be combined into far from obvious but completely transparent calculations of the probable strength of causal effects. But I will not pursue such analysis here.

4. Some Background on Development Economics

The paper I have chosen for the example is profoundly influential. I presented some evidence of its prestige in the introduction, but one needs an idea of the background arguments to see how important the issues are.

Here is how the matter is characterised by Dani Rodrik – something of a gadfly, but one of our most prominent development economists – writing in the *Journal of Economic Literatures*:

> 'Daron Acemoglu, Simon Johnson, and James A. Robinson's (2001) important work drove home the point that the security of property rights has been historically perhaps the single most important determinant of why some countries grew rich and others remained poor. Going one step further, Easterly and Ross Levine (2003) showed that policies (i.e., trade openness, inflation, and exchange rate overvaluation) do not exert any independent effect on long-term economic performance once the quality of domestic institutions is included in the regression' (Rodrik, 2006).

It is a discouraging conclusion, for 'institutions' in this context means almost by definition things that are difficult to change and nearly impossible to change quickly. To economists, institutions are the rules of economics games, and they don't really change except as part of a game-theoretic equilibrium including widespread beliefs (North, 1994). A mere constitutional amendment may sometimes be called 'institutional change', but the change is not really considered to have happened until enforcement is widely credible.

The security and clarity of property rights are institutional in this sense of the word. In the United States these rights are codified in constitutional provisions prohibiting governments from seizing private property even should it be the case that all elected representatives believe some people have too much of it; and this 'formal institution' is backed by a widespread belief in the legitimacy of private property rights and of the idea of limiting government. The latter, 'informal' part of that institutional arrangement is widely shared among the English-speaking nations, and it has long been speculated that this accounts for a large part of the prosperity characterising those nations. Where property is thus secured, it is

argued, people are more willing to invest in the improvement of land, the building of factories, the organising of business, and other such productive ventures which, in other societies, might be taken by politicians or neighbouring mobs once they proved valuable.

Such institutions were the central focus in AJR. The institutional variable on which they concentrate is 'protection against expropriation risk', measured by a firm called 'Political and Risk Services' for the benefit of global investors. They argue that this is also the best available measure of property rights security for domestic agents. (Some alternative institutional measures are considered as 'robustness checks'.) The main causal effect they seek to measure is that of protection from risk around 1990 on GDP per capita in 2000.

This may at first seem quite odd. The biological metaphor embedded in the phrase 'economic growth' encourages us to think of a national economy as something that must be constructed, like a growing plant or population, from the energies of its own constituent parts. And this is how models in the Solow (1956) tradition work: the population saves some part of current income and turns it into 'capital' such as factories and roads. More capital per person allows more income per person. Such a process leads naturally to *exponential* growth – that is, a kind of speed limit expressible as a fraction of current income. Thus, it may seem natural to think that countries that were poor centuries ago should still be relatively poor, even if all policies and institutions were everywhere the same, rather than viewing relative income today as 'current performance'.

In Solow's model, however, income differentials do not persist. This is because each addition to a nation's capital stock adds less to its income than did earlier additions – that is, there are decreasing returns to capital. The result is that if nations have the same rates of population growth and technical progress and the same propensity to save, those which are poorer just because they are further behind in the capital accumulation process will grow faster. Incomes ought to converge. Parameterised versions of the Solow model indicate that the poorest countries should converge to near equality with the rich in just a few decades, if indeed all these other things are equal (Dowrick and Rogers, 2002). Further, a poor country can achieve faster technological progress by learning from the rich (Durlauf et al., 2005).

Consider as well that countries are part of a world economy. If they wall themselves off from the rest that is a policy decision, or perhaps an institutional outcome – in any case, not a fundamental natural limit. And if they are *not* somehow walled off, then there is no reason their capital stock growth should be limited to what they can save. Quite the opposite: in a world of perfect markets, a nation with lots of labour and land will be a very attractive location for those thinking to build factories and roads – the return on capital investment will be higher than elsewhere – and capital should flood in rapidly until that return has fallen to the prevailing global level. If labour and land are the same everywhere, that rapid flow will continue until the prices of all inputs to production are everywhere equal. In particular, wages will be equal. Capital will flow in as fast as investors can organise themselves until GDP per capita in the countries starting poor is just as high as in those countries starting rich. Such equilibration ought not to take many decades. If this is indeed the whole burden of the past, then it is *only* the slow-to-change nature of institutions that prevents a nation from leaping all the way from poor to rich in a few decades. And historically there are cases of unusually fast institutional change followed by growth that at least doubled income in a decade: Chile after democracy was restored under a constitution giving unprecedented protection to property, or China after Deng Xiaping decollectivised agriculture and opened industry to private firms. The economy of Shenzhen, a city designated for the most ambitious reform experiments, grew a thousand-fold in 30 years, and almost seven-fold *per capita*.

Such notions are in the background of AJR: current levels of GDP per capita are thought of as the 'current performance' of the economy, influenced by current institutions.

That 'current' is a long-run concept encompassing the final decades of the 20th century, or perhaps even the whole time since WWII. The informative data are the cross-section of national income levels. Take that for the moment as given, and accept also that institutions such as protection from expropriation risk are likely important causes of that performance. Generations of work by economic historians and economists debating the limits of the Solow growth model tradition render this at least a reasonable point of departure. And accept (for the whole of this discussion) that to measure 'institutional quality' as a number is meaningful (although this seems unlikely to stand up to sceptical reading) and the variables used in AJR measure it accurately. It is then tempting to interpret the correlation between performance and institutions as measuring the causal effect of interest.

For two reasons that is unacceptable. The first is that institutional quality itself is likely to be strongly affected by economic performance. This effect could go in either in direction, although positive is the usual concern. 'It is quite likely that rich countries choose or can afford better institutions' is how it is phrased in AJR (p. 1369). Of course, not all that is good is protective of property and it certainly is not out of the question that social welfare motivations, an unwillingness to massacre protesters and honest elections could lead the propertied to feel insecure. But whether wealth causes growth-enhancing or growth-retarding institutional change, the implied reverse causality means that correlation does not measure the impact of institutions. So too do the many reasons institutions and income may have common causal ancestors.

Both problems are overcome by a valid instrument. The heart of AJR is the claim that the mortality rate of settlers in early European colonies will serve. The core argument is that where many colonists were killed by disease, European powers set up 'extractive' institutions to transfer resources rapidly to Europe, whereas in healthier environments they settled and recreated the institutional environment of the home countries. The latter included more secure property rights as well as some other desirable features within the category we have come to call 'institutional'. Thus, they argue, the early mortality rates serve as a randomising agent to measure the impact of an institutional measure on long-run economic performance. This is the claim we must scrutinise as we give the paper a close reading, a task to which I now turn.

5. Reading 'Colonial Origins'

Start as with any close reading: the role of graphs will be stressed, but the goal is to unpack all the language, prose too, and be clear about what to believe.

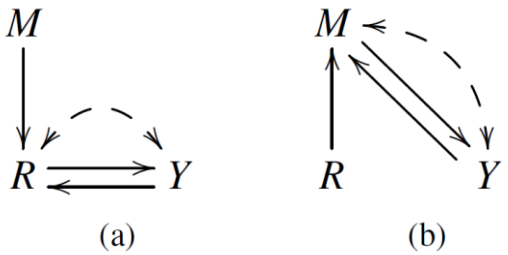
'What are the fundamental causes of the large differences in income per capita across countries?' So begins 'Colonial origins'. Economists often motivate regressions with such questions, even though they do not correspond well to econometric theory. Instrumental variable methods, unlike ordinary least squares, are about causality, but they are not designed to determine what is a cause and what is not. Rather, the analyst is supposed to know the causal structure (at least as analytical fiction), and then measure the strength of the causal effects. The motivating question as posed would seem to call for methods under development in the tradition of Spirtes et al. (2000).

The appearance of the word 'fundamental' in that opening sentence provides ambiguity that rescues the work from what would otherwise be misuse of econometric method. They shall, we may infer, measure causal effects conditional on causal assumptions as authorised by the textbooks; and then characterise some of those effects as large enough to be 'fundamental', thus allowing us a take-home message easier to memorise than the whole set of estimates.

That this is indeed the mission is made clear in the following paragraphs. Institutional variables will be the central focus; their importance 'receives some support from cross country correlations', but 'we lack reliable estimates of the effects'. Attention is then drawn to the problems of simultaneous causation and common causal ancestry that disallow interpretation of the cross-country correlations as causal measures (text quoted in last section).

This is the place to start graphing. Let Y be income (actually, log income per capita, but in the graphical view such details are put aside) and R the institutional variable. We have been told the problems to be solved included two-way causality, as seen in the bottom part of Figure 4(a); and common causal ancestry, represented by the confounding arc.

Figure 4 Minimal graphs for AJR identification strategy: the assumed causal structure (a) and corresponding fatal graph (b). M is settler mortality, R is protection against appropriation and Y is log per capita GDP around 2000/



Using the conventional language of econometricians, the authors point out that 'we need a source of exogenous variation in institutions' in order to estimate the impact of R on Y. As noted above, in graphical terms this suggests a variable from which an arrow points into R, which is not linked to Y directly or by any unblocked back-door path.

The language with which AJR begin to justify these missing arrows is classic econometric jargon. 'The exclusion restriction implied by our instrumental variable regression is that, conditional on the controls included in the regression, the mortality rates of European settlers more than 100 years ago have no effect on GDP per capita today, other than their effect through institutional development.' Here an indisputably causal phrase, 'no effect on', is blurred by a clause with only statistical meaning, 'conditional on the controls included in the regression'. Crucially, this blurring helps to understate the strength of the assumptions that need to be accepted, and this understatement will be repeated several times.

Figure 4 makes it rather simpler to see what needs to be justified. A graph representing justified causal assumptions must be reducible to one rather like (a) – more precisely, one in which all of the arrows in the fatal graph (b) are either absent or blocked.[7] It may be granted without much discussion that neither GDP in 2000 (Y) nor property rights protection circa 1990 (R) caused settler mortality (M), which is mostly from before 1848. Thus the two upward pointing arrows in Figure 4(b) may be swiftly crossed out. What remains is a close reading of the language in which they claim that the remaining two arrows are absent or blocked.

The description in AJR emphasises the absence of the arrow $M \rightarrow Y$. The possibility of blocking the confounding arc between M and Y is apparently what they meant by stating that the direct effect of M must not exist 'conditional on the controls'.

[7] To *reduce* a graph while retaining the needed causal content, proceed as follows: when a node A is removed, all the children of A became children of all of the parents of A, unless there were no parents of A represented, in which case the children are linked to each other with confounding arcs. Repeat until only the desired nodes remain.

But to say that an effect is conditional may mean something quite different. The effect of a trigger pull on the lifespan of a rabbit is conditional on the presence of a shell in the shotgun, on the hunter's aim, etc. If this is what AJR had in mind, they simply forgot the danger of the confounding arc altogether.

Of course, that is not the problem. Despite the fuzziness of the language, these authors (and many other economists) clearly have a pretty good idea of the issues. But by failing to express clearly and consistently that their control variables are required to block the path $M \leftarrow - - - \rightarrow Y$, they distract us from a critical failure of the purported instrument.

Similar deflections of attention occur at several other key points. The first mention of the identifying assumption, in the introduction, is quite similar to the quote above, although it is followed by a clearer acknowledgment of what that 'conditional' means (discussed below). Much later, the presentation of ordinary least squares coefficients relating R to Y concludes with a reminder of why they cannot be given causal interpretation, and concludes 'All of these problems could be solved if we had an instrument for institutions. Such an instrument must be an important factor in accounting for the institutional variation that we observe, but have no direct effect on performance' (p 1380).

The point is stated similarly in terms the error term defined by their Equation (1)

$$\log y_i = \mu + \alpha R_i + \mathbf{X}_i'\gamma + \epsilon_i$$

> 'This identification strategy will be valid as long as $\log M_i$ is uncorrelated with ϵ_i – that is, if mortality rates of settlers between the seventeenth and nineteenth centuries have no effect on income today other than through their influence on institutional development' (p. 1383).

Again there is the unacknowledged shift from correlational to causal talk. The first part of the sentence is exactly right, but it needs causal content to be useful; and when they give it that content, they remember only one part of the causal assumption they are making, i.e., that M is not a causal ancestor of Y. Again, the assumption of no confounding arc is brushed aside. But let us consider what they have to offer by way of an argument against the fatal arrow $M \rightarrow Y$.

The defence of their instrument is, we are told (following the above quote from page 1380) contained in their Section I. Its structure is shown in Figure 5, which represents their entire system of equations. If Figure 5 is reduced to the three variables of Figure 4 it contains an arc connecting M to Y. But as long as X is observed this path is blocked, so if Figure 5 is accepted as representing a true linear causal system the instrument is valid. The most questionable assumptions are shown in the partial fatal graph Figure 6.

The identification defence in AJR Section I chiefly argues for the likely presence of key arrows – the second condition for instrument validity listed above – dedicating one subsection each to $M \rightarrow S$, $S \rightarrow C$ and $C \rightarrow R$.

The first gives evidence that high mortality rates discouraged settlers, which I find hard to doubt.

The second documents a distinction between 'types' of colonisation – settler versus extractive – which is shown to have a substantial pedigree among historians. Taken at face value, this implies a binary classification of C, which seems inconsistent with the linear causal impact required for instrumental variables estimation. And most of the discussion indicates that face value is the right way to take it: the two colonial types were developed for two very different purposes, with no apparent motivation of intermediate types. But for present

purposes, I will let that pass – it will only be an issue if we are persuaded that all the links in Figure 6 can be crossed out.[8]

Figure 5

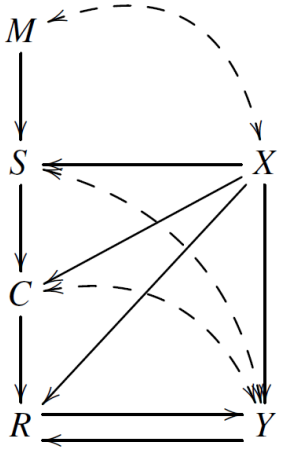
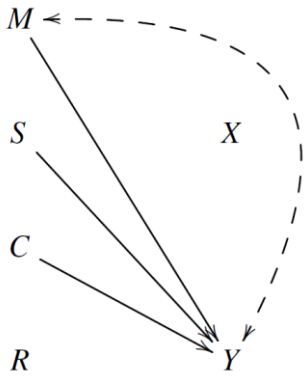


Figure 5 shows Acemoglu et al.'s equations (1)-(4). Their identification claim is valid if this graph represents the true causal mechanism. M is European settler mortality prior to 1900. S is European population in 1900. C is a measure of institutional quality around 1900 or at the time of independence. R is a measure of institutional quality near the end of the 20th century. Y is log GDP per capita in 2000. X varies across treatments, including various combinations of latitude, continent, coloniser, legal origin, temperature, distance from coast and disease indicators.

Figure 6 Partial fatal graph for the identification strategy in Figure 5. Fatal arrows, the absence of which is easily accepted, are not shown.



The third of these links, that institutions persist ($C \rightarrow R$, subsection I.C) is the most historically surprising. Several authorities are cited to support the claim that in general, even when anti-colonial forces kicked the European populations out of newly independent countries, they adopted the inherited institutions. This is of course crucial if differences are to be attributed to institutions rather than ethnic or cultural identities of the elites: the former must sometimes

[8] The causal logic of instrumental variables has been applied to discretely measured causes by Abadie et al. (2002), but a distinct measure is required.

have persisted even where the latter did not. Footnote 10 contains several examples from Africa where this seems to be the case. The Latin American examples do not provide the same differentiation of institutional from cultural persistence, but they do support the central claim that institutions persist – i.e., $C \rightarrow R$. The point is also supported by compelling explanation – for example, that it is probably much easier for successful rebels to step into existing institutional roles than to redesign government and law (the first of three numbered points on p. 1376).[9]

On the whole, I find these arguments persuasive. But as noted, they only claim that the arrows on the left side of Figure 5 are present. That is, the section we were told 'suggests that settler mortality during the time of colonisation is a plausible instrument' (p. 1380) contains no argument at all against the arrows fatal for that claim.

Some of those arrows can be easily dismissed as implying causes that occur long after the effects. The rest are shown in Figure 6. The procedure I advocate is that, having come to this point, we first ask whether intuition or our own knowledge allows any of these arrows to be crossed out, then re-read the paper seeking arguments to cross out those remaining. At the same time, of course, we want to consider arguments in the contrary direction – i.e., are there strong reasons to believe that any of these arrows actually does exist? In the present case I also examined a working paper version (Acemoglu et al., 2000) in case space constraints had kept important points out of the journal.

Essentially all discussion of the fatal arrows is in the penultimate section, entitled 'Robustness'. This opens with a fifth instance (I have not discussed them all) in which the identifying assumption is stated as though $M \nrightarrow Y$ were sufficient. The issue of back-door paths (M← −−→Y) is then incorporated with a strong implication that it is no more troublesome than that of M → Y, consideration of which is rather going the extra mile for credibility:

> 'The validity of our 2SLS results in Table 4 depends on the assumption that settler mortality in the past has no direct effect on current economic performance. Although this presumption appears reasonable (at least to us), here we substantiate it further by directly controlling for many of the variables that could plausibly be correlated with both settler mortality and economic outcomes, and checking whether the addition of these variables affects our estimates' (p. 1388).

That 'presumption' shifts the burden to the sceptic. Perhaps this is not unreason able: if, with some effort, we cannot think of some good reasons why M might cause Y through non-institutional channels, then $M \nrightarrow Y$ should be accepted. But it does not take much effort to come up with a list such channels: genes, traditions, social networks, language and financial wealth are all inherited and not institutions. So let's concentrate on whether that arrow might be negated.

The case for this missing arrow is made by considering alternative paths that would be blocked by some observable variable, including that variable in the regression, and observing that the key result of interest changes little. This is a legitimate argument provided that all the paths likely to exist can be blocked, although it raises some subtle issues. If the blocking assumptions are correct, then the regressions including the blocking variables are the unbiased ones – the 'baseline' is not. Presumably, the justification for the baseline-

[9] 'Forced labour' makes its first appearance in this section, as an institution alleged to be 'persisting' – although in fact, it is 'reintroduced', and no mention is made of its ubiquity outside of these colonies, or indeed its presence in the United States at least through the 1860s (with some reintroductions thereafter through vagrancy laws (Wilson, 1933; Glenn, 2009)).

robustness structure is about statistical power, an important issue on which graphical analysis is silent. There is also the danger that the blocking assumption is wrong in such a way that adding the control causes bias, as illustrated in Figure 7(b). And note, in support of my broader brief that economists should study causal graphs, there is no comparably simple way to summarise this situation without them. For example, even the causal structure Figure 7(a), in which the AJR argument is justified, violates the traditional advice that control variables be 'quite unaffected by the treatment' (Cox, 1958 quoted in Pearl, 2009).

Figure 7

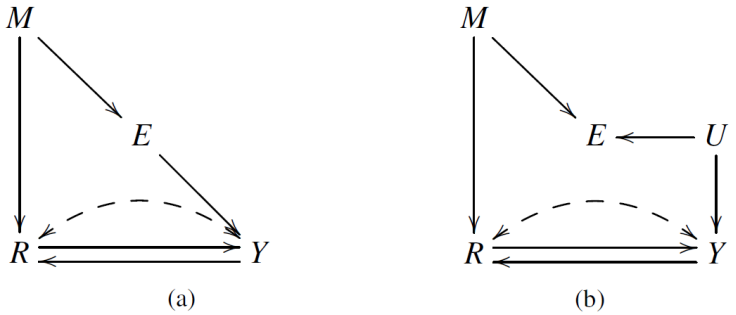


Figure 7 shows causal structures that justify AJR's claim that E (say, percent European population in 1975) can block a path that would otherwise be fatal for their identification strategy (a); and a plausible alternative (b) in which that fatal path does not exist, and wrongly including E in the regression causes bias. U is some unobservable variable, perhaps from the early 20th century, that affected both later economic development and net migration of Europeans from nations in the sample.

For present purposes, the key question is whether it is plausible if all the relevant paths are blocked – i.e., whether the node E in Figure 7(a) can be interpreted as any subset of the observed candidates. The candidates discussed are legal system, religion, European population, and 'Ethnolinguistic fragmentation', which is a measure of how many languages are spoken in a country.

Of these, the most informative would seem to be the fraction of the population that was European in 1975. This could conceivably capture all the non-institutional channels that occurred to me above, and its inclusion has little impact on the main causal measure of interest. In the baseline specification, a one-point improvement in protection against expropriation is estimated to cause 0.94 log points of income increase (that is, it about doubles per capita income). With the percent of Europeans considered as an alternate channel the estimate increases to 0.96 – no difference. So let us concentrate further on that variable.

It is not what it may seem. Throughout Latin America, Mestizos are classified as not of European decent, but nearly all speak Spanish and many of their cultural practices and about half their genes are European. Thus Mexico is supposed to be 15% European, but at least 70% of the population has European ancestors. Some of those Mestizo families have been in constant cultural contact with Spain and the rest of the world through literature, travel, receipt of guests, etc. – less so, no doubt, than the 15% classified as 'European', but much more so than the Guatemalan ethnic plurality, which is Mayan. Yet Guatemala was given a score of 20% European in 1975 – it is apparently more European than Mexico. And this same 20% score is assigned to Costa Rica, where 90% would be closer to the mark. This is because rather than actual 1975 data by nation, all of Central America was given the same figure, 'assuming unchanged from Central American proportions of 1800' (Acemoglu et al., 2000, Appendix Table A5).

Because I study Latin America, it is easy for me to see these faults and their significance. I don't know, but it is hard to believe that the zeros assigned to 23 of the 27 countries in Africa, and to all of South Asia and the Malay Archipelago, capture the actual imprint of European culture, genes, connections and asset portfolios on the post-colonial populations. Bits of casual knowledge as well was intuition suggest otherwise. India still has a large English-language educational system, so large that is has absorbed much of the American phone-based software support business. Unless all these people learned English because India's high Protection Against Expropriation Risk[10] encouraged enterprise, this is the sort of non-institutional channel we are supposed to be ruling out. Yet India is given a zero on the European population variable.

Thus, it is not surprising and not terribly informative that adding 'Percent of European descent in 1975' to these regressions alters nothing. The variable appears almost independent of the whole system, just as a meaningless variable would be. It is a very messy measure for the purpose.

The other variables that might be candidates for blocking this path add little. Take my intuitive list of alternative channels – genes, traditions, social networks, language and financial wealth – and ask whether each is likely to be well captured by anything in the list of candidate blocking variables – whether British or French law is used; what is the dominant religion; how many languages are spoken. Recall the way causation is defined in this tradition: what we want to know is whether, if we could independently manipulate the candidate blocking variable, would we be able to neutralise each of those proposed non-institutional channels? For example, if we could travel back to 1900 and make a country Catholic, would this wipe out the impact of the transnational social networks centred on England? Ask Graham Greene. From my intuition, I come up with zero of the five channels blocked. I can hardly imagine anyone believing that all five are blocked; and I can easily imagine that someone else's intuition adds to the list of five.

The treatment in AJR leaves a very different impression, arising partly from a different structure. Rather than trying to lay out all the reasons the alternative channel might exist, then what variables might be used to block them, the discussion in AJR is organised around the variables. The sequence opens with, 'La Porta et al. (1999) argue for the importance of colonial origin (the identity of the main colonizing country)...', then 'von Hayek (1960) and La Porta et al. (1999) also emphasize the importance of legal origin...', and so on. This is the rhetorical convention in econometrics, and it has this insidious aspect, that the variables no authority has championed – perhaps because they cannot be observed, and thus lead to no authority-building publications – are neglected. The result is a sort of argument by selective refutation, which is sometimes classified as a logical 'fallacy of omission'. It would be wrong to accuse these authors of that fallacy, which generally involves evasions or distortions of points an opponent has explicitly made in a debate. Quite probably, not all the alternative channels I have listed have been brought to their attention. But in terms of our ability to cross out an arrow in the fatal graph, the effect is the same.

It is possible, of course, that these alternative channels are simply not important. That seems to be what AJR implied by relegating consideration of the issue to 'Robustness', and by remarking that their 'presumption appears reasonable'. It does not appear that way to me, but all we can do with the tools at hand is to bring the key questions into the foreground – the reader may fully grasp my points and still agree with Acemoglu, Johnson and Robinson. Rather than stating that the prosecution rests, therefore, let us consider the back-door paths.

The case against M← ←→Y takes the same form as that against M → Y, with selected channels considered and alleged to be blocked. The candidates for blocking the back-door

[10] Measured at 8.27, more than a standard deviation above the base sample average.

path are the identity of the colonising country,[11] latitude, temperature, humidity, soil quality, whether land-locked, natural resource endowments and several indicators of disease burden.

Some readers may be impressed by the sheer number of control variables to which these results appear robust. There are, however, many more variables that might have been used. Sala-i Martin et al. (2004) examined 67 correlates of long-run growth, finding 18 of these correlations to be 'robust'. About half of those 18 or close correlates have been incorporated into AJR, and some others could be seen as mediators in channels AJR include. But since it is all but certain that regressions on a sample of 64 countries will not be robust to the inclusion of 67 covariates, and there are many other things measurable that Sala-i Martin et al. (2004) did not examine (including at least two in AJR – yellow fever and distance to the coast) a high degree of scepticism is warranted regarding the process by which covariates were chosen for presentation. A good process would be to graph likely causal channels and choose one variable sufficient to block each.

Most of the variables listed above do not strike me as especially likely to constitute back-door channels. The land-locked status, for example, and natural resource endowments have obvious impacts on current income, but they do not stand out as likely also to have affected settler mortality. In a setting with plenty of observations that would be no reason not to throw them into the regressions – although since borders themselves were determined after the settler mortality it is not safe to assume that they could not cause bias – but this is not a case with plenty of observations. Mostly, this choice of control variables seems rather arbitrary, and no argument is presented that would make these the top priority for inclusion.

However, the authors do seem to share my intuition about the story in this channel most likely to be powerful, the next target of concentrated scrutiny. Disease reservoirs would have killed lots of European settlers and might continue to burden economic development. There is appropriately much discussion of this issue. It begins in the introduction, with some language that makes clear the authors do indeed refer to blocking back-door paths when they say 'conditional on the controls included...' in the language I criticised above as causally ambiguous: 'The major concern with this exclusion restriction is that the mortality rates of settlers could be correlated with the current disease environment, which may have a direct effect on economic performance' (p. 1371).

The case against that fatal arrow commences immediately with a preview then gets its own dedicated subsection (III.A). Some 80 percent of European deaths were due to two diseases, yellow fever and malaria, with another 15 percent due to gastrointestinal disorders. Those top two diseases do not kill many adults in the indigenous populations, who have high rates of immunity both due to childhood exposure and genetic inheritance. From these observations, AJR conclude 'settler mortality is a plausible instrument for institutional development: these diseases affected European settlement patterns and the type of institutions they set up, but had little effect on the health and economy of indigenous people' (p. 1382).

But that is not what they showed. Their evidence only shows that the top two diseases had little effect on adult mortality. That is not the same as 'health and economy'. The description they give of children developing immunity is also a description of bodies burdened during years crucial to brain development. It must also be a burden on parents, as is the mourning and burial of those children who do not survive. The genetic immunity, too, has side effects including sickle cell anaemia. On what basis are we to assume that none of this has an impact on what people can produce with a day at work, and on how much they can develop

[11] This variables was incorporated both as dummy variables for each coloniser and as a single dummy variable for 'French legal origin' applying to all the colonisers that had maintained Napoleonic civil law reforms. The latter choice has only the advantage of statistical power.

themselves in a day of study? And on what basis do we conclude that the 15 percent of deaths due to gastrointestinal disorder is too little to matter?[12]

No arguments are presented. Among the plethora of interesting points, this one does not stand out as requiring special attention, until something like a graphical analysis accentuates the crucial role it plays. As in the case of the direct effect through other channels M → Y it is easy for the reader to be impressed by how much has been done, but if you plod through asking when the case was made for crossing out a fatal arrow, a critical deficiency remains .

There is one final claim that may seem to deal with all these objections, and most others that might be made. It involves a set of 'overidentification' tests presented in last half of the 'Robustness' section. This is advertised in the introduction with the acknowledgement, 'Naturally, it is impossible to control for all possible variables that might be correlated with settler mortality and economic outcomes...' (It is also neither necessary or desirable to do so – see Figure 7(b) – but they meant was 'all variables that could block back-door paths'.) Then they acknowledge they might be measuring 'the effect of settler mortality on economic performance, but working through other channels'. Thus the next sentence is unmistakably a promise to address both the fatal arrows I still cannot cross out:

> 'We deal with these problems by using a simple overidentification test using measures of European migration to the colonies and early institutions as additional instruments. We then use overidentification tests to detect whether settler mortality has a direct effect on current performance' (p. 1372).

Perhaps it is too much to make of a small thing, but in the spirit of McCloskey's call for close reading we must wonder about that odd repetition of 'overidentification test', which seems to say the say same test will be done twice in succession. Of course, no one will take it as meaning such a thing. But for the reader who breezes through, the repetition is a powerful devise for emphasis (a diacope). For most readers, what is emphasised is that the authors will make use of a very advanced technique vaguely recalled from an econometrics class but never fully grasped (in part because the class muddled causal and statistical language, in part because the topic was treated late and hastily, perhaps as optional.) And the advanced technique is going to be powerful: it will perform a task that better understood techniques leave impossible; it will allow us to infer a causal pathway from statistical tests.

Whether or not such an impression was made, the subsequent discussion of the tests appears to be directed at readers too much in awe to pay attention. What is actually done is that C and S (see Figure 5) are used as additional instruments. The regression is then said to be 'overidentified' since there is only one endogenous variable for which an instrument is needed (R) and there are multiple instruments. The resulting estimate is the average of the simple covariance ratios.[13] What is tested is whether the estimated impact of R differs between the overidentified and just-identified regressions.

The idea is that if (say) S is independent of M, and S is a valid instrument, then adding M as a second instrument will change the estimate if and only if M is not a valid instrument. Thus, they test the null hypothesis that adding M does not change the estimate in order to test the validity of M. As they note, however, this 'may not lead to a rejection if all instruments

[12] It is noted in Footnote 12 that malaria immunity is highly local – 'a person may have immunity to the local version of malaria, but be highly vulnerable to malaria a short distance away'. This acts as a labour-force immobiliser, another possible back-door channel.

[13] The algorithm is not, even in the one-instrument case, to calculate covariances and divide. It is the 'Two stage least squares' procedure: first R is regressed on all instruments and controls, then Y is regressed on the predicted value of R and the controls.

are invalid, but still highly correlated with each other. Therefore, the results have to be interpreted with caution' (p. 1393).[14]

But look again at Figure 5, which represents not my creative effort but AJR's Equations (1)-(4). The quote above makes it sound as though correlation among the candidate instruments would be some sort of unfortunate coincidence. But it is clear from the graph that, until now, links correlating M, S and C have been an indispensable justification for the whole strategy. Arguments for the presence of those links occupied most of their Section I. Their Table 1, which includes several C and S measures by quartile of M, leaves no doubt that they are indeed highly correlated. Thus to interpret "with caution" a test that assumes low correlation is close to dismissing it entirely.

Instead, the paper proclaims that 'subject to the usual problems of power associated with overidentification tests, we can rule out' not only the fatal arrows $C \rightarrow Y, S \rightarrow Y$ and $M \rightarrow Y$ but also heterogeneity of the treatment effect (i.e., of the impact of R on Y). This takes us beyond causal imprecision into plain abuse of statistical language. The failure to reject a null hypothesis does not 'rule out' anything, and when the test is low-power and the sample size in the 60s such a failure really says nothing at all. Despite the low power, in some specifications the tests rejected their overidentifying restrictions at 10% significance. This point is relegated to a footnote, with the comment that there in fact 'good reasons' to believe the restrictions are false – i.e., that C and S in fact impact Y. No explanation is offered for how these observations can be consistent with the sentence footnoted: 'The data support the overidentifying restrictions implied by our approach' (p. 1393).

Thus, these tests do nothing to help us cross out the fatal arrows. To recap, the tests are invalid unless the alternative instruments are implausibly uncorrected with settler mortality; were they valid they could only yield evidence in favour of those arrows, not against them; and so interpreted, they do yield such evidence, although the evidence is weak.

Concluding Remarks

I noted in the introduction that AJR was singled out for praise as the macro/development example of Angrist and Pischke's 'credibility revolution'. That revolution has by now produced something of a backlash. In September 2015, for example, the economics blogosphere buzzed briefly over a post on 'Kids Prefer Cheese' entitled 'Friends don't let friends do IV' (Instrumental Variables).[15] The post warned young economists against treating IV methods as a sort of magic, forgetting how narrow are the limits that this theory places on their interpretation. It ended,

> 'I pretty much refuse to let my grad students go on the market with an IV in the job market paper. No way, no how. Even the 80-year-old deadwoods in the back of the seminar room at your job talk know how to argue about the validity of your instruments. It's one of the easiest ways to lose control of your seminar.

[14] This note of caution was not sounded in the working paper version (Acemoglu et al., 2000). It seems likely that it was added at the insistence of a referee, and that the authors themselves were unaware of how thoroughly this caveat undermines their claims.

[15] I believe the author is Kevin Grier, using the screen name Angus, see http://mungowitzend.blogspot.mx/2015/09/friends-dont-let-friends-do-iv.html (accessed September 2015).

> We've had really good luck placing students who used Diff in diff (in diff), propensity score matching, synthetic control, and even regression discontinuity...'

This was surely to be read with something of a knowing wink, but even so it referred to a real phenomenon, and one contrary to a healthy scientific culture. Many know how to see the limits in our most venerable tool of (non-experimental) causal inference; and our response, in part, is to urge our students to newer stuff that fewer people understand. What I have advocated here is a rather healthier response. We must recognise that many of the causal claims made by economists can only be understood as conditional on the assumption that some potential confounding factors are trivial. (Rarely can we believe they are absent altogether.) Given this, surely we want these assumptions stated as simply as possible. It ought to be considered unambiguous success if it is easy to specify the reasons to doubt a causal claim.

This discussion points again to disfunction in our empirical tradition sufficient to warrant exploration of new tools. I have sought to demonstrate that graphical analysis can be a powerful tool for the close reading of econometric papers, in the hope that this will help persuade economists to learn these methods, as well as to aid in the digestion of such papers. In the process, I have concluded that an extremely influential paper is deeply misleading. Students of comparative development should take heed.

I want to make it clear that I am not accusing the authors of any intent to deceive. McCloskey called on economists to 'become more self-conscious about their rhetoric', not because many of us are liars, but because we absorb these methods in the same way a child learns syntax – through exposure, imitation and response to the rewarding sensation of success in persuading our mentors and audiences. By the time one is professional, successful rhetorical methods are often persuading the writer himself, and the audience feedback reinforces that persuasion immensely. I am sure that is the case here.

Of course, had I found nothing to quibble with this would have been a poor example to advertise the procedure. But the extent to which this paper misleads is not unusual. Less prestigious examples abound, but their dissection is intrinsically less interesting and does not carry the clear indictment of econometric tradition I believe merited. The point is this: the conventional language in which we discuss causal identification does not protect us from raising highly dubious claims to the status of received wisdom.

Our tradition is not without merit. The habit of thinking about common ancestors and alternative causal channels as a single threat to an exclusion restriction may in some cases be efficient. All of the issues I raised can be discussed in words, as evidenced by how many were, in some way. But prose is forgiving in ways that graphs are not. It is too easy to address quite a few potential problems, leaving the reader with a vague sense of thoroughness, without in fact covering all the questions that need coverage. The reader who has drawn a fatal graph for each claim will not be misled.

In this case, I found two paths, each sufficient to be fatal for causal interpretation, that seemed unlikely to be blocked by the observed variables. That observation destroys causal interpretation, but it does not destroy the paper. It is quite interesting that democracy, prosperity, property security, etc. are all strongly predicted by this centuries-ago settler mortality. It is, perhaps, a stylised fact of history as awesome as the Kuznets curve or the stability of labour's share in US income. The story proposed to explain it (to which I do not claim to have done justice here) is interesting and backed by equally interesting snips of narrative. And with all the talk from theorists and historians about how important institutions are, someone had to go and try the pseudo-experiment of creating an instrument. This one was really clever. That is reason enough for it to be in a top economics journal.

Had the paper been received entirely in that spirit it could have done little harm. But judging from subsequent discussion, it was not. Dani Rodrik's remarks quoted above (Section 4) come very close to saying that, thanks to this instrument, we know most macroeconomic policy choices don't much matter for poor countries.[16] Rodrik has very clearly distanced himself from this view in subsequent writings, but if this was indeed the orthodoxy of 2006, an empirical approach that should have been seen as just interesting became a primary driver of the advice our profession offered on questions that matter to mass poverty.

It would be wrong to leave the impression that economists have swallowed this instrument whole. The paper has attracted the criticism its prestige warrants. The crucial points I have raised have probably all been discussed somewhere, some prominently (Glaeser et al., 2004; Albouy, 2012). But this wide-ranging conversation is no substitute for the rigorous procedure of close reading that I am advocating here. The fatal graph approach gives the reading a transparent structure, and it narrowed my focus to the right set of questions. For example, I did not give other variables the scrutiny I gave to Percent of European in 1975, because none of the others seemed to hold the same potential to block a set of fatal paths. The extensive discussion of those paths in Glaeser et al. (2004) does not even mention this variable. In the traditional process, the questions of whether there are other paths and what variables might block them get scattered over multiple papers, and often over discussions within a single paper. It is easy to get lost.

As a systematic framework for addressing such questions there is no substitute for graphs.

Acknowledgements

I thank Sunil Mitra Kumar and Dimitris Sotiropoulos for their comments posted on the *Economic Thought* Open Peer Discussion forum, as well as Judea Pearl and two anonymous referees from the *Journal of Causal Inference* for helpful comments.

References

Abadie, A., J. Angrist, and G. Imbens (2002) 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings,' *Econometrica*, 70(1), pp. 91–117.

Acemoglu, D., S. Johnson, and J.A. Robinson (2000) 'The colonial origins of comparative development: an empirical investigation,' Technical Report, National Bureau of Economic Research.

Acemoglu, D., S. Johnson, and J.A. Robinson (2001) 'The colonial origins of comparative development: an empirical investigation,' *American Economic Review,* 91(5), pp. 1369-1401.

Albouy, D.Y. (2012) 'The colonial origins of comparative development: an empirical investigation: comment,' *American Economic Review*, 102(6), pp. 3059-3076.

[16] The claim is not quite that blunt, and its implications can be criticised on grounds other than instrument validity. Easterly and Levine (2003), he said, showed policies 'do not exert any independent effect on long-term economic performance once the quality of domestic institutions is included in the regression'. That is a poor characterisation of statistical insignificance. The words were literally true, of course; but since no one really cares what happens in the regression they will have been read by many as indicating policy ineffectiveness in the world, which would be invalid even if based on unassailable causal assumptions.

Angrist, J. D. and J.-S. Pischke (2010) 'The credibility revolution in empirical economics: how better research design is taking the con out of econometrics,' *Journal of Economic Perspectives*, 24(2), pp. 3-30.

Angrist, J. D., G.W. Imbens, and D.B. Rubin (1996) 'Identification of causal effects using instrumental variables,' *Journal of the American Statistical Association*, 91(434), pp. 444-455.

Babula, R., D. Bessler, J. Reeder, A. Somwaru et al. (2004) 'Modeling US soy-based markets with directed acyclic graphs and Bernanke structural VAR methods: The impacts of high soy meal and soybean prices,' *Journal of Food Distribution Research*, 35(1), pp. 29-52.

Bryant, H.L., D.A. Bessler and M.S. Haigh (2009) 'Disproving Causal Relationships Using Observational Data,' *Oxford Bulletin of Economics and Statistics*, 71(3), pp. 357-374.

Cartwright, N. (2007) *Hunting Causes and Using Them: Approaches in Philosophy and Economics*, Cambridge: Cambridge University Press.

Chen, B. and J. Pearl (2012) 'Regression and causation: A critical examination of econometrics textbooks,' *Mimeo., UCLA Cognitive Systems Laboratory*.

Cox, D.R. (1958) *The Planning of Experiments*, John Wiley & Sons.

Dawid, A.P. (2000) 'Causal inference without counterfactuals,' *Journal of the American Statistical Association*, 95(450), pp. 407-424.

Demiralp, S. and K.D. Hoover (2003) 'Searching for the causal structure of a vector autoregression,' *Oxford Bulletin of Economics and statistics*, 65(s1), pp. 745-767.

Dowrick, S. and M. Rogers (2002) 'Classical and technological convergence: beyond the Solow-Swan growth model,' *Oxford Economic Papers*, 54(3), pp. 369-385.

Durlauf, S.N., P.A. Johnson and J.R.W. Temple (2005) 'Growth econometrics,' *Handbook of Economic Growth*, *1*, pp. 555-677.

Easterly, W. and R. Levine (2003) 'Tropics, germs, and crops: the role of endowments in economic development,' *Journal of Monetary Economics*, January, 50(1).

Eichler, M. (2006) 'Graphical modeling of dynamic relationships in multivariate time series,' in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, p. 335 Björn Schelter, Matthias Winterhalder, Jens Timmer (eds.) John Wiley and Sons **(**ISBN 3-527-40623-9).

Glaeser, E.L., R. La Porta, F. Lopez de Silanes, and A. Shleifer (2004) 'Do institutions cause growth?,' *Journal of Economic Growth*, 9(3), pp. 271-303.

Glenn, E.N. (2009) *Unequal Freedom: How Race and Gender Shaped American Citizenship and Labor*, Cambridge MA: Harvard University Press.

Glymour, C., R. Scheines, P. Spirtes, and K. Kelly (1987) *Discovering Causal Structure,* San Diego, CA: Academic Press Inc.

Granger, C.W.J. (1969) 'Investigating causal relations by econometric models and cross-spectral methods,' *Econometrica*, pp. 424-438.

Greene, W.H. (2003) *Econometric analysis*, Fifth edition, Pearson (ISBN-10: 0130661899).

Haavelmo, T. (1943) 'The statistical implications of a system of simultaneous equations,' *Econometrica*, pp. 1-12.

Haavelmo, T. (1944) 'The probability approach in econometrics,' *Econometrica* Vol. 12 (supplement) pp. 1-118.

Haigh, M.S. and D.A. Bessler (2004) 'Causality and price discovery: An application of directed acyclic graphs', *The Journal of Business*, 77(4), pp. 1099-1121.

Hayek, Frederich A. von (1960) The Constitution of Liberty University of Chicago Press (reprinted 2011, ISBN 0226315398).

Hoover, K.D. (1991) 'The causal direction between money and prices: An alternative approach,' *Journal of Monetary Economics*, 27(3), pp. 381-423.

Hoover, K.D. (2005) 'Automatic inference of the contemporaneous causal order of a system of equations,' *Econometric Theory*, 21(01), pp. 69-77.

Huang, Y. and M. Valtorta (2006) 'Pearl's calculus of intervention is complete,' in 'Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence', AUAI Press, Arlington, Virginia, pp. 217–24.

Kima, J.W. and D.A. Besslerb (2007) 'The causal modelling on equity market innovations: fit or forecast?,' *Applied Financial Economics*, 17(8), pp. 635-646.

Koopmans, T.C., H. Rubin, and R.B. Leipnik (1950) 'Measuring the equation systems of dynamic economics,' *Statistical Inference in Dynamic Economic Models*, Cowles Commission for Research in Economics Monograph No. 10*,* John Wiley and Sons, New York*.*

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert W. Vishny (1999) "The Quality of Government," *Journal of Law, Economics and Organization*, 15, pp. 222-279.

Leamer, E.E. (2008) *Macroeconomic Patterns and Stories*, Springer Science & Business Media, Berlin (ISBN: 978-3-540-46388-7).

Marschak, J. (1950) 'Statistical inference in economics,' *Statistical Inference in Dynamic Economic Models*, Cowles Commission for Research in Economics Monograph No. 10*,* John Wiley and Sons, New York, pp. 1-50.

Matzkin, R.L. (2008) "Identification in nonparametric simultaneous equation models," *Econometrica*, 76(5), pp. 945-978.

McCloskey, D.N. (1998) *The Rhetoric of Economics*, Madison, Wisconsin: University of Wisconsin Press.

Neyman, Jerzy. (1923) [1990]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Statistical Science 5(4), pp. 465–472. Translated by Dorota M. Dabrowska and Terence P. Speed.

North, D.C. (1994) 'Economic Performance through Time,' Lecture to the Memory of Alfred Nobel December. *American Economic Review,* 84(3) 359-68.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Burlington, MA: Morgan Kaufmann.

Pearl, J. (1995) 'Causal Diagrams for Empirical Research,' *Biometrika*, 82(4), pp. 669-688.

Pearl, J. (2009) *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Pearl, J. and T. Verma (1991) *A Formal Theory of Inductive Causation*, University of California (Los Angeles). Computer Science Department.

Pearl, J. et al. (1991) *A Theory of Inferred Causation*, Burlington, MA: Morgan Kaufmann.

Queen, C.M. and C.J. Albers (2008) 'Intervention and causality in a dynamic Bayesian network,' *Open University Statistics Group Technical Report*, 8(01).

Rodrik, Dani "Goodbye Washington Consensus, Hello Washington Confusion? A Review of the World Bank's "Economic Growth in the 1990s: Learning from a Decade of Reform", *Journal of Economic Literature,* Vol. 44, pp. 973-987

Robins, J. (1986) 'A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect,' *Mathematical*

Robins, J.M. (2003) 'Semantics of causal DAG models and the identification of direct and indirect effects,' in Green, P., I-ljort, N., Richardson, S. (eds) *Highly Structured Stochastic Systems*, Oxford University Press, Oxford (ISBN 0-19-851055-1) pp. 70-8.

Rubin, D.B. (1990) 'Comment: Neyman (1923) and causal inference in experiments and observational studies,' *Statistical Science*, 5(4), pp. 472-480.

Rubin, D.B. (2005) 'Causal inference using potential outcomes,' *Journal of the American Statistical Association*, Vol. 100(469) pp. 322-331

Sala-i-Martin, G. Dopplehofer, and R.I. Miller (2004) 'Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach,' *American Economic Review*, 94(4), pp. 813-835.

Simon, H.A. (1977) *Models of Discovery*, Netherlands: Springer, pp. 53-80.

Solow, R.M. (1956) 'A contribution to the theory of economic growth,' *Quarterly Journal of Economics*, pp. 65-94.

Spirtes, P., C.N. Glymour and R. Scheines (2000) *Causation, Prediction, and Search*, Lecture Notes in Statistics (Book 81), Springer-Verlag, New York

Spohn, W. (2000) *Bayesian Nets are all there is to Causal Dependence*, in Maria Carla Galavotti, Patrick Suppes, and Domenico Costantini (eds.) *Stochastic Causality* CSLI Publications, Stanford, California, pp 157-172.

Stock, J.H. and F. Trebbi (2003) 'Retrospectives Who Invented Instrumental Variable Regression?,' *Journal of Economic Perspectives*, Vol. 17(3), pp. 177–194.

Swanson, N.R. and C.W.J. Granger (1997) 'Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions,' *Journal of the American Statistical Association*, 92(437), pp. 357-367.

Tan, Z. (2006) 'A distributional approach for causal inference using propensity scores,' *Journal of the American Statistical Association*, 101(476), pp. 1619-1637.

Tinbergen, J. (1940) 'Econometric business cycle research,' *Review of Economic Studies*, *7*(2), pp. 73-90.

White, H. (2006) 'Time-series estimation of the effects of natural experiments,' *Journal of Econometrics*, *135* (1-2), pp. 527-566.

Wilson, E. B. (1927) 'Probable inference, the law of succession, and statistical inference,' *Journal of the American Statistical Association*, 22, pp. 209-212.

Wilson, W. (1933) *Forced Labor in the United States*, International Publishers Co.

Wold, H. (1954) 'Causality and Econometrics,' *Econometrica*, 22(2), pp. 162-177.

Wright, S. (1921) 'Correlation and causation,' *Journal of Agricultural Research,* 20(7), pp. 557-585.

Wyatt, G.J. (2004) *Macroeconomic Models in a Causal Framework*, Edinburgh, U.K.: En Exempla Books (Harmony House).

Zhang, J., D.A. Bessler and D.J. Leatham (2006) 'Does consumer debt cause economic recession? Evidence using directed acyclic graphs,' *Applied Economics Letters*, 13(7), pp. 401-407.

SUGGESTED CITATION:

Margolis, M. (2017) 'Graphs as a Tool for the Close Reading of Econometrics (Settler Mortality is not a Valid Instrument for Institutions)' *Economic Thought*, 6(1), pp. 56-82.
http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-6-1-Margolis.pdf