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
is a set of correlations, then we can try to describe these correlations ‘parsimoniously’ in terms of restrictive models. But that is about all we can do. There is no LISREL method of theory construction. (p. 372)

Others have expressed distress much more bluntly: for example Ling’s (1982) review of David Kenny’s book Correlation and Causation. Rather than cheering for Freedman’s specific technical critique of Hope, I’ll use my space to focus on two general themes: (i) the importance of modeling the process that generates the data and (ii) the formidable difficulties in making as-if-by-experiment conclusions from nonexperimental data.

My own view and (limited) historical perspective does not see causal models as inherently a negative force. The introduction of path analysis into sociology, psychology, and education 15-20 years ago did have the very positive effect of stimulating the translation (some might say transmogrification) of often vague verbal statements of theory into forms suitable for empirical investigation. The primary hurdle for this process was the construction of the path analysis diagram: simply multiple regression with pictures, where the pictures are presumed to be meaningful. Statements of theory in the form of path analysis diagrams do provide a statistical model for the empirical research. Alas, a statistical model is not necessarily a scientific model; statistical models may have little substantive meaning or interpretability even when their technical assumptions are satisfied by the data. In the spirit of the path analysis discussion I use the following picture:

| no model | statistical model | scientific model |

Freedman makes a related distinction between statistical models of associations and the “structural equations” that would support as-if-by-experiment conclusions. My informal use of “scientific model” is a bit broader and is intended to describe serious representations of the substantive (psychological, sociological, etc.) processes that generate the data but are not always intended to support explanatory or as-if-by-experiment conclusions. Building scientific models for the social sciences is very hard work and requires orders of magnitude more thought, preliminary empirical research, careful data analysis, and creativity in statistical modeling than is now evident. My own view is that, at best, the path analysis (causal) models may be useful until a field acquires some insight into what’s going on and then moves to appropriate models and methods.

Another relevant distinction is that between illustrations of a statistical method (even using actual data) and applications of the statistical model and method. I first came to appreciate this distinction in the context of causal models when I was asked to review the volume of collected papers Advances in Factor Analysis and Structural Equation Models (Jöreskog & Sorbom, 1979). I found the volume loaded with detailed illustrations of LISREL and covariance structure analyses, but found little that went beyond the illustrations to reveal something important about psychological processes. At the same time I realized that I wasn’t aware of any covariance structure analyses that I felt were serious applications. Since then I’ve come to feel that models for relations among variables are fundamentally off the mark, and almost preclude successful applications. I come back to this idea in my closing section.

My own personal history includes some interest/skepticism with causal models, including a chapter (Rogosa, 1979) in the Baltes-Nesselroade volume on longitudinal research. My technical work over the last 7 years on statistical methods for longitudinal research serves as a useful context for explaining in some detail my disillusionment with causal models. Besides being an area I know well, longitudinal panel data are commonly analyzed using path analysis and covariance structure methodological, and longitudinal data often serve as the basis for expositions of these methods.

Causal Models and Longitudinal Data Analysis

The first part of this section attempts to illustrate, in the context of longitudinal methodology, some implications of using statistical models to represent the process generating the data. The basic tools I use are statistical models for individual growth and individual differences in growth. Then I use the growth curve representation to show the unattractiveness of three related causal modeling approaches to the analysis of longitudinal data: path analysis, latent-variable (LISREL) regressions, and simplex models. These three procedures all use the between-wave covariance matrix as the starting point for the statistical analysis. My main message is that the between-wave covariance matrix provides little information about change or growth. Thus, regardless of the sophistication of the modeling of the relations between manifest or latent variables, the causal model analysis is fatally flawed. Finally, I discuss the popular method of cross-lagged correlation which has an even weaker justification than path analysis yet attempts to make as-if-by-experiment inferences from nonexperimental, longitudinal data.

Statistical Models for Longitudinal Panel Data

The basic model for longitudinal data is a two-part representation using a pair of statistical models: (a) a model for individual growth and (b) a model for how the parameters of the individual growth curves vary over individuals. The individual growth curves are functions of true score over time, \( \xi(t) \). Research questions about growth, development, learning, and the like center upon the systematic change in an attribute over time, and thus the individual growth curves are the natural foundation for modeling the longitudinal data. The simplest example is straight-line growth, which specifies a constant rate of change denoted by \( b \). The straight-line growth curve for individual \( p \) is written:

\[
\xi(t) = \xi(0) + bt. \tag{1}
\]
Individual differences in growth exist when the individual growth curves have different values of rate of change \( \dot{y} \). Systematic individual differences in growth exist when individual differences in a parameter of the individual growth curve (\( \dot{y} \)) can be linked with exogenous individual characteristics. The constant rate of change model in (1) is used in this discussion for simplicity in the exposition and mathematical results. Moreover, straight-line growth often serves as a useful approximation to more complex growth trajectories; for example, if the growth curve is a second-order polynomial \( Y \) has the interpretation of an "average rate of change" (Hui & Berger, 1983; Seigel, 1975). Rogosa and Willett (1985a) consider a variety of more complex models for individual growth and individual differences in growth.

The most prevalent type of longitudinal data in the behavioral and social sciences is longitudinal panel data. Longitudinal panel data consist of observations on many individual cases (persons) on relatively few (2 or more) occasions (waves) of observation. An observation on a variable \( X \) at time \( t \) for individual \( p \) is written as \( X_{pt} \) where \( i = 1, \ldots, T \) and \( p = 1, \ldots, n \). The \( X_{pt} \) are presumed to be composed of a true score \( \xi_{pt} \) and an error of measurement \( \epsilon_{pt} \). The value of the growth curve (Equation 1) at a discrete time \( t \) yields the \( \xi_{pt} \), and the \( X_{pt} \) are formed by the addition of measurement error according to the classical test theory model: \( X_{pt} = \xi_{pt} + \epsilon_{pt} \).

Path Regressions

Path analysis models for longitudinal data use the temporal ordering of the measurements to delimit the possible paths between the variables. Consider the example of a three-wave design with measures on \( X \) at times \( t_1, t_2, t_3 \). The path regressions for the unstandardized variables are:

\[
\begin{align*}
X_2 &= \alpha_0 + \beta_0 X_1 + \epsilon_2 \\
X_3 &= \alpha_3 + \beta_0 X_2 + \beta_1 X_1 + \epsilon_3.
\end{align*}
\]

(2)

Thus the path analysis model includes direct paths from \( X_t \) to \( X_s \) and to \( X_r \) (parameters \( \beta_0 \) and \( \beta_1 \), respectively) and from \( X_r \) to \( X_s \) (parameter \( \beta_2 \)). The path coefficients are functions of the entries of the between-wave covariance matrix. An example of the use of this model is Goldstein's (1979) in which \( X \) is a reading test score obtained on a nationwide British sample with measurements of ages 7, 11, and 16. This simple 3-wave path model was also discussed in a number of the early expositions of path analysis in the social sciences. However, the comparison below between the path analysis and the mathematical results for straight-line growth illustrates some of the perils of summarizing the longitudinal data by the analysis of the between-wave covariance matrix of the \( X_t \) or even the \( \xi_{pt} \), thereby ignoring the analysis of individual growth.

The properties of the path coefficients in (2) illustrate the meaningfulness of models for relations among variables. To take the simplest situation let the true scores \( \xi_{pt} \) (\( i = 1, 2, 3 \)) be determined by a straight-line growth curve for each individual (c.f. Equation 1) and assume perfect measurement of the \( X_t \). For this specification the population partial regression (path) coefficients from (2) are:

\[
\begin{align*}
\beta_0 &= \frac{\bar{X}_2 - \bar{Y}}{\bar{X}_1 - \bar{Y}} < 0 \\
\beta_1 &= \frac{\bar{X}_3 - \bar{Y}}{\bar{X}_1 - \bar{Y}} > 0.
\end{align*}
\]

(3)

[Causal Models]

[Note: the results in (3) are easily verified by using (1) to substitute \( X_t = X_1 + \xi_{pt}(t-1) \) into the \( X_t \) equation in (2). Substituting the values for \( \beta_0 \) and \( \beta_1 \) from (3) and collecting terms yields \( X_t = X_1 + \xi_{pt}(t-1) \).]

The implications of (3) for the path analysis model in (2) are devastating. Remarkably, the parameters depend only on the times at which the observations were taken; thus neither the path regression coefficient contains any information about growth! One might think that because \( X_t \) is perfectly predicted from \( X_1 \) and \( X_2 \) the analysis of relations among variables would be informative. Yet, under this simple structure estimates of either parameter are totally independent of the information in the data. Exclusion of measurement error does not mitigate the import of this example; after all, a statistical procedure that works poorly with perfect measurement can hardly be expected to perform better with fulfillable measurement. (Different forms of the individual growth curve will alter somewhat the results for the coefficients in (2); exponential growth to an asymptote for each individual does produce results similar to Equation 3.)

For estimation of (2) using the reading test data, Goldstein obtains the following estimates: \( \beta_0 = 1.11, \beta_1 = -1.47 \) (and also \( \beta_2 = .841 \)). Goldstein's analysis employs complex transformations of the measures to straighten the between-wave scatterplots of the \( X_t \) prior to the estimation of (2) and also uses discontinuation procedures in an attempt to remove the effects of measurement errors of the sample regression coefficients. The results in (3) agree with Goldstein's negative value of \( \beta_0 \), with the magnitude likely affected by deviations from straight-line individual growth curves in (1), the data transformations and the errors of measurement in the test scores. Also the results in (3) are consistent with Goldstein's positive value for \( \beta_1 \). The negative estimate for \( \beta_0 \) causes considerable discomfort, summarized by Goldstein (1979, p. 139):

This is difficult to interpret and may indicate that non-linear or interaction terms should be included in the model, or perhaps that the change in reading growth between seven and 11 years is more important than the seven-year score itself. However, the addition of non-linear terms does not change this picture to any extent.

[Latent Variable (LISREL) Regression Models]

Latent variable regression models are a more sophisticated, but equally...
flawed, approach to the analysis of longitudinal data. These structural equation models incorporate regression relations among latent variables (i.e., Z(t)) with measurement models relating the observed indicators (X) to the latent variables. Estimation of these models is based on fitting the covariance structure implied by the structural equation model to the between-wave covariance matrix of the observations. Consider the simple structural regression model with one latent variable, ξ, observed at times t₁ and t₂ and a latent exogenous measure, W, illustrated in Figure 1. Each latent variable has two indicators. This model is equivalent to the model for change in alienation that appears frequently as an example in Jöreskog’s papers. In Jöreskog’s examples, ξ is alienation and W is socioeconomic status. The path from W to ξ₂ represents the exogenous influence on change. The structural parameter for that path is the regression coefficient for the latent variable at time 2 on the exogenous variable, with the latent variable at time 1 partialled out, β_{W2,W1}. In terms of the simple growth model in Equation 1, parameters of interest for the relationship between the exogenous variable and change are the correlation between true rate of change and the exogenous variable, ρ_w, or the analogous regression parameter β_w. What does the regression parameter β_{W2,W1} reveal about exogenous influences on growth? Not very much. For the simple case of a collection of straightforward growth curves, this structural parameter has a complicated functional form that depends strongly upon the time chosen for the initial measurement. Rogosa and Willett (1985a, Section 3.2.2) give mathematical results for the form of the structural regression parameter. For a specified relation between the exogenous variable and the individual growth parameter θ, the structural parameter may be positive, negative, or zero depending upon the choice of time of initial status. Also, the structural parameter increases with the length of the interval between measurements. Numerical examples of the bizarre properties of the structural regression parameter are given in Rogosa (in press).

FIGURE 1. Two-wave structural regression model with exogenous variable

Simplex Models

A third example of longitudinal analyses based on the between-wave covariance matrix is the simplex model, which specifies a first-order autoregressive process for true-scores. The numerical example of Rogosa and Willett (1985b) in this journal seeks to caution against the propensity to base many analyses of longitudinal data on a simplex structure without careful consideration of the longitudinal data or of alternative growth models. Expositions of covariance structure analyses have encouraged such thinking; for example, Jöreskog states “For one measure administered repeatedly to the same group of people, an appropriate model is a simplex model” (Jöreskog, 1979, p. 54). Moreover, Werts, Linn, and Jöreskog (1977, p. 745) assert: “The simplex model appears to be particularly appropriate for studies of academic growth.”

Rogosa and Willet (1985b) present an example of a 5 × 5 covariance matrix for observed scores X_v over five occasions of observation. To the eye, the correlation matrix corresponds extremely well to a simplex. Our analysis shows that a simplex covariance structure marvelously fits this covariance matrix although it was generated by growth curves that maximally violate the assumptions of the simplex growth model. The consequences are far from benign because even when the simplex model fits wonderfully, the results of the covariance structure analyses can badly mislead. The covariance structure analyses usually go on to compute growth statistics and reliability estimates based on the simplex model, and these growth statistics (such as the correlation between true change and true initial status) estimated from the LISREL analysis can differ markedly from the actual values. Covariance structure analyses provide very limited information about growth in the sense that covariance matrices arising from very different collections of growth curves can be indistinguishable. Therefore, analyses of covariance structures cannot support conclusions about growth. Analysis of the collection of growth curves cannot be ignored.

Cross-Lagged Correlation

It seems appropriate to make the point that social scientists frequently have been attracted to methods for the analysis of nonexperimental data that are far more flawed and less justified than path analysis and relatives. A most vivid example is provided by the method of cross-lagged correlation, which remains a very popular procedure for the analysis of reciprocal effects from nonexperimental, longitudinal data. Cross-lagged correlation purports to answer the question—Does Y cause X or does X cause Y?—by a simple comparison of the lagged correlations between X and Y (i.e., the correlations between X₁ and Y₁ and Y₁ and X₂ for two time points). A remarkable attribution of as-if-by-experiment is provided by Crano and Melton (1978): “With the introduction of the cross-lagged panel correlation method, . . . causal inferences based on correlational data obtained in longitudinal studies can be made and enjoy the same logical status as those accompanied by Causation. Chapter III: SOURCE: Rosenberg (1971).
derived in the more standard experimental settings” (p. 41). In other words, the use of cross-lagged correlation dispenses with the need for experiments, statistical models, or careful data analysis; a quick comparison of a few correlation coefficients is all that’s required to study reciprocal effects.

Rogosa (1980) was only one in a tradition of papers, starting with Duncan (1969), Goldberger (1971), and Heise (1970), sharply critical of cross-lagged correlation. Even Cook and Campbell (1979, Chap. 7) are unenthusiastic about the usefulness of cross-lagged correlation; yet most advocates and users of this procedure remain undaunted. Rogosa (1980) exposes a number of simple statistical models for reciprocal effects between two variables—path analysis models, continuous-time feedback models, and multiple time series models. The mathematical results in Rogosa (1980) demonstrate the inability of the method of cross-lagged correlation to recover the structure of the reciprocal effects specified by these models. Results and numerical examples are presented for two-wave and multiwave data. Rogosa (1985) provides a non-technical overview and extensive references on approaches to the analysis of reciprocal effects.

The mathematical and numerical demonstrations of the failures of cross-lagged correlation in Rogosa (1980) had the following simple, limited structure. Start with a basic path-analysis regression model for two variables \( X \) and \( Y \) measured at times 1 and 2 (the popular two-wave, two-variable panel design)

\[
\begin{align*}
X_2 &= \beta_0 + \beta_1 X_1 + \gamma_1 Y_1 + \nu_1 \\
Y_2 &= \gamma_0 + \beta_2 X_1 + \gamma_1 Y_1 + \nu_2
\end{align*}
\]

(4)

In the context of the statistical model in (4) the parameters \( \beta_i \), \( \gamma_i \) represent the influence of a variable on itself over time. The parameters \( \beta_2 \) and \( \gamma_1 \) represent the lagged, reciprocal effects between \( X \) and \( Y \); thus the relative magnitudes of \( \beta_2 \) and \( \gamma_1 \) are presumed to indicate the nature of the reciprocal causal effects. In Rogosa (1980) combinations of \( \beta_2 \) and \( \gamma_1 \) values are compared with the results of the method of cross-lagged correlation. The major (and perhaps only) virtue of the path analysis model (4) is the identifiability of specific parameters believed to represent the reciprocal effects. If this model of the reciprocal influences between \( X \) and \( Y \) were valid, then estimation of \( \beta_2 \) and \( \gamma_1 \) would inform about reciprocal effects. Perhaps the best way to think about (4) and the related structural regression models is that these comprise a simple statistical model for reciprocal effects that, however, may be a far-from-satisfactory scientific model of the psychological (etc.) process.

The real moral about the analysis of reciprocal effects is that you can’t estimate something without first defining it, and statistical models at least allow definition of key parameters. Regrettably, the seductive simplicity of cross-lagged correlation has inhibited serious work on the complex question of reciprocal effects. Despite the complexity of research questions about reciprocal effects, empirical research has attempted to answer the oversimplified question, Does \( X \) cause \( Y \) or does \( Y \) cause \( X \)? by casually comparing a couple of correlations.

Model Building in the Social and Behavioral Sciences

The material on longitudinal panel data has a broader purpose than the direct message of “Don’t use path analysis or covariance structure analysis to analyze longitudinal data.” The basis of my own work on longitudinal data analysis is to begin by constructing models for the individual time trajectory and then to represent individual differences by differences over individuals in the values of the model parameters (or even the form of the individual time trajectory). This type of modeling for longitudinal data is also used in the so-called “hierarchical models” of Bryk and Raudenbush (1987), and in random coefficient models for longitudinal data at least since Rao (1965). Furthermore, my work on statistical models and methods for behavioral observations (Rogosa, Floden, & Willett, 1984; Rogosa & Ghadour, in press) is based on the same approach, using a renewal process model for the behavioral observations on an individual unit and allowing the parameters of the model to vary over individuals. Similarly, in Item Response Theory models, the probability of a correct response by an individual to a given item is modeled in terms of an individual ability parameter \( \theta \), and individual differences in ability are represented by the distribution of \( \theta \) over individuals. The common idea is that it is necessary first to represent the individual process by a statistical model and be able to identify the parameters of the individual process model. I feel you have to understand individual processes before attempting to understand individual differences, and this modeling strategy allows a commonplace understanding and interpretation of the model parameters.

The critical distinction is between models that start with the individual process as opposed to models for relations among variables, of which path analysis, covariance structure analysis, and other causal modeling strategies are prominent examples. I see these models for relations among variables as statistical models without a substantive soul. Substantive processes happen to or act on individual units (persons), not to variables. And a useful model should be a representation of the relevant phenomena. Otherwise, as Freedman finds with path analysis, “the technology tends to overwhelm common sense,” and I feel that’s because the technology has little or no link to common sense.

I’ve come to view the phrases causal inference, explanatory research, and so forth, as often deceptive and polarizing ways of thinking. My view is that investigators set out to address research questions, and not all important research questions are explanatory in the sense of seeking as-if-by-experiment inferences. Attempts to answer experimental (i.e., causal, explanatory, etc.) research questions with nonexperimental data seem fundamentally askew. And it seems this mismatch generates some adversarial
feelings between statisticians and social scientists. Statisticians can’t work
magic or can they be effective policemen. The divisions in thinking about
nonexperimental research may be summarized by use of a familiar format:
Causal inference from non-experimental research is
(a) made easy with LISREL
(b) possible if some attention is given to standard assumptions of the
statistical analysis
(c) nearly impossible, unless an exhaustive set of theoretical and statisti-
cal assumptions (often untestable) are satisfied
(d) a grand oxymoron
Observing how causal model methodology has progressed (?) in the last two
decades has made me increasingly partial to (d)
What can be done with nonexperimental data is to address important
research questions that do not require as-if-by-experiment inferences. In-
vestigating carefully the research questions that are appropriate with non-
experimental data seems far superior to the house of cards (or chicken wire)
provided by the path analyses. Successful research will be based on substanti-
ate insight and careful modeling of the processes generating the data (not
path analysis or LISREL). An example from my own research is the question
about correlates of charge—for example “What kind of persons grow (learn) fastest?” (see Rogosa & Willett, 1985a, p. 203). Rogosa and Willett
(1985a) develop models for systematic individual differences in growth that
are based on the individual processes generating the data. Certainly, an-
swers to this question do not provide the prescriptive information that
would be provided by a causal explanation of the factors influencing student
learning. Yet those answers should provide a basis for knowledge to accu-
ulate, and eventually, causal explanations may be attained by building a
dependable base of empirical knowledge. Statisticians can be very produc-
tive partners in addressing focused research questions by development and
application of a technology for the collection and summary of data.
That is my optimistic view. My pessimistic view is that regardless of
presentations of the vivid shortcomings of path analysis and related pro-
duces, its proselytizers and practitioners will pay little heed.

References
models to assessing change. Psychological Bulletin, 101, 147-158.
on children’s academic performance. A cross-lagged panel analysis. Journal of
Educational Psychology, 70, 39-49.
Psychometrika, 59, 371-375.

David Rogosa

Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analy-
sis. Psychological Bulletin, 72, 177-182.
Goldberger, A. S. (1971). Econometrics and psychometrics: A survey of commu-
nalities. Psychometrika, 36, 83-105.
Academic Press.
Joreskog, K. G. (1971). Analyzing psychological data by structural equation mod-
Joreskog, K. G., & Sorbom, D. (1979). Advances in factor analysis and struc-
Rao, C. R. (1965). The theory of least-squares when the parameters are stochastic
Rogosa, D. R. (1979). Causal models in longitudinal research: Rationale,forma-
tion, and interpretation. In J. R. Neuenschwander & P. B. Baltes (Eds.), Longitudi-
88, 245-258.
R. T. Campbell, W. M. Meredith, & S. M. Rawlings (Eds.), Methodological issues
in aging research. New York: Springer.
the measurement of change. Psychological Bulletin, 90, 726-748.
observations. Journal of Educational Statistics.
Rogosa, D. R., & Willett, J. B. (1985a). Understanding correlates of change by
Rogosa, D. R., & Willett, J. B. (1985b). Satisfying a simplex structure is simpler
than it should be. Journal of Educational Statistics, 10, 99-117.
Seigel, D. G. (1975). Several approaches for measuring average rates of change
for a second degree polynomial. The American Statistician, 29, 36-37.

Author
DAVID ROGOSA. Associate Professor, School of Education and Dept. of Statis-
tics, Stanford University, Stanford, CA 94305. Specializations: statistical meth-
ods, longitudinal research, behavioral observations.