TETRAD and SEM

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

The authors have given us a lucid, concise, and accurate account of the TETRAD project, an account likely to clear much of the confusion over the philosophy of graph-based causal discovery methods. As one who has been active in the exploration and formulation of this philosophy, I am very pleased to see TETRAD arrive at this stage of development and now presented as a practical modeling tool for the research community. In this commentary, I wish to expand an aspect of TETRAD which the authors, by using the analogy with ordinary statistical estimation, have played down: the importance of TETRAD to the foundations of structural equations modeling (SEM). The assumptions underlying statistical estimation are of fundamentally different character from those underlying causal modeling tasks, and TETRAD, by fostering an understanding of these differences, can help establish not only its own legitimacy but the legitimacy of the entire SEM enterprise. Thus, I will first add a few arguments in support of the assumptions underlying TETRAD and then attempt to show that, even if these assumptions are not fully embraced, the general lessons learned from graphical causal modeling are destined to have a profound impact on SEM's practice and philosophy.

1 On Faithfulness

The operation of TETRAD is based on two assumptions: "causal independence" and "faithfulness."[1] While the assumption of causal independence is rarely challenged, the faithfulness assumption has drawn quite a few objections and is a likely target for further criticism from SEM researchers. I would like therefore to elaborate some additional rationale for faithfulness.

Because TETRAD relies primarily on nonexperimental data, causal claims are issued with guarantees weaker than those obtained through controlled randomized experiments. Pearl and Verma (1991) expressed these guarantees in terms of two orthogonal notions: minimality and stability.[2] Minimality guarantees that any alternative structure compatible

[1]Causal independence, also known as "Reichenbach's Principle" or "no correlation without causation," is rooted in the principle of no action at a distance [Arntzenius, 1990]. The notion of faithfulness (also called "DAG-isomorphism" and "nondegeneracy" [Pearl, 1988, p. 391]) was termed "stability" by Pearl and Verma (1991) to emphasize the invariance of certain independencies to functional form.

[2]These guarantees were advanced in connection with the discovery algorithm IC (for "Inferred Causation") which was developed in parallel with and on the same principles as the PC algorithm of Spirtes et al. (1993).

with the data is necessarily less specific, and hence less falsifiable and less trustworthy, than the inferred structure(s). Stability ensures that any alternative structure compatible with the data must be less stable than the inferred structure(s): slight fluctuations in experimental conditions will render the alternative structure no longer compatible with the data.

Minimality can easily be illustrated using the graph G in Figure 17. The fact that G is unique in its covariance-equivalent class still does not make C a genuine cause of D because the specific data at hand, summarized by a covariance matrix Σ, could in fact be generated by another graph, say G', that is not covariance-equivalent to G and in which an arrow is directed from D to C. For example, one choice of G' would be a *complete* DAG (i.e., one containing a link between every pair of nodes) rooted at D, which can be made (with a clever choice of parameters) to fit any covariance matrix whatsoever, including Σ. Is there a rationale, then, for preferring G to G', given that both fit Σ precisely? There is. Having the potential to fit any data means that G' is empirically nonfalsifiable, that Σ is overfitted, and hence that G' is less trustworthy than G. This preference argument holds against any DAG G' that can be made to fit a set of covariance matrices broader than the set fitted by G. Indeed, it can be shown that any DAG G' that fits Σ and contains an arrow from D to C would fit a broader set of matrices than the set fitted by G.

This minimality argument would fail if latent variables were permitted. For example, the DAG $A \leftarrow \epsilon_1 \rightarrow C \leftarrow \epsilon_2 \rightarrow B$ implies the same set of zero partial correlations among the observed variables A, B, C as the structure $A \rightarrow C \leftarrow B$, yet the former does not present A as a cause of C. The remarkable thing about minimality is that it uniquely determines the directionality of some edges even when we allow for the presence of latent variables. The arrow from C to D in Figure 17 is an example of such an edge. Among all DAGs that fit Σ, including DAGs containing unobserved variables, those that do not include an arrow from C to D are nonminimal, namely, capable of fitting a superset of the matrices fitted by G. This notion of minimality (measured by specificity) is more compelling than simplicity (measured by the number of free parameters). The latter connotes syntactic artificialities (e.g., what appears complex in a Cartesian system may turn simple in polar coordinates), while the former is a semantical notion that is language independent. Note also that considerations of minimality enter into standard model-selection procedures in SEM through the use of Akaike-like measures or cross-validation tests. However, while such considerations are normally used to score entire models in the final stage of the analysis, TETRAD uses minimality considerations to guide intermediate decisions about orienting or deleting individual edges in the construction of any one model.

I now address the notion of stability. The arguments usually advanced to justify faithfulness [Spirtes et al., 1993] appeal to the fact that strict equalities among products of parameters have zero measure in any probability space in which parameters are allowed to vary independently of one another. However, similar arguments can be advanced for any set of parameters, say the correlation coefficients or the eigenvalues of Σ or the entries of Σ^{-1}, with each leading to different conclusions as to which equalities should be considered accidental or structural. What then gives the zero partial correlations induced by DAGs a privileged status over zeros induced in other systems of encoding probability distributions?

The answer, I believe, rests on the notion of *autonomy* [Aldrich, 1989], a notion at the heart of all causal models. Researchers prefer causal models to other types of models in order to exploit the former for purposes of control and policy evaluation. The distinctive feature of causal models is that each variable is determined by a set of other variables through a functional relationship (called "mechanism") that remains invariant when those other variables are subjected to external influences. Only by virtue of this feature do causal models allow us to predict the effect of changes and interventions, including even such unanticipated changes as raising the value of an endogenous variable X by n units without perturbing any of X's causes. This invariance means that mechanisms can vary independently of one another, which implies that it is the set of structural coefficients (i.e., the mathematical representation of mechanisms in linear systems), rather than other types of parameters, that will vary independently when experimental conditions change.

For this reason, I believe that correlation-based model-searching schemes like TETRAD will find their greatest potential in analysis of data from longitudinal studies conducted under slightly varying conditions, namely, where accidental independencies are destroyed and only structural independencies are preserved. This assumes, of course, that under such varying conditions the parameters of the model will be perturbed yet its structure remain intact – a delicate balance that might be hard to control. Still, considering the alternative of depending only on controlled, randomized experiments, such longitudinal studies are an exciting opportunity.

2 Beyond Faithfulness: Reaching the Soul of SEM

Busy SEM practitioners might be tempted to dismiss the TETRAD experiment altogether, on the score that faithfulness is an unwarranted, untestable assumption.[3] That, in my opinion, would be a mistake, because TETRAD, even stripped of the faithfulness assumption and made to operate in a SEM-like mode of causal inference rather than its current mode of causal discovery, still offers many advantages over conventional SEM methods. In fact, the deeper message that TETRAD holds for SEM practitioners lies in its use of graphs as a language for causal modeling, and this message, I am convinced, bears on the very foundations of SEM (see [Pearl, 1998]).

2.1 Identifiability and Causal Authenticity

A prominent SEM researcher once asked me, "Under what conditions can we give causal interpretation to identified structural coefficients?" I thought this colleague was joking. As a faithful reader of Wright (1921) and Haavelmo (1943), I had come to believe that the answer is simply, "Always! The conditions that make a set of equations *structural* and a specific equation $y = \beta x + \epsilon$ *identified* are precisely those that make the causal connection between X and Y have no other value but β." Little did I know at that time that the teachings of Wright and Haavelmo have all but disappeared from the literature on SEM (in both econometrics and the social sciences) and that what has resulted is one of the more bizarre confusions in the history of science. I would now like to address this issue, because it is in the resolution of this confusion that I believe programs such as TETRAD will, as part of the general movement of causal modeling, have their greatest impact.

[3] Freedman (1997), for example, objects to the stability argument on the grounds that mechanisms need not, in principle, fluctuate independently of one another. Others might argue that simplicity is a dangerous criterion to use in model selection because successful theories often turn out to be more complex than expected.

My colleague has not been the only SEM researcher who thought (and I believe still thinks) that some extra ingredients are necessary for the conclusions of a SEM study to be turned into legitimate causal claims. Bollen (1989, p. 45), for example, states that a condition called "isolation" or "pseudo-isolation" is necessary: "When pseudo-isolation does not hold, the causal inferences regarding x's impact on y_1 are jeopardized."[4] Bullock, Harlow, and Mulaik (1994) likewise reiterate the necessity of isolation. After admitting that "confusion has grown concerning the correct use of and the conclusions that can be legitimately drawn from these [SEM] methodologies," they note that "three basic requirements [to establish causality] have been cited: association between variables, isolation of the effect, and temporal ordering." Evidently, the simple assumptions that underly structural equations have been obscured so badly through the years, that SEM researchers now need to import or invent new requirements to justify the causal interpretation of the conclusions.

Social scientists are not alone in this predicament; the econometric literature has no lesser difficulty dealing with the causal reading of structural parameters. LeRoy (1995), for example, states: "It is a commonplace of elementary instruction in economics that endogenous variables are not generally causally ordered, implying that the question 'What is the effect of y_1 on y_2' where y_1 and y_2 are endogenous variables is generally meaningless." According to LeRoy's latest proposal, causal relationships cannot be attributed to any variable whose causes have separate influence on the effect variable, thus denying causal reading to most of the structural parameters that economists and social scientists labor to estimate.

Cartwright, a renowned philosopher of science, raises a related question: "*Why* can we assume that we can read off causes, including causal order, from the parameters in equations whose exogenous variables are uncorrelated?" [Cartwright, 1995]. Recognizing that causes cannot be derived from statistical or functional relationships alone, she launches a search for a set of causal assumptions that would endow the parameter β in a regression equation $y = \beta x + \epsilon$ with a legitimate causal meaning, and then labors to prove that the assumptions she proposes are indeed sufficient. Cartwright does not consider the obvious answer, however, one that applies to models of any size and shape, including ones with correlated exogenous variables: the license to draw causes from parameters comes from precisely the standard (causal) assumptions that make equations "structural" and their parameters identifiable. Moreover, those (causal) assumptions are encoded in the syntax of the equations and can be read off the associated graph as easily as a shopping list[5]; they need not be searched elsewhere, nor do they require specialized proofs of sufficiency.

These examples echo an alarming tendency among economists and SEM researchers to view SEM as an algebraic object that carries functional and statistical assumptions but is void of causal content.[6] A causality-free conception of SEM may explain both Cartwright's

[4]Bollen (1989, p. 44) defines pseudo-isolation as the orthogonality condition $cov(x, \epsilon) = 0$, where ϵ is the error term in the equation $y = \beta x + \epsilon$. This condition is neither necessary (as seen, for example, in the analysis of instrumental variables [Bollen, 1989, pp. 409–413], and in Figure 6 (c, e) of [Pearl, 1995]) nor sufficient (e.g., [Cartwright, 1995, p. 50]) unless causal meaning is already attached to β.

[5]Specifically, if G is the graph associated with a causal model that renders a certain parameter identifiable, then the assumptions sufficient for authenticating the causal reading of that parameter can be read off G as follows: Every missing arrow, say between X and Y, represents the assumption that X has no causal effect on Y once we intervene and hold the parents of Y fixed. Every missing bi-directed link between X and Y represents the assumption that there are no common causes for X and Y, except those shown in G [Pearl, 1995].

[6]Perhaps the boldest expression of this trend has recently been voiced by Holland (1995, p. 54): "I am

search for causal assumptions outside the model and the urge of SEM researchers to fortify the equations with extra conditions (e.g., isolation) or ban the natural causal readings of the equations [LeRoy, 1995].

2.2 Back to Structuralism, with Graphs

Strangely, the founders of SEM expressed no such trepidation. Wright (1923) did not hesitate to declare that "prior knowledge of the causal relations is assumed as prerequisite " in the theory of path coefficients, and Haavelmo (1943) explicitly interpreted each structural equation as a statement about a hypothetical controlled experiment. One wonders, therefore, what has happened to SEM over the past 50 years.

What I believe has happened is that the causal content of SEM has been allowed to gradually escape the consciousness of SEM practitioners mainly for the following reasons:

1. SEM practitioners have sought to gain respectability for SEM by keeping causal assumptions implicit, since statisticians, the guardians of respectability, abhor such assumptions because they are not directly testable.

2. The algebraic, graphless language that has dominated SEM research, primarily through the influence of econometricians, lacks the notational facility needed for making causal assumptions, as distinct from statistical assumptions, explicit.

The SEM community has paid dearly, in both stature and substance, for abandoning the causal conception of Wright and Haavelmo. For example, when Freedman (1987, p. 114) challenged SEM's definition of direct effects as "self-contradictory," none of the eleven discussants was able to point out the correct, noncontradictory reading of the path diagram he used. And when authors of SEM textbooks (e.g., [Bollen, 1989, p. 376]) discuss effect decomposition, they invariably overlook the operational significance of direct and indirect effects, and make the mistake of equating the total effect with a power series of coefficient matrices. The results are erroneous expressions in models with feedback.[7]

This, I believe is where graphs come in and where the troubled soul of SEM can find rescue and revival.[8] Properly interpreted, a graph can serve both as a checklist for modeling assumptions and as a mathematical language for deriving the statistical and causal implications of those assumptions. Complex algebraic tasks, such as identifying the partial correlations that the model compel to vanish or recognizing the set of structural coefficients that the model

speaking, of course, about the equation: $\{y = a + bx + \epsilon\}$. What does it mean? The only meaning I have ever determined for such an equation is that it is a shorthand way of describing the conditional distribution of $\{y\}$ given $\{x\}$." Holland's interpretation stands at variance with the structural reading of the equation above [Haavelmo, 1943], which is: "In an ideal experiment where we control X to x and any other set Z of variables (not containing X or Y) to z, Y is independent of z and is given by $a + bx + \epsilon$" [Pearl, 1995, p. 704].

[7]For instance, given the pair of equations $\{y = \beta x + \epsilon,\ x = \alpha y + \delta\}$, the total effect of X on Y is simply β, not $\beta(1 - \alpha\beta)^{-1}$ as stated in [Bollen, 1989, p. 379]. The latter has no operational significance worthy of the phrase "effect of x." This error was noted by Sobel (1990) but, perhaps because autonomy was presented as a new and extraneous assumption, Sobel's correction has not brought about a shift in practice or philosophy.

[8]My characterization of the state of SEM may seem like an exaggeration, yet my recent informal survey of well over a hundred SEM scholars and practitioners indicates that very few SEM experts can either resolve the contradictions presented in [Freedman, 1987, p. 114] or apply SEM to simple policy-related problems.

renders identifiable (regardless of the data) can now be accomplished by simple visual inspection [Pearl, 1998]. The equation-deletion semantics for interventions [Strotz and Wold, 1960] and the graphical representation of these semantics [Spirtes et al., 1993] now enable us to ask direct questions about the effects of policies and interventions [Balke and Pearl, 1995], rather than simply chasing coefficients whose meanings may easily be forgotten. With this semantics, both the parameters and the error terms in SEM obtain unambiguous causal readings,[9] and one can show that the standard procedures for estimating parameters are indeed sound procedures for estimating the effects of interventions, hence, the strength of causal connections.

Furthermore, if one takes Haavelmo's view seriously and accepts that structural equations and structural coefficients are merely statements about hypothetical controlled experiments, then Cartwright's question translates into a more subtle philosophical question: "*Why* can we read off the outcome of one experiment from statements about other experiments that are run under totally different conditions?" This revised question, though it concerns entities of the same kind, is still far from being trivial, because license for deriving causes from other causes is not handed out automatically, and the rationale and conditions for obtaining such license have not been seriously explored in the philosophical literature.

This neglect is somewhat surprising. Given that progress in the empirical sciences requires the transfer of knowledge from one experiment to another, why is it that science has so far not found a language or a mathematical machinery for facilitating such transfers? For example, suppose we conduct a controlled experiment and find that X has a marked effect on Y but ceases to have an effect on Y once we hold Z fixed (by intervention). Can we legitimately conclude that in a new experiment, where we have no control over X, varying Z will have an effect on Y? Intuition supports such inference, but can it be derived mathematically in some formal theory of experimentation? To legitimize such inferences, we need a formal logic in which action phrases (e.g., "having no effect on," "holding Z fixed"), as distinct from observational phrases (e.g., "being independent of," "conditioning on Z"),[10] are given formal notation, semantical interpretation, and axiomatic characterization. Moves toward the development of such logic (e.g., [Pearl, 1995], [Galles and Pearl, 1997]) reveal that, again, autonomy-based SEM provides the most natural semantics for the language of causation and experimentation.[11]

The new power unleashed by graphical analysis is a double-edged sword, however. On the one hand, SEM researchers now have the assurance that the standard causal interpretation of (identified) parameters is authenticated by hard logic; no additional assumptions are needed beyond the input assumptions embedded in the structural reading of the equations (see footnote 5). On the other hand, this assurance entails a new responsibility: to substantiate the input assumptions. No longer can the blurring of the distinction between tested and

[9]The meaning of β is simply $\frac{\partial}{\partial x}E[Y|do(x)]$, namely, the rate of change (in x) of the expectation of Y in an experiment where X is held at x by external control [Pearl, 1995, p. 685]. This interpretation holds regardless of whether ϵ and X are correlated (e.g., via another equation $x = \alpha y + \delta$). Moreover, this interpretation provides an operational definition for the mystical error-term, ϵ; unlike errors in regression equations, $\epsilon = y - \beta x$ measures the deviation of Y from its controlled expectation $E[Y|do(x)]$ and not from its conditional expectation $E[Y|x]$ [Pearl, 1998].

[10]Philosophers, statisticians, and SEM researchers have often confused "holding Z constant" with "conditioning on a given Z."

[11]Earlier steps toward such a logic were taken by Gibbard and Harper [Gibbard and Harper, 1981], based on Lewis' closest-world semantics of counterfactuals.

untested assumptions serve as an excuse for thoughtless studies[12] – the untested causal assumptions that underlie a given model now are made vividly explicit (see footnote 5) and must be attended to with extreme seriousness. TETRAD makes it clear that no statistical test can ever confirm or refute those causal assumptions; the best one can hope for is to test whether the entire covariance-equivalence class of any given model is compatible with the data. However, the ability to merely display the elements of an entire equivalence class now makes discussions of assumptions and their alternatives more meaningful and accessible and, for this feature alone, TETRAD should become an indispensable tool in SEM.

References

[Aldrich, 1989] J. Aldrich. Autonomy. *Oxford Economic Papers*, 41:15–34, 1989.

[Arntzenius, 1990] F. Arntzenius. Physics and common causes. *Synthese*, 82:77–96, 1990.

[Balke and Pearl, 1995] A. Balke and J. Pearl. Counterfactuals and policy analysis in structural models. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 11–18. Morgan Kaufmann, San Francisco, 1995.

[Bollen, 1989] K.A. Bollen. *Structural Equations with Latent Variables.* John Wiley and Sons, New York, 1989.

[Bullock et al., 1994] H.E. Bullock, L.L. Harlow, and S.A. Mulaik. Causation issues in structural equation modeling research. *Structural Equation Modeling*, 1(3):253–267, 1994.

[Cartwright, 1995] N. Cartwright. Probabilities and experiments. *Journal of Econometrics*, 67:47–59, 1995.

[Freedman, 1987] D. Freedman. As others see us: A case study in path analysis (with discussion). *Journal of Educational Statistics*, 12:101–223, 1987.

[Freedman, 1997] D.A. Freedman. From association to causation via regression. In V.R. McKim and S.P. Turner, editors, *Causality in Crisis?*, pages 113–161. University of Notre Dame Press, Indiana, 1997.

[Galles and Pearl, 1997] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

[Gibbard and Harper, 1981] A. Gibbard and L. Harper. Counterfactuals and two kinds of expected utility. In W. L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 153–190. D. Reidel, Dordrecht, 1981.

[Haavelmo, 1943] T. Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11:1–12, 1943.

[Holland, 1995] P.W. Holland. Some reflections on Freedman's critiques. *Foundations of Science*, 1(1):50–57, 1995.

[12]McDonald (1997) surveys prevailing SEM practices where uncorrelated errors are assumed as a matter of mathematical convenience, void of substantive thought.

[LeRoy, 1995] S.F. LeRoy. Causal orderings. In K.D. Hoover, editor, *Macroeconometrics: Developments, Tensions, Prospects*, pages 211–228. Kluwer Academic Publishers, Boston/Dordrecht/London, 1995.

[McDonald, 1997] R.P. McDonald. Haldane's lungs: A case study in path analysis. *Multivariate Behavioral Research*, 32(1):1–38, 1997.

[Pearl and Verma, 1991] J. Pearl and T. Verma. A theory of inferred causation. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452. Morgan Kaufmann, San Mateo, CA, 1991.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligence Systems.* Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl, 1995] J. Pearl. Causal diagrams for experimental research. *Biometrika*, 82(4):669–710, 1995.

[Pearl, 1998] J. Pearl. Graphs, causality, and structural equation models. To appear in *Sociological Methods and Research*, 27(2).

[Sobel, 1990] M.E. Sobel. Effect analysis and causation in linear structural equation models. *Psychometrika*, 55(3):495–515, 1990.

[Spirtes et al., 1993] P. Spirtes, C. Glymour, and R. Schienes. *Causation, Prediction, and Search.* Springer-Verlag, New York, 1993.

[Strotz and Wold, 1960] R.H. Strotz and H.O.A. Wold. Causal models in the social sciences. *Econometrica*, 28:417–427, 1960.

[Wright, 1921] S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–585, 1921.

[Wright, 1923] S. Wright. The theory of path coefficients: A reply to Niles' criticism. *Genetics*, 8:239–255, 1923.