12

Exploring Causal Structure with the TETRAD Program

*Clark Glymour, Richard Scheines and Peter Spirtes**

1. INTRODUCTION

This paper describes some aspects of a new approach to the search for adequate causal explanations of nonexperimental and quasi-experimental data. The ideas are embodied in a computer program, TETRAD, and are described in much more detail in a recent book, *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science and Statistical Modeling* (Glymour et al. 1987). In addition, some new results about our methods are reported in what follows.

For quantitative sociologists and psychometricians, the methods should have a familiar heritage. They are computerized extensions of ideas that were proposed in the early days of correlation analysis by followers of Charles Spearman. In more recent decades, Herbert Simon, Hubert Blalock, Herbert Costner, O. D. Duncan, and many others have worked on model specification through the analysis of constraints on covariances; Simon not only contributed to some of the central ideas connecting multivariate analysis with causal explanation, he is also the most articulate exponent of the role of *heuristic search* in scientific discovery. The TETRAD program uses new mathematical results to combine these lines of work into a computer package to aid in model specification. In what follows, we briefly describe the kind of

* Carnegie Mellon University.

problem to which the program is addressed, the mathematical results that make it possible, and the search strategy used, and we give a number of illustrations of the application of the procedures.

2. MODEL SPECIFICATION PROBLEMS

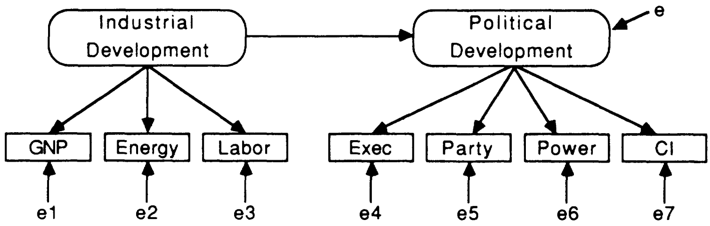
Social science research papers that use causal modeling techniques usually contain an explicit *specification problem*: Something about causal relations among measured or unmeasured variables is unknown and is to be discovered. Examples abound:

1. Maruyama and McGarvey (1980) are concerned with whether variation in a latent variable interpreted as student peer popularity causes variation in a latent variable interpreted as disposition for academic achievement, or whether the causal relation is in the other direction, or whether the variables have no direct effects on one another.
2. Miller, Slomczynski, and Schoenberg (1981) and Schoenberg and Richtand (1984) are concerned with whether correlations among responses to a survey questionnaire can be accounted for by a single latent variable, or whether further causal dependencies are needed to explain the data, and if so, what dependencies.
3. McPherson, Welch, and Clark (1977) are concerned with whether correlations among responses to a questionnaire administered to the same cohort at two different times can be explained as the result of a single latent factor acting at each time, or whether further causal dependencies are present.

In each of these examples, and in many other modeling problems, a partial model is specified, and the researchers are interested in how, if at all, it should be extended. Sometimes the number of a priori possible extensions is very large, even in cases such as these. Costner and Schoenberg (1973) discuss a model of industrial and political development to which they recommend adding two additional causal connections (see Figure 1). Two directed edges or correlated errors can be added to their initial model in more than 9,900 different ways.

There is often no well-confirmed justification even for the initial, partial model, which means that a serious scientific case requires the consideration of alternative initial models. A recent study by

FIGURE 1. Costner and Schoenberg's initial model.



Timberlake and Williams (1984), for example, treats data for four variables concerning foreign investment and economic and political development in "peripheral" nations. Parameter estimates from a particular regression model are used to support their causal conclusions, even though there are 262,144 alternative specifications of the connections among four variables (including models with correlated errors, but not counting latent variable models), which include several plausible models that explain patterns in the Timberlake and Williams data, that are testable, that show excellent fit, and that support causal hypotheses contrary to the ones Timberlake and Williams favor. The number of alternatives grows exponentially with the number of measured variables. Jay Magidson (1977, p. 413) put the problem this way:

> The problem we face is that there is an infinite number of ways to formulate a causal model, and it is not a straightforward matter to determine how to go about doing it, particularly when the causes are unknown and/or unobserved. It is important for researchers to formulate not one but many models so that they can determine whether their conclusions may differ if they accept a different set of assumptions. It is also important to follow some general guidelines in building (formulating) models when the researcher has limited information about the causal process.

Similar concerns have been voiced by many other authors. Clearly, in many cases there are a lot of alternative causal models, even granted the assumption of linearity and the usual statistical assumptions made in structural equation modeling. Any real scientific effort, therefore, has to make a case for the causal hypotheses that are proposed, and it looks as though that may be a very difficult thing to do. Exactly for this reason, several critics claim that causal modeling in practice amounts to a kind of pseudoscience. There are at least three ways in which a case could be made for the causal assumptions of a model:

1. Prior knowledge or *well-justified* theory could imply a unique set of causal assumptions, or at least reduce the alternatives that need to be considered to a very small number. That is not usually the case in social scientific work. It is certainly not enough to appeal to "theory" when the theory to which appeal is made has not been subjected to severe tests and is not well confirmed.
2. Experimental controls could be systematically introduced to isolate causal effects. That is rarely feasible in social scientific work and has limited feasibility even in medical science, epidemiology, and other areas.
3. Using whatever prior knowledge may be available, one could conduct a systematic search for alternative models, showing that the proposed explanation provides, or is likely to provide, the best available explanation of the data.

A considerable body of work in social science methodology has focused on the third of these alternatives. Although a variety of approaches to model specification have been sketched, two approaches have dominated investigations in sociology and related subjects. One of them, associated with Jöreskog and Sörbom (1984) and the LISREL program, uses maximum likelihood estimation and any of several associated fitting statistics; the other, illustrated in the work of Blalock, Costner, Duncan, Simon, and others, compares the *constraints* implied by models with constraints satisfied by the data.

2.1. *Maximum Likelihood Methods*

Maximum likelihood specification techniques focus on elaborations of an initial model. Several techniques have been employed, all of which share a dependency on properties of maximum likelihood estimates of the parameters in an initial model or in elaborations of the initial model. Jöreskog and Sörbom use a general procedure in the more recent editions of the LISREL program. Essentially, the method is to consider partial derivatives of the LISREL fitting statistic taken with respect to each of the fixed parameters of the model. Provided the difference in chi-square values is significant, the method recommends freeing that fixed (usually at zero) parameter with the largest partial derivative (and the right sign for the second derivative). McPherson et al. (1977), who are concerned with whether or not a certain measurement model is stable over time, examine the *size* of the coefficients connecting a latent variable at two different times with the respective measures of an observed variable at two different times. Many authors compare correlations computed from an estimated model with the empirical correlations. Maruyama and McGarvey (1980), interested in the direction of a hypothetical causal relation between latent variables, modify their initial model in two ways, once by adding the hypothetical causal relation in one direction and estimating the coefficient with LISREL, and a second time by adding instead a hypothetical causal relation in the other direction and estimating the coefficient with LISREL. They then infer that the numerically larger coefficient is associated with the real direction of causation.

2.2. *Analyzing Constraints on Correlations*

Blalock's (1961) method analyzed the constraints on partial correlations that are implied by path models. Two models such as those in Figure 2 can in principle be discriminated empirically because they *robustly imply* different constraints on the population covariances. Model I robustly implies that $\rho_{23.1} = 0$, but model II does not. A model robustly implies a constraint if it implies that the constraint is satisfied no matter what the values of the free linear coefficients of the model

FIGURE 2.

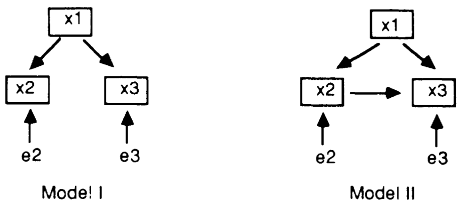
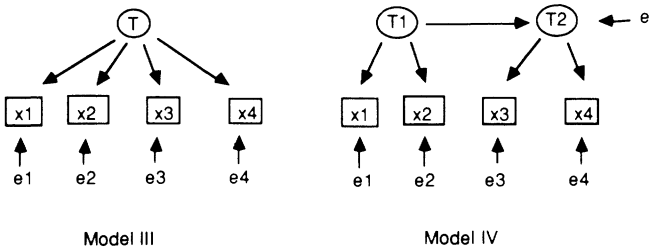


FIGURE 3.



may be. Blalock called such constraints *prediction equations* and emphasized that they may be used to test a model. The Spearman school, which originated the analysis of covariance constraints, emphasized that models that robustly imply empirically correct constraints are to be preferred to models that are merely consistent with the constraints for particular values of their linear coefficients. Similar methodological ideas were developed for latent variable models by Costner, Duncan, and others. Multiple indicator latent variable models do not robustly imply vanishing partial correlations for their measured variables, but they can in principle be discriminated by the *vanishing tetrad differences* they imply. Thus, in Figure 3, model III implies the constraints

$$\rho_{12}\rho_{34} = \rho_{13}\rho_{24} = \rho_{14}\rho_{23},$$

while model IV implies only that $\rho_{13}\rho_{24} = \rho_{14}\rho_{23}$. Costner and Schoenberg (1973) proposed a kind of mixture of the LISREL and the constraint analysis approaches. Given a multiple indicator model, their method for model revision proposed to examine those submodels like IV and to test them by maximum likelihood chi-square procedures. If a submodel failed the test, the larger model was to be modified by correlating an indicator of one of the latent variables in the submodel with an indicator of the other latent variable in the submodel.

2.3. *Comparisons of Maximum Likelihood and Constraint Analysis Methods*

Each of these approaches has advantages and disadvantages, and the limitations also affect combined methods such as Costner and Schoenberg's. The disadvantages of Jöreskog and Sörbom's modification procedure in LISREL are that it leads to a narrow "beam" search through the space of possible elaborations of a model, it may stop too soon, and the derivatives of the fitting statistic may miss plausible modifications. The considerable advantage is that it is computerized and fast and can therefore be carried out without considerable theoretical effort or excessive computational demands. Choosing elaborations on the basis of correlation residuals is known in many cases to be unreliable (see Herting and Costner 1985). The other methods described lack any convincing theoretical justification, although they sometimes work in the sense that they improve model fit. None of the methods offers any guidance in constructing an initial model.

The advantage of the analysis of correlation constraints is that it appeals to intuitions about scientific explanation that go back to the very beginnings of covariance analysis and that have played a major role in the historical successes of the natural sciences (see Glymour et al. [1987] for a discussion of the historical background). The intuition is that, other things equal, those models that explain patterns in the data without having to specify particular nonzero values for various parameters are best. Kepler used the same intuition to argue for Copernican theory, and Eddington used it to argue for general relativity. Cannizarro used the same principle to argue for his system of atomic weights, and it was the fundamental idea behind all the statistical work done by the Spearman school in the 1920s.

The strategy of searching for models by analyses of constraints on correlations promises to *localize* errors in model specification in a principled way, since particular submodels of a given model may be responsible for the implication of false constraints. And in principle, the approach promises guidance in initial model specification. Yet the disadvantages of the approach may seem overwhelming:

1. No statistical test for vanishing tetrad differences is reported or used in the sociological literature.
2. Blalock, Costner, and the many other researchers who followed their lead published no general algorithm for determining the constraints robustly implied by an arbitrary model. There are scores of papers, and some books, showing how to use constraints on covariances or correlations to distinguish between particular sets of alternative models, but there is no statement of a general algorithm, let alone a computer implementation. Nothing in the literature, for example, shows whether or not the property of robustly implying a vanishing

partial correlation or a vanishing tetrad difference depends on the distribution assumptions of the model. Heise (1975) gives the basis for an algorithmic procedure for computing implied constraints from the graphs of acyclic standardized models, but it is left open how to calculate the implications when the model is not acyclic, a problem that we have only partially solved.

3. Save for Costner and Schoenberg's (1973) mixed and partly informal procedure, no scheme has been developed for organizing the search for the elaborations of an arbitrary initial model that will best explain the constraints satisfied empirically.

3. THE TETRAD PROJECT

In the last several years we have attempted to answer the technical questions just raised concerning the foundations of the constraint analysis strategy and to apply the answers we have obtained to form a computerized aid for model specification. Our procedures rest on strong methodological sensibilities, but they are free of many of the metaphysical concerns that are the focus of other papers in this volume.

3.1. *Causality*

We assume only two things about causal relations. Suppose A and B are quantities that can take any of a continuum of values in a population, and assume that the population is unbounded in size. We assume, first, that if A causes B, then B is a function of A and perhaps of other variables. Our specific methods apply when the function is linear. Second, if A does not cause B and B does not cause A and there are no common causes of A and B, then A and B are statistically independent in the population.

This minimal understanding of causality gives a reading to the graphs that often accompany or are implicit in linear models: The graphs encode not only the linear structural equations of the model but also the *independence assumptions* of the model.

We do not take any more detailed stand on the nature of causality because our methods do not depend on these issues.[1]

[1] For comments on Holland's views about causality, see Glymour (1985).

Because our methods make only these assumptions, they apply to cases that some would insist do not involve causal relations at all. Thus, some factor analysis procedures distinguish "measurement models" from "causal models" or "theoretical models," and some hold, for reasons we do not pretend to understand, that relations between a latent variable and its measured indicators, or relations between different measured indicators, are not causal relations. From our perspective, this is a purely verbal dispute: As long as a model is associated with a directed graph that represents functional dependencies and statistical independence assumptions, our methods apply.

Neither do we have any objection to causal models with "latent" variables.[2] Since data sets throughout the social and behavioral sciences usually measure only a fragment of the variables that might be relevant to one another, unmeasured variables should be expected in plausible causal explanations of many data sets in sociology, economics, social psychology, and other areas. Even models that contain latent variables that are not associated with some factor known to be directly measurable—e.g., psychometric models—are preferable to models that do not contain them, provided the latent variable models have better explanatory power. In Glymour and Spirtes (forthcoming), we have described one statistically determinable respect in which in some cases latent variables can have superior explanatory power.

3.2. *Mathematical Results*

The constraint analysis strategy could be carried out generally and automatically provided that some mathematical results can be established.

First, does whether a model implies a vanishing tetrad difference or a vanishing first-order partial correlation for all values of its linear coefficients depend on the *variances* of the independent variables and of the error terms? If not, then the constraints of these two kinds implied by a linear model are determined *entirely by the directed graph associated*

[2] We do, however, have various objections to the interpretations often given to such variables (cf. Blalock 1982).

with the model. The following theorem has been proved:

Theorem 1. For all linear models, the implied tetrad and first-order partial correlation constraints are independent of the variances of the independent variables and error terms.

In view of theorem 1, it is proper to speak of the constraints implied *by a graph*, since all models that share the same graph of causal relations will imply the same vanishing tetrad differences and vanishing partial correlations.

Second, given that the constraints implied by a model are determined by its associated graph of causal relations, what local graph theoretic properties determine whether or not any particular constraint is implied? If such properties could be found, then they might be used in a general algorithm for computing the constraints implied by a model.

The answers we have found to the second question are a little complicated, and we give only the simplest case.

Define a *trek* between vertices $v1$ and $v2$ in a directed graph G to be a pair of acyclic paths, one terminating in $v1$ and one terminating in $v2$, having the same origin, called the *source* of the trek, and intersecting nowhere save at the origin. We allow one of the paths to be empty, so that a single path between $v1$ and $v2$ also counts as a trek between $v1$ and $v2$.

Then we have the following local graph theoretic characterization of necessary and sufficient conditions for any graph, whether cyclic or acyclic, to imply a vanishing first-order partial correlation:

Theorem 2. For any directed graph G and any three distinct variables x, y, and z that are vertices of G, the following two conditions are equivalent:

1. G implies that $\rho_{xz} - \rho_{xy}\rho_{yz} = 0$.
2. Every trek between x and z contains y, and either every trek between y and z is an acyclic path from y to z or every trek between x and y is an acyclic path from y to x.

A comparably geometric but more complicated necessary and sufficient condition for the implication of a vanishing tetrad difference has recently been proved (see Spirtes 1988).

One can take advantage of these results to construct an algorithm that will compute the vanishing first-order partial correlations implied by an arbitrary graph and the vanishing tetrad differences implied by an arbitrary acyclic graph.

Theorem 3. There is an algorithm for computing the vanishing first-order partial correlations implied by any directed graph and the vanishing tetrad differences implied by any acyclic graph such that the number of steps required by the computation increases by no more than the cube of the number of vertices of the graph.

An easy but important result is that adding additional causal relations or correlated errors to a graph of causal relations *never results in the implication of additional constraints not implied by the initial graph.*

Theorem 4. If graph G is a subgraph of graph G' and if G' implies a vanishing tetrad difference or vanishing first-order partial correlation among variables occurring in G, then G also implies that vanishing tetrad difference or vanishing first-order partial correlation.[3]

3.3. *Statistics and Computation*

In the 1920s, Spearman's followers pursued a model specification strategy that started, in every case, with a single latent common cause of every measured variable, compared the implied tetrad constraints with the constraints satisfied empirically, and then added additional latent factors until the modified model implied all and only the constraints found to be satisfied in the data. The results of the previous subsection, together with the well-known fact that a one-factor model implies every possible tetrad equation among the measured variables, explain why the strategy worked, insofar as it did. In practice, Spearman's strategy was limited by the substantive assumptions associated with his own psychological theory and by computational difficulties. Guilford (1936) gave the difficulty of computing the implied tetrad equations as the chief argument for abandoning constraint analysis in favor of factor analysis.

[3] Because of a typographical error, this result is misstated in one place in Glymour et al. (1987), but the result stated here is proved there.

The work of the Spearman school prompted John Wishart (1929) to investigate the variance of the sampling distribution of tetrad differences under the assumption that the variables are jointly multinormally distributed. Wishart's results enable us to construct an asymptotic test of the hypothesis that any particular tetrad difference vanishes. Tests of hypotheses about vanishing first-order partial correlations can of course be based on the sampling distribution of the correlation coefficient (see Anderson 1984). For a given data set, a test of this kind can be automatically conducted separately for each possible constraint, and the TETRAD program carries out such a procedure either automatically at a series of significance levels or at a user-specified significance level. Since in general the tests are not independent, the procedure is not ideal, but it avoids a vast series of complicated simultaneous testing problems (see Miller 1981). We regard it as an example of Simon's recommendation to "satisfice" in those cases in which the ideal solution is infeasible.

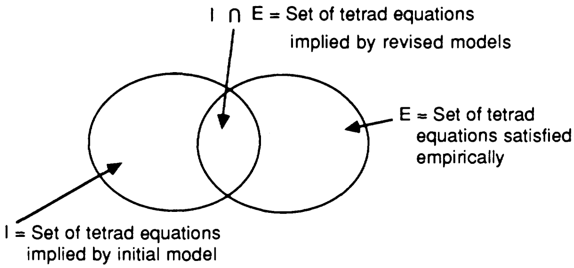
3.4. *Automatic Search Strategy*

The mathematical results make feasible an automatic search strategy that shares some features of the strategy of the Spearman school.

1. Start with a simple model but not necessarily with a one-factor model.
2. Add directed edges and correlated errors to the graph of the model until, insofar as possible, you obtain a modified graph (or graphs) that imply all and only the empirically correct constraints implied by the initial model.

The strategy is illustrated in Figure 4. TETRAD contains an automatic procedure that carries out a version of this strategy. To make the program run in tolerable time on personal computers widely available in 1987, the automatic search is restricted to latent variable multiple indicator models. New hardware developments and algorithm improvements should permit us to relax this restriction in subsequent versions of the program.

FIGURE 4.



3.5. *Power*

The strategy makes sense only if the constraints considered are sufficient in many cases to distinguish between alternative models. That question can be investigated systematically using the TETRAD program itself and a little mathematics to generalize the results. We report here only the simplest relevant results that we have obtained.

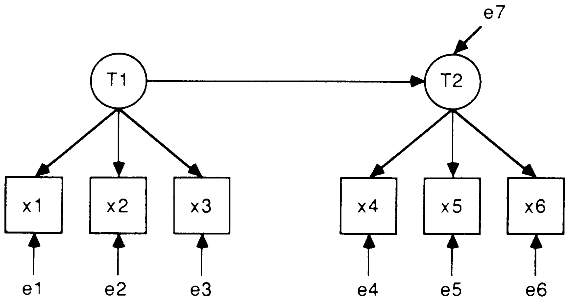
Say that a model is *skeletal* if every measured variable depends on one and only one latent variable, if no measured variable has a direct effect or correlated error with any other measured variable, and if the model is acyclic.

Theorem 5. Consider any set of models consisting of a skeletal model and all models that can be obtained by adding at most one directed edge or correlated error to the skeletal model. Assume further that each latent variable has at least three measured indicators and that there are at least five measured variables.

Then, all models in such a set imply distinct collections of vanishing tetrad differences save that

1. if A and B are measured indicators of the same latent variable, then the models formed by adding to the skeleton exactly one of "A causes B," "B causes A," and "A and B have a correlated error" are indistinguishable from each other;
2. if A is a latent variable and B is a measured indicator of a different latent variable, then the models formed by adding to the skeleton

FIGURE 5. Generating skeleton.



exactly one of "*A* causes *B*," "*B* causes *A*," and "*A* and *B* have a new common cause" are indistinguishable from each other.

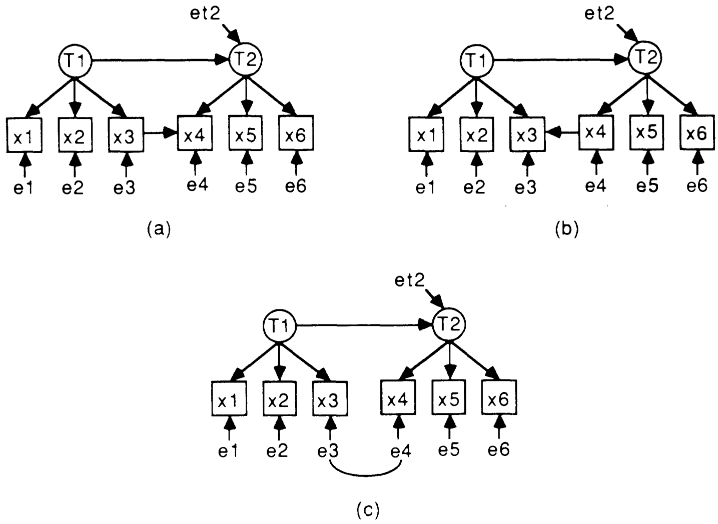
Similar but more complicated results have been obtained for two edge additions to skeletal models. The power of the result stated here can be illustrated by a simulation study. Consider the skeletal model in Figure 5. Models with this sort of structure are typical in studies that repeat measurements for the same cohort at different times. Such simple skeletal models are often incorrect, and the effect of the common cause may be confounded with direct effects of earlier measures on later measures, or with other common causes that produce correlated errors.

Each indicator of the first latent variable can be connected with each indicator of the second latent variable in three different ways. We illustrate the possibilities in Figure 6. Each of these elaborations of the initial model implies a distinct set of tetrad constraints. The inequivalence of these models is shown in Table 1. Every possible tetrad equation among the six measured variables is listed in a row. A row contains a *y* in a column just in case the model corresponding to that column implies the equation listed in that row.[4]

It should, therefore, be possible to distinguish data sets generated by the three models in Figure 6. In fact, the point is much more general. Rather than focus on the three different ways of connecting $x3$

[4] The tetrad equation $\rho_{x1,x2}\rho_{x3,x4} = \rho_{x1,x3}\rho_{x2,x4}$ is abbreviated $x1\,x2, x3\,x4 = x1\,x3, x2\,x4$.

FIGURE 6. Three elaborations of the generating skeleton.



and $x4$, consider the 27 different ways of connecting one indicator of $T1$ and one indicator of $T2$, and consider as well the skeletal model of Figure 5: *Each of these models implies a distinct set of tetrad constraints.* Thus, in principle it should be possible to distinguish each of these models from the others based on the set of tetrad constraints satisfied by the data.

We generated data sets from models of these kinds using SYSTAT BASIC. We produced data for the exogenous variables with a pseudorandom number generator distributed normally with mean 0 and variance 1. All other variables are linear combinations of their respective immediate ancestors in the directed graph of the model. The linear coefficients were also chosen at random, although their values for each particular model are nonstochastic constants. $T1$ and $T2$ are to be interpreted as unmeasured latent variables, and $e1$–$e7$ as unmeasured error terms. $T1$ and $e1$–$e7$ are exogenous, and $x1$–$x6$ are our measured variables. The sample size is 2,000. Choosing at random 5 of the 27 elaborations of the initial model, we generated five data sets, each of which, along with the skeletal model shown in Figure 5, was given to a TETRAD user. The user did not know which of the 27

TABLE 1
TETRAD Implications of Initial Model and its Modifications

	Implied by			
Tetrad Equation	Initial Model	Model a $x3 \rightarrow x4$	Model b $x4 \rightarrow x3$	Model c $x3 \rightarrow x4$
$x1\,x2,\ x3\,x4 = x1\,x3,\ x2\,x4$	*y*			
$x1\,x2,\ x4\,x3 = x1\,x4,\ x2\,x3$	*y*			
$x1\,x3,\ x4\,x2 = x1\,x4,\ x3\,x2$	*y*	*y*	*y*	*y*
$x1\,x2,\ x3\,x5 = x1\,x3,\ x2\,x5$	*y*	*y*		*y*
$x1\,x2,\ x5\,x3 = x1\,x5,\ x2\,x3$	*y*	*y*		*y*
$x1\,x3,\ x5\,x2 = x1\,x5,\ x3\,x2$	*y*	*y*	*y*	*y*
$x1\,x2,\ x3\,x6 = x1\,x3,\ x2\,x6$	*y*	*y*		*y*
$x1\,x2,\ x6\,x3 = x1\,x6,\ x2\,x3$	*y*	*y*		*y*
$x1\,x3,\ x6\,x2 = x1\,x6,\ x3\,x2$	*y*	*y*	*y*	*y*
$x1\,x3,\ x4\,x5 = x1\,x4,\ x3\,x5$				
$x1\,x3,\ x5\,x4 = x1\,x5,\ x3\,x4$				
$x1\,x4,\ x5\,x3 = x1\,x5,\ x4\,x3$	*y*			
$x1\,x3,\ x4\,x6 = x1\,x4,\ x3\,x6$				
$x1\,x3,\ x6\,x4 = x1\,x6,\ x3\,x4$				
$x1\,x4,\ x6\,x3 = x1\,x6,\ x4\,x3$	*y*			
$x1\,x4,\ x5\,x6 = x1\,x5,\ x4\,x6$	*y*		*y*	*y*
$x1\,x4,\ x6\,x5 = x1\,x6,\ x4\,x5$	*y*		*y*	*y*
$x1\,x5,\ x6\,x4 = x1\,x6,\ x5\,x4$	*y*	*y*	*y*	*y*
$x2\,x3,\ x4\,x5 = x2\,x4,\ x3\,x5$				
$x2\,x3,\ x5\,x4 = x2\,x5,\ x3\,x4$				
$x2\,x4,\ x5\,x3 = x2\,x5,\ x4\,x3$	*y*			
$x2\,x3,\ x4\,x6 = x2\,x4,\ x3\,x6$				
$x2\,x3,\ x6\,x4 = x2\,x6,\ x3\,x4$				
$x2\,x4,\ x6\,x3 = x2\,x6,\ x4\,x3$	*y*			
$x2\,x4,\ x5\,x6 = x2\,x5,\ x4\,x6$	*y*		*y*	*y*
$x2\,x4,\ x6\,x5 = x2x6,\ x4\,x5$	*y*		*y*	*y*
$x2\,x5,\ x6\,x4 = x2\,x6,\ x5\,x4$	*y*	*y*	*y*	*y*
$x3\,x4,\ x5\,x6 = x3\,x5,\ x4\,x6$	*y*			
$x3\,x4,\ x6\,x5 = x3\,x6,\ x4\,x5$	*y*			
$x3\,x5,\ x6\,x4 = x3\,x6,\ x5\,x4$	*y*	*y*	*y*	*y*

models had generated which data set. His task was to identify the model from the data. The chance of randomly choosing the correct sequence of models is less than 1 in 14 million. In five minutes a TETRAD user identified the sequence perfectly.

It should be noted that the inferences TETRAD makes with such simulated data are much more demanding than the application to longitudinal measurement models requires. In such applications, one knows that the measurements taken at the later time cannot cause the measurements made at an earlier time, and thus one can rule out a priori models in which, say, $x4$ has a direct effect on $x3$. Under appropriate conditions, TETRAD can discriminate among such models even without the information provided by time ordering. This means that in empirical cases in which there are additional direct effects between measurements made at different times, TETRAD can be used to infer the time order from the correlations. Such inferences may sometimes provide a useful nonstatistical test of a model.

3.6. *The TETRAD Program*

The TETRAD program lets the user specify an initial model by means of a *graph of proposed causal relations*. No equations or distribution assumptions need to be described. The error terms do not have to be explicitly entered (the program infers that they are present). An alternative initial model can be specified simply by adding or deleting edges from whatever graph the program is previously given. Thus, a model that would require a page of specifications in an easily used maximum likelihood estimation package, such as EQS, can be described in TETRAD simply by giving a list of the variables between which there are direct causal connections.

The simplification of input and operation is possible because of the purpose of the TETRAD program and because of some mathematical facts. The purpose of the program is not to estimate values of free parameters in a model, since there are already plenty of programs that do that. The purpose of the TETRAD program is instead to explore causal specifications and to discover those causal models that have mathematical properties that are important in explaining the data.

TETRAD helps the user search in the following way:

1. The user provides covariance data and an initial model.
2. TETRAD determines all the vanishing partial correlations and all the vanishing tetrad differences that pass a statistical test at a significance level specified by the user.
3. TETRAD determines all the vanishing partial correlations and all the vanishing tetrad differences implied by the initial model.
4. TETRAD compares the implied constraints with the constraints that hold empirically.
5. TETRAD provides the user with a similar comparison for every elaboration of the initial model that adds one causal connection or correlated error to the initial model.
6. If the initial model is a multiple indicator model, the automatic TETRAD search strategy will find the elaborations of the initial model that imply the same empirically correct constraints as the initial model while implying as few empirically incorrect constraints as possible. This information is given in the form of suggested trek additions to the initial model, and the user must judge how best (if at all) to realize the suggested trek additions by directed edges or correlated errors.

In addition, the information the program provides can be used to search for other models not suggested by the automatic search procedure. In what follows, we describe such searches as *TETRAD-aided.*

We imagine TETRAD to be used in the following way:

1. The researcher uses prior knowledge (e.g., about what the variables mean, how they were measured, how they cluster, etc.) to formulate a class of alternative initial models. The TETRAD program may help in this process by locating constraints satisfied in the data that perhaps ought to be explained.
2. The initial models are elaborated with TETRAD, either using the automatic search component or step by step using the information the program gives at each stage about the properties of one-step additions to a model and using substantive knowledge.
3. The investigator applies whatever is known about the domain to rule out models suggested by TETRAD that contradict established principles or are nonsensical.

4. The remaining models are subjected to statistical test, or if possible to nonstatistical testing.

4. EXAMPLES

In what follows, we present a series of studies of empirical data using the TETRAD program. Most, but not all, of these cases are described in more detail in Glymour et al. (1987), where a variety of other cases are considered in detail. Here, we will simply state the task and the TETRAD results. The examples are chosen because they illustrate a variety of ways in which the program may be used, but they do not begin to exhaust the wealth of kinds of applications for the TETRAD procedures.

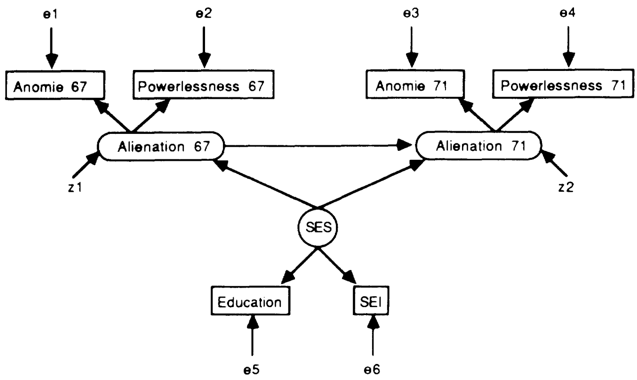
In few of these cases do we mean to endorse the substantive claims of the particular models we consider. The point of the exercise is to illustrate the power of the program and its heuristics in finding alternative elaborations of real models for real data, alternatives that are at least as plausible as published models, that provide comparable or better fit, and that explain constraints found to be satisfied by the data. Details of the models, their interpretation, and the procedures by which the data were obtained can be found in the references given in each case.

4.1. *Finding Omitted Correlated Errors*

A study by Wheaton et al. (1977) on the stability of alienation has become a standard example in manuals for computer programs that perform statistical analyses of structural equation models. Jöreskog and Sörbom (1984) discuss the example in the LISREL manuals, and Bentler discusses the example in the EQS manual. The initial causal model considered by these authors is shown in Figure 7. Alienation is a latent construct measured by anomie and powerlessness at two different times, 1967 and 1971. SES is the latent construct interpreted as socioeconomic status and measured by Duncan's index (SEI) and by an index of educational achievement . The probability of this model's chi-square statistic (with 6 degrees of freedom) is less than 0.01.

Using the LISREL technique for model revision, Jöreskog and Sörbom revise the initial model by freeing two parameters that corre-

FIGURE 7. Alienation: Original model.



spond to correlations for error terms for the same indicator measured at different times. The revised model, shown in Figure 8, has 4 degrees of freedom and $p = 0.335$. The revised model is plausible enough and has acceptable fit, by the usual standards. But are there other models, based on the same initial causal hypotheses, that are substantively plausible and have comparable or better fit?

The automatic portion of the TETRAD program finds one such model, illustrated in Figure 9. The model has one less degree of freedom than the previous revision, but the p value of its chi-square statistic is 0.91. One other model with comparable fit ($p = 0.8$, same degrees of freedom) can be located by a TETRAD-aided search.

4.2. *Alternatives to Regression Models*

Timberlake and Williams (1984) claim that foreign investment in Third World or "peripheral" nations causes the exclusion of various groups from the political process within a peripheral country. Put more simply, foreign investment promotes dictatorships and oligarchies. They also claim that "foreign investment penetration increases government repression in noncore countries" (p. 144). It is clear that such theses, if true, have important policy implications. Timberlake and Williams try to support their first claim by means of a simple regression model.

FIGURE 8. Alienation: Amended model.

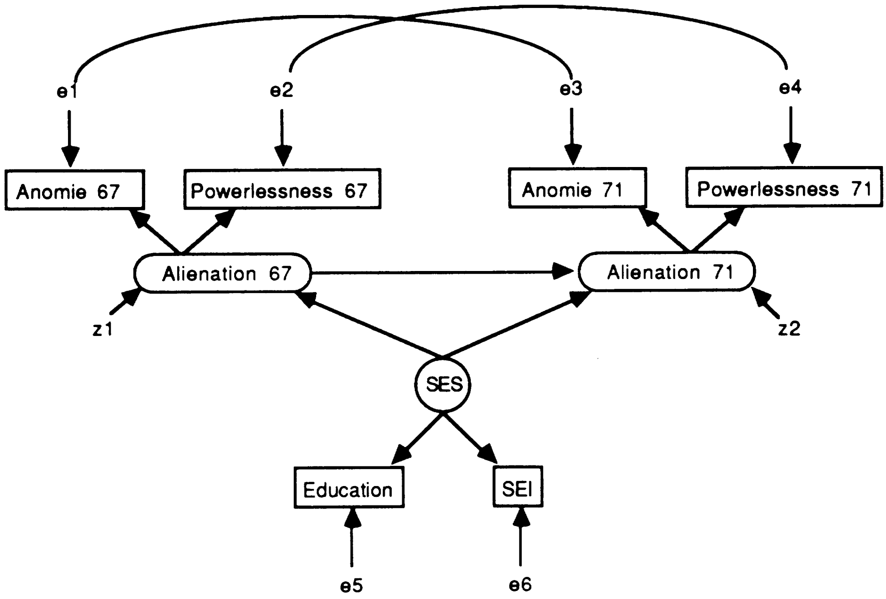
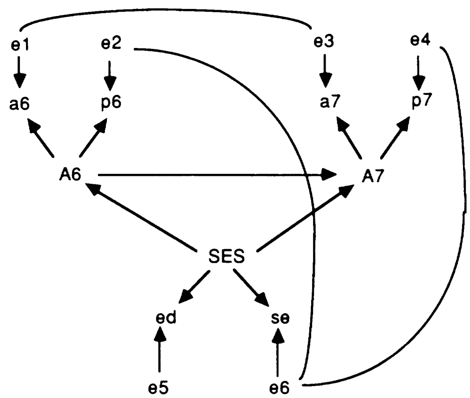


FIGURE 9. TETRAD revision.



Their more complicated argument for the second thesis depends on the correctness of the regression model they propose. We will concentrate on their regression model and its alternatives. In this case, TETRAD does not itself construct alternative models, but it provides the user with information and heuristics that make it easy to find plausible alternatives.

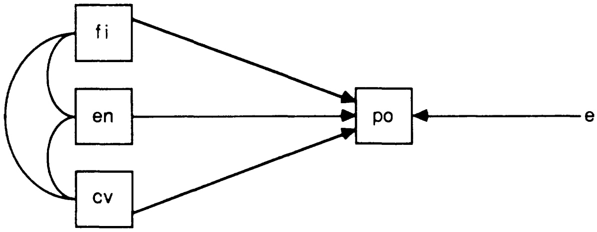
Timberlake and Williams develop measures of political exclusion (*po*), foreign investment penetration (*fi*), energy development (*en*), civil liberties (*cv*), population, and government sanctions and political protests. The last two variables, which do not figure in our analysis, were measured over two time spans (1968–72 and 1973–77), but the other variables were measured for one and the same period; therefore, time of occurrence cannot be used to restrict the causal relations among those variables. Timberlake and Williams correlate these measures for 72 "noncore" countries. All the variables, save population, have substantial positive or negative correlations with one another; absolute values range from 0.123 to 0.864. It should be noted that their investment data concern a period preceding the increase in petrodollars loaned to Third World countries following the dramatic OPEC increases in oil prices.

A straightforward embarrassment to the theory is the finding that political exclusion is *negatively* correlated with foreign investment penetration and that foreign investment penetration is *positively* correlated with civil liberties and negatively correlated with government sanctions. Everything appears to be just the opposite of what the theory requires. The gravamen of the Timberlake and Williams argument is that these correlations are misleading, and when other appropriate variables are controlled for, the effects are reversed.

To sustain their first hypothesis, they regress the political exclusion variable on foreign investment penetration together with energy development and civil liberties (measured on a scale whose increasing values measure decreases in civil liberties) (see Figure 10). They find a statistically significant positive regression coefficient for foreign investment penetration and conclude that their first hypothesis is supported.

Timberlake and Williams thus claim to have found evidence that foreign investment in Third World nations causes governments to be unrepresentative and undemocratic. Their conclusion suggests that the development of democracy and human rights would have been furthered in the early 1970s if international corporations, private

FIGURE 10. Timberlake and Williams's first hypothesis.



banks, and other organizations based in industrial countries had not invested in Third World nations.

There are some puzzling features of the data, which we might expect a good theory to explain. For example, there are in the data some relations among the correlations that hold much more exactly than we expect by chance. Using TETRAD, we find that the following relations hold almost exactly in the sample data:

$$\rho_{po,fi} - \rho_{po,en}\rho_{en,fi} = 0, \tag{1}$$

$$\rho_{en,cv} - \rho_{en,po}\rho_{po,cv} = 0. \tag{2}$$

These equations are interesting exactly because *they are the kind of relationship among correlations that can be explained by causal structure*. Equation (1) can be explained by supposing that the only effects of political exclusion on foreign investment, or of foreign investment on political exclusion, or of any third factor on both political exclusion and foreign investment, are mediated by per-capita energy consumption; one variable affects another only through its effect on energy consumption. More visually, equation (1) will be explained provided the causal connections between political exclusion and foreign investment are as illustrated in Figure 11.

In the same way, equation (2) can be explained by supposing that any correlations between energy consumption and absence of civil liberties are due to the effects of political exclusion; e.g., if increases in per-capita energy consumption cause an increase in civil liberties, they do so because of their direct effect on totalitarianism.

Timberlake and Williams's model does not provide any causal explanation of relations (1) and (2), but it is easy to find assumptions that do explain these patterns, and explain them rather neatly. We

FIGURE 11. Causal explanations of equation (1).

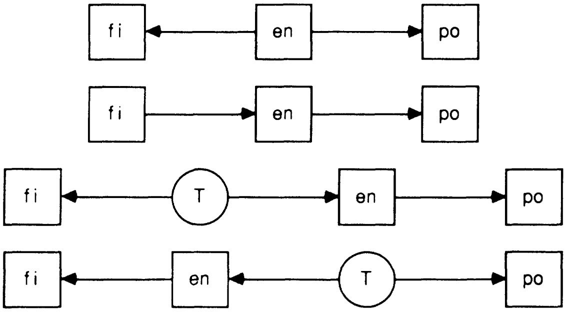


exhibit some alternative explanations in Figure 12. T signifies a latent common cause. The causal hypotheses in all alternatives, under the assumption of linearity, imply that both (1) and (2) hold in the population, *no matter what the values of the linear coefficients may be.*

Model I, for example, gives a chi-square statistic for 2 degrees of freedom with $p = 0.94$. If we accept model I, then we conclude that foreign investment in peripheral nations neither promotes nor inhibits the development of democracy and civil liberties but that raising the energy consumption per capita promotes both foreign investment and more representative government and, through representative government, increases respect for civil liberties. On this data, and given the alternatives, we would not argue that model I should be accepted. We do claim that it, and very likely the alternatives suggested here, are preferable to Timberlake and Williams's regression model.

4.3. *The Stability of Measurement Models*

McPherson et al. (1977) consider a model of responses to a four-item scale assumed to measure indicators of political efficacy or, more clearly, the respondents' judgments of their political influence. Measures of the same four items were obtained from a chart of 978 persons in 1956 and in 1960. Their initial model is shown in Figure 13.

After considerable discussion, the authors conclude that there is a factor acting on $v6$ and $v0$ and some other factor acting on $c6$ and $c0$. The conclusion is based on the fact that the linear coefficients

FIGURE 12. Alternatives to the regression model.

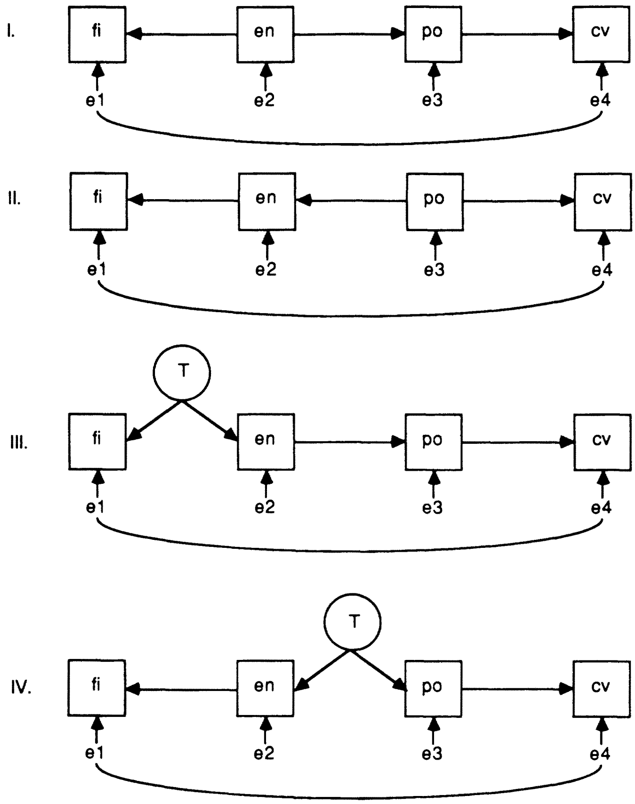
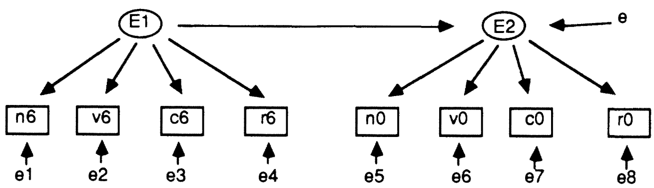


FIGURE 13. McPherson, Welch and Clark's initial model.



connecting these variables with their parent latent variables are the smallest of the eight and the difference between estimated and empirical correlations is the largest for the $v6$–$v0$ and $c6$–$c0$ pairs.

TETRAD automatically suggests that $v6$ and $v0$ must have a further common cause. The program also tells the user that a further common cause of $c6$ and $c0$ will improve fit with a minimal reduction in the number of empirically correct tetrad constraints implied by the model.

The basis for TETRAD's discriminations in this case has already been partly demonstrated in the study with simulated data described in a previous section.

4.4. *Determining Causal Order from Correlations in One-Factor Models*

Kohn (1969) describes several large studies that investigate the relationships among social class, attitudes, and personality structure. For data from five questions intended to measure an authoritarian-conservative personality trait, Kohn suggests a simple factor model, shown in Figure 14. Schoenberg and Richtand (1984) suggest the revision shown in Figure 15. The sample size is larger than 3,000.

One reason for residual correlations in measurement models for survey data may be that responses to earlier questions set a mood or create a desire for consistency and thus, independently of the value of the latent variable the items are intended to measure, influence the responses given to later items (see Campbell et al. 1966). On this assumption, we gave the data and Kohn's initial model to TETRAD and considered only revisions that postulate direct effects between

FIGURE 14. Initial authoritarian-conservatism measurement model.

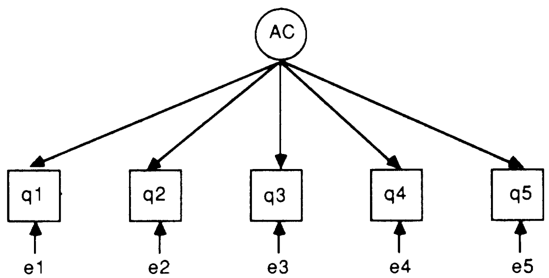
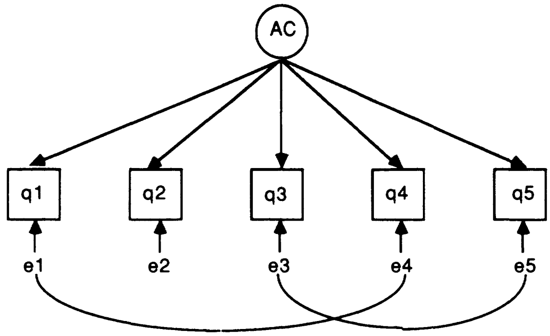


FIGURE 15. Schoenberg and Richtand's revised model.



measured variables. For reasons to be explained shortly, we did so in ignorance of the order in which the five questions had been asked. We found two "best" revisions of the initial model under these assumptions, namely, models that add to the model in Figure 14 either the edges

$$q1 \rightarrow q3, q2 \rightarrow q5, q5 \rightarrow q3$$

or the edges

$$q1 \rightarrow q3, q5 \rightarrow q2, q5 \rightarrow q3.$$

Each of these models gives a chi-square statistic with $p = 0.994$. The hypothesis that the data was generated by one or the other of these models implies restrictions on the order in which four of the five questions were asked. There are 24 possible orderings of four questions, and only eight of them are consistent with the hypothesis, namely,

1-5-3-2
5-1-3-2
1-5-2-3
5-1-2-3
5-2-1-3
1-2-5-3
2-1-5-3
2-5-1-3.

The actual order of the five questions on Kohn's survey is 2-1-4-5-3, which is consistent with the seventh ordering in the list permitted by our hypotheses.

The case provides an unintended illustration of the power of TETRAD's methods to distinguish causal order and illustrates a kind of *nonstatistical* prediction that can sometimes be derived from a model or small set of alternative models. In this case, we took the data from Schoenberg and Richtand's (1984) study. They number the questions in an order different from the order of the questions on Kohn's questionnaire. We discovered, initially to our dismay, that the best TETRAD models were *inconsistent* with the order of the questions given by Schoenberg and Richtand. After consulting Kohn's book, we discovered that the program and the substantive hypothesis about anchoring had led us unwittingly to a correct prediction of the order of the questions.

5. OBJECTIONS AND QUESTIONS

The very idea of using heuristic search in applied statistics creates a variety of misgivings. Many of the objections are mutually inconsistent: No methodology could satisfy all of them. Others are misplaced and suppose that heuristic search procedures impose some condition they do not. Often the misgivings depend on the strictest application of principles that are unscientific in a straightforward sense: If the principles were used in the natural sciences, modern physical science would never have emerged.[5] Since the objections we consider are never given as full, clear arguments, we are forced to reconstruct what we take to be the implicit arguments behind brief remarks, and we therefore avoid attributions. The most frequent objections have no explicit argument at all and amount to simple name calling: Heuristic computerized search is "data dredging" or "ransacking." In so far as any argument lies behind these epithets, we hope one or another of the following considerations may capture it.

5.1. *Data Separation*

Many people believe that a theory or model should always be tested on samples that are distinct from the sample used to generate the model, and they may even hold that the data used to generate the

[5] Most of the objections are considered in more detail in Glymour et al. (1987).

model provides no support for it. While we think the *general* principle erroneous, it is in any case irrelevant to the appropriateness of the TETRAD program. TETRAD automatically separates the data in one respect, since model generation is based on only an *aspect* of the data—the covariance constraints—which contain less information than the full set of covariances. But those who insist on complete sample separation can use the TETRAD program with equanimity simply by searching on one sample and testing the models that result on another sample. We have done exactly that in our studies of the Head Start program.

5.2. *Sample-Dependent Generation*

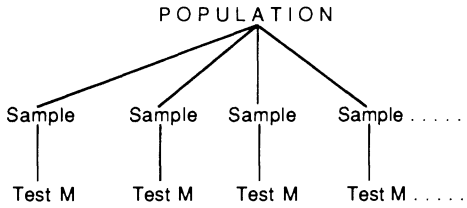
Some writers hold that any procedure in which the model or models generated depends on the sample obtained, and will vary with different samples from the same population, is unacceptable. The prejudice seems to be founded on two very different arguments, which we will treat separately.

Argument 1. No model-generation procedure should be sample-dependent, because in the worst case, sample-dependent procedures will with high probability find spurious dependencies. For example, if one searches the data for correlations, there are cases (with appropriate sample sizes and numbers of independent variables) in which a random sample of values of independent variables will with high probability exhibit a significant correlation for *some* pair of variables. A sample-dependent procedure would therefore with high probability incorrectly conclude that the variables are correlated even though in reality they are independent. Therefore, such procedures should not be used.

For several reasons, the conclusion of this argument does not follow from the case that is imagined:

1. The difficulty can be avoided simply by refusing to make inferences when the sample size is inappropriately small for the number of variables considered. The generation procedure can still be sample-dependent.
2. The difficulty can be avoided by testing any generated model on an independent sample.
3. The worst-case argument against a general procedure fails to consider the *expected* case.

FIGURE 16. Procedure 1.



4. The argument considers only the risk of drawing a false conclusion; it fails to consider the risk of failing to draw a correct conclusion. No argument of such a form for a methodological restriction can be valid.

Argument 2. Model-generation procedures should always be sample-independent; otherwise, test statistics are "meaningless." That is, if model generation is sample-dependent, then the p values of test statistics such as chi square cannot be given a long-run frequency interpretation. The usual long-run frequency interpretation of the p value of a statistic is the frequency with which a value more extreme than the computed value of the statistic would be obtained in a long sequence of tests on samples of the same size drawn at random from a population truly described by the model tested (see Figure 16). The objection is that if probabilities are long-run (or limits of) frequencies, the chi-square statistic for a model obtained for a sample is not the long-run frequency of the following sequence of sampling, model generation, and statistical testing procedures (see Figure 17).

The last point is correct, but the argument contains a logical blunder. The claim of the argument is that no long-run frequency interpretation can be found for the p value of a test statistic of a model obtained by a sample-dependent generation procedure. But the argument shows only that the sequence in Figure 17 does not provide such an interpretation. That is well short of showing that no frequency interpretation can be given. One can. There is a frequency interpretation for the chi-square probability obtained when a model generated by a sample-dependent procedure is tested, and it is the interpretation associated with the sequence shown in Figure 18.

FIGURE 17. Procedure 2.

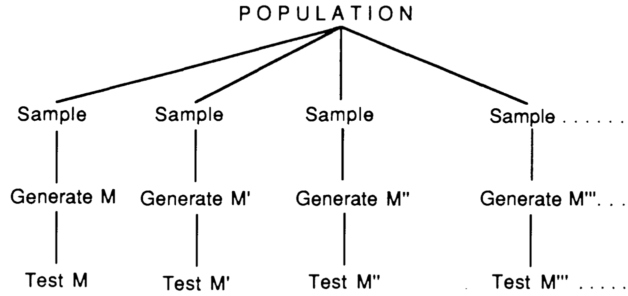
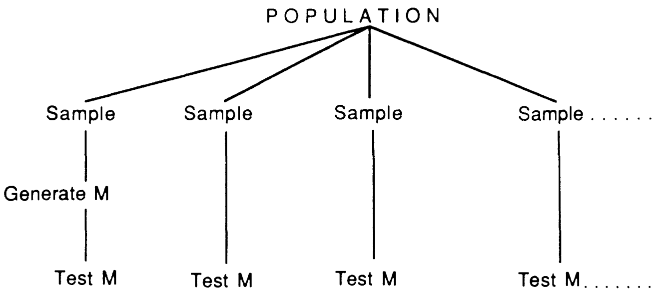


FIGURE 18. Procedure 3.



This sequence gives the correct probability interpretation in the only relevant sense; namely, the long-run frequency corresponds to the p value. Any continued argument that this interpretation is not the "right" frequency interpretation is bootless or tacitly changes the argument completely. The argument, recall, was that there is no frequency interpretation that can be given to test statistics for models obtained with a sample-dependent procedure. But there is such an interpretation, and there is no relevant objection to be made to it that does not apply to frequency interpretations generally.

5.3. *Is TETRAD Fit for Bayesians?*

One question is whether TETRAD is consistent with the philosophical viewpoint of Bayesian statisticians and econometricians. TETRAD does not do explicitly Bayesian calculations; it does not assign prior probabilities and change them by conditioning on the evidence, and that may give Bayesian statisticians qualms. Yet a Bayesian statistician would have no qualms about using a pocket calculator, even though the pocket calculator does not work explicitly on Bayesian principles. Like the pocket calculator, TETRAD determines *mathematical* relationships, relationships that humans cannot conveniently and reliably calculate for themselves. The mathematical relationships are, moreover, relationships that *ought* to matter to Bayesians. A Bayesian statistician may entertain a set of alternative hypotheses to account for a body of data, but if he or she is wise, some prior probability is usually reserved for the catch-all hypothesis that none of the explicitly considered hypotheses is correct. Since the mathematical structure and properties of whatever theories may be contained in the catch-all hypothesis are generally unknown, the Bayesian cannot determine much about the likelihood such hypotheses give to the data or to features of the data. One way of viewing the TETRAD program is that it searches the vast space of alternatives encompassed by the catch-all hypothesis; in its search, TETRAD looks for models that give a high likelihood to constraints on population covariances that are suggested by the data. What it reports are mathematical facts about the existence and properties of such models. In that respect, it is no different from the pocket calculator. The user, whether a self-conscious Bayesian or not, is free to use these mathematical facts in many ways in the light of prior belief, but one would be unwise to ignore them altogether.

5.4. *Hypothesis Testing Only*

Some people seem to think that the only meaningful comparisons of alternative models are those that test one model against the other in Neyman-Pearson fashion. They have no use for comparisons of p values of test statistics of different models on the same data, for comparisons of explanatory power, or for simplicity. But comparisons of the p values of alternative models are perfectly meaningful; they are comparisons of likelihoods. And if the only methodological comparisons permitted were hypothesis tests, most of the history of science would have to be dismissed, along with a great deal of contemporary natural science.

5.5. *Should We Use Theories Suggested by a Computer?*

Someone might take the view that we need to consider only the hypotheses explicitly proposed *by people*: Any hypotheses discovered by computer search can be ignored and thus in effect given zero prior probability. This prejudice may derive from the conviction that, after all, people know a lot more than computers do. They know about how things were measured and what they mean; they know about social practices, about time order, and about prior theorizing. All of that is true, but not really relevant. The special knowledge humans have can interact with the special computational powers computers have in at least two ways. First, humans can use their knowledge to restrict the range of search the computer conducts and to edit and choose from among the results of the computer search. Second, humans can explicitly represent their knowledge within the computer program and let the computer automatically apply that knowledge in conducting a search. The TETRAD program relies on the first procedure rather than the second, but the second procedure is perfectly feasible in searching for causal models.

Sometimes, *part* of discovering a theory can be reduced to combinatorial analysis, and when it can, why shouldn't we have a computer do it? Nothing guarantees us (or even makes it likely) that the truth about some subject matter must be contained in one of the theories entertained by some human at some particular time. There is a world of mathematical possibilities, and we humans cannot very well search through that world unaided; if computers can aid us in the search, why shouldn't we use them? In using the computer to help us search for models with special mathematical properties or special relationships to the data, we are simply using a computational device to help us do the sort of thing that has been fundamental to theory development in the natural sciences. Newton, for example, did not just posit the inverse square gravitational force law without searching for alternatives. A major theorem in Newton's *Principia* characterizes features of the orbits of a body about a central source for *every* inverse power law for the attractive force. For decades, physicists have carried out the same kind of systematic search for alternatives to the general theory of relativity. If, in the course of such a search, a computer were needed to provide numerical approximations to solutions of differential equations, no one would hesitate to use it. In applied statistics, the space of possible linear models is usually so large that humans cannot search it for models that have interesting mathematical relationships to the data. If a computer can help in that search, and TETRAD shows that it can, it would be unscientific not to make use of it.

5.6. *Varieties of Conservatism*

A related doubt about using computer aids for search rests on a kind of conservatism. Someone might think that we should not search for alternative theories unless the most popular current theory has run into trouble. Such a view defies the practice throughout the natural sciences, in which, for example, research programs have for many years attempted systematically to search for alternative theories to quantum mechanics and to general relativity.

Some people think that a body of evidence should never be used to search for more than a single theory. We may be very glad that our scientific ancestors did not think the same way. If they had, we would not have had Copernican astronomy, or Kepler's laws, or modern tables of atomic weights, or indeed most of physical science.

5.7. *Will Computerized Search Make Modeling Practice Better or Worse?*

Any technical innovation can be used badly. If the possibility for misuse were grounds for dismissing a technical innovation, then we should have to do without the telescope, the micrometer, the photographic plate, the cyclotron, indeed virtually every scientific instrument. All have at times been poorly used to produce erroneous conclusions. The same is true of statistical innovations and computational aids. The general principle that says that a technical innovation ought not to be used if it can be used badly is a policy for ending science, not furthering it.

All considered, programs such as TETRAD should produce better, not worse, scientific practice. Perhaps the most striking defect of causal modeling is the difficulty researchers have in considering alternative explanations for their data. That is the chief point in criticisms of causal modeling that point out that investigators often fail to make any reasoned, persuasive case for the structural equations they assume. The difficulty people have in constructing alternatives has two unfortunate consequences. First, we sometimes rush to embrace a causal explanation when other, better explanations exist. Second, the case for a particular causal explanation may sometimes be weaker than it need be, because even when a researcher has found the very best explanation, he or she cannot provide an argument that there are not better alternatives that have been overlooked. TETRAD makes a small contribution toward removing both difficulties. Scientific papers routinely end with a call for further research, but what further research is relevant depends in large part on what alternative explanations need to be distinguished empirically. That kind of question is likely to become more focused if systematic search procedures find wide use.

5.8. *Reliability*

There is no bound to the number and variety of bad arguments and red herrings that can be contrived, and no matter how many are addressed, others will spring up in their place. It is better to focus on the central issue about the TETRAD procedures and about any other discovery procedure, automated or not. The central issue is always this: Where can the procedure be relied upon, and exactly what can it be relied upon to do? That question cannot be answered by methodological dogmas. It can be addressed through systematic proofs about the classes of models whose members are asymptotically distinguishable by constraints of various kinds. It can be addressed by detailed studies of the behavior of the search strategy on simulated data samples of varying sizes, generated from known structures with a variety of probability distributions. And it can be addressed by comparative studies of alternative search procedures, including search procedures that employ no computational aids. Some of the results of such studies have been reported here, and we are conducting further studies of the same kind.

6. FUTURE DEVELOPMENTS

Consider a system that could do all of the following:

1. The user specifies whatever is known about a domain, including the knowledge that specific variables are or are not causally connected, or do or do not have common causes; that connections between

specific variables must be of a specific sign, positive or negative; that specific variables are lags of one another; that specific variables may have a common cause; that specific variables must have symmetric relationships to other variables; that no cycles occur, etc. The information might be as detailed as a specific initial model, or it might be much less definite.
2. The program finds a variety of initial models consistent with the user's specifications and lets the user add to them or eliminate any number of them.
3. The program finds the elaborations of the initial models that best explain patterns in the data, that are simple, and that fit the covariances.
4. The program compares the elaborations of different initial models, eliminating those that are redundant or inferior.
5. The program gives maximum likelihood estimates and chi-square tests for every remaining model, using whatever data set the user specifies.

A program of this kind would make it genuinely feasible for researchers to explore a large space of alternative models. It would leave people free to focus on the aspects of theory construction that good researchers are really good at—finding restrictions on models based on prior knowledge, providing substantive constraints on model specification, assigning a plausible interpretation to latent variables, wondering about measurement procedures, putting a causal story in a more general framework, making policy inferences from a causal explanation—and it would leave to the computer what the computer is really good at—computing mathematical relationships. A program of this kind would be elementary to use, and while it might require considerable computer time to operate, it would require very little of the *user's* time.

TETRAD is not a program of this sort. It is very restrictive in the kind of prior information it uses, and it gives output that is hard to interpret without considerable practice. Nonetheless, we believe the TETRAD program provides the core of a system of the kind just outlined, and we are in the process of carrying out its development and implementation. Thus far, we have done the following:

1. We have written an extension of TETRAD that automatically writes an input file to EQS for any model the user wishes to have

evaluated; the model is given simply by a list of causal edges or correlated errors. The input files for EQS are collected in a batch file. To run EQS on any number of models, the user need only construct the graphs of the models in TETRAD, tell TETRAD it wishes EQS files, and type "run." The result is EQS data on all the models considered. A similar extension can be written for LISREL.

2. We have written a program to implement an algorithm that considers additional constraints besides tetrad equations and vanishing partial correlations. Whether a model implies that a tetrad difference is positive or negative, or that a partial correlation is positive or negative, depends only on the graph of the model and on the *signs* of the linear coefficients. Peter Spirtes has implemented an algorithm that computes these features for a wide class of models, given the graph and the signs. This feature will permit us to modify TETRAD to do a more discriminating search than hitherto possible.

We expect that over the next two years, these pieces can be expanded to a fully automatic system of the kind described and that the system will run successfully on relatively inexpensive personal workstations.

REFERENCES

Anderson, T. 1984. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Blalock, H. M. 1961. *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

———. 1982. *Conceptualization and Measurement in the Social Sciences*. Beverly Hills: Sage.

Campbell, D., R. Schwartz, L. Sechrest, and E. Webb. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.

Costner, H., and R. Schoenberg. 1973. "Diagnosing Indicator Ills in Multiple Indicator Models." Pp. 168–99 in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan. New York: Seminar.

Glymour, C. 1985. "Statistics and Metaphysics." Educational Testing Service Technical Report 85–63. Princeton: Educational Testing Service.

Glymour, C., R. Scheines, P. Spirtes, and K. Kelly. 1987. *Discovering Causal Structure*. San Diego: Academic Press.

Guilford, J. 1936. *Psychometric Methods*. New York: McGraw Hill.

Heise, D. 1975. *Causal Analysis*. New York: Wiley.

Herting, J., and H. Costner. 1985. "Respecification in Multiple Indicator Models." Pp. 321–93 in *Causal Models in the Social Sciences*. 2d ed., edited by H. M. Blalock, Jr. New York: Aldine.

Jöreskog, K., and D. Sörbom. 1984. *LISREL VI User's Guide*. 3d ed. Mooresville, IN: Scientific Software.

Kohn, M. *Class and Conformity*. Homewood, IL: Dorsey.

Magidson, J. 1977. "Toward a Causal Model Approach for Adjusting for Preexisting Differences in the Nonequivalent Control Group Situation." *Evaluation Quarterly* 1:399–420.

Maruyama, G., and B. McGarvey. 1980. "Evaluating Causal Models: An Application of Maximum Likelihood Analysis of Structural Equations." *Psychological Bulletin* 87:502–12.

McPherson, J. M., S. Welch, and C. Clark. 1977. "The Stability and Reliability of Political Efficacy: Using Path Analysis to Test Alternative Models." *American Political Science Review* 71:509–21.

Miller, J., K. Slomczynski, and R. Schoenberg. 1981. "Assessing Comparability of Measurement in Cross-National Research: Authoritarian-Conservatism in Different Sociocultural Settings." *Social Psychology Quarterly* 44:178–91.

Miller, R. 1981. *Simultaneous Statistical Inference*. New York: Springer.

Schoenberg, R., and C. Richtand. 1984. "Application of the EM Method." *Sociological Methods and Research* 13:127–50.

Spirtes, P. 1988. "Fast Geometrical Calculations of Overidentifying Constraints." Unpublished manuscript, Department of Philosophy, Carnegie Mellon University.

Spirtes, P., and C. Glymour. Forthcoming. "Latent Variables, Causal Models and Overidentifying Constraints." *Journal of Econometrics*.

Timberlake, M., and K. R. Williams. 1984. "Dependence, Political Exclusion, and Government Repression: Some Cross-National Evidence." *American Sociological Review* 49:141–46.

Wheaton, B., B. Muthén, D. Alwin, and G. Summers. 1977. "Assessing Reliability and Stability in Panel Models." Pp. 84–136 in *Sociological Methodology 1977*, edited by D. Heise. San Francisco: Jossey-Bass.

Wishart, J. 1928. "Sampling Errors in the Theory of Two Factors." *British Journal of Psychology* 19:180–87.