

Low-Tech Causal Modeling

"I had a different reaction, however, when I reread the entire book in 1976. It was the same curious reaction I get to a morning of listening to *Revolver*, *Blonde on Blonde*, and *American Beauty*: 'here they are folks, the wonderful sixties, recollected in tranquility.'" [Neil Henry, reviewing H. M. Blalock's *Causal Models in the Social Sciences*.]

1. Background

The analytical approach called causal modeling (also, "path analysis" or "structural equation modeling") has become a distinctive feature of U.S. sociology. Since its earliest sociological applications in the 1960s, causal modeling held out the promise of bridging the island worlds of sociological theory, data and statistics. Over the first two decades development was energetic, attracting some of the best minds in the field (early papers in Blalock 1971). By the mid-1970s, key texts had been published (Duncan 1975; Heise 1975; Namboodiri et al. 1975), and causal modeling assumed its place as new dominant paradigm. For a sociology-of-sociology account see Mullins (1973); Bielby and Hauser (1977) provide a more technical review.

Some of the logic behind causal modeling had been explored by earlier sociologists using huge survey datasets (e.g., Berelson, Lazarsfeld and McPhee 1954). Their ideas and findings laid foundations for modern social research, even though the available technology—mainly, punched-card sorting machines—limited data analysis to multiway crosstabulations. The published tables became data treasures that later were mined by a new generation of log-linear modelers, notably Leo Goodman (1978). But the causal modeling breakthroughs of the late 60s and 70s hinged on the spread of mainframe computers, which made more practical the calculation of continuous-variable statistics. In the earliest causal-modeling articles, the statistics chosen were correlational (e.g., the now-almost-extinct concept of partial correlation), borrowed from psychometrics. As the superiority of regression-type statistics became apparent, more sociologists became aware of the highly-developed regression traditions of econometrics (e.g., Theil 1971; Johnston 1972). The correlation/psychometric influence lingered, however, in sociologists' preference for standardized regression coefficients or "beta weights" (which econometric texts disdain).

Two Swedish statisticians, Karl Jöreskog and Dag Sörbom, led the next step forward beyond regression (Jöreskog and Sörbom 1979). The Jöreskog/Sörbom method was codified and marketed in Sörbom's LISREL (Linear Structural RELations) computer program. LISREL combined the estimation of measurement and structural parameters, so that models that would previously have required a number of separate factor analyses and multiple regressions could now be fit—and more importantly, tested—in one unified step. Whereas least-squares regression focuses on how well the model fits individual-case data, LISREL takes a different approach. It uses maximum-likelihood estimation to find, for a specified model, the parameters that give the best possible fit not to individual cases, but to the variance-covariance matrix (basically an

unstandardized correlation matrix; see footnote 5). Although a major advance, LISREL at first was a challenge to use—it required users to start out by writing their model in matrix algebra form. Gradually, more user-friendly versions of LISREL, and LISREL competitors, emerged. The general approach, now accessible through a variety of software, is variously termed “covariance structure analysis,” “full information maximum likelihood estimation,” or most often simply “structural equation modeling.” By whatever name, such maximum-likelihood methods have superceded least squares as the preferred way to do causal modeling in elite U.S. sociology journals. Among the introductions to this inherently technical topic are books by Bollen and Bollen (1989), Hoyle (1995), Kline (2004), Loehlin (1998) and Schumacker and Lomax (1996).

Paddling crosswise to the currents of modern sociology, this paper does not present the maximum-likelihood methods of structural equation modeling. Instead, it briefly describes how to apply the earlier low-tech approach, based on ordinary least squares (OLS) regression. That choice is expedient, because a low-tech approach is far simpler to teach and learn. A further rationale appears in section 7.

2. Causal Diagrams

The signature device of causal modeling is a causal diagram, which visualizes our ideas about the order of causation (**Figure 1**). Causality flows left to right, so one variable could be a cause for any other variable that is shown farther to its right. A variable cannot be a cause for variables to its left, or in the same vertical plane.¹ In this diagram, X_1 , X_2 and X_3 are shown as possible causes of Y . X_1 and X_2 also are possible causes of X_3 , which makes X_3 an “intervening variable” between X_1 , X_2 and Y . Straight arrows depict all the possible causal effects. We will eventually have data-derived estimates and tests for these effects. Some effects might turn out to be statistically indistinguishable from zero, which would allow us to simplify our diagram by erasing those arrows.

¹ Sometimes we believe there should be a feedback or “reciprocal causation”: for example, X_1 and X_2 both cause each other. Although reciprocal causation can sound reasonable in theory, and texts explain how to do it (for a low-tech version based on instrumental variables, see Heise 1975:168–172; or very briefly, Hamilton 2009:175), the practical and interpretive challenges are greater than they seem.

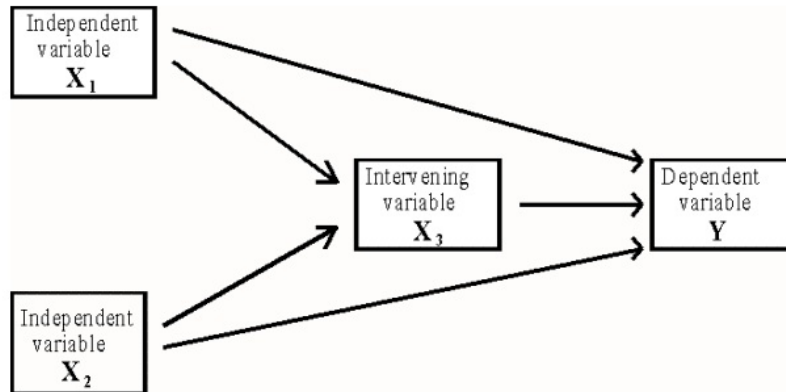


Figure 1

The assumed causal ordering of variables is critical: it underlies the whole enterprise of causal modeling. Put another way, causal analysis does not “prove” causality, but rather *assumes* a particular causal structure. It then applies statistical techniques to fill in details, perhaps fine-tuning the specification in some respects. Figure 1 assumes that X_3 is determined after X_1 and X_2 , but prior to Y . Statistical analysis of cross-sectional data cannot tell us whether a particular causal ordering is correct. Instead, we must rely on external knowledge or theory to specify the causal order. If that knowledge or theory is shaky, then so are all the analyses that follow.

A major point of causal modeling is that it allows for the possibility of intervening variables—and hence, *indirect effects*. In this example, X_1 and X_2 have direct effects on Y . But they also affect X_3 , which in turn affects Y , creating additional indirect pathways from the independent variables to Y . As we will see, the *total effect* of an independent variable on the dependent variable equals the sum of its direct and indirect effects.

Sketching out one or more causal diagrams is the first step in causal modeling. Often the diagrams help (force) us to clarify fuzzy thinking. They also aid in communicating our ideas to students, colleagues and professors.

3. Multiple Regression as a Causal Model

Once we have our causal diagram, the next step is to fit this theoretical construction to data. That can be done most simply by regression. Regression is not inherently “about” causality, but it can be used in the context of causal models. **Figure 2** shows a causal diagram that corresponds to an ordinary multiple regression with three independent variables.² Any causal relations among the

² In Stata, for example, we could obtain the correlations and standardized regression coefficients needed for Figure 2 with two commands:

```

regress Y X1 X2 X3, beta
correlate X1 X2 X3
  
```

X variables remain unanalyzed, although the simple correlations among them are shown by double-headed curved arrows. There are no intervening variables here, which makes the causal diagram unnecessary—although still useful for purposes of illustration.

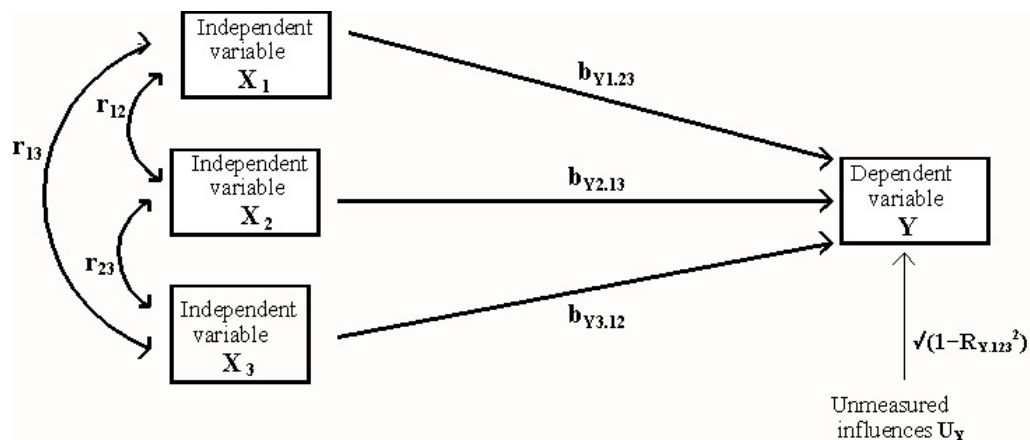


Figure 2: Causal diagram corresponding to an ordinary multiple regression.

Statistical analysis can supply estimates called *path coefficients*, or numbers indicating the strength of the effects. For Figure 2, these estimates would be:

- r_{12} , r_{13} , r_{23} Pearson correlations among the X variables, taken from their correlation matrix. These correlations might exist because the X variables have some common causes, or because there exist unanalyzed causal relations among the X variables themselves..
- $b_{Y1.23}$, etc. Regression coefficients (usually, standardized regression coefficients or “beta weights”) taken from the multiple regression of Y on X_1 , X_2 and X_3 .
 $b_{Y1.23}$, for example, denotes the regression of Y on X_1 , controlling for X_2 and X_3 .
 $b_{Y2.13}$ denotes the regression of Y on X_2 , controlling for X_1 and X_3 .
- $R_{Y.123}^2$ The coefficient of determination, R^2 , from the multiple regression of Y on X_1 , X_2 and X_3 . The square root of the quantity $\{1 - R_{Y.123}^2\}$ yields an estimate of the effect of unmeasured influences. Sometimes this is called a “disturbance” or “error” term. *The disturbance encompasses everything else in the world, besides the linear effects of X_1 , X_2 and X_3 , that affects the observed values of Y.*³

There are no intervening variables in Figure 2; all X variables are predetermined in this model world. If some of the X variables actually do cause other X variables, then the Figure 2 model is unrealistically simple, and overlooks the possibility of indirect effects. Note that this issue only arises if we care about causality. If our goal is simply prediction, then indirect effects are less important.

³ The disturbances, by assumption, are uncorrelated with the X variables. Since U_Y explains all of the variance of Y that is not explained by the X variables, and U_Y is uncorrelated with those X’s, the squared correlation between Y and U_Y must equal $1 - R_{Y.123}^2$. Therefore the correlation between Y and U_Y , which also is the path coefficient, equals $\sqrt{1 - R_{Y.123}^2}$.

Because Figure 2 corresponds to an ordinary regression, all of the usual regression assumptions and pitfalls apply here: we assume normal, independent and identically distributed (i.i.d.) errors, and should watch out for sample evidence of problems such as nonlinearity, influential cases, heteroscedasticity and multicollinearity. One aspect of the i.i.d. errors assumption deserves special mention, in the context of causal modeling: we are assuming that the disturbances are uncorrelated with any of the X variables. Another way to put this is that *we assume we have not omitted any relevant X variables*, i.e. variables related to both Y and the X's.

4. Estimating Path Coefficients with Regression

Figure 3 makes a small modification in Figure 2: X_3 appears now as an intervening variable. (This makes it similar to Figure 1, but with path coefficients written in.) Diagrams of this general sort illuminate articles in our flagship journals. Many of the symbols in Figure 3 are identical to those used in Figure 2, and have the same statistical meaning and interpretations:

r_{12}	Pearson correlation between X_1 and X_2 .
$b_{Y1.23}$, etc.	Regression coefficients (standardized) taken from the multiple regression of Y on X_1 , X_2 and X_3 .
$R_{Y.123}^2$	The coefficient of determination, R^2 , from the multiple regression of Y on X_1 , X_2 and X_3 .

The only new elements in Figure 3 pertain to the causes of the intervening variable X_3 :

$b_{31.2}$, $b_{32.1}$	Regression coefficients (standardized) taken from the multiple regression of X_3 on X_1 and X_2 .
$R_{3.12}^2$	The coefficient of determination, R^2 , from the multiple regression of X_3 on X_1 and X_2 .

In other words, the coefficients in Figure 3 come from the same multiple regression as those in Figure 2—plus a second multiple regression of the intervening variable on the independent variables.⁴

⁴ Three Stata commands produce all the coefficients needed for Figure 3:

```
regress Y X1 X2 X3, beta
regress X3 X1 X2, beta
correlate X1 X2
```

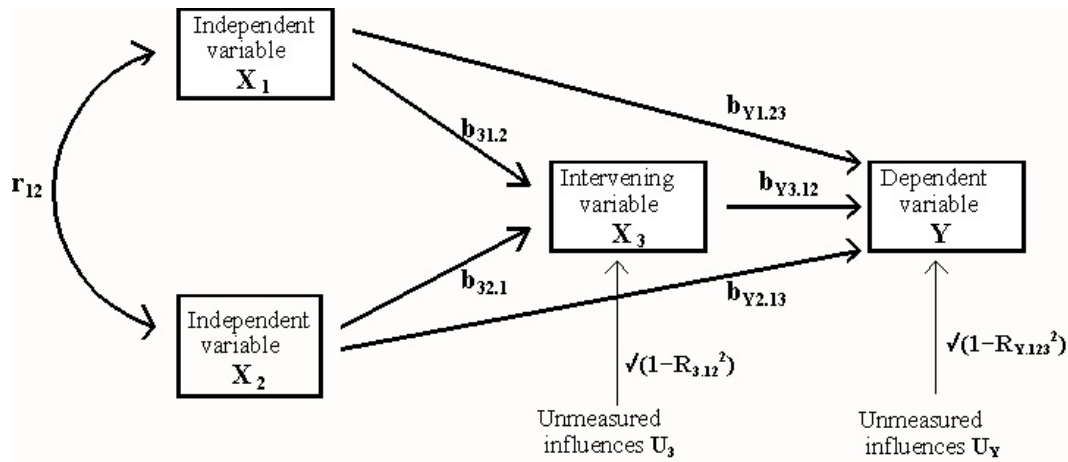


Figure 3: A simple causal model.

The Figure 3 model is isomorphic with these two *structural equations*:

$$\hat{Y} = b_{Y1.23}X_1 + b_{Y2.13}X_2 + b_{Y3.12}X_3$$

$$\hat{X}_3 = b_{31.2}X_1 + b_{32.1}X_2$$

In structural equations, both the variables and their coefficients are understood to be in standardized form.⁵

The two multiple regressions that estimate the coefficients in these structural equations are depicted as causal diagrams in **Figure 4** and **Figure 5** below.

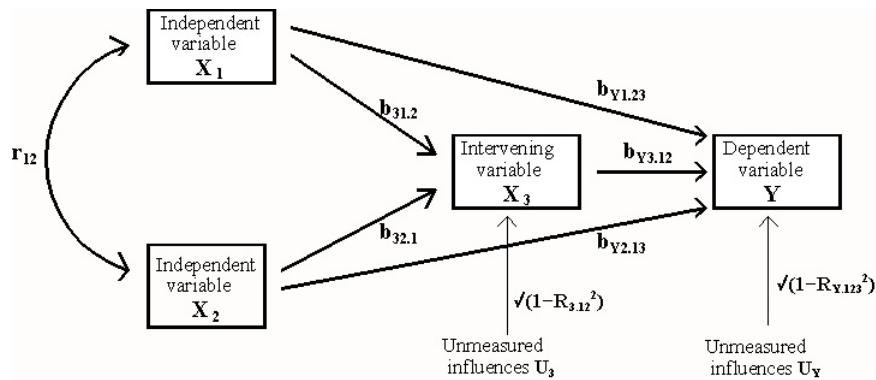


Figure 4: Regression of Y on X₁, X₂ and X₃.

⁵ To standardize a variable we subtract its mean, then divide by the standard deviation: $Y^* = (Y - \mu)/\sigma$. Y^* then has a mean of zero, and a standard deviation of one. Standardized variables are measured in units of standard deviations from their means. Standardization has no effect (normalizing or otherwise) on distribution shape. Standardized regression coefficients equal the coefficients we would obtain if all variables are standardized. A correlation matrix is a standardized variance–covariance matrix.

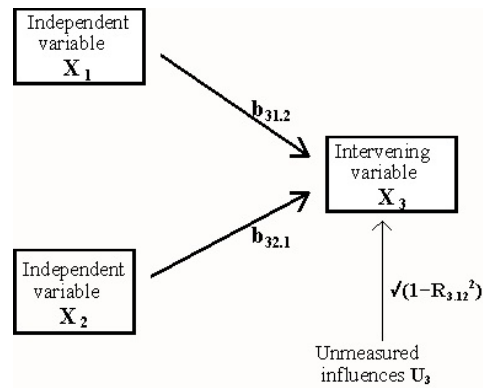


Figure 5: Regression of X_3 on X_1 and X_2 .

Causal models can be more complicated, including several waves of intervening variables. Whether they are simple or complex, the same method is used to estimate their path coefficients: perform a series of multiple regressions:

- (1) Regress the dependent variable on all of the variables to its left in the causal diagram.
- (2) Then do further multiple regressions, one for each intervening variable, regressing that variable on all of the variables to its left in the diagram.

Thus a model with one dependent and 10 intervening variables would be estimated through 11 multiple regressions. That model, and the 11 regressions, correspond to 11 structural equations. Some work with a pencil is necessary to write out these equations, or fill in the diagram's path coefficients, with numbers read from a succession of regression tables.

For all causal modeling regressions, the usual regression assumptions and cautions apply. And we continue to assume that there are no omitted variables that influence more than one intervening or dependent variable in our model. If this assumption is false, our conclusions could be mistaken. Unfortunately, it often *is* false, so analysts need to think carefully about how, and with what likely consequences.

5. Direct, Indirect and Total Effects

Causal modelers sometimes are willing to make heroic assumptions, because they prize the insights that causal models can offer. Those insights primarily involve indirect effects. *Indirect effects* equal the product of coefficients along any series of causal paths that link one variable to another. *Total effects* equal the sum of all direct and indirect effects linking two variables.

These rules follow from the structural equations. Using the example in Figure 3:

If

$$\hat{Y} = b_{Y1.23}X_1 + b_{Y2.13}X_2 + b_{Y3.12}X_3$$

and

$$\hat{X}_3 = b_{31.2}X_1 + b_{32.1}X_2$$

then by substituting \hat{X}_3 for X_3 in the first equation, we obtain a version containing both direct and indirect effects:

$$\begin{aligned}\hat{Y} &= b_{Y1.23}X_1 + b_{Y2.13}X_2 + b_{Y3.12}\hat{X}_3 \\ &= b_{Y1.23}X_1 + b_{Y2.13}X_2 + b_{Y3.12}(b_{31.2}X_1 + b_{32.1}X_2) \\ &= b_{Y1.23}X_1 + b_{Y2.13}X_2 + b_{Y3.12}b_{31.2}X_1 + b_{Y3.12}b_{32.1}X_2\end{aligned}$$

Going one step further yields the *reduced equation*, in which the coefficients on X_1 and X_2 are their total effects:

$$\hat{Y} = (b_{Y1.23} + b_{Y3.12}b_{31.2})X_1 + (b_{Y2.13} + b_{Y3.12}b_{32.1})X_2$$

The indirect and total effects can be calculated most easily by multiplying coefficients along any sequence of paths in a causal diagram, and then adding the results for all paths connecting each independent variable to Y . Figure 3 contains the following effects.

Direct effects

X_1 to Y :	$b_{Y1.23}$	(regression coefficient of Y on X_1 , controlling for X_2 and X_3)
X_2 to Y :	$b_{Y2.13}$	(regression coefficient of Y on X_2 , controlling for X_1 and X_3)
X_3 to Y :	$b_{Y3.12}$	(regression coefficient of Y on X_3 , controlling for X_1 and X_2)
X_1 to X_3 :	$b_{31.2}$	(regression coefficient of X_3 on X_1 , controlling for X_2)
X_2 to X_3 :	$b_{32.1}$	(regression coefficient of X_3 on X_2 , controlling for X_1)

Indirect effects

X_1 to Y , through X_3 :	$b_{31.2} \times b_{Y3.12}$
X_2 to Y , through X_3 :	$b_{32.1} \times b_{Y3.12}$

Total effects

X_1 to Y :	$b_{Y1.23} + (b_{31.2} \times b_{Y3.12})$
X_2 to Y :	$b_{Y2.13} + (b_{32.1} \times b_{Y3.12})$

If the indirect effects are theoretically or empirically important, then we can learn something new from our causal model. It is possible that most of an independent variable's total effect is indirect. For example, classical status attainment research found that parents' occupational prestige has only a weak direct effect on respondents' occupational prestige. It has stronger indirect effects, through education: high-prestige parents make sure that their offspring go to college, and graduating from college increases anyone's chances of landing a high-prestige job. There can also be opposite signs among direct effects and one or more indirect effects, so that they cancel out or weaken the total effect.

On the other hand, if relations among our variables are weak, then the indirect effects are likely to be very weak, and perhaps substantively unimportant. For the same reason, indirect effects involving more than two steps tend to be very weak. For example, suppose that the path from X_1 to X_3 , and also the path from X_3 to Y , both have standardized coefficients around 0.30. This could be "statistically significant" even in a smallish sample. It means that a one-standard-deviation increase in X_1 leads directly to a 0.30-standard-deviations increase in the predicted

value of X_3 , if X_2 does not change. Similarly, a one-standard-deviation increase in X_3 leads directly to a 0.30-standard-deviations increase in predicted Y , if X_1 and X_2 do not change. But the *indirect* effect from X_1 to Y is then only $0.30 \times 0.30 = 0.09$. That is, a one-standard deviation increase in X_3 leads indirectly to 9/100ths of a standard deviation increase in predicted Y . For a three-step indirect effect where each path was 0.30, the magnitude of the indirect effect would be $0.30 \times 0.30 \times 0.30 = 0.027$, or less than 3/100ths of a standard deviation. Of course, when a total effect adds together several weak indirect effects, it could be more substantial.

6. Variables and Measurement

Any measurement variables suitable for OLS regression analysis could be used in a causal model. This includes variables re-expressed by nonlinear transformations, and interaction terms formed by multiplying variable pairs. Interpretation with such variables becomes trickier, and requires careful reasoning with the structural equations. Regression diagnostic tools (residual plots, influence statistics, etc.) remain useful, as do conditional effects plots for interpretation.

Techniques such as robust regression or quantile regression are not well suited to causal modeling. Robust regression, for example, would weight cases differently for each regression, so their coefficients could not reasonably be combined to find indirect effects. If outliers are a problem, it is better to deal with them through transformation rather than trying to apply an outlier-resistant fitting technique within the causal modeling framework. Categorical-variable methods such as logit or probit regression also are difficult to adapt, because they predict probabilities or odds rather than actual values of the dependent variable. Dummy variables work fine as independent variables, but are less desirable as intervening or dependent variables.

Any of the variables could be scales, formed by combining other variables. The usual concerns about validity, reliability and approximate normality apply. If the scales are factor scores, formed through factor analysis, then our causal model contains both a “measurement model” (the factor analysis) and a “structural model” (causal relations between factor scores and other variables). LISREL-type methods can estimate complex measurement/structural models in a single step, but similar results often can be achieved in practice through a low-tech concatenation of factor analysis with OLS. What the low-tech approach cannot do, and LISREL can, is estimate and test all the measurement and structural parameters at once. Because LISREL uses full information from the variance–covariance matrix, instead of performing analyses one at a time, it might have enough degrees of freedom also to estimate or test some of the correlations between the unmeasured influences. This amounts to a test of whether relevant variables have been omitted.

7. Why Use Low-Tech Causal Modeling?

Given the “true” causal ordering, and well-measured, multivariate-normal data, LISREL-type methods are clearly superior to the low-tech approach described above. In particular, they offer

much greater flexibility in estimating and testing possible correlations among the unmeasured influences or error terms. This enhances the potential for realism in our modeling efforts.

As usual in statistics, we pay for this additional potential by making additional assumptions. If our assumptions are false—we have guessed wrongly about the true causal structure, or do not have well-measured, multinormal variables—then maximum-likelihood estimation could yield results as misleading as least-squares regression, if not worse.⁶ Some progress has been made in efforts to diagnose problems and “robustify” LISREL-type methods against non-normality and other troubles (e.g. Bollen and Arminger 1991; West et al. 1995), but the vast literature on diagnostics and robustness for the simpler least-squares case shows how complicated these issues are (for introductory discussions see Hoaglin, Mosteller and Tukey, 1983, 1985; Cook and Weisberg 1982, 1994; Hamilton 1992, 2009). Moreover it appears that most published LISREL-type analyses, like most OLS regression articles, uncritically accept the standard—and commonly false—assumptions (Micceri 1989; Breckler 1990).

A more subtle difficulty arises from the way that structural equation modeling can insulate the analyst from direct contact with actual data. If we focus on how well a model fits the covariance matrix, as LISREL encourages, incautious analysts might neglect more basic questions about how the covariance matrix fits the raw data; and lose sight of the fit between model and raw data—i.e., the individual-case predictions and residuals. Structural equation modeling provides fewer opportunities and tools than ordinary regression for testing basic assumptions, noticing unusual data points, or thinking through the substantive implications of univariate and bivariate distributions. Besides aiding the multivariate analysis, simple diagnostic work sometimes yields the most unexpected, interesting and replicable findings from our research.

The chief advantages of a low-tech approach to causal modeling are its practical simplicity, less demanding assumptions, and the toolkit of diagnostic and exploratory methods that accompany OLS regression. The main advantages of causal modeling itself, whether low or high-tech, lie in the clarifying and communicating device of the causal diagram, and its analysis of indirect effects.

⁶ David Freedman (1991) takes a lucid and cheerfully pessimistic look at the problems associated with most nonexperimental analyses of social causality. Equally worth reading and in the same book are the responses to Freedman by Berk, Blalock and Mason.

8. References

- Bielby, William T. and Robert M. Hauser. 1977. "Structural equation models." Pp. 137–161 in Alex Inkeles (ed.) *Annual Review of Sociology 1977*. Palo Alto, CA: Annual Reviews Inc.
- Blalock, Hubert M. Jr. (ed.). 1971. *Causal Models in the Social Sciences*. Chicago: Aldine.
- Bollen, Kenneth A. and G. Arminger. 1991. "Observational residuals in factor analysis and structural equation models." Pp. 235–262 in P. Marsden (ed.) *Sociological Methodology 1991*. Oxford: Blackwell.
- Bollen, Kenneth A. and William Bollen. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, Kenneth A. and J. Scott Long (eds.). 1993. *Testing Structural Equation Models*. Thousand Oaks, CA: Sage.
- Breckler, S. J. 1990. "Application of covariance structure modeling in psychology." *Psychological Bulletin* 107:260–273.
- Cook, R. Dennis and Sanford Weisberg. 1982. *Residuals and Influence in Regression*. New York: Chapman and Hall.
- Coo Cook, R. Dennis and Sanford Weisberg. 1994. *An Introduction to Regression Graphics*. New York: Wiley.
- Duncan, Otis D. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.
- Freedman, David A. 1991. "Statistical models and shoe leather." Pp. 291–313 in Peter V. Marsden (ed.) *Sociological Methodology 1991*. Oxford: Blackwell.
- Goodman, Leo A. 1978. *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis*. Cambridge, MA: Abt.
- Hamilton, Lawrence C. 1992. *Regression with Graphics*. Pacific Grove, CA: Brooks-Cole.
- Hamilton, Lawrence C. 2009. *Statistics with Stata*, updated for version 10. Pacific Grove, CA: Cengage.
- Heise, David R. 1975. *Causal Analysis*. New York: Wiley-Interscience.
- Henry, Neil W. 1976. Review of *Causal Models in the Social Sciences*, edited by H. M. Blalock, Jr. *Contemporary Sociology* 5(6):777–778.

Hoaglin, David C., Frederick Mosteller and John W. Tukey (eds.). 1983. *Understanding Robust and Exploratory Data Analysis*. New York: Wiley-Interscience.

Hoaglin, David C., Frederick Mosteller and John W. Tukey (eds.). 1983. *Exploring Data Tables, Trends, and Shapes*. New York: Wiley-Interscience.

Hoyle, Rick H. (ed.) 1995. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage.

Johnston, J. 1972. *Econometric Methods* (2nd edition). New York: McGraw-Hill.

Jöreskog, Karl G. and Dag Sörbom. 1979. *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt.

Kline, Rex B. 2004. *Principles and Practice of Structural Equation Modeling*, 2nd edition. New York: Guilford Press.

Loehlin, John C. 1998. *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Micceri, T. 1989. "The unicorn, the normal curve, and other improbable creatures." *Psychological Bulletin* 105:156–166.

Mullins, Nicholas. 1973. *Theories and Theory Groups in Contemporary American Sociology*. New York: Harper and Row.

Namboodiri, N. Krishnan, Lewis F. Carter and Hubert M. Blalock, Jr. 1975. *Applied Multivariate Analysis and Experimental Design*. New York: McGraw-Hill.

Schumacker, Randall E. and Richard G. Lomax. 1996. *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum.

Theil, Henri. 1971. *Principles of Econometrics*. New York: Wiley-Interscience.

West, Stephen G., John F. Finch and Patrick J. Curran. 1995. "Structural equation models with nonnormal variables." Pp.56–75 in Rick H. Hoyle (ed.) *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage.