DOCUMENT RESUME

ED 204 392 TM 810 387

AUTHOR Wolfle, Lee M.
TITLE Is Causal Modeling Really Helpful?
PUB DATE Apr 81
NOTE 15p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Mathematical Models; Multiple Regression Analysis; *Path Analysis; Research Methodology
IDENTIFIERS *Causal Models

ABSTRACT

Hierarchial causal models are described as pictorial representations of multiple regression equations. These models are particularly helpful for three reasons: (1) the formulation of problems in a path analytic framework forces a degree of explicitness that is often not present in research reports that rely solely on regression; (2) they provide a powerful aid to the substantive interpretation of results; and (3) they aid in the interpretation of relationships between unmeasured variables. Though causal modeling techniques are very powerful, important prerequisites are a thorough knowledge of one's subject matter and a stylish appreciation of alternative explanations. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204392

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

IS CAUSAL MODELING REALLY HELPFUL?

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L. M. Wolfle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Lee M. Wolfle

Virginia Polytechnic Institute and State University

TM 810 387

This paper was prepared for a symposium, "Is Causal Modeling in Educational Research and Evaluation an Art Form?" with Robert L. Baker, William W. Cooley, Susan E. Whitely, and Peter M. Bentler, presented at the annual meetings of the American Educational Research Association, Los Angeles, April 13 - 17, 1981.

ERIC

IS CAUSAL MODELING REALLY HELPFUL?

Lee M. Wolfle

The charge I received from Bob Baker was to answer the question, "Is causal modeling really helpful?" My answer is a definite but qualified, "yes." I would first like to tell you why I would answer in the positive; then I want to say why the answer has to be qualified.

In the past year I have had several people ask me, "What is path analysis, anyway?". Several times I have answered, "It's just multiple regression with pictures." That line usually produces at least a chuckle, because they think I am kidding. But I'm not.

In hierarchical models (all that means is that the causal influences all flow in one direction without any feedback), the estimation of effects is accomplished by multiple regression. Of course, things can get more complicated if there are feedback loops; then identification of the model becomes a problem, and multiple regression no longer produces the proper estimates. Also, if the model contains unmeasured or latent variables, then some other procedure is required for estimation, such as factor analysis, canonical correlation, or the analysis of covariance structures (e.g., LISREL). But let us temporarily focus our attention just on hierarchical causal models, estimable with multiple regression.

If hierarchical causal models are merely pictorial representations of multiple regression equations, why are they so helpful? The answer obviously does not lie in the method of estimation. I believe there are at least two reasons why hierarchical causal models are helpful. The first is that the formulation of problems in a path analytic framework forces a degree of explicitness that is often not present in research reports that rely solely on regression. The drawing of the picture, the arrangement of the variables, the connections with causal arrows, all force the researcher to confront his or her model of reality. Once that arrow has been drawn, the researcher knows, and the researcher knows the reader will know, that the model specifies one variable to be a cause of another.

How is one to decide which variable is the cause, and which the effect? Clearly, the data will not tell us; such decisions must be made prior to data analysis. There are some simple rules. For X to cause Y, X must precede Y in time. For X to cause Y, they must be functionally related, which is only to say that knowing X allows one to predict Y with greater accuracy than if X was not known. Also, for X to cause Y, there must not be a third variable Z that causes both X and Y in such a way that the association between X and Y disappears once Z is controlled, which is only to say that the relationship between X and Y is not spurious. Beyond these simple rules, one determines causal relationships by knowing one's subject matter; but that is a topic I want to postpone for a few minutes.

So, one draws the arrow on the basis of what one knows _a priori_, or theoretically if you will, about the subject being investigated. Once the arrows have been drawn, the researcher knows which variables are conceived to be the causes, and which variables the effects. And the reader knows. The two (the author and the reader) have clearly communicated, and there should be no misunderstanding. The researcher may, of course, be wrong; but at least he or she won't be misunderstood. It is this degree of explicitness in causal modeling that led me to conclude an article published last year in AERJ (Wolfle, 1980a) that explicit communication was the most important strategy of path analysis.

I would like to share with you a couple of examples. Last year I published an article in Sociology of Education (Wolfle, 1980b) which addressed the enduring effect of educational attainment on adult knowledge. Previous research indicated a strong causal relationship; that is, the acquisition of more schooling causes people to possess more knowledge as adults. I thought the relationship was probably spuriously due to intelligence. I developed a causal model which incorporated (among other variables) childhood intelligence, educational attainment, adult intelligence, and a measure of adult knowledge of vocabulary words. I specified the model in such a way that education was a cause of adult intelligence, and both were causes of the vocabulary score. I have since then heard from Ward Keesling (1980) that in his opinion adult intelligence would be better considered a prior cause of education. This is no small matter, because in my specification of the model the effect from education to adult intelligence and hence to vocabulary is interpreted as part of the total effect of education on vocabulary. In Keesling's preference, only the direct effect would be included in the interpretation of the effect of education on vocabulary; the rest would be interpreted as a spurious effect due to antecedent variables, including adult intelligence. The point, however, is not who is right and who is wrong (although I have an opinion on the matter). The point is that neither of us can be misinterpreted in how we stand on the issue. The arrow points in either one direction or the other, but it does point. It is explicit. In contrast, if I had merely regressed a vocabulary score on educational attainment, adult intelligence, and some other variables, no one would know, including me, how substantively to interpret the relative effects of education and intelligence, beyond a rather simplified measure of education's net (direct) effect on knowledge.

I would like to offer another example. A year or so ago I read an article in the Journal of Reading Behavior (Yap, 1979), which addressed the question of whether a child's comprehension of a written passage resulted from an understanding of the words used in the passage, or whether the child's understanding of words resulted from a comprehension of the passage. In other words, does vocabulary cause comprehension, or does comprehension cause vocabulary? It seems clear to me that such a question must be framed in terms of the same passage, but this is not what the author did. He chose to analyze the data with cross-lagged correlations, which implied a causal structure in which vocabulary at time 1 caused comprehension at time 2, and comprehension at time 1 caused vocabulary at time 2. I disagreed with that specification (Wolfle and McGee, 1979). It seems to me that the causal relationships between vocabulary and comprehension should be specified at the same point in time. But once again, the point is not who is right and who is wrong. The point is that specifying the analysis in causal terms led to an unambiguous statement of one's belief. Personally, I would rather be wrong than misinterpreted, and setting up analyses in the framework of causal models helps one to be explicit both in their thinking and in their communication to others.

There is a second reason why I believe hierarchical causal models are helpful. They provide a powerful aid to the substantive interpretation of results. Causal models not only allow the assessment of direct causal links, hierarchical models also allow the researcher to obtain estimates of the extent to which intervening variables account for relationships between predetermined and subsequent variables. These are interpreted as indirect causal effects. In addition, the researcher may obtain estimates of the extent to which antecedent variables account for relationships between subsequent variables. These may be interpreted as spurious effects.

How a causal model is constructed determines the kind of interpretations one can draw from it. (This is the nature of the disagreement over whether adult intelligence should precede educational attainment or come after it.) It was one of the lessons in "Strategies of Path Analysis" (Wolfle, 1980a) that the kind of model one builds depends on the kind of research questions being asked. If a researcher's analytic goal is to assess the extent of intervening causal effects, a hierarchical causal model permits its realization.

I would like to offer a simple but elegant example, not because it is simple but because it represents one of the very first applications of causal models to a substantive problem in educational research (broadly considered). Duncan and Hodge (1963) had three analytic questions in mind: (1) What is the zero-order association between the socioeconomic status of sons and their fathers? (2) How is this association mediated by the intervening factor of educational attainment? and (3) What is the net association of education and the socioeconomic status of sons, apart from its dependence on father's status? To answer these questions, they developed and evaluated the first causal model of status attainment. They found a correlation of about .30 between the socioeconomic statuses of fathers and sons. When they decomposed this association into direct and indirect causal components, they found in every age cohort that the indirect effect of father's status manifested through education was more important than the direct causal link of father's status and son's status. Third, they found that education was a more important determinant of occupational achievement than was father's status. This was the analysis that led us to understand that the reason statuses of fathers and sons are correlated is not because sons inherit their father's status, but because father's status helps to determine the amount of education the son acquires, which in turn helps to determine the son's status. It is a classic example of going from a set of substantive questions to a causal model designed to address the questions, and hence to the interpretation of results.

There is yet another advantage to causal modeling that I would like to mention only in passing, because Peter Bentler (1981) and Susan Whitely (1981) have already reminded us of the analytic power of latent-variable models. There are many concepts and constructs that form powerful conceptual mechanisms for understanding social relationships, and yet these concepts and constructs are not directly measureable. All one can do is measure the effects of these unobservable variables on manifest indicators. Here I am speaking of such unobserved concepts as intelligence, for which one has manifest scores of, say, a reading test, a vocabulary test, and a mathematics test; or socioeconomic status, measured by occupational prestige, earnings, and years of education; or anomie; or Protestant ethic; or sex stereotype. This list could be considerably expanded.

Causal models with unmeasured variables have a long history, and go back at least as far as Sewell Wright's (1925) analysis of the fluctuations in corn and hog prices. His model included an unmeasured hog breeding variable. Paralleling the development of path analysis, there is a long history of factor analysis, which has, as you know, been concerned with finding unmeasured factors which can explain why manifest variables are intercorrelated. But until recently, the factors themselves were not conceived of as being causally related, only intercorrelated.

Among others, it was Jöreskog (see Jöreskog and Sörbom, 1979) who wedded the techniques of causal modeling to factor analysis, and has thus provided us with a powerful new analytic tool. The procedure, commonly known as LISREL (Jöreskog and Sörbom, 1978), provides the advantages of causal modeling, which I have mentioned above in the context of hierarchical models, as applied to latent variables, which are often the variables of real theoretical interest. LISREL thus makes possible the rigorous testing of theories that have until now been very difficult to test adequately (Kerlinger, 1977).

I would like to add one example to those already mentioned by Bentler (1981) and Whitely (1981). This example is taken from some of the recent work of Zajonc (1980). He was basically interested in whether affective reaction had to occur after cognitive recognition, or whether affective judgments occur independently of cognitive encoding. Part of his analysis was based on a causal model with latent variables, for neither affect nor cognition were measured perfectly or by single manifest variables. He was able to show that these latent variables were related to each other in a way that suggested affective reaction had a stronger effect on cognitive recognition than vice versa. It is a good piece of work (an affective judgment), and was awarded the Distinguished Scientific Contribution Award by the American Psychological Association. I think it represents a good example of how new research tools, such as causal modeling with latent variables, can open up new avenues of inquiry, or in this case enlighten an area of speculation that waited 100 years for a methodology to emerge that could answer the questions of interest.

Thus, in answer to the question, is causal modeling helpful, I'd say, "yes." While the methods employed in hierarchical models of manifest variables are not new, the application of them in a causal framework both permits and demands a degree of explicitness which is good for scientific communication. In addition, causal models aid in the interpretation of results by permitting the decomposition of associations into direct, indirect, and spurious causal effects. The substantive interpretations of these components provide some powerful insights into social processes. Moreover, new techniques which combine both factor analytic and causal modeling techniques into a single package have provided some real breakthroughs in areas of substantive interest.

But I said in the beginning that my answer would be a qualified one, and now I would like to stipulate what that qualification is.

I think that the scientific fields that move forward the fastest are not those that generate the greatest number of hypotheses, but rather those that discard the greatest number. Platt (1964) called such methods "strong inference." One devises alternative hypotheses, decides upon the specific experiment which will exclude one or the other of the hypotheses, and then carries out the experiment to its logical conclusion. One then reformulates competing hypotheses among the possibilities that remain, and thus moves forward by excluding those avenues of thought that are unsupported by evidence.

Of course, in the social sciences in general and education in particular, we are not allowed to apply crucial experiments. We must deal with intact groups; we must live with self-selection. So be it; I would not want to live in any other kind of society. But as analysts of such data, we must deal with such matters as self-selection. To do that requires a fine appreciation of one's subject matter.

In the nineteenth century, Louis Pasteur was able to solve several biological and medical puzzles by applying the logic of strong inference. He didn't know more about fermentation, or anthrax, or rabies than others of his contemporaries. But he applied a method which allowed him to exclude alternative explanations. Today, in the social sciences, there are a few people who know the techniques of causal modeling very well. But that does not necessarily mean they can make advances in any particular field of study just by applying causal modeling techniques to the field. In the absence of the experimental method, it is vitally important to include in one's analysis the variables that control for alternative explanations. If one is to decide if a treatment works, or an effect exist, one must first know why some people selected themselves for treatment and others did not. Without such knowledge, and the means to control for it, it is impossible to exclude alternative hypotheses. Thus, the most important prerequisite for good causal models is a thorough knowledge of one's subject matter, and a stylish appreciation of alternative explanations. Without these, neither you nor I, using any analytic procedure, can make advances in our respective fields.

Bill Cooley (1978) made this point more eloquently than I, when he noted that Paul Lazarsfeld could 'o better causal analyses with contingency tables than Cooley could with all his number crunching, because Lazarsfeld know which variables to control for.

It is easy enough to say that alternative explanations must be taken into account. It is quite another thing actually to do it. The parameters of human behavior are numerous and complex, much more numerous and complex that the analytic tools we have available to us. Duncan (1975) made the distinction between the easy part of structural equation models, by which he meant the essential tools of matrix algebra and mathematical statistics, and the hard part of causal models, by which he meant the formulation of creative ideas that are necessary to construct causal models. So, at the risk of sounding presumtuous, do not undertake to apply causal models to areas of inquiry in the hope that the technique alone will automatically yield good research. To do so confuses the tool with the aim. Causal modeling is an analytic tool, but the aim is a thriving line of research with theories that have scope and coherence, and that yield predictions of unexpected new facts. While I think I know at least something about the tool, my record of substantive breakthroughs is a thin one. But then so is the record of most of the social sciences.

Thus, the techniques of causal modeling are indeed helpful. But they are merely tools to be used by those who first know their subject matter. Without knowledge of one's subject matter, causal modeling becomes merely a faddish analytic technique used to disguise one's ignorance. The advantage of causal models is that they force a degree of explicitness which reveals good analyses as quickly as the implausible ones. Toward that end, I would indeed say, "Yes, causal modeling is really helpful."

REFERENCES

Bentler, Peter M. Latent variables and observed measures. Paper presented to the American Educational Research Association, Los Angeles, 1981.

Cooley, William W. Explanatory observational studies. Educational Researcher, 1978, 7(9), 9-15.

Duncan, Otis Dudley. Introduction to structural equation models. New York: Academic Press, 1975.

Duncan, Otis Dudley, and Hodge, Robert W. Education and occuptational mobility: A regression analysis. American Journal of Sociology, 1963, 68, 629-644.

Jöreskog, Karl G., and Sörbom, Dag. LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: National Educational Resources, 1978.

Jöreskog, Karl G., and Sörbom, Dag. Advances in factor analysis and structural equation models. Cambridge: Abt Books, 1979.

Keesling, Ward. Personal communication, December 1980.

Kerlinger, Fred N. The influence of research on educational practice. Educational Researcher, 1977, 6(8), 5-12.

Platt, John R. Strong inference. Science, 1964, 146, 347-353.

Whitely, Susan E. Hypothesis testing and drawing inferences. Paper presented to the American Educational Research Association, Los Angeles, 1981.

Wolfle, Lee M. Strategies of path analysis. American Educational Research Journal, 1980a, 17, 183-209.

Wolfle, Lee M. The enduring effects of education on verbal skills. Sociology of Education, 1980b, 53, 104-114.

Wolfle, Lee M., and McGee, Lea M. On the determination of causal ordering between vocabulary and comprehension. Journal of Reading Behavior, 1979, 11, 273-277.

Wright, Sewell. Corn and hog correlations. Washington, D.C.: U.S. Department of Agruculture Bulletin 1300, 1925.

Yap, K. O. Vocabulary -- building blocks of comprehension? Journal of Reading Behavior, 1979, 11, 49-59.

Zajonc, R. B. Feeling and thinking: Preferences need no inferences. American Psychologist, 1980, 35, 151-175.