Logic, Methodology and Philosophy of Science

Proceedings of the Thirteenth International Congress

edited by

Clark Glymour,

Wang Wei

and

Dag Westerståhl

Probability and Structure in Econometric Models[1]

Kevin D. Hoover

Duke University

kd.hoover@duke.edu

ABSTRACT. The difficulty of conducting relevant experiments has long been regarded as the central challenge to learning about the economy from data. The standard solution, going back to Haavelmo's famous "The Probability Approach in Econometrics", involved two elements: first, it placed substantial weight on *a priori* theory as a source of structural information, reducing econometric estimates to measurements of causally articulated systems; second, it emphasized the need for an appropriate statistical model of the data. These elements are usually seen as tightly linked. I argue that they are, to a large extent, separable. Careful attention to the rôle of an empirically justified statistical model in underwriting probability explains puzzles not only in economics, but more generally with respect to recent criticisms of Reichenbach's principle of the common cause, which lies behind graph-theoretic causal search algorithms. And it provides an antidote to the pessimistic understanding of the possibilities for passive observation of causal structure in econometrics and related areas of Nancy Cartwright and others.

1 Econometrics and the Problem of Passive Observation

For nearly two centuries – at least since Mill [23, p.327] – philosophers have observed, and economists have lamented, the barriers to turning economics into an experimental science. At one point, the lack of scope for controlled experiments was seen as a serious barrier to the application of modern, probability-based statistics to economics. The situation was saved – or, at least, economists were comforted – with the publication of Trgyve Haavelmo's " The Probability Approach in Econometrics" [7] and the subsequent development of the theory of econometric identification by the Cowles Commission ([11, 21]; see Bouman's [3] excellent history of these developments).

There were two critical ideas in the new approach. The first is that statistical controls, accounting for covariates, could take the place of experimental controls

[1]The author is with Departments of Economics and Philosophy, Duke University, Durham, North Carolina, U.S.A. This paper is an abridgment of Hoover (2007). An earlier draft was presented to the 13^{th} International Congress of Logic, Methodology, and the Philosophy of Science at Tsinghua University, Beijing, China 8-16 August 2007. I thank Julian Reiss for comments on that draft.

[24, chap.8, esp. pp.246–248]. Haavelmo proposed that an economic process could be partitioned into a deterministic and a random part. If the causal structure of the deterministic part were articulated fully and accurately enough, the random part would conform to the laws of probability.

The second critical idea was that the causal structure of the deterministic part of the economic process had to be articulated accurately. Haavelmo suggested that *a priori* economic theory could do the job. This second idea was the necessary prop of the first. The Cowles Commission, which took up Haavelmo's project, made the crucial discovery that causal structure is richer than, or (at least) distinct from, the probability structure [11, 21].

Take a very simple example. Suppose that A causes B, where A and B are two stochastic variables. Their relationship can be represented graphically as $A \rightarrow B$ and algebraically, with some additional structure, as:

$$A \Leftarrow \varepsilon_A \tag{1}$$

$$B \Leftarrow \alpha A + \varepsilon_B \tag{2}$$

where ε_A and ε_B are random error terms; for convenience we assume that each is distributed independent normal with mean zero and variances σ_A^2 and σ_B^2 (notated $\varepsilon_A \sim$ independent $N(0, \sigma_A^2)$ and $\varepsilon_B \sim$ independent $N(0, \sigma_B^2)$). Each is independent across successive draws and independent from each other (which implies that the covariance of ε_A and ε_B is zero $(cov(\varepsilon_A, \varepsilon_B) = 0)$. The coefficient α is a fixed parameter. And the arrowhead on the equal sign turns it into an assignment operator, a reminder that the model incorporates the asymmetry of causation.

The causal structure of Eq.1 and Eq.2 determines its probability structure. Substituting Eq.1 into Eq.2 yields what econometricians refer to as reduced forms, which completely characterize the probability structure of the variables:

$$A = \varepsilon_A = E_A \tag{3}$$

$$B = \alpha\varepsilon_A + \varepsilon_B = E_B \tag{4}$$

E_A and E_B are themselves distributed $E_A \sim$ independent $N(0, \sigma_A^2)$ and $E_B \sim$ independent $N(0, \alpha^2\sigma_A^2 + \sigma_B^2)$. But they are not independent of each other; in fact, $cov(E_A, E_B) = \alpha\sigma_A^2 = \Sigma \neq 0$.

The sense in which the causal structure is richer than the probabilistic structure is that the implication runs one-way: Eq.1 and Eq.2 imply Eq.3 and Eq.4, but not the other way round. In fact, if $B \rightarrow A$ with a causal structure analogous to Eq.1 and Eq.2, instead of $A \rightarrow B$, we can generate the reduced forms:

$$A = \varepsilon'_A + \beta\varepsilon'_B = E_A \tag{5}$$

$$B = \varepsilon'_B = E_B \tag{6}$$

The important thing is despite the difference in causal structures (reflected in the difference in the middle terms in the two sets of equations), both equations Eq.3 and Eq.4 and equations Eq.5 and Eq.6 define the same random terms, E_A and E_B. And these terms have the same interdependence in each case: $cov(E_A, E_B) = \beta\sigma_{B'}^2 = \Sigma$. They define the same probability distribution; they

are observationally equivalent; or, in the argot of econometrics, they are not identified.

The observational equivalence of the two sets of equations means that we can work backwards to form estimates of the parameters only if we are willing to commit to a particular causal structure. If we believe that theory (or some other extra-statistical source) tells us that Eq.1 and Eq.2 constitute the correct causal structure, then we can use observations on A and B to estimate the parameter α. But what if, as economists often believe, causation is mutual $(A \leftrightarrow B)$? For example, suppose that the causal structure is

$$A \Leftarrow \beta' B + \varepsilon''_A \tag{7}$$

$$B \Leftarrow \alpha' A + \varepsilon''_B \tag{8}$$

Then there are infinite combinations of causal strengths connecting them, so that the equivalence class is itself infinite.[2] There is no way to recover estimates of α' or β' without further non-empirical assumptions (e.g., about their relative strengths). This is the classic *identification problem in econometrics.*

The classic solution is to imagine that A and B are subject to independent experimental control. Suppose

$$X \rightarrow A \leftrightarrow B \leftarrow Y \tag{9}$$

where X is a means of intervening on A independent of B or Y, and Y is a means of intervening on B independent of A or X. The Cowles Commission showed that, in such circumstances, unique estimates of α' or β' could be recovered. If we have independent reasons for thinking that the world is structured like Eq.9 and that X and Y can somehow be observed, then passive observations could replace controlled experiments.[3] In macroeconomics, the analogy with controlled experiments is the basis for techniques of causal inference based on patterns of invariance and noninvariance [12, chap.8–10]; while in microeconomics, it motivates the search for "natural experiments" [1]. The Cowles Commission itself and econometric orthodoxy in the second half of the 20^{th} century downplayed experimental analogues, emphasizing instead the role of *a priori* economic theory in selecting the warranted causal structure such as Eq.9, in which X and Y are just additional observed variables, now christened *instrumental variables* (or just *instruments*).

The Cowles Commission's strategy opened an era of optimism about the possibilities for passive observation and the articulation of causal structure. Soon, however, pessimism set in: where does our confidence in the causal structure come from? Sims [31] famously stigmatized the assumed causal order as relying on "incredible" identifying restrictions (i.e., assumptions about which variables are *not* causally connected) – cf.[22]. For a while, some economists were resigned to using reduced forms only, but one can say so little about policy problems without a causal understanding that a whiff of *a priori* theory was soon reintroduced; and,

[2] The reduced forms are $A = \left(\frac{1}{1-\alpha'\beta'}\right)(\varepsilon''_A + \beta'\varepsilon''_B) = E_A$ and $B = \left(\frac{1}{1-\alpha'\beta'}\right)(\alpha'\varepsilon''_A + \varepsilon''_B) = E_B$. Once again, as indicated by the right-hand terms, they define the same probability distribution as Eq.3 and Eq.4 or Eq.7 and Eq.8.

[3] Scheines [29] provides a careful exposition not only of the logic of such inference from natural experiments, but also of the close analogy with the logic of controlled experiments.

for those who were still skeptical of *a priori* theory, natural experiments became the lodestone.[4]

Behind the alternating optimism and pessimism lies perhaps the biggest questions in empirical economics: how, and exactly what, can economists learn from passively data?

2 Probability Models: Function and Inference

In clarifying the identification problem, however, the Cowles Commission raises issues that go beyond economics. Cartwright [4] maintains that econometrics provides the clearest example of how probability should function in physical, as well as social, sciences. What impresses Cartwright are the detailed theoretical assumptions that inform theoretically identified econometric models. These correspond, in experimental contexts to experimental controls and shielding, which she argues allow "nature's capacities" to display themselves without conflating interference, just as the set of instrumental variables allow the strengths of causal connections to be measured in econometric models. Cartwright [5, p.173] emphasizes the stringency of the conditions needed to throw capacities into clear relief:

> My claim is that it takes hyperfine-tuning ... to get a probability. Once we review how probabilities are associated with very special kinds of models before they are linked to the world, both in probability theory itself and in empirical theories like physics and economics, we will no longer be tempted to suppose that just any situation can be described by some probability distribution or other. It takes a very special kind of situation with arrangements set just right – and not interfered with – before a probabilistic law can arise.

And while she is grateful to econometricians for clarifying the logic of the problem, she is pessimistic with respect to the project of applied econometrics, seeing econometricians as having themselves laid the groundwork for showing it to be a quite hopeless undertaking [5, chap.6–7].[5]

Cartwright's pessimism about econometrics is based in what it teaches her about the application of probability. First, probabilities are not simply there in the world to be invoked whenever it suits our inferential purposes. Rather, following Hacking [8], she argues that probabilities are ways of codifying the propensities of physical (and perhaps social) machinery to display behavior with frequencies that follow certain patterns. The propensity of a coin to display heads half the time and to provide no evidence of dependence between successive tosses (in short, to follow a binomial probability model with the key parameter set to $\frac{1}{2}$), requires that the coin, the flipping device, and the environment to be constituted in highly particular ways, and only then can we expect the probability model to apply. Physical experiments are examples of such *nomological machines*; while what an applied economic theory describes is a *socioeconomic machine*.

Second, Cartwright asks, how do probabilities attach to the world? Her answer is that probabilities attach to the world through models. It is the actual

[4]See [16, 18] for fuller accounts.

[5]Hoover [13, 14] shows that Cartwright's position is genuinely and unnecessarily, if only *implicitly*, pessimistic.

success of the binomial probability model in describing the frequencies generated in the coin-flipping nomological machine that warrants claims about probabilities – for example, claims about how often five heads in a row ought to be expected.

Two of the theses that I shall elaborate and defend are reactions or qualifications of the lessons that Cartwright draws from econometrics. First, her position that probabilities arise only in chance setups is correct up to a point. The generation of stable frequencies is a property of real-world (not just physical) systems appropriately configured. But consistent with Hacking [8, pp.24–25], it is not the frequencies directly but the frequencies on particular kinds of trials that exhibit chance behavior. The same data may be viewed on different kinds of trials: "there is nothing unusual about regarding one event under several aspects" [8, p.25]. Different probability models may be applied to the same data without conflict or contradiction.

And this connects to my second thesis: Cartwright is correct to stress the role of models, but their role is not merely to attach probabilities to the world, but to create probabilities. Without the models, there are no probabilities to discuss. This is not an anti-realist thesis. For I suppose that that some models are better than others when judged in relation to actual frequencies from a particular aspect or point of view and that different points of view may lead to different, but not contradictory, probability models. The upshot of my theses is that a (perhaps, *the*) central problem of econometrics is to establish appropriate probability models. While there are plenty of statistical tools devoted to specification testing and specification search, the logical role of probability models in econometric inference is a relatively neglected topic.[6]

I also want to defend a third thesis that is well illustrated by the equivalent probabilistic implications of the causal structures $A \rightarrow B$, $A \leftarrow B$ and $A \leftrightarrow B$ (see equations Eq.1-Eq.8 above): namely, probability models do not in general require causal presuppositions. In saying this, I do not wish to contradict another of Cartwright's [4] well-known principles that an output of causal knowledge is delivered only by inputs of causal knowledge. Rather I want to defend the weaker claim that, while prior causal knowledge may be useful and, sometimes at least, essential, some causal claims may be supported by facts about probability models that do not depend on assumptions about the truth of these very same causal claims.

To illustrate, consider a simple causal structure: $A \rightarrow C \leftarrow B$, where

$$A \Leftarrow \varepsilon_A \tag{10}$$
$$B \Leftarrow \varepsilon_B \tag{11}$$
$$C \Leftarrow \alpha A + \beta B + \varepsilon_C \tag{12}$$

and $\varepsilon_i \sim$ independent $N(0, \sigma_i^2), i = A, B, C$ and $cov(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$. The arrangement is one with a *common effect* of two independent causes, sometimes known as an *unshielded collider* [36, p.10]. The causal structure itself cannot be directly observed, but we can observe realizations of the variables.

Now suppose that we want to infer the causal structure from the data. We begin by trying to establish a probability model of the data. A normal distribution is often a good place to start. There are standard statistical tests for normality. The joint normal model of three variables can be described by nine

[6]However see [10, esp. chs.1, 15], [19, 20, 34, 35].

parameters: three means (call them $\bar{A}, \bar{B}, \bar{C}$), three variances ($\sigma_A^2, \sigma_B^2, \sigma_C^2$), and three covariances or, equivalently, population correlations ($\rho_{AB}, \rho_{AC}, \rho_{BC}$).[7] A particularly simple model of the data assumes that the three variables are distributed independent normal with constant means and variances:

Model 1 $(A, B, C) \sim N(\bar{A}, \bar{B}, \bar{C}; \sigma_A^2, \sigma_B^2, \sigma_C^2; 0, 0, 0)$

where the last three arguments indicate that each of the population correlations is zero. Probabilistic independence can be defined formally as occurring when $P(X,Y) = P(X)P(Y)$. Informally, it means that the probability distribution of a variable is the same whatever realization is taken by another variable. Probabilistic independence implies that the corresponding population correlation is zero.

Each probability model sees the data from a point of view. Is Model 1 a good model? The answer, of course, depends in part on our purposes. If we, subscribe to an inferential scheme that requires judgments about probabilistic dependence (for example, the various causal-search algorithms in Spirtes and Pearl [26, 36]), then it is not a good model, since it *assumes* that there is no probabilistic dependence. A better model would be

Model 2 $(A, B, C) \sim N(\bar{A}, \bar{B}, \bar{C}; \sigma_A^2, \sigma_B^2, \sigma_C^2; \rho_{AB}, \rho_{AC}, \rho_{BC})$

There is no loss from taking this point of view, since Model 2 nests Model 1: if we decide that on our best estimates $\rho_{AB} = \rho_{AC} = \rho_{BC} = 0$, then Model 2 collapses to Model 1.

Whether a model is good also depends on its relationship to the data. Various statistical techniques allow us to estimate the parameters of Models 1 and 2 and whether one encompasses or nests the other, as well as to test their errors for normality, randomness and probabilistic independence against various alternatives. It is not to our purpose to discuss them in any detail.

The assertion that a probability model truly describes some portion of the world is a conjecture from which we can deduce that the model accounts for the co-occurrences of the data (both observed and yet-to-be-observed) except for some random residual. Conjectures about probability models, like all scientific conjectures, are accepted because they are supported by the right kind of data. They are never deductively certain, and they always remain open to doubt and criticism. Serious criticisms must be adjudicated in the light of the data and may lead to a reassessment of the appropriateness of a probability model. The crucial point is that the probability model is not a directly observable fact about the frequencies displayed by the data; rather it is a conjecture, the support for which depends on a complex of statistical inferences.

Returning to our illustration, suppose that we have obtained estimates for the parameters of Model 2, we have tested it for normality, and we have tested and rejected Model 1 as a special case. (These by no means exhausts all that the statistician might do to convince himself that Model 2 is a good model.) Now, here is a principle invoked in many causal search algorithms:

[7] The relationship of the population correlation to the covariance is, for example, $\rho_{AB} = \dfrac{cov(A,B)}{\sigma_A \sigma_B}$.

Principle of the Common Effect if X and Y are probabilistically independent conditional on some set of variables (possibly the null set) excluding Z, but are probabilistically dependent conditional on Z, then Z is the common effect of X and Y (or Z forms an unshielded collider on the path XZY).

I do not wish to defend this principle here, but instead consider the logic of its application.

We start with an estimate of Model 2. Suppose that a statistical test tells us that we cannot reject $\rho_{AB} = 0$ and that the correlation of A and B conditional on C does not equal zero ($\rho_{AB|C} \neq 0$). Then, we are working with a particularization of Model 2, call it

Model 2' $(A, B, C) \sim N(\hat{\bar{A}}_A^2, \hat{\bar{B}}_B^2, \hat{\bar{C}}_C^2; \hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\sigma}_C^2; 0, \hat{\rho}_{AC}, \hat{\rho}_{BC})$

where, in the custom of econometricians, the "hats" indicate estimated values and the estimates $\hat{\rho}_{AC}$ and $\hat{\rho}_{BC}$ are constrained to fulfill the condition $\hat{\rho}_{AB|C} = 0$.[8]

Probabilistic dependence is a property of probability distributions and not of realized data. The important judgments here are $P(AB) = P(A)P(B)$ and $P(AB|C) \neq P(A|C)P(B|C)$. The crucial point is that these are deductive consequences of Model 2' and are not unmediated consequences of observed data. This is easily seen by noting that Model 1 is, for other purposes and from other points of view, a perfectly acceptable model of the data; and Model 1 does not imply $P(AB|C) \neq P(A|C)P(B|C)$. Given Model 2', we can *deduce* that the antecedents of the Principle of the Common Effect are fulfilled and, therefore, conclude that the data, through the mediation of Model 2', imply $A \rightarrow B \leftarrow C$, which we know by assumption is the causal structure that generated the data.

There are two points to take away from this illustration. The first is that an inference such as the one from the Principle of the Common Effect is a two-step process. Step 1 establishes the probability model through statistical inferences; step 2 deduces the probabilistic (in this case, causal) consequence from the inferential principle applied to the probability model. Commentators on causality frequently seem confused on the two-step nature of the inference because the parameters of common probability models frequently have easily calculated analogues among descriptive sample statistics. For example, Pearson's sample correlation coefficient r_{XY} is analogue to ρ_{XY}.[9] One cannot, however, work directly with the sample correlation coefficient or other descriptive statistics without reference to the probability model:

(a) parameterization is distribution-specific; some distributions may have no parameter closely related to ρ_{XY} in the normal distribution and, so, nothing to which r_{XY} can serve as an analogy;

[8] Which requires in particular that $(\hat{\rho}_{AC}^2 - \hat{\rho}_{AC}\hat{\rho}_{BC})\left(\sqrt{1-\hat{\rho}_{AC}^2}\sqrt{1-\hat{\rho}_{BC}^2}\right)^{-1} = 0$, so that, given $\hat{\rho}_{AB} \neq 0$, any nonzero values for both $\hat{\rho}_{AC}$ and $\hat{\rho}_{BC}$ are sufficient.

[9]
$$r_{XY} = \sum_{j=1}^{N}(X_j - \bar{X})(Y_j - \bar{Y}) / (\sum_{j=1}^{N}(X_j - \bar{X})^2 \sum_{j=1}^{N}(Y_j - \bar{Y})^2)^{\frac{1}{2}}$$

and

$$\rho_{XY} = E((X - E(X))(Y - E(Y))) / (E(X - E(X))^2 E(Y - E(Y))^2)^{\frac{1}{2}}$$

(b) even when the analogy holds, r_{XY} may not coincide with the best estimate of ρ_{XY}, since the parameters of a probability distribution are typically estimated jointly (e.g., by maximum likelihood methods) rather than individually;
(c) sample descriptive statistics are calculated without the aid of a probability distribution, and it is only through one or other distribution that they can have any bearing at all on probability – to act otherwise is to commit a category mistake.

The principal interest of most researchers with respect to causal search is in the second step of inferring causal structure from patterns of probabilistic dependence. They frequently take for granted that the problem of justifying a particular probability model from the data has been (or can easily be) solved.

The second point to take away from the illustration is that, in inferring the pattern of causal dependence from which causal order is itself inferred, we nowhere refer to the facts about causal structure that form the endpoint of our inferential chain (namely, the connection of A and B to their common effect C). That is not to say that we do not use causal knowledge at all. In restricting our model to three variables, we have implicitly judged that none of the other causal connections that our three variables has is structured in such a way as to interfere with the appropriateness of Models 2 or Model 2'. Such a judgment may, of course, be challenged, suggesting further investigation. That we cannot step out of a causal context notwithstanding, the key point is that we have begged no question.

People often intuitively think of probabilistic dependence as a causal notion. Hacking [8, p.20] provides one formulation:

> Two events are commonly said to be independent of each other if neither causes the other, and if knowledge that one occurred is no aid in discovering if the other occurred.

But Hacking also agrees to a second formulation: X and Y are independent if $P(XY) = P(X)P(Y)$. Not only is no assumption about causation cited in this second, *standard* formulation, the statistical tests of independence are based on patterns of co-occurrence without causal reference. Whether knowledge of one aids in discovering whether the other occurred depends importantly on what knowledge we have in mind and what we mean by "aids".

Generally, two watches give knowledge of each other: if I know that my watch says 2:39 PM, then it is a fair bet that my neighbor's watch is pretty close. My watch even gives me knowledge of what a watch in Australia is likely to read, knowing the difference in time zones between the east coast of the United States and, say, the west coast of Australia. And although we do not have a common reference point, I suspect that, if there were Martians and Martians had watches, then an hour passed on my watch would be an hour passed on a Martian's watch (suitably adjusting Martian units to our own). But generally, it is a well-supported conjecture that my watch is probabilistically independent of my neighbor's, the Australian's, and the Martian's watches;(cf.[37, section 2], [28, p.181]).

For example, take the standard time signal provided by the U.S. National Institute of Standards and Technology and the U.S. Naval Observatory as a reference time. Define the random variable A = the difference between the time on my watch and the reference time and B = the difference between the time on my neighbor's watch and the reference time. I maintain that typically we will find a well-supported probability model in which $P(AB) = P(A)P(B)$. More directly, if b is a particular realization of B, then we will find $P(A|B = b) = P(A)$. That watches convey knowledge about the likely behavior of other watches and clocks explains their widespread use. That watches are typically probabilistically independent of each other explains why, when the power has been cutoff, we can usefully look to our wristwatches to reset the clock on the microwave.[10]

Contrary to Hacking, it would be more accurate to say that

> *two random variables are independent of each other if the realization of one conveys no information about the distribution of the other.*[11]

Formulated this way, probabilistic independence does not invoke causal order conceptually, nor do statistical tests of independence presuppose causal order.

3 The Principle of the Common Cause

The importance of clarity with respect to the two-step inferential process – from data to probability model, from probability model to causal structure – is thrown into relief by recent discussions of the Principle of the Common Cause, a version of which lies at the heart of the graph-theoretic analysis of causal structure and related search algorithms. Hans Reichenbach [27, p.157] provides the original statement:

Principle of the Common Cause (Reichenbach) "If an improbable coincidence has occurred, there must exist a common cause."

Reise [28, p.184] gives a version, which he attributes to Hoover [15, p.548], that is clearer for the issues to hand:

Principle of the Common Cause (Hoover) "If variables X and Y are probabilistically dependent ..., then either X causes Y or Y causes X, or X and Y are the joint effects of a common cause."[12]

Reiss states the principle mainly to criticize it.

The background for Reiss's criticisms is Sober's [32, 33] putative counterexample.[13] In Sober's example, bread prices in England and sea levels in Venice are both rising and *ex hypothesi* not causally connected. In a sense that is less than

[10] I say "typically," because, for example, two old-fashioned electric clocks on the same circuit that use the cyclicality of the household electricity to control the speed of their motors may be probabilistically dependent after all.

[11] And two *events* are independent of each other if each is a realization of a mutually independent random variable.

[12] Reiss states he needs to modify my statement of the principle to make it consistent with his own paper. I agree that this formulation is better than my original formulation, which was specific to a particular context.

[13] Hoover [15] offers a detailed refutation of Sober's counterexample. While I remain convinced of its argument, at some key points the exposition apparently misled some readers about its essence. I hope to be clearer here.

perfectly clear, Sober maintains that bread prices and sea levels are correlated and, therefore, probabilistically dependent.

Reiss categorizes various reactions to Sober's counterexample as following one of two strategies: the first strategy argues that Sober's claim that there is a genuine probabilistic dependence between bread prices and sea levels is defective; the second proposes to defuse the counterexample by showing that it fails to apply to data when they are appropriately prepared.[14] Reiss treats these two strategies separately, but observes that they may be complementary. I would put the point more strongly: if "data preparation" (second strategy) is understood appropriately and if the first strategy is stated in its positive form (patterns of probabilistic dependence may support causal inference when genuine), then there is no legitimate way to separate the two strategies, for the second is part of establishing the *bona fides* of the probability model necessary for the first.

Although Sober does not offer any formal measures of correlation between bread prices and sea levels, he does provide some cooked data and notes that "higher than average sea levels tend to be associated with higher than average bread prices" [33, p.332,334]. Unlike Reiss, Sober [33, p.343] acknowledges the two-step inferential process involved in applying the Principle of the Common Cause; so, we are entitled to ask what the sample association of bread prices and sea levels says about probabilistic dependence. Note that on a common measure of sample association, Pearson's correlation coefficient (r), which was previously defined in footnote 8, bread prices (B) and sea levels (S) are highly correlated ($r_{BS} = 0.99$, where $-1 \leq r \leq 1$ and $|r| = 1$ indicates perfect correlation, whereas $r = 0$, indicates no correlation.). We cannot interpret this high correlation in terms of probability without a probability distribution. This is obvious, since it is exceedingly rare to find a correlation coefficient that is *exactly* zero; we must judge whether it is *effectively* zero or not relative to an assumed probability distribution.[15]

The stationary, multivariate normal distribution is the workhorse of statistics. It has the nice property that r is an analogue for its population parameter ρ, and that it can be shown that, as the sample size increases, the expected value of r converges to ρ. Roughly speaking, a distribution is stationary when its moments (mean, variance, and higher moments) are constant through time. But a stationary distribution is not a good model of Sober's data. A stationary distribution implies that a time-series will cross and re-cross its sample mean frequently. Sober's data cross their sample means only once. While this is a nice clue, there are also formal tests for non-stationarity. There are a number of non-stationary alternatives to the stationary, multivariate normal distribution – none of which display the correspondence between the sample correlation coefficient r and a fixed population parameter, like ρ.

One alternative is the random-walk in which the best expectation of the value of a time series at $t + 1$ is its value at time t. If the data were generated by two probabilistically independent random walks, then r would be a worse-than-useless measure of probabilistically dependence; for the expectation of r as the sample size grows converges not to a single value but to a uniform distribution over the interval -1 to 1 [10, p.128]. This means that when the world is populated

[14] Reiss offers [15] as an example of the first strategy and [2, 36, 37] as examples of the second.

[15] Notice that this is true whether we accept classical statistical testing or a decision-theoretic approach.

by random walks that it is easy (and meaningless) to find high levels of sample correlation among some of them.

The paradigm random walk can be expressed as:

$$X_{t+1} = X_t + \varepsilon_t \tag{13}$$

or equivalently as

$$\Delta X_{t+1} = \varepsilon_t \tag{13'}$$

where ε_t is a stationary random error term (e.g., normal). The form (13') suggest to some commentators (e.g,. [6, 9, 25]) a quick fix. If we difference the non-stationary time series X, it becomes stationary, and the correlation coefficient between two such differenced, non-stationary time series is an indicator of probabilistic dependence. The problem with this approach is that, while the differences of nonstationary variables may be probabilistically dependent, so may the levels (even when the differences are not), and differencing the data eliminates the information about this relationship between the levels. Nonstationary variables that display probabilistic dependence in levels are said to be *cointegrated.* If the nonstationary random walk is sometimes illustrated by the path of drunk stumbling aimlessly as he leaves the bar, then cointegration is the situation in which the drunk has a faithful friend who follows at a discreet distance to make sure that he comes to no harm.[16]

Sober's counterexample "works" only because he trades on our implicitly judging probabilistic dependence against a probability model in which casual measures of association have a natural interpretation. But even at a casual level, it is obvious that a stationary probability model is not a good characterization of the data. And in any non-stationary model, the sample association of bread prices and sea levels is both natural and not indicative of probabilistic dependence. The situation is exactly like the association between time kept on two watches: rising bread prices in England give some indication of sea levels in Venice, but the distribution of sea levels is the same whether the current realization of bread prices is a rise or a fall.

The point is not that any particular non-stationary model fits Sober's counterexample. Rather it is that we must establish the probability model before we can make any judgment of probabilistic dependence. Sober may acknowledge the two-step process, but he fails to do the work – either statistically by testing the data or hypothetically by establishing the true distribution in his thought experiment – necessary to move from the first to the second step. What Reiss thinks of as data preparation is integral to establishing the probability model from which probabilistic dependence is ascertained. On the one hand, the probability model must be appropriate to the data; on the other hand, the probability model helps to guide the meaningful preparation of the data. This is crucial work for statistics or econometrics, though it is typically neglected in discussions of the Principle of the Common Cause.

In contrast to my analysis of Sober's counterexample, which is not so much a defense of the Principle of the Common Cause as a demonstration that it does not fail for Sober's particular reasons, Reiss offers a defense of the principle – a defense which falls into the category of "destroying-the-village-in-order-to-save-it." Reiss agrees with Cartwright and Hacking that probabilities (i.e., frequencies

[16] See [15, section 4–5], for a more technical exposition.

that are correctly described within the canons of a axiomatization of the behavior of random variables) arise only in well-constructed *chance setups*. At the same time, he objects to the two-step inferential process: he claims that my strategy "deprives the principle of much of its inferential power and to some extent betrays the motivation behind it" [28, p.185]. Reiss defends the principle as a sometimes useful heuristic, providing what he refers to as an "epistemological reading" as opposed to the "metaphysical reading" implicit in the two-step inferential process.

The contrast between an epistemological and a metaphysical reading is spurious. The two-step process is about inference (that is, classically epistemological) and says nothing about what causation actually is.

Reiss's strategy is explicitly analogous to Patrick Suppes' [38] well-known probabilistic analysis of causation. Suppes begins by defining *prima facie* cause as the case in which $P(A|B) > P(A)$. He then tries to catalogue the cases in which *prima facie* causes fails to correspond to actual cause and to suggest appropriate corrections. In parallel, Reiss takes the Principle of the Common Cause as providing a rule for inferring *prima facie* cause, and then catalogues a (partial) list of exceptions to the rule. Reiss's heuristic rule is not stated as a relationship of causal structure to probabilistic dependence, but as a relationship of sample association (or frequency) to causal structure, thus short-circuiting the first step of the two-step inferential process:

Principle of the Common Cause (Reiss) "The proposition e = 'Random variables X and Y are (sample or empirically) correlated' is *prima facie* evidence for the hypothesis h = 'X and Y are casually connected.' If all alternative hypotheses h_i^a (e.g., 'the correlation is due to sampling error,' 'the correlation is due to the data-generating process for X and Y being non-stationary,' 'X and Y are logically, conceptually, or mathematically related') can be ruled out, the e is genuine evidence for h." [28, p.193]

It is instructive to see how the two-step inferential process handles Reiss's exceptions. Reiss considers seven specific exceptions (non-stationary nonsense correlations, colliders, mixing, stationary nonsense correlation, homoplasies, non-statistical nonsense correlations, and laws of coexistence) and suggests that the list is actually open-ended. All six are easily treated using the two-step inferential process (see [17] for details). We have already discussed the case of non-stationary nonsense correlations and, in the interest of space, we consider only one other case here, that of *colliders*:

Reiss [28, pp.187–188] takes the causal configuration $A \rightarrow C \leftarrow B$ as an exception to the Principle of the Common Cause, because the correlation of A and B conditional on C does not indicate their causal connection (see equations Eq.10-Eq.12 above for the structure of the probability model). At first this seems clearly wrong: the Principle of the Common Cause begins with the claim that A and B are *correlated* (whereas the Principle of the Common Effect begins with them *uncorrelated*); hence the antecedent of the Principle of the Common Cause is not fulfilled, so it fails to provide a counterexample. What Reiss has in mind, however, is that the data on A and B may be collected in such a way that, without knowing it, we observe them only conditional on C, so that they appear to be unconditionally correlated.

A real example illustrates Reiss's concern. [30] Data on child molestation and exposure to child pornography was collected from prisoners in jail for possession of child pornography. Eighty-five percent admitted to having molested children. Consider three binary $(0, 1)$ variables: $A = 1$ means viewed child pornography, $B = 1$ means molested children, and $C = 1$ means incarcerated for possession of child pornography. A concern immediately expressed by various critics of the study amounts to asserting the possibility that the variables form an unshielded collider and that A and B are correlated only because the mode of data collection implicitly conditions on C, the fear being that those who *both* view child pornography and molest children are more likely to be incarcerated for possession than those who merely view child pornography and that viewing it and child molestation may be unconditionally independent.

The intuition of the critics of the study can be interpreted with the two-step inferential process as the requirement that we get the right probability model, which means seriously entertaining the criticism and widening the scope of the data collection, so that the alternative hypothesis of an unshielded collider can be assessed. This strategy is suggested by a wider understanding of the world. But that is not an objection; there is nothing in the two-step process that suggests that a probability model is a black box for processing statistics without reference to their nature and provenance nor that only statistical criteria can be used to support a particular probability model. It is possible that sometimes we may make a mistake and do not notice accidental conditioning on an unobserved variable. Reiss persistently confuses the epistemic with the practical. It is nonsense to attack an inferential principle as metaphysical and not epistemic because we sometimes make errors in practice. The right response to accidental conditioning is to try to use all our knowledge to anticipate situations that give rise to such errors, to criticize research and to respond to criticism, and to test, test, test.

Reiss's remaining cases (mixing, stationary nonsense correlation, homoplasies, non-statistical nonsense correlations, laws of coexistence) are all answered in much the same way as his case of colliders (see [17]).

4 Cartwright's Pessimism Confounded

We have not offered a general defense of the Principle of the Common Cause. Instead, we have demonstrated that certain ways of attacking it illustrate the general proposition that we cannot neglect the need to provide a convincingly supported probability model if we wish to draw probabilistic conclusions from data. Reiss's "defense" of the Principle of the Common Cause goes wrong because he accepts simultaneously two premises: 1) Cartwright's view that only tightly controlled nomological (or socio-economic) machines generate frequencies that can be modeled probabilistically; and 2) that something like the Principle of the Common Cause is used – as a matter of fact – in actual research. Omitting the step of establishing a probability model, the conditions for which, Cartwright has argued, are too severe to be met in many practical contexts, is a fairly desperate move. Reiss's defense ends with the "irony" that, once his open-ended list of exceptions has been taken on board, we had better stick to experiments and eschew passive observation [28, p.194]. In the name of strengthening the practical applicability of the principle, Reiss kills it. His first premise is the mortal enemy of his second premise. In the end, he is left exactly where he started with Cartwright's view that the scope of probability models is very narrow and almost entirely restricted to experimental contexts.

Cartwright's position is deeply pessimistic with respect to economics and other sciences that rely on passive observation. Is it justified? Cartwright's argument points to a substantial disanalogy between experiments and passive observation. The disanalogy is real enough; it is the bane of empirical economics. But there is also a strong analogy between experiments and passive observation, which was central to Haavelmo's analysis sixty years ago.

Cartwright stresses that experiments generate probabilistically well behaved data (and, generally, display nature's capacities) only when they are arranged "just so". There is a danger of overstating the case. In a discussion of coin-flipping machines, Cartwright [5, p.166] says:

> Imagine ... that we flip the coin a number of times and record the outcomes, but that the situation of each flip is arbitrary. In this case we cannot expect any probability at all to emerge.

Approach this claim empirically: take any coin in general circulation and sit down anywhere and flip it how you will, recording heads and tails. Having done the experiment, I am sure that over the course of, say, 200 flips your data will conform – as judged by standard statistical tests – to a binomial probability model with a probability 1/2 for heads.

It matters that on a typical flip, the coin rotates at least once or twice. It matters that you do not wait to decide how to record the coin until you can see its resting position clearly. (For example, if a coin is leaning against the leg of table showing heads, and you decide after you see it that the rule for that particular flip will be to turn its visible side down before reading it, then the implicit preference for tails – if it persists in other such cases – may skew the results.) It does not matter whether you catch the coin in midair and turn it over on your wrist before reading it or let it fall to the floor or pick it out of the crack between the cushions on the sofa or fish it out from under the table. It does not matter whether the coin is new or worn or clean or dirty. The frequencies displayed by coins are very robust.

The point is not that every capacity nor every probability is similarly robust. It is quite difficult to construct a machine that will robustly deliver any probability other than zero for the frequency of a coin falling on its edge. Rather, the point is that it is a mistake to assert that a *very high* degree of control is an *a priori* requirement of frequencies conforming to stable probability models. Whether they do or do not is just something that we have to learn about the world in particular cases.

Even in controlled experiments, the range of factors that we attempt to control are frequently quite limited. Partly because we judge that certain factors are irrelevant and partly out of ignorance, many factors are left to nature's whims. And when we (or some other researcher) "replicates" our experiment, we cannot set every control in precisely the same way, and whimsical nature picks different values for factors that we have left uncontrolled. To paraphrase Heraclitus: you cannot perform the same experiment twice. At best each experiment is a model of its fellows. As with all models, we have to ask, is it a good or successful model? We learn from experience and from diagnostic tests whether we have successfully implemented controls of the right type and what we may neglect or ignore. Sometimes we find out later that we were wrong and that a neglected or

overlooked factor is essential to the result, so that our experiment needs to be reinterpreted, redone, or set aside as uninformative.

The situation is not different in kind from what we face in formulating probability models for passively observed data. We may construct a chance setup and discover that it is reliable. Equally, we may observe the world and discover that there is a way of modeling it that reliably acts like a fabricated chance setup would. Hacking [8, p.1] introduces the notion of a chance setup and promptly illustrates it with a passively observed example: "the frequency of traffic accidents on foggy nights in a great city is pretty constant." The controls (day or night, foggy or clear, in the city or not) are of the same nature as the ones that find their ways into econometric models. The controls may be represented in coarse or finely delineated categories, as may the category of traffic accidents itself. Whether we need finer controls or more controls or controls of different kind and what kind of probability model should tie them together is a matter – exactly as it is for physically controlled experiments – for experience and testing to reveal.

Passive observation is, in many respects, at a disadvantage in comparison to active experimentation. That fact poses serious challenges for empirical economics. Nevertheless, that the inferential logic of passive observation is not of a radically different kind and that statistics provides many useful tools that help us to specify and test appropriate probability models is grounds for optimism.

Bibliography

[1] Angrist, J., Krueger, A. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4):69–85, 1999.

[2] Arntzentius, F. Reichenbach's common cause principle. In Zalta, E. N. (Ed.), *Stanford Encyclopedia of Philosophy*. Spring, 2005 edn., 2005. URL `http://plato.stanford.edu/archives/spr2005/entries/physics-Rpcc/`.

[3] Boumans, M. *The Problem of Passive Observations*. University of Amsterdam, 2007, unpublished manuscript.

[4] Cartwright, N. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press, 1989.

[5] Cartwright, N. *The Dappled World*. Cambridge: Cambridge University Press, 1999.

[6] Forster, M. Sober's principle of the common cause and the problem of comparing incomplete hypotheses. *Philosophy of Science*, 55(4):538–559, 1988.

[7] Haavelmo, T. The probability approach in econometrics. *Econometrica*, 12 (supplement), 1944. July.

[8] Hacking, I. *The Logic of Statistical Inference*. Cambridge: Cambridge University Press, 1965.

[9] Hausman, D. M., Woodward, J. Independence, invariance, and the causal Markov condition. *British Journal for the Philosophy of Science*, 50(4):521–583, 1999.

[10] Hendry, D. F. *Dynamic Econometrics*. Oxford: Oxford University Press, 1995.

[11] Hood, W. C., Koopmans, T. C. (Eds.). *Studies in Econometric Method*, Cowles Commission Monograph 14. New York: Wiley, 1953.

[12] Hoover, K. D. *Causality in Macroeconomics.* Cambridge: Cambridge University Press, 2001.

[13] Hoover, K. D. *The Methodology of Empirical Macroeconomics.* Cambridge: Cambridge University Press, 2001.

[14] Hoover, K. D. Econometrics and reality. In Mäki, U. (Ed.), *Fact and Fiction in Economics: Models, Realism, and Social Construction.* Cambridge: Cambridge University Press, 2002.

[15] Hoover, K. D. Nonstationary time series, cointegration, and the principle of the common cause. *British Journal for the Philosophy of Science*, 54(4):527–551, 2003.

[16] Hoover, K. D. The past as future: The marshallian approach to post-walrasian econometrics. In Colander, D. (Ed.), *Post Walrasian Macroeconomics: Beyond the Dyanamic Stochastic General Equilibrium Model*, 239–257. Cambridge: Cambridge University Press, 2006.

[17] Hoover, K. D. Probability and structure in econometric models. unabridged working paper, downloadable from, 2007. URL `http://econ.duke.edu/~kdh9/Source%20Materials/Research/Probability%20and%20Structure%2019%20September%202007.pdf`.

[18] Hoover, K. D. Economic theory and causal inference. In Mäki, U. (Ed.), *Handbook of the Philosophy of Economics.* Amsterdam: Elsevier/North-Holland, forthcoming. In Gabbay, D., Thagard, P. and Woodsthe, J. (Ed.): *Handbook of the Philosophy of Science.*

[19] Johansen, S. Confronting the economic model with the data. In Colander, D. (Ed.), *Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, 287–300. Cambridge: Cambridge University Press, 2006.

[20] Juselius, K. Models and relations in economics and econometrics. *Journal of Economic Methodology*, 6:259–290, 1999.

[21] Koopmans, T. C. *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. New York: Wiley, 1950.

[22] Liu, T. Underidentification, structural estimation, and forecasting. *Econometrica*, 28(4):855–865, 1960.

[23] Mill, J. S. On the definition of political economy. In Robson, J. M. (Ed.), *Essays on Some Unsettled Questions of Political Economy*, 309–340. Toronto: University of Toronto Press, 1844, reprinted in 1967. Collected Works of John Stuart Mill, vol.4, Essays on Economics and Society: 1824-1845.

[24] Morgan, M. S. *The History of Econometric Ideas.* Cambridge: Cambridge University Press, 1990.

[25] Papineau, D. Can we reduce causal direction to probabilities. In Hull, D., Okruhlik, K. (Eds.), *PSA 1992*, vol. 2, 238–242. East Lansing, MI: Philosophy of Science Association, 1992.

[26] Pearl, J. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press, 2000.

[27] Reichenbach, H. *The Direction of Time.* Berkeley: University of California Press, 1956.

[28] Reiss, J. Time series, nonsense correlations and the principle of the common cause. In Russo, F., Williamson, J. (Eds.), *Causality and Probability in the Sciences*, 179–196. London: College Publications, 2007.

[29] Scheines, R. The similarity of causal inference in experimental and non-experimental studies. *Philosophy of Science*, 72(5):927–940, 2005.

[30] Sher, J., Carey, B. Child-porn study raises question of behavior. *New York Times*,

2007. URL `http://www.nytimes.com/2007/07/19/us/19sex.html?_r1`. Online edition, 19 July.

[31] Sims, C. A. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

[32] Sober, E. The principle of the common cause. In *From a Biological Point of View*, 158–174. Cambridge: Cambridge University Press, 1994.

[33] Sober, E. Venetian sea levels, british bread prices, and the principle of the common cause. *British Journal for the Philosophy of Science*, 52(2):331–346, 2001.

[34] Spanos, A. On theory testing in econometrics: Modeling with nonexperimental data. *Journal of Econometrics*, 67(1):189–226, 1995.

[35] Spanos, A. *Probability Theory and Statistical Inference : Econometric Modeling with Observational Data.* Cambridge: Cambridge University Press, 1999.

[36] Spirtes, P., Glymour, C., Scheines, R. *Causation, Prediction, and Search.* Cambridge, MA: MIT Press, 2nd edn., 2000.

[37] Steel, D. Making time stand still: A response to sober's counter-example to the principle of the common cause. *British Journal for the Philosophy of Science*, 54(2):309–317, 2003.

[38] Suppes, P. A probabilistic theory of causality. *Acta Philosophica Fennica*, Fasc. XXIV, 1970.