To Path Analyze or Not To Path Analyze: Is There an Alternative Approach

Isadore Newman, The University of Akron
and
Joseph R. Marth, Bluefield College

During the past twenty years there has been a tremendous increase in the frequency of social scientists attempting to investigate phenomena that can not be studied in a laboratory. Since the ideal is to be able to explain complicated relationships in the causal sense, these social scientists have been highly attracted to sophisticated multivariate causal modeling.

Much has been written on the problems of modeling techniques such as path analysis. The concept that any research based upon ex post facto design can not assume causation (post hoc fallacy), that is correlation does not imply causation, has been widely accepted. However, some social scientists are more frequently wondering why not accept causal modeling assumptions? Do the advantages out-weigh the disadvantages? Are the concerns voiced by many statisticians really nitpicking (Cliff, 1983; Daggett & Freedman, 1985; Freedman, 1989; Huber, 1985; Kenny, 1979)?

Purpose

The purpose of this paper is to examine the underlying assumptions of path analysis and to discuss some theoretical concerns. This paper will also suggest an alternative approach that the authors believe to be more robust to the violation of some of the underlying assumptions and still is very effective in testing the overall "goodness of fit" of a theory.

Before beginning, however, a caveat is necessary. There are a number of uses for which researchers employ path analytic procedures that this paper does not deal with. For example, we are not dealing with situations where researchers use path analysis analogous to almost a stepwise model building in which the computer identifies the best fitting models. From a theoretical point of view, this has virtually all of the problems (and maybe even more) of a stepwise regression procedure, and has received much criticism because of its antitheoretical and unstable nature. This paper also does not discuss the use of path analysis for the purpose of determining which alternative models are better. Rather, discussion here is focused on the traditional intent and most conservative approach of path analysis, that of theory testing and model confirmation.

The Assumptions

The underlying premise of path analysis is that if one can meet all of the assumptions, it is justifiable to presume "causation." Therefore, this paper begins with a discussion of these assumptions. The following is a summary of the basic assumptions of path analysis identified by Bollen (1989), Freedman (1987), Dillon and Goldstein (1984), Kenny (1979), and Williams (1978):

1. Requires a theory and nomological net;
2. There is significant relationship between the variable that is assumed to be the cause and the variable that is assumed to be the effect;
3. Causal variable precedes the effect variable in time;
4. Spuriousness has been controlled...all meaningful relationships are included in the model;
5. Variables are additive and no interaction exists;
6. The weights are stable (paths), therefore no multicollinearity;
7. The distribution of residuals are the same no matter what the value of the independent variable;
8. The mean of the residual values is zero;
9. The variance of the residual values is finite;
10. The residuals of each of the variables are independent of all the other variables in the system; and
11. Endogenous variables have at least interval scale properties.

An added concern is that totally different path analytic models can produce a sufficient amount of statistical verification to justify a variety of theoretical explanations for the same variables. Also, there are concerns about the use of latent variables (Cliff, 1983), similar to the concerns of virtually all factor analytic procedures. That is, concern that latent traits, when used, are stable, meaningful, interpretable, and valid. Finally, we should further note that little is known about the effects of heteroscedasticity or autocorrelated disturbances for latent variables (Bollen, 1989).

A discussion of some of the these crucial assumptions and related concerns is presented below, followed by an alternate approach to path analysis should the researcher be unable to meet the assumptions.

The Need for Theory

In path analysis and structural equation modeling (SEM), one builds analytic diagrams that are reflective of the nomological net exposited by the theory it is intended to reflect. Therefore, one of the key underlying assumptions before doing any path analysis or SEM, traditionally, has been the necessity of theory (Bollen, 1989; Borgatta, 1969; Duncan, 1975, 1969; Heise, 1974, 1975, 1977; Williams, 1978). The purpose of theory is to explain and help understand the occurrence of natural phenomena (Kerlinger, 1973). Theory explains the causal effects among and between variables (constructs). Further, since one of the original purposes of path analysis and SEM is to assume "causal" relationships between variables which are frequently, if not always, nonmanipulable (Newman & Newman, 1992; Kerlinger, 1973), one is required to assume causation from correlational-type data. However, this does not mean you can not use path analytic procedures on experimental data. Thusly, theory is an essential component to this process. If one assumes causation which is consistent with a nomological net, one is standing on firmer ground than if one were assuming causation merely because phenomena were correlated.

Happily, when reading research which uses path analysis, there tends to be a much greater explanation of theory and the derivation of its hypotheses, and we strongly support such approaches. This is more likely to require the researcher to know the literature, to know the theory, and to think about the possible logical interrelationships of the variables.

It should also be noted that in the use of path analysis for testing theory, there are goodness of fit indices to help estimate how well the model fits the theoretically predicted relationships. Chi square and the absolute size of the residuals were initially the most frequently used goodness of fit indices. Bentler and Bonnett (1980) and Tanaka and Huba (1985), have developed goodness of fit indices, indicating that they are robust to N size. However, an article by Marsh, Balla, and McDonald (1988) mathematically demonstrates that all of the indices are really dependent to differing degrees on N size.

Time-precedence and Non-spuriousness

Kenny (1979) identified two requirements for path analysis: time precedence and non-spuriousness. These requirements tend to be design concerns in which time precedence indicates that the independent variable, which is the presumed cause of the dependent variable (endogenous variable), logically has to precede the dependent variable. For example, in a causal sense, one would expect IQ to logically precede GPA, but GPA would be less likely to logically precede IQ. Non-spuriousness can be thought of as an underlying assumption of the path analysis design, in that it assumes that the path analytic model contains all of the relevant causal variables.

Interaction

An intriguing aspect for and against the use of path analysis is that, with very few exceptions, little has been said about the issue of interaction. The underlying regression structures of path analysis are analysis of covariance regression models. One of the most important assumptions of analysis of covariance, which can not be violated with impunity, is that there is no significant interaction between the independent variable and the covariates. This means that anyone testing a simple or complex path analytic model which represents a nomological net, is making the assumption, consciously or unconsciously, that there is no interaction. One merely has to think of the social science theories and ask how many of them make that assumption.

In situations where interaction is found, for example between sex and motivation in predicting achievement, one suggested procedure for handling such interactions would have the researcher run separate analyses for males and females. It is likely that a complex path analytic design will have more than one simple first-order interaction. Actually, one would probably expect more than one second-order interaction (which is an interaction between at least two first-order interactions) or third-order interaction (which is an interaction between at least two second-order interactions) to exist in a complex path analytic design. The implications of these interactions for interpretation of path analysis is that researchers will have to consider many subset designs which can become so conditional that they become complex beyond understanding.

For non-linear second-order types of relationships a similar solution has been suggested: that a two-stage least square procedure be incorporated. However, it is interesting to ask individuals who are using path analysis to test a theory if they are in fact assuming that there is neither interaction nor a curvilinear relationship.

To the extent that the path analytic models do not reflect interactions that exist in the theoretical conceptualization, the researcher is actually committing a Type VI Error. That is, there is an inconsistency between the research question of interest and the statistical model which was written to reflect the research question (Newman, Deitchman, Burkholder, Sanders, & Ervin, 1976).

Beta Weight Interpretation

It has been well established in the statistical literature that beta weights are either non-interpretable (Kerlinger & Pedhazer, 1973; McNeil, 1993, 1992, 1991; Ward & Jennings, 1973) or are misleading and should be interpreted with extreme caution. Beta weights are more likely to be interpreted correctly if there is zero multicolinearity between the independent variables. The higher the correlation between these variables, holding everything else constant, the higher the standard deviation and the greater the instability of the weights.

The causal interpretation of a path analytic model needs predictor variables that are low or zero correlated and/or sample sizes that are very large in relation to the number of variables. If the sample sizes are so large, such as the High School and Beyond data set with 58,000 subjects, they can be considered virtual populations. That is, the more subjects per variable, the more stable these weights tend to be. Unfortunately however, when the sample size is very large, traditional tests of significance become virtually meaningless, because any slight difference will be statistically significant. (The proportion of variance accounted for can be considered or the model can be used in a more descriptive manner.)

Some approaches have dealt with the multicolinearity problem by employing measurement models along with statistical models. The measurement model uses a set of indicator variables that are conceptually factor analyzed. These factors, sometimes called latent traits, are assumed to be better measures of the underlying construct than any individual item. These underlying traits are often assumed to be stable or at least more stable than the individual items they are composed of, and therefore are thought to be more reliable and valid. However, one must also keep in mind that these factors are sample specific and may be in turn highly unstable.

Some path analytic users think that using latent traits (factors) decreases or eliminates the multicolinearity problem and reduces measurement error. This is not necessarily the case. For example, if five indicator variables for achievement are factor analyzed and five for ability level, and the ten indicator variables are not factor analyzed together, each set of five items can produce factor solutions that are highly correlated (multicolinear). In addition, five indicator variables may produce three factors when factor analyzed but only the first factor is usually used because this approach assumes the other factors are not meaningful or useful. There may be no justification for such an assumption. Another approach sometimes employed is to only use the first non-rotated factor which maximizes the variance accounted for by that one factor, but also tends to disregard the empirically identified multidimensionality on the construct.

If the sample is virtually a population size or is a population, then the model, even if not causal, can definitely be used descriptively to help explain potential relationships without ever assuming causal effect. There appears to be much less criticism of such an application of path analysis, but there is less interest in using it in this way. For example, while major economic forecasting models that have used path analysis have not held up well (McNees, 1986; Zarnowitz, 1979), they have been found to be useful in a more descriptive sense.

Testing for Underlying Statistical Assumptions

Applied statisticians and sophisticated users of path analysis such as Bollen (1989), Bentler (1987), and Freedman (1985) have pretty much agreed that one should test for certain underlying assumptions and do a pre-analysis of the data related to these assumptions before path-analysis or any statistical treatments are used. Berkane and Bentler (1987) state that BMDP provides a test for multivariate normality, detecting or eliminating outliers for EQS, and Berkane and Bentler (1987) developed a test for homogeneity of kurtosis. In addition, before doing any analysis, one should look at plots of residuals and should always cross validate to establish the stability of the prediction from sample to sample.

Some underlying assumptions are more robust than others. For example, certain assumptions of normality and homogeneity can be violated with virtual impunity if the N is large enough. However, certain assumptions of linearity, no interaction between the independent variables, and no multicolinearity are assumptions to which covariant structural models are highly sensitive (not robust). The question is, how frequently does the literature report the use of these procedures to check underlying assumptions, and why not make this a requirement of the data analysis for publication?

Corrections for Violations of Assumptions

Bollen (1989) and others (Bentler, 1987; Bentler & Dijkstra, 1985; Bentler & Lee, 1983; Freedman, 1985; Johnson, 1984; Joreskog & Sorbon, 1981; Tukey, 1954) have dealt with violations to the assumptions and have suggested solutions. For example, the use of alternate estimators such as General Least Squares (GLS), Unweighted Least Squares (ULS), Elliptical Generalized Least Squares (EGLS), Two-Stage Least Squares (2SLS), Three-Stage Least Squares (3SLS), Instrumental-Variable Estimators (IVE), and Full-Information Maximum Likelihood (FIML) are discussed by Bollen (1989). However, these techniques themselves tend to have assumptions about what the data truly look like in the population. If the researcher is correct about the nature of the distribution of data in the population and s/he picks a statistical procedure that is most appropriate for that distribution, it is obvious that his/her analysis is most likely to produce the most accurate parameter estimates. Unfortunately, however, the researcher frequently does not know what the data look like in the population and/or is unaware of what is "causing" abnormalities in the distribution. Further, while a statistical technique may allow one to correct for anomalies, the researcher must make the assumption that the anomalies are in fact errors. Otherwise, the very corrections themselves create greater errors than no correction at all. What we are arguing is that statistical corrections for anomalies in the distribution, without considering the causes of the anomalies, is a fatal flaw in the research study. Therefore, one has to be aware of the assumptions one is making about the anomalies when one is making a correction. There is no correction which is a panacea that will replace understanding one's data.

A Simple Alternative Approach to Path Analysis for Testing Theoretical Relationships

The following is a suggested approach that is methodologically much simpler and is more robust to some of the devastating assumptions such as linearity and no interaction that are underlying assumptions of path analysis, and yet has many of the same advantages for testing a nomological net. This approach starts with theory that produces a nomological net, then identifies the logically derived hypothesis to be tested. For example, let's assume that 15 hypotheses are produced from the nomological net. Some can be interactional, repeated measures, time lagged, multiple wave, curvilinear, main effects or direct effects. Let's further assume that 13 of the hypotheses are significant in the predicted direction. One can then get an estimate, by using a Sign test, of how well these hypotheses support the overall theory (nomological net). Depending upon one's productivity and situation specifics, one may choose to do a Sign test on the directions of each individual hypothesis with no concerns for the tests of significance. Or, one can do the Sign test only on the number of significant hypotheses and compare it to the total number of hypotheses. In either case, this nonparametric test can be used to estimate the overall support of the theory. In addition, this test of significance is not dependent upon the N size, but rather on the number of hypotheses generated. It is apparent how this approach can fit well into a meta-analysis. As Pedhazer (1990) and Ward and Jennings (1973) suggest, researchers should keep their analysis simple but well thought out and have hypotheses that are derived from previous research and theory.

The authors believe much path analysis research gets lost in the complexity of the models and the sophistication of the analyses. In cases where more sophisticated analysis may be required, based upon the theory and the derived hypotheses which may infer, for example, underlying latent structures, the suggested approach would be to do:

1. A factor analysis of the variables of interest;
2. A cross validation of the factor structures to estimate stability;
3. A factor regression using the factors as predictor and criterion variables where appropriate; and
4. Cross validation on the regression equations to estimate their stability.

Needless to say, before doing any type of analysis, it is always desirable to first look at your means, standard deviations, frequencies, correlations, and residual plots before proceeding. It is this pre-analysis that helps to identify potential errors in the data, to what degree underlying assumptions have been violated, and if and what data transformations are needed or desirable. We think it is appropriate to end with a quote from Rogosa (1987): "[t]he transition of substantive theory into methods for data collections and analysis is where I think the fertile interaction between statistician and social scientist lies[,] rather than in arguing 'thumbs up' or 'thumbs down' on path analysis" (p. 185).

References

Bentler, P. M. (1987). Structural modeling and the scientific method: Comments on Freedman's critique. *Journal of Educational Statistics, 12*(2), 151-157.

Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaish (Ed.), *Multivariate analysis VI,* Amsterdam: North Holland.

Bentler, P. M., & Lee, S. Y. (1983). Covariance structures under polynomial constraints: Application to correlation and alpha-type structural models. *Journal of Educational Statistics,* 8, 207-222.

Berkane, M., & Bentler, P. M. (1987). Distributions of kurtosis, with estimators and tests of homogeneity of kurtosis. *Statistics & Probability Letters,* 5, 201-207.

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

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley & Sons.

Borgatta, E. F. (Ed.). (1969). *Sociological Methodology 1969,* San Francisco: Jossey-Bass.

Cliff, N. (1987). Comments on Professor Freedman's paper. *Journal of Educational Statistics, 12*(2), 161-164.

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

Crano, W. D., & Mellon, P. M. (1978). Causal influence of teachers' expectations on children's academic performance: A cross-lagged panel analysis. *Journal of Educational Psychology, 70*, 39-49.

Daggett, R., & Freedman D. (1985). Econometrics and the law: A case study in the proof of antitrust damages. In L. LeCam & R. Olshen (Eds.), *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer* (Vol. I, pp. 126-175). Belmont, CA: Wadsworth.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: John Wiley & Sons.

Duncan, O. D. (1975). *Introduction to structural equations models*. New York: Academic Press.

Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, 72, 177-182.

Duncan, O. D., Featherman, D. L., & Duncan, B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.

Freedman, D. (1981). Some pitfalls in large econometric models: A case study. *Journal of Business, 54*, 479-500.

Freedman, D. (1985). Statistics and the scientific method. In W. Mason & S. Fienberg (Eds.), *Cohort analysis in social science research: Beyond the identification problem* (pp. 345-390). New York: Springer.

Fox, J. (1987). Statistical models for nonexperimental data: A comment on Freedman. *Journal of Educational Statistics*, 12(2), 161-164.

Heise, D. R. (Ed.). (1974). *Sociological methodology 1975*. San Francisco: Jossey-Bass.

Heise, D. R. (1975a). *Causal analysis*. New York: Wiley.

Heise, D. R. (Ed.). (1975b). *Sociological methodology 1976*. San Francisco: Jossey-Bass.

Heise, D. R. (Ed.). (1977). *Sociological methodology 1977*. San Francisco: Jossey-Bass.

Huber, P. (1985). Projection pursuit. *Annals of Statistics, 13*, 435-475.

Johnson J., (1984). *Econometric methods*. New York: McGraw Hill.

Joreskog, K. G., & Sorbom, D. (1981). *LISREL - Analysis of linear structural relationships by the method of maximum likelihood.* Chicago: International Educational Services.

Kenny, D. (1979). *Correlation and causation*. New York: Wiley.

Kerlinger, F. N. (1973). *Foundations of behavioral research (2nd edition)*. New York: Holt, Rinehart, and Winston.

Marsh, H., Balla J., & McDonald R. P. (1988). Goodness of fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

McNees, S. (1986). Forecasting accuracy of alternate techniques: A comparison of U.S. macroeconomic forecasts. *Journal of Business and Economic Statistics, 4*, 5-24.

McNeil, K. (1991). *Reasons for not interpreting regression coefficients.* Paper presented at the meeting of the Southwest Education Research Association, San Antonio, Texas.

McNeil, K. (1992). Conditions for interpretation of regression weights. *Multiple Linear Regression Viewpoints, 19*(1), 37-43.

McNeil, K. (1993). Cautions and conditions for interpreting weighting coefficients, *Midwestern Educational Researcher, 6*(1), 11-14.

Newman I., Deitchman R., Burkholder J., Sanders R., & Ervin L. (1976). Type VI Error: Inconsistencies between the statistical procedure and the research question. *Multiple Linear Regression Viewpoints, 6*(4), 1-19.

Newman, I., & Newman, C. (1992). *Conceptual statistics for beginners* (2nd edition). Lanham, MD: University Press of America.

Noonan R., & Wold, H. (1983). Evaluating school systems using partial least squares. *Evaluation in Education, 7*, 219-364.

Pedhazer E. J. (1982). *Multiple regression in behavioral research.* New York: Holt, Rinehart, and Winston.

Rogosa, D. R. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press.

Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88*, 245-258.

Rogosa, D. R. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12(2)*, 185-195.

Simon, H. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association, 49*, 467-479.

Tanka, T. S., & Huba, G. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 38*, 197-201.

Tukey, J. W. (1954). Causation, regression, and path analysis. In O. Kempthorne, T.A. Bancroft, J. W. Gowen, & J. L. Lush (Eds.), *Statistics and mathematics in biology.* Ames: Iowa State College Press.

Ward J. H., & Jennings, E. (1973). *Introduction to linear models* Inglewood Cliffs, NJ: Prentice Hall

Williams, J. D. (1978). Path analysis from a regression perspective [Monograph]. *Multiple Linear Regression Viewpoints*, 9(2), 1-81.

Wold, H. (1985). Partial least squares. In J. Kotz & N. L. Johnson (Eds.), *Encyclopedia of social sciences* (Vol. 6). New York: Wiley.

Wold, H. (1987). Response to D. A. Freedman. *Journal of Educational Statistics, 12(2)*, 202-204.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557-585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161-215.

Wright, S. (1960). Path coefficients and path regressions. *Biometrics, 16*, 189-202.

Zarnowitz, V. (1979). An analysis of annual and multiperiod quarterly forecasts of aggregated income, output, and price level. *Journal of Business, 52*, 1-34.