# On The Foundation Of Structural Equation Models or When Can We Give Causal Interpretation To Structural Coefficients? *

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

**Abstract**

The assumptions underlying statistical estimation are of fundamentally different character from the causal assumptions that underly structural equation models (SEM). The differences have been blurred through the years for the lack of a mathematical notation capable of distinguishing causal from equational relationships. Recent advances in graphical methods provide formal explication of these differences, and are destined to have profound impact on SEM's practice and philosophy.

## 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 $\beta$." 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 believe that the resolution of this confusion is essential for understanding the foundation of SEM.

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

*Portions of this report are included in a commentary prepared for *Multivariate Behavioral Research*.

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."[1] 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." Given that a host of (causal) assumptions, including isolation-like assumptions of uncorrelated errors, are required to establish the identifiability of structural parameters, it seems odd that a new requirement would still be called for, on top of identifiability. Or is "isolation" merely an inaccurate summary of those assumptions?

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, considers the pair of equations

$$n = \alpha_0 + \alpha_1 p + \alpha_2 r \tag{1}$$

$$p = \beta_0 + \beta_1 n + \beta_2 c \tag{2}$$

and concludes that "the question 'What is the effect of pesticides ($p$) on nematodes ($n$)?' is seen to be ambiguous: pesticide use is affected by both rainfall ($r$) and pesticide cost ($c$), and the effect on nematodes depends on which caused the change in pesticide use (because rainfall affects nematodes directly as well as through its effect on pesticide use, whereas cost does not)." According to LeRoy's theory, even if we can estimate $\alpha_1$ with infinite precision, we are still forbidden from equating $\alpha_1$ with "the effect of $p$ on $n$" or even from asking what that effect is. Causal relationships cannot be attributed to any variable whose causes have separate influence on the effect variable – an extreme version of "isolationism" which rules out most policy variables in economics [Balke and Pearl, 1995].

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 $\beta$ 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 regression 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[2]; they need

[1]Bollen (1989, p. 44) defines pseudo-isolation as the orthogonality condition $cov(x, \epsilon) = 0$, where $\epsilon$ 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 $\beta$.

[2]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 $G$.

not be searched elsewhere, nor do they require specialized proofs of sufficiency.

I do not think Cartwright is less observant than students of graphical models, but I believe her analysis is a reaction to 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.[3] A causality-free conception of SEM may explain both Cartwright's 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 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.[4]

This, I believe is where graphs come in and where the troubled soul of SEM can find rescue and revival. Properly interpreted, a graph can serve both as a checklist for modeling assumptions and as a mathematical language for distinguishing and combining causal and statistical

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

[4]For instance, given the pair of equations $\{y = \beta x + \epsilon,\ x = \alpha y + \delta\}$, the total effect of $X$ on $Y$ is simply $\beta$, 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.

information. 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 [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,[5] 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 causation.

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 $Z$"),[6] are given formal notation, semantical interpretation, and axiomatic characterization. Moves toward the development of such logic (e.g., [Pearl, 1995], [Galles and Pearl, 1996]) reveal that, again, autonomy-based SEM provides the most natural semantics for the language of causation and experimentation.[7]

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 2). On the other hand, this assurance entails a new responsibility: to substantiate the input assumptions. No longer can the blurring of the distinction between

[5] The meaning of $\beta$ is simply $\frac{\partial}{\partial x}E(Y|\hat{x})$, 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 $\epsilon$ 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, $\epsilon$, which is clearly a causal, rather than a statistical, entity.

[6] Philosophers, statisticians, and SEM researchers have been notoriously sloppy about confusing "holding $Z$ constant" with "conditioning on a given $Z$."

[7] Earlier steps toward such a logic were taken by Gibbard and Harper [Gibbard and Harper, 1981], based on Lewis' closest-world semantics of counterfactuals. That basis turns out to be too weak to sanction the variety of action sentences we find in natural discourse, e.g., "If $X$ has no effect on $Y$ when $Z$ is held constant and no effect on $Z$ when $Y$ is held constant, then $X$ has no effect on either $Z$ or $Y$."

tested and untested assumptions serve as an excuse for thoughtless studies[8] – the untested causal assumptions that underlie a given model now are made vividly explicit (see footnote 2) and must be attended to with extreme seriousness. Graphical model make 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 probability-equivalence class of any given model is compatible with the data. However, the ability to merely identify models that are statistically indistinguishable makes discussions of assumptions and their alternatives more meaningful and, for this feature alone, graphical models should become an indispensable tool in SEM.

## References


[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.

[Galles and Pearl, 1996] D. Galles and J. Pearl. Axioms of causal relevance. Technical Report R-240, Department of Computer Science, University of California, Los Angeles, 1996. Short version in *Proceedings of the Fourth International Conference on mathematics and AI*, For Lauderdale, FL, 64–67, January 1996.

[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.

[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.


[8] McDonald (1996) surveys prevailing SEM practices where uncorrelated errors are assumed as a matter of mathematical convenience, void of substantive thought.

[McDonald, 1996] R.P. McDonald. Haldane's lungs: A case study in path analysis. Technical report, Department of Psychology, University of Illinois, Urbana, 1996.

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

[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.