Journal of Educational Statistics
Summer 1987, Vol. 12, No. 2, pp. 178–184

Response to Freedman's Critique of Path Analysis: Improve Credibility by Better Methodological Training

Bengt O. Muthen
University of California, Los Angeles

Introduction

Before I started to read the Freedman contribution, I feared that it was going to be another criticism by a statistician of social science statistical methodology, where little interest is shown in the specific problems of social science applications and their overwhelming complexities, and where any attempt by social scientists to use complex modeling rather than exploratory data analysis is considered naive. I wondered if I would find factor analysis, path analysis, latent variables, and structural equation modeling discounted as tools for the social sciences. As an applied statistician, I have become bored with such an attitude, which I believe puts social scientists and statisticians in an adversarial relationship, fostering the social scientists' opinion that statisticians are only concerned with irrelevant problems.

For a long time, I have been frustrated by misuses of path analysis and more general structural equation modeling in the social and behavioral sciences. I still see very few good applications and I definitely feel that the way path analysis is presently practiced is often depressingly bad. Therefore, I think that Freedman's contribution is very necessary and I welcome the opportunity to discuss these matters. After I read Freedman's paper, I found that it is not totally of the type described above and that I cannot disagree with the details of his critique from a statistical perspective. Omitted variables and poor measurement are probably the most important and common sources of misspecifications, certainly leading to very distorted results. Infatuation with expansive modeling, with or without latent variables, cannot cure such fundamental misspecification. Like Freedman, I too think that much more exploratory data analysis should precede rigorous modeling and that results should be interpreted more modestly.

This research was supported by Grant No. SES-8312583 from the National Science Foundation. I wish to thank Norman Cliff, Linda K. Muthen, and an anonymous reviewer for helpful comments on earlier drafts.

I would, however, draw a completely different conclusion from the one Freedman seems to favor and the one already drawn by disillusioned statisticians. I think bad applications can be improved by better education of students of social science methodology. I do not think that there is something inherently wrong with path analysis that warrants it being discarded. I agree with Fienberg's views as stated in his opposition to Freedman's chapter, "Statistics and the Scientific Method" (Mason & Fienberg, 1985). Natural and social science analyses have very much in common and good and bad examples can be found in both areas. There is, however, a greater opportunity for hindsight when examining successful natural science analyses. I find that there are bad applications of all statistical methods, not only regression analyses, and statisticians themselves certainly do not go free from criticism of sloppy practice. Think of all the new approaches to data analyses by stepwise or setwise regression or ANOVA on minute samples without good possibilities to check the strong assumptions. Are the path analysts, whom we assume here are non-statisticians, behaving any differently? Well, I would say that their sins are along the same dimension (a factor analysis term, I'm afraid), but probably can be said to be further along the bad direction of that dimension.

In my response, I want to discuss the state of the art of path analysis and more general structural equation modeling and then comment on the Freedman critique of the Hope analysis. I would also like to discuss potential educational strategies that I believe can improve the state of the art.

On the Quality of Path Analyses

I would like to speculate about some reasons for the present situation. In the early days of path analysis, the bad applications were most likely due to the fact that little experience and guidance was available. Since the beginnings, bad applications have also been seen because people wanted to show off new emerging developments of the methods, with real data analyses mostly used as (bad) illustrations. Now, the poor quality of many applications exists, in my opinion, for different reasons. One obvious reason has to do with the pressure to publish based on the data collected. We statisticians, who can write interesting papers without real data, perhaps do not always fully appreciate the dilemma of the social scientist who has spent great efforts on data collection and now is trying to make the most out of it, even when he or she realizes that, for example, not all relevant variables are measured or the sampling did not turn out as well as it should. Another reason has to do with the fact that there is not enough pressure on the ultimate users of these new statistical techniques to learn the methodological part of their complex social science research trade well. There are a great number of users attempting very complex modeling using advanced software who have very poor training in the underlying statistical principles.

#17

The Credibility Problem

Improvement is mandatory, I believe, because there is a growing risk that many research results deduced from path analyses or more general structural equation models are not really taken seriously. There is an emerging credibility problem. This is well exemplified in Freedman's, and other statisticians', disdain for path analysis. Several of us interested in psychometric methods are concerned. To quote from the thoughtful assessment by Cliff (1983):

> Initially, these methods seemed a great boon to social science research, but there is some danger that they may instead become a disaster, a disaster because they seem to encourage one to suspend his normal critical faculties. Somehow the use of one of these computer procedures lends an air of unchallengable sanctity to conclusions that would otherwise be subjected to the most intense scrutiny. These methods have greatly increased the rigor with which one can analyze his correlational data, and they solve many major statistical problems that have plagued this kind of data. However, they solve a much smaller proportion of the interpretational—inferential in the broader sense—problems that go with such data. These interpretational problems are particularly severe in those increasingly common cases where the investigator wishes to make causal interpretations of his analyses.

I agree that the credibility is further strained by an unhealthy insistence on strong causal inferences from the analyses. In my view, these statistical analyses have very little to do with causality. A similar opinion was expressed in de Leeuw (1985). Although the researcher may find that the analyses offer some limited plausibility for his or her causal speculations, in my view these techniques are not devices for rigorous testing of causal theories, but merely a powerful way of analysing covariance structures. It would be very healthy if more researchers abandon thinking of and using terms such as *cause* and *effect*. Instead they should work in terms of regression relations with predictors and outcomes that may lead to parsimonious models that have restrictions on the covariance matrix. The analyses may still have great interpretational value.

There is definitely an unhealthy infatuation with the beautiful statistical framework of path analysis and structural equation modeling with latent variables, for example, as presented in the sophisticated LISREL package. Many inexperienced users wanting to publish "state-of-the-art" analyses are tempted to believe that the complicated computing machinery will take care of model specification and measurement problems. I do not regard this as a problem of the software, but of the user. I find that it is absurd for a person to seriously formulate a path model or any other complex model of a phenomenon without years of prior analyses by him or her or others using more exploratory means. I often find that the number of equations a model has is inversely related to the quality and thoroughness of the model specification. (Hence, I see little need for technical improvements enabling the

estimation of larger systems as is sometimes brought up as desirable.) Researchers often do not know how to limit their models to a manageable marginal part that is still correctly specified given the larger picture.

On Freedman's Critique of Hope (1984), Chapter 1

I think it is unfortunate that our discussion has to be made concrete at the expense of any particular researcher. The Hope example, however, is of general interest in this discussion inasmuch as it is, in my view, quite typical of the style of path analysis in certain areas of social science research. Freedman criticizes the Hope path model comparison between Scotland and the United States because of measurement problems, the reliance on the "autonomy coefficient," nonlinearity, and omitted variables. Freedman's general critique of group comparisons based on standardized coefficients could be added here too.

I would agree with these pieces of criticism, stressing the issues of measurement and omitted variables. As a layman in the area of social stratification, I must say that I would be skeptical about a model that for such detailed interpretations as Hope's uses such limited measurements and is so simple in terms of number of predictors in each of the two relations. Consider, however, the possibility of important omitted variables such as geography, as mentioned by Freedman. We statisticians who can easily call a model naive because it seems to include too few predictors relative to the level of complexity of the phenomenon studied, sometimes tend to overlook that simple models may follow from years of previous model fitting attempts using many other variables. Consider the similar modeling attempts that improve on Blau and Duncan (1967), for example, Sewell, Haller, and Ohlendorf (1970) and Hauser, Tsai, and Sewell (1983), formulating "the Wisconsin model." If I understand these modeling attempts correctly, an array of demographic variables have failed to improve the model and Sewell et al. found a large degree of parameter invariance across subjects from farms, villages, and different sized cities.

Nevertheless, Hope's analysis does not seem convincing to me. I find it particularly worrisome when it comes to Hope's reliance on details such as the "autonomous effect." This is simply the part of the correlation between education and occupation that is due to their common influence by the residual in the education relation, obtained as the product of the regression coefficient of occupation on education and the residual variance in the education equation. Hope furthermore describes this residual as what education itself contributes to occupation over and above IQ and father's occupation. In the Wisconsin model, many more predictors are included in the education and occupation equations: academic performance, level of occupational and educational aspiration, and significant others' influence. If inclusion of such variables is warranted in Hope's case, it would seem that the interpretation of the education residual and the value of the autonomy coefficient could change drastically. It is probably not wise to make so much of such details for which the chances for biases due to omitted variables seems so large.

Given Hope's relatively limited data, particularly in terms of number of variables measured, I think it would have been better to not do path analysis. If regressions were to be attempted, detailed sensitivity analyses should have accompanied, considering the many possibilities for misspecifications of the equations, for example, based on experiences in others' research. In my opinion, this type of analysis could have been done in a speculative way for this data set. They should clearly not be used for conclusive statements, merely as a side interest that would focus on measurements and modeling that, given our interest in such model formulations, could be improved in the next round of data collection.

Teaching for Improvement

The above discussion leads to my main message: We need better methodological education in the social and behavioral sciences. This education involves two important components: technical knowledge of the statistical methods and the ability to translate conceptual theories into statistical models in a credible way.

I believe that the necessary statistical methods and application strategies can be taught to serious social scientists not rigorously trained in statistics. What is needed are solid research methods programs at the graduate level, post-doctoral programs, and training sessions at professional meetings that stress the skills of model building in practice, showing by exemplary social science applications how painfully slow and difficult such a process necessarily is.

In my interactions with social and behavioral science researchers involved in a path analysis or more advanced latent variable structural equation modeling, I have found much less difficulty in conveying the technicalities of how to estimate and test a given model than how to make the transition from conceptual modeling ideas to statistical ones. Social science researchers often have very little understanding of this transition and I think it should be a major focus of an attempt to improve the poor standard of the field.

Conceptually, the social science researcher may feel a strong affinity for a statistical model that includes, say, an intervening variable, or a certain latent variable construct that is measured with multiple indicators. However, the statistical translation may be largely ignored. For instance, in the intervening variable example, the researcher's conceptual model may only concentrate on a certain criterion "being intervened" by a particular predictor and hence may not measure or include other important predictors correlated with the first one. Taking this further, I know of examples where researchers have found that the inclusion of a new set of explanatory variables is important and are excited about a new data set that measures these reasonably well. They then happily go on to write articles on the new

"effects" using models that exclude the already accepted variables, since they were not measured well at all, without discussing the obvious resulting misspecification problem. Using the more complex latent variable example, the researcher may not fully realize that the questionnaire format used or the particular phenomenon intended to be measured causes complications for the latent variable modeling to be carried out in a standard psychometric framework with standard software (such as LISREL). The indicators may be nonlinearly related to the latent variable; they may, by the question format, have certain direct dependencies; or measurement errors may be likely to be correlated with the latent variable and have strongly heteroscedastic variances.

The common problem is that measurement issues and statistical assumptions that are incidental to the researchers' conceptual ideas become stumbling blocks that invalidate the statistical modeling. Again, I think the solution lies in better methodological education. We need to better teach the conceptual-statistical translation. In terms of the latent variable situation, we need to point out—and study further in applied methodological research—how to design multiple measurements to enable our standard latent variable modeling with its specific set of statistical assumptions. Or, if this proves unnatural to the particular measurement application, take on the statistical task of attempting to generalize the statistical model.

I do not believe the above is unrealistic. At UCLA I teach a structural equation modeling class as part of the Graduate School of Education's Research Methods Specialization that covers various applications of covariance structure modeling. The students that I teach have as prerequisite courses basic methodological training in regression analysis, introductory path analysis, and multivariate statistics including elementary matrix algebra and introductory factor analysis. In 10 weeks, I am able to cover the basic elements of the necessary statistical methods and the conceptual-statistical translation and to discuss both poor and exemplary applications. Students who consider serious use of these methods in the future are encouraged to try them out extensively on real data and then come back for more penetrating studies in terms of special seminars. A part of my teaching is the emphasis on self-critique in any papers that are produced. A greater acceptance of self-critique and attempts at sensitivity analyses addressing possibilities of misspecifications are important in avoiding bad applications of one kind mentioned above in the section "On the Quality of Path Analyses," having to publish based on limited, existing data. I believe, although I may eventually find otherwise, that the students leave this program with a healthy respect for statistical modeling and that they will not be among those who contribute to the credibility problem discussed earlier.

General Conclusions

To respond generally to Freedman's last paragraph questioning the value of path analysis, I do think such activities in several instances can clarify aspects of social process. However, I also think avoiding detailed causal inferences and increasing the level of self-critique is much needed. By not refraining from modeling completely in situations with limited theoretical knowledge, I think that I depart from Freedman. The final conclusion is, however, the important one. I think the field can produce interesting and useful studies using path analysis and more general structural equation models, but only when carried out by skillful practitioners. The practitioners need better methodological training and statisticians should contribute to this process.

References

Blau, P., & Duncan, O. D. (1967). *The American occupational structure.* New York: John Wiley and Sons.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115–126.

de Leeuw, J. (1985). Review of books by Long, Everitt, Saris and Stronkhorst. *Psychometrika, 50,* 371–375.

Hauser, R. M., Tsai, S-L., & Sewell, W. H. (1983). A model of stratification with response error in social and psychological variables. *Sociology of Education, 56,* 20–46.

Hope, K. (1984). *As others see us: Schooling and social mobility in Scotland and the United States.* New York: Cambridge University Press.

Mason, W., & Fienberg, S. (Eds). (1985). *Cohort analysis in social research: Beyond the identification problem.* New York: Springer.

Sewell, W. H., Haller, A. O., & Ohlendorf, G. W. (1970). The educational and early occupational status attainment process: Replications and revision. *American Sociological Review, 35,* 1014–1027.

Author

BENGT O. MUTHEN is Associate Professor, Graduate School of Education, University of California, Los Angeles, CA 90024. *Specializations:* measurement modeling, categorical data, latent variables.

Journal of Educational Statistics
Summer 1987, Vol. 12, No. 2, pp. 185–195

Casual Models Do Not Support Scientific Conclusions: A Comment in Support of Freedman

David Rogosa
Stanford University

Overview

A critical distinction in methodological work is between (a) building (and applying) statistical models for the processes that generate the social science data and (b) tossing the data at available statistical methods. In my own work I strive for (a) and discourage others from settling for (b). Regrettably, expositions and applications of the popular causal modeling methods (under the various names path analysis, structural equation models, LISREL, etc.) contain much of (b) and little of (a). In fact, my favorite typographical error "casual models" (which I've suffered in print) is enjoyable in large part because of its accidental accuracy. And an argument can be made that the methodological proselytizing for and dominance of causal models has retarded the much more useful methodological work of (a).

A similar theme is present throughout Freedman's paper, as in the last paragraph of his conclusion which begins "My opinion is that investigators need to think more about the underlying social processes . . . ". Earlier in the paper Freedman requires that the "as-if-by-experiment" conclusions "must depend on a theory of how the data came to be generated." The translation of substantive theory into methods for data collection and analysis is where I think the fertile interaction between statisticians and social scientists lies (rather than in arguing a "thumbs up" or "thumbs down" on path analysis).

My subtitle "in support of Freedman" is to congratulate him for his energy and courage in assuming the role of point person in what I feel is an attempt to stimulate serious and critical discussion of the proper role of these causal models in behavioral and social science. Freedman is not the first to voice serious concerns, nor should he be the last. For example, de Leeuw's (1985) review essay of causal model texts does a good job of discussing the casual attention given to model construction and the indefensibility of "cause-effect" (i.e., as-if-by-experiment) conclusions:

> It seems to me that the use of cause-effect terminology cannot be defended, except in those rare cases (such as Mendelian genetics) in which information is available from other sources. If all the information we have