

## LATENT VARIABLES, CAUSAL MODELS AND OVERIDENTIFYING CONSTRAINTS

Clark GLYMOUR

*Carnegie-Mellon University, Pittsburgh, PA 15213, USA*  
*University of Pittsburgh, Pittsburgh, PA 15260, USA*

Peter SPIRITES

*Carnegie-Mellon University, Pittsburgh, PA 15213, USA*

When is a statistical dependency between two variables best explained by the supposition that one of these variables causes the other, as opposed to the supposition that there is a (possibly unmeasured) common cause acting on both variables? In this paper, we describe an approach towards model specification developed more fully in our book *Discovering Causal Structure*, and illustrate its application to the aforementioned question. Briefly, the approach is to determine *constraints* satisfied by the variance–covariance matrix of a sample, and then to conduct a quasi-automated search for the causal specifications that will best explain those constraints.

### 1. Introduction

One of the most immediate and elementary questions about causal relations in non-experimental or quasi-experimental data is this: when is a statistical dependency between two variables best explained by the supposition that one of these variables acts as a cause of the other, and when is such a dependency instead best explained by the supposition that there is a common cause acting on both variables? When should a common cause that is itself unmeasured (or ‘latent’) be postulated? If the dependency is to be explained by the assumption that one of the measured variables has a causal influence on the other, how in the absence of prior information about that order can it be determined which of the measured variables is the cause and which is the effect?

These questions are implicit in many theoretical and applied studies in econometrics. When, for example, a study of the British economy neglects variables that were unmeasured in British studies but were known to be relevant in American studies, one wonders whether such variables should not appear as latent in econometric models of the British case [see Klein (1961) for an example]. Granger’s work offers theoretical examples. He provides an analysis of causation for time-series data in terms of predictability [Granger (1969)]. For two stationary series,  $x(t)$  and  $y(t)$ , and background knowledge

$D$ , exclusive of the series  $x(t)$ , Granger's analysis is that  $x$  causes  $y$  if and only if the history of the  $x$  series together with  $D$  provides a least-squares predictor of the series  $y$  with smaller variance than the best least-squares predictor conditioned on  $D$  alone. Granger's proposal is reminiscent of the notion of *prima facie* causality found in Suppes (1970):  $x$  is a *prima facie* cause of  $y$  if  $x$  precedes  $y$  in time and  $\text{Prob}(y/x) > \text{Prob}(y)$ . *Prima facie* causality is not causality, as Suppes noted, because the statistical dependence of  $y$  on  $x$  may be due to the action of some third factor,  $z$ , and when conditioned on  $z$ ,  $x$  and  $y$  may be statistically independent. Granger's analysis specifically includes the background  $D$ , but if  $D$  does not include relevant variables, Granger causality may be as spurious as Suppes' *prima facie* causality. Granger noted this feature of his own definition:

'The definition of causality is now relative to the set  $D$ . If relevant data has not been included in this set, then spurious causality could arise. For instance, if the set  $D$  was taken to consist only of the two series  $X_t$  and  $Y_t$ , but in fact there was a third series  $Z_t$  which was causing both within the enlarged set  $D' = (X_t, Y_t, Z_t)$ , then for the original set  $D$  spurious causality between  $X_t$  and  $Y_t$  may be found. This is similar to spurious correlation and partial correlation between sets of data that arise when some other statistical variable of importance has not been included.'

Evidently it would be important and useful to have statistical criteria that indicate when correlations or autocorrelations among measured variables are best explained by common causes, even by unmeasured or latent common causes.

Granger did not claim that his account of causality for time series specifies the *only* context in which causal explanations are appropriate. Although he argues persuasively (see his contribution to this issue of the *Journal of Econometrics*) that instantaneous causality has no role in econometrics, it remains the case that for some bodies of economic and other social scientific data no time order is known. Granger's or Suppes' or some other account of causality that makes use of time order may truly describe the process that generates such data, but if the time order is not known, the accounts themselves cannot be used to determine the causal relationships among the variables. Even in these cases, a causal explanation is generally wanted. For example, questions about the direction of a causal relation arise in the use of regression models for data not ordered by time. In these cases regression of measured variables on measured variables often presupposes a causal ordering that is usually not tested even indirectly, and it is natural to seek strategies for determining the direction of causality from data and from background assumptions. The same questions arise if one rejects probabilistic analyses of causality altogether and instead endorses the sort of counterfactual analyses of

causality that have for some time been popular in the philosophical literature [see Lewis (1973)] and are gaining currency among statisticians [see Holland (1985)].

The aim of this paper is to describe an approach towards model specification that we have developed more fully in our book (with R. Scheines and K. Kelly, *Discovering Causal Structure*), a strategy that is partially implemented in the TETRAD program which accompanies that book and to illustrate its application to the kinds of questions that have just been raised. In briefest form, the approach is to determine *constraints* satisfied by the variance-covariance matrix of a sample and then to conduct a quasi-automated search for the causal specifications that will best explain those constraints. We apply to social scientific data and theory construction the same sensibilities about good scientific explanation that have contributed to the historical successes of the natural sciences.

This paper describes some patterns of measured correlations or autocorrelations that indicate when apparent causal dependencies in static or longitudinal data may be spurious and due to neglected variables. We also describe some circumstances in which background information and data determine that statistical dependencies are best explained by direct effects among measured variables and determine the order of the causal relationship. The larger purpose of this paper is to describe a program of research aimed at searching for other patterns of the same kind. To that end we raise a number of central mathematical questions that remain open. While our illustrations are not from time-series models, we describe the problems and prospects for extending our methods to such models.

## 2. Scientific explanation and latent variables

Any theory based on the conclusion that correlations or autocorrelations of measured variables are due to some unmeasured cause will contain unmeasured, or latent, variables other than error terms. Although latent variables have found a role in econometric discussions [Aigner (1977)], their value is often disputed by social scientists. If correlations are attributed to some latent variable or variables, one typically will not know what those variables are, or how to measure them directly or manipulate them. Models that contain such variables may therefore seem of limited use in forecasting or in controlling dependent variables.

Nonetheless, provided there are no computational difficulties involved, one is always better off in forecasting and control with a correct model than with an incorrect one, or with a better approximation than with a worse. If measured correlations or autocorrelations are due to unmeasured common causes, one cannot be the worse off for knowing as much, or for considering

models that are in accord with this conclusion. In the natural sciences, the discovery that a regularity can be explained by introducing an unmeasured factor is a basis for scientific progress; the factor is investigated, methods of measuring it are pursued, generalizations are developed about its properties and relations. The same procedures deserve a place in econometrics and other social sciences.

The importance of unmeasured variables in the natural sciences is in the first instances for *explanation*. Sometimes the introduction of latent or 'theoretical' quantities enables us to explain regularities among measured variables in a powerful way. The history of science is full of examples. Atoms and their weights enabled 19th century chemists to explain empirical laws about chemical combination, about vapor densities, about specific heats, and a variety of other phenomena. The motions of the planets and the earth in three-dimensional space around the sun could not be observed by Copernicans, but the hypothesis of such motions enabled them to explain a variety of regularities about the apparent motion of the planets on the celestial sphere. In the long run, good explanation leads to good prediction. Copernican theory, introduced originally for its explanatory properties, led to new and correct predictions of the phases of Venus, the apparent motion of 'fixed' stars, and other phenomena. By the end of the 19th century the atomic theory had generated a wealth of new predictions and further explanations.

There is nothing mysterious about 'latent' variables: they are simply variables that, for whatever reason, one failed to measure. They are not necessarily 'unobservable', whatever that means; they are variables that were not in fact *observed*, for whatever reason. Empirical studies in the social sciences are almost always confined to a restricted set of variables, and one can almost always think of potentially relevant variables that were not measured in any given study. Any policy that rejects all theories containing latent variables is a policy that ignores the reality of social scientific data collection and confines us to explanations that are often known to be erroneous. The only real objection to theories with latent variables is that researchers are unclear as to when such variables should or should not be introduced. The natural sciences provide some guidance about that question.

### 3. Explanation and parameters

What makes for a good scientific explanation? We suggest that one important criterion is that a theory be able to account for patterns in the data and that it do so without specifying particular, accidental values of adjustable parameters. Such a criterion has had a powerful role in the history of the natural sciences. The principal argument Kepler gave for the superiority of Copernican to Ptolemaic astronomy was that Ptolemy's theory requires specific

values of adjustable parameters to save the phenomena and Copernican theory does not. Arthur Eddington gave exactly the same argument for the superiority of General Relativity to Newtonian celestial dynamics. Richard Feynman's autobiography describes his abandonment of a theory of his own, even though it saved various phenomena, because it did so only by adjusting too many arbitrary parameters. The same view has a long history in psychometrics; it was the principal methodological tool of one of the founders of the subject, Charles Spearman.

A more philosophical version of the same idea is that explanation consists on the reduction of *contingency*. A good explanation is one that shows that feature of the data that appears arbitrary, and seems as though it could easily have been otherwise, could not easily have been otherwise.

We apply this same sensibility to linear models of non-experimental data. The theories we consider include path analytic models, 'structural equation' models, factor models, ARMA models, and linear transfer models. We locate patterns of correlations or autocorrelations that can be explained by *some models* no matter what the values of the free parameters in these models. The patterns are in fact traditionally described in the econometrics literature as *overidentifying constraints*. We claim that models that explain overidentifying constraints in this robust way are preferable to models that can only be fitted to the constraints by specifying particular values of the linear coefficients, variances, etc. that occur in their equations. In various cases, the models that robustly explain patterns in the data *must* contain latent variables.

It turns out that models that robustly explain overidentifying constraints generate those constraints entirely from the assumption of linearity and the causal structure the model assumes.

Unfortunately, our techniques do not work on non-linear models, and hence should only be applied after a researcher has some reason to believe that the relationships between variables are at least approximately linear.

#### 4. Linear models and directed graphs

It is a common practice in econometrics to represent linear models by graphs. See, for example, Ancot and Duru (1984), Boutillier (1984), Gaberly and Gilli (1977), Fontela and Gabus (1974), Fontela and Gilli (1977), Gilli and Rossier (1981), and Warfield (1976). The methods we use to establish connections between latent variables and overidentifying constraints are graph-theoretic, and so we begin by describing the relations between directed graphs and linear stochastic models. We consider the 'static' case, which may contain lagged variables, but in which all variables, lags included, are treated as random variables from a single multivariate distribution.

Consider any linear model given by a simultaneous set of linear stochastic equations for the  $k$ th population unit by

$$(I) \quad X_i = \sum a_{ij} X_j + e_i,$$

where the matrix  $a_{ij}$  is random, possibly with some components constant, and all components are independently distributed. Many additive statistical models used in the social sciences are specializations of this general model, obtained by adding constraints of the  $a_{ij}$  matrix, the vector  $e_i$ , the covariances of  $X_j$  and  $e_i$ , and on the distributions of these several parameters for distinct individuals in the population.

We consider models that extend (I) not only by adding further distribution assumptions of various kinds, but also by adding a further independent mathematical structure, a *directed graph*. The vertices of the graph are variables of the model, and a directed edge  $\langle X_i, X_j \rangle$  from variable  $X_i$  to  $X_j$  is understood as the causal hypothesis that a variation in  $X_i$ , with all other variables (save  $X_j$ ) held constant, produces a linear variation in  $X_j$  that is not mediated by any other variable in the system. We also permit the graph to contain undirected edges, which signify a covariance that is given no particular causal interpretation. We assume classes of linear models to be paired with graphs in such a way that much of the structure shared by linear models of the class can be recovered from the directed graph alone. In particular, the pairing is such that the structure of the matrix  $a_{ij}$  and any zero covariances of the  $e_i$  variables can be recovered from the directed graph. Such a pairing can be obtained from two simple principles that use only very elementary graph-theoretic ideas.

We say that the *indegree* of a vertex or variable in a graph is the number of edges directed into it, the *outdegree* the number of edges directed out of it. Two vertices connected by a directed edge are said to be *adjacent*. A *path* is a sequence of directed edges such that the second vertex in the  $n$ th member of the sequence is the first vertex of the  $n + 1$  member of the sequence, if there is an  $n + 1$  member. The *source* for a path is the first vertex of the first directed edge in the path; the *sink* of a path is the second vertex in the last directed edge in the path. A path is *cyclic* if it contains a subpath whose first and last vertices are the same. Otherwise it is *acyclic*. A graph is cyclic or acyclic accordingly as it does or does not contain a cyclic path. A *trek between  $x$  and  $y$*  is either an acyclic path from  $x$  to  $y$ , an acyclic path from  $y$  to  $x$ , or a pair of acyclic paths with the same source, whose sinks are  $x$  and  $y$ , respectively, and which intersect only at the source. The *digraph form* of a graph containing one or more undirected edges is the graph obtained by replacing each undirected edge with a new vertex and a directed edge from that new vertex to each of the vertices connected by the original undirected edge. Two vertices in a graph are

*trek-connected* if and only if the digraph form of the graph contains a trek between them.

Now the pairing of graphs and linear models is obtained by two principles that permit one to associate a class of linear models with the digraph form of a graph:

1. Every variable  $X_i$  has  $a_{ij} = 0$  and constant if and only if there is no directed edge from  $X_j$  to  $X_i$ .
2. Two variables that are not trek-connected are statistically independent.

These principles do not determine whether the non-zero  $a_{ij}$  are random or constant, they do not determine whether the error variables are heteroscedastic, nor do they determine whether a variable for one population unit is correlated with that variable for another population unit. Thus a graph is consistent with many alternative further specifications.

Henceforth, when we speak of a 'model' we will mean a graph, and any further set of specifications consistent with the graph and the two principles just stated. We will consider only graphs with the following property:

Every vertex not of zero indegree is adjacent to a vertex of zero indegree and unit outdegree.

This is the graph-theoretic version of the assumption that every endogenous variable has a unique exogenous source of variance, or 'error' source.

## 5. Explanation, resiliency and overidentifying constraints

Our purpose is to describe relations between overidentifying constraints and the existence of latent variables. By an overidentifying constraint we shall mean any constraint on the population correlation or autocorrelation matrix. A linear model  $M$  *implies* an overidentifying constraint provided the constraint is entailed by every set of values of the random coefficients in every model  $M'$  differing from  $M$  at most in the values of the coefficients  $a_{ij}$  that are not specified to be zero and constant in  $M$ . We will say that a model *positively implies* an overidentifying constraint if it implies the constraint when only positive values of the coefficients are considered.

If a model implies an overidentifying constraint that does in fact hold in the population, and is satisfied approximately in a sample, then the model provides an explanation of that feature of the population and the sample. We have two fundamental methodological assumptions:

- Other things being equal, a model that implies empirically satisfied constraints is preferable to a model that does not have such implications.

- Other things being equal, a model that does not imply (or if the coefficients are known to be positive, positively imply) constraints that are not satisfied by the data is preferable to a model that does imply (or positively imply) such constraints.

Some models may yield correlations that agree with their constraints only for particular values of their non-zero coefficients, while other preferable models may imply the constraints for all values of their linear coefficients. Other things equal, we regard the second sort of explanation as preferable. Our view implies that in linear models zero coefficients, which indicate no direct causal connection between a pair of variables, have a special status. The justification of that view is natural but not rigorous. On the one hand, if zero were not regarded as special, then any system of linear equations would be indefinite, because to any equation an unbounded sequence of dependencies on other variables, but with zero coefficients, could be added. Physical theories are not regarded as having such commitments, and it therefore seems unreasonable and unpromising to make social theories so indefinite. Further, a zero coefficient represents more than just a particular value on a scale; in linear causal models it marks an important qualitative distinction between pairs of variables that have a direct effect one on the other and pairs of variables that are not so connected. Finally, as we noted earlier, the preference is justified by a history of successful practice in the natural sciences.

We will consider several kinds of overidentifying constraints on the population covariance matrix, including vanishing partial correlations, vanishing tetrad differences, positive partial correlations, and positive tetrad differences. (The only type of partial correlation constraints that TETRAD calculates are those in which only one variable is held constant, i.e., of the form  $\rho_{ij.k}$ . We will also denote the correlation of  $x_i$  and  $x_j$  as  $\rho_{ij}$ .)

The partial correlation of  $X_i$ ,  $X_j$  with respect to  $X_k$  is

$$\rho_{ij.k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{(1 - \rho_{ik}^2)^{1/2}(1 - \rho_{jk}^2)^{1/2}}.$$

A tetrad difference is just the determinant of a  $2 \times 2$  submatrix of the covariance matrix:

$$\rho_{ij}\rho_{kl} - \rho_{ik}\rho_{jl}.$$

Two forms of constraint on the covariance (or correlation) matrix are obtained by specifying either that a tetrad difference or partial correlation vanishes. Constraints in the form of inequalities can be obtained by specifying that a partial correlation or tetrad difference is positive, or that one partial correlation is greater than another, or that one tetrad difference is greater than another. Constraints of each of these kinds can be implied, or positively



implied, by appropriate linear models. There are many other forms of constraint that we shall not consider here.

Partial correlations are distributed as correlations. Fisher's  $z$  transformation therefore permits an asymptotic test, based on the normal distribution, of the hypothesis that a partial correlation vanishes, and likewise that it is positive. The exact distribution of tetrad differences does not seem to be known, but the formula for the sampling variance of a tetrad difference was obtained by Wishart (1928/9). Wishart's formula permits an asymptotic test of the hypothesis that a tetrad difference vanishes and of the hypothesis that a tetrad difference is positive.

The fundamental mathematical fact upon which our methods rely is that the tetrad equations and vanishing partial correlations implied by a model can be computed from the graph of the model alone. One does not need to know the variances or other distributional properties of the variables in the model, only the causal structure hypothesized. The same is true for the positive tetrad differences and positive partial correlations positively implied by a model. Some of these facts have been known in a more or less tacit way by many social scientists since the twenties; the first proof and actual algorithm for computing the equalities is presented in Glymour et al. (1987).

## 6. The basic idea

Consider four measured variables ( $x_1$ ,  $x_2$ ,  $x_3$  and  $x_4$ ) and suppose the variables take their values at different times. Since the variables are ordered by time, the graph of causal relations among them cannot be cyclic. Suppose further that a tetrad difference vanishes, for example:

$$\rho_{13}\rho_{24} - \rho_{14}\rho_{23} = 0.$$

Now we appeal to a mathematical fact.

*Any acyclic graph of causal relations among just four variables (and their associated error variables) that implies a vanishing tetrad difference also implies a set of vanishing partial correlations.*

For example, the graph in fig. 1 implies that the tetrad difference vanishes and does not imply that any other tetrad difference vanishes, but it also implies the following set of vanishing partial correlations:

$$\rho_{13.2} = \rho_{14.2} = \rho_{14.3} = \rho_{24.3} = 0.$$

Now suppose that no partial correlations among the variables just listed does in fact vanish, and that the only tetrad difference that vanishes is  $\rho_{13}\rho_{24} - \rho_{14}\rho_{23} = 0$ . Then among models with just the four variables listed

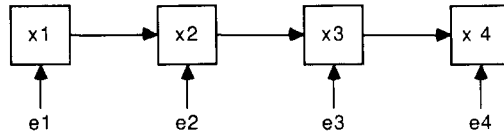


Fig. 1

above (and their associated error variables) there is no robust explanation of the vanishing tetrad difference that does not falsely imply that a partial correlation vanishes. Hence, if there is a model that does provide a robust explanation of the vanishing tetrad difference that holds in the data, and does not imply any vanishing tetrad differences or vanishing partial correlations that do not hold in the data, it must contain a latent variable. And as a matter of fact, there are models containing latent variables that do imply the vanishing tetrad difference that hold in the data, and do not falsely imply any tetrad or partial constraints that don't hold in the data.

### 7. Heuristics for model construction

The preceding section illustrates a combination of correlation constraints whose respective presences and absences in the population (or sample) can only be explained by linear causal models that contain additional variables beyond those involved in the constraints. This information does not itself tell us much about just what sort of causal structure will imply various tetrad constraints but not imply various vanishing partial correlations. Fortunately, there are further considerations that do.

It is easily shown that if a model implies an appropriate set of vanishing partial correlations, it also implies a vanishing tetrad difference. If the graph of a model implies that for some variable  $v$ :

$$\rho_{ij.v} = \rho_{kl.v} = \rho_{ik.v} = \rho_{jl.v} = 0,$$

then the graph also implies that

$$\rho_{ij}\rho_{kl} - \rho_{ik}\rho_{jl} = 0.$$

The converse of this claim is not true. There are models that imply a tetrad equation without implying a set of vanishing partial correlations that imply the tetrad equation. The cyclic graph in fig. 2 (in which the depiction of error variables is omitted for simplicity) implies a vanishing tetrad difference  $\rho_{15}\rho_{23} - \rho_{25}\rho_{13} = 0$ , but does not imply  $\rho_{ij.k} = 0$  for any distinct vertices  $i, j, k$ .

There are also acyclic models that imply a tetrad equation without implying a set of vanishing partial correlations that imply the tetrad equation. The graph in fig. 3 implies the tetrad equation  $\rho_{14}\rho_{23} - \rho_{24}\rho_{13} = 0$ , but does not imply that  $\rho_{24.k} = 0$ ,  $\rho_{13.k} = 0$ ,  $\rho_{14.k} = 0$ , and  $\rho_{23.k} = 0$ , for any  $k$ .

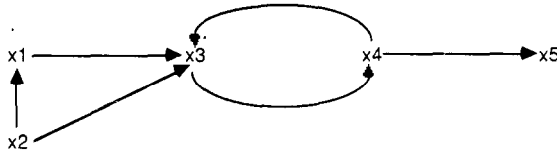


Fig. 2. A counterexample to the converse of Theorem 4.

Note that in both of the counterexamples and in all of the graphs that we have examined the following holds true: a graph  $G$  implies a tetrad equation of the form

$$\rho_{ij}\rho_{kl} - \rho_{ik}\rho_{jl} = 0,$$

if and only if there exists a vertex  $q$  such that every pair of treks between  $i$  and  $j$ ,  $k$  and  $l$ ,  $i$  and  $k$ , and  $j$  and  $l$ , respectively, intersect at  $q$ . We conjecture that this is true for every graph.

We have elsewhere proved the following theorem:

*In an acyclic graph, if all of the equation coefficients are positive and every trek between  $i$  and  $j$  contains  $k$ , then  $\rho_{ij,k} \leq 0$ .*

If the previous conjecture is true, then every graph that implies a tetrad equation also implies that there exists a vertex  $q$  such that  $\rho_{ij,q} \leq 0$ ,  $\rho_{k_l,q} \leq 0$ ,  $\rho_{ik,q} \leq 0$ , and  $\rho_{jl,q} \leq 0$ .

Thus we propose a heuristic principle, when the equation coefficients are known to be positive and when a tetrad difference vanishes but no measured partial correlations are less than or equal to 0:

*Introduce latent variables that imply that there exists a variable  $q$  such that  $\rho_{ij,q} \leq 0$ ,  $\rho_{kl,q} \leq 0$ ,  $\rho_{ik,q} \leq 0$ , and  $\rho_{jl,q} \leq 0$ .*

Of course, the partial inequalities in this case will not be observed; measured correlations occurring in the tetrad equation will be partialled on some latent variable.

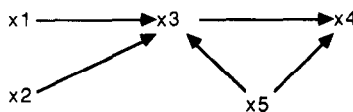


Fig. 3

## 8. An example

Consider data from a longitudinal study of the performances of 799 school girls on the Scholastic Aptitude Test. The same cohort of students took the test in the 5th, 7th, 9th, and 11th grades [see Jöreskog in Magidson (1979)]. The variance-covariance matrix is:

$q_5$	67.951			
$q_7$	71.01	141.578		
$q_9$	85.966	134.748	249.748	
$q_{11}$	97.153	151.068	218.757	300.669

Using the TETRAD program, we compute for each possible vanishing tetrad difference and each possible vanishing partial correlation the probability of the sample difference on the hypothesis that the population difference is zero. The current version of TETRAD does not calculate whether a partial correlation is greater than zero in the population. However, since all of the sample partial correlations have a probability of zero given that the partial correlation in the population is zero, and they are all positive, we conclude that each partial correlation is greater than zero in the population. The relevant output is:

Tetrad equation	Residual	$P(\text{diff.})$
$q_5 q_7, q_9 q_{11} = q_5 q_9, q_7 q_{11}$	0.0953	0.0000
$q_5 q_7, q_{11} q_9 = q_5 q_{11}, q_7 q_9$	0.0917	0.0000
$q_5 q_9, q_{11} q_7 = q_5 q_{11}, q_9 q_7$	0.0036	0.7580
Partial	Residual	$P(\text{diff.})$
$q_5 q_7. q_9$	0.4806	0.0000
$q_5 q_7. q_{11}$	0.4542	0.0000
$q_5 q_9. q_7$	0.2931	0.0000
$q_5 q_9. q_{11}$	0.2655	0.0000
$q_5 q_{11}. q_7$	0.3177	0.0000
$q_5 q_{11}. q_9$	0.3379	0.0000
$q_7 q_9. q_5$	0.4595	0.0000
$q_7 q_9. q_{11}$	0.3208	0.0000
$q_7 q_{11}. q_5$	0.4742	0.0000
$q_7 q_{11}. q_9$	0.3819	0.0000
$q_9 q_{11}. q_5$	0.6347	0.0000
$q_9 q_{11}. q_7$	0.5763	0.0000

The discussion of the previous sections suggests that any adequate linear causal model for this data must contain a latent variable. If, for example, we

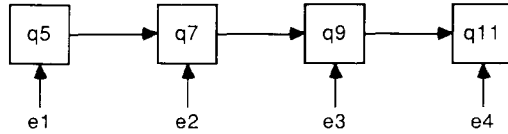


Fig. 4. Path model.

attempt to model the data by supposing that each measurement is a direct cause of the succeeding measurement (e.g., the path model in fig. 4), then the model implies the third tetrad equation and no other vanishing tetrad differences, but it also implies several vanishing partial correlations, for example, that the correlation of  $q5$  and  $q11$  vanishes when partialled on  $q9$ . If the model is modified to avoid these incorrect constraints, for example by correlating the error terms or by introducing further direct effects between earlier and later measurements, then the third tetrad equation is no longer implied.

If we attribute the correlations to the action of a latent variable or variables, then no vanishing partial correlation among the four observed variables will be implied (as long as the partial correlation involves three distinct variables). Unless, however, the latent structure is chosen carefully, the wrong tetrad constraints will be implied. For example, if the data are explained by postulating a single latent variable ('test-taking ability' or whatever) and permitting it to have different linear effects on the several administrations of the SAT, then we obtain the model shown in fig. 5. This model implies all three tetrad equations for the four measured variable, and judged from the sample, two of these implications are incorrect.

The heuristic principle tells us more about the latent variable structure. There must be a latent variable such every trek between  $q5$  and  $q9$ , between  $q7$  and  $q11$ , between  $q5$  and  $q11$  and between  $q7$  and  $q9$  passes through that variable. But, in the simplest case, there must be treks between  $q5$  and  $q7$ , and between  $q9$  and  $q11$ , that do not pass through any latent variable common to

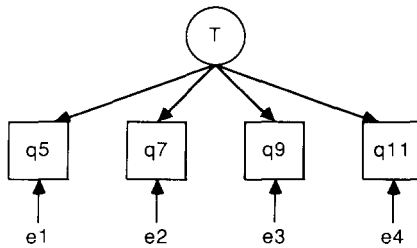


Fig. 5. Factor model.

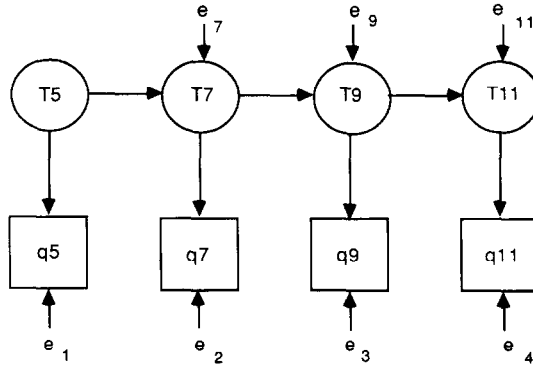


Fig. 6. Single, self-lagged latent variable model.

all treks among the other pairs. An adequate model is obtained if we treat the latent variable as itself lagged (see fig. 6). This model implies the tetrad equation found to hold in the sample, and only that tetrad equation. It does not imply that as long as the equation coefficients are positive any partial correlations among the measured variables are less than or equal to 0. These implications are unaltered even if we specify that the linear coefficients connecting the latent variable lags are all equal to one another, and even if, in addition, we specify that the linear coefficients from the latent variable lags to the measured variables are all equal to one another. Note finally that a model with the correct implications, and none of the incorrect ones, could equally be obtained by using only a latent variable and one lag, although the result intuitively seems less plausible (see fig. 7).

Although the example that we have discussed involved only four measured variables, the heuristic can be applied to models with any number of measured variables.

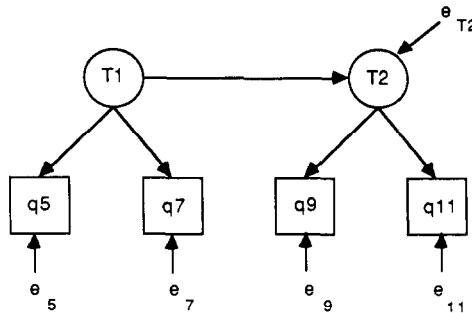


Fig. 7

### 9. Causal order and correlation

The question with which we began was: when is a statistical dependency between two variables,  $X$  and  $Y$ , due to a direct effect of  $X$  on  $Y$ , a direct effect of  $Y$  on  $X$ , or a common cause acting on both  $X$  and  $Y$  or some combination of these causes? In general the statistical dependency of  $X$  and  $Y$  alone does not provide sufficient information to answer this question. If, however, there is appropriate prior knowledge about the causal structure in which  $X$  and  $Y$  are imbedded, knowledge which itself does not answer the question at issue, then the statistical dependencies of  $X$  and  $Y$  on each other and on other variables can provide an unambiguous answer. Such prior knowledge may be either structural or substantive. For example, if  $X$  measures GNP in 1967 and  $Y$  measures GNP in 1975, clearly  $Y$  cannot cause  $X$ . Prior substantive knowledge might also be based on common sense; for example, occupation is a cause of income.

Suppose one correctly assumes that the causal relations among six measured and two unmeasured variables include the relations shown in fig. 8. In this case it does not matter whether variables  $T1$  and  $T2$  are measured or latent. Suppose further that there is an unknown relation between  $x2$  and  $x4$ : either  $x2$  causes  $x4$  or  $x4$  causes  $x2$  or there is a further common cause of both. If appropriate sample data are available for the  $x_i$ , the unknown relation can be reliably discovered, because each alternative elaboration of the initial model above produces a model that implies a distinctive set of vanishing tetrad differences.

Suppose that there is an unknown causal relation between *some indicator of  $T1$*  and *some indicator of  $T2$* , but one does not know which. Again, with appropriate sample data, the unknown causal relation can be identified unambiguously. In simulation studies with normal distributed variables and sample sizes of 2000, using the TETRAD program we have correctly identified the missing relation five times out of five. The probability of such an identification by chance is less than one in fourteen million.

The technique works with real data quite as readily as with simulated data. McPherson et al. (1977) consider a causal model of responses to a four-item

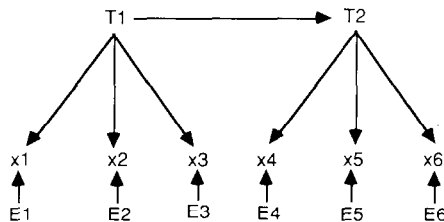


Fig. 8

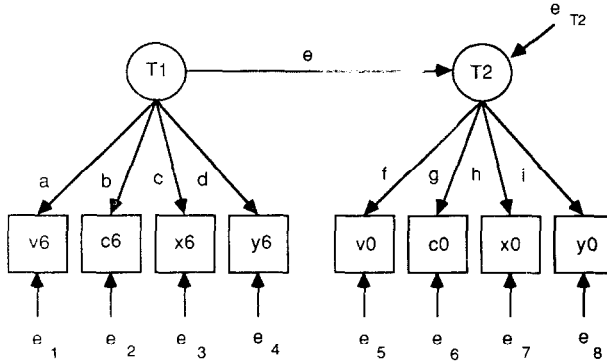


Fig. 9

scale assumed to measure 'political efficacy', or more exactly, the respondents' judgments of their political influence. Measures of the same four items were obtained from a cohort of 978 persons at a four-year interval, once in 1956 and once in 1960. The initial model looks like fig. 9. After considerable discussion the authors conclude that there is some other factor acting on  $c6$  and  $c0$  and some other factor acting on  $v6$  and  $v0$ ; the conclusion is based on the fact that the linear coefficients 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.

The TETRAD program reaches the same conclusions automatically from a comparison of overidentifying constraints satisfied by the data, constraints implied by the initial model, and constraints implied by various modifications of the initial model. Moreover, the program distinguishes the preferred elaboration, which postulates further common causes of  $v6$  and  $v0$  and of  $c6$  and  $c0$ , respectively, from models that postulate direct causal connections (which might arise if recollection of previous responses was a determinant of later responses) between the responses given at the two times.

## 10. Alternatives to regression

Regression methods are widely used in the causal analyses of non-experimental data, often without any consideration of alternative models. The analysis of overidentifying constraints can lead to the identification of alternative models that afford a better explanation of the data, and such alternatives may sometimes contain latent variables. We will illustrate the point with a recent study of the social effects of international economic policies.



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 in third-world nations. They support their claim by means of a simple regression model. Timberlake and Williams develop measures of political exclusion, foreign investment penetration (in 1973), energy development, civil liberties, population, government sanctions in two years (1972 and 1977) and political protests in those same years. They correlate these measures for 72 'non-core' countries. All of the variables, save population, have substantial positive or negative correlations with one another, with absolute values ranging 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 that political exclusion is *negatively* correlated with foreign investment penetration, and 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.

Timberlake and Williams 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 fig. 10. They find a statistically significant positive regression coefficient for foreign investment penetration and conclude that their hypothesis is supported. Their conclusion implies that the development of democracy and human rights would have been furthered in the early 1970s if international corporations, private banks and other organizations based in industrial countries had not invested in third world nations.

The analysis *assumes* that political exclusion is the effect of the absence of civil liberties, of energy development and of foreign investment. They further

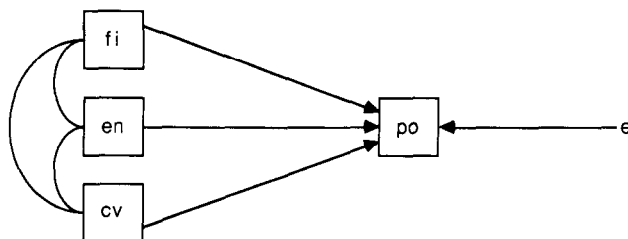


Fig. 10. Foreign investment.

assume that these causes act independently, that their effects are additive, and that nothing else has an effect on any of the independent variables and on political exclusion. They give no particular reasons for these assumptions, and one might have thought otherwise. For example, one might have thought that unrepresentative government causes an absence of civil liberties, or each causes the other.

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:

$$(A) \quad \rho_{po,fi} - \rho_{po,en}\rho_{en,fi} \approx 0,$$

$$(B) \quad \rho_{en,cv} - \rho_{en,po}\rho_{po,cv} \approx 0.$$

We find, again using TETRAD, that if the constraints (A) and (B) hold, then the probability of obtaining a difference at least as large as that found in the sample for (A) is 0.868 and the probability of obtaining a difference at least as large as that found in the sample for (B) is 0.800. These numbers help convince us that (A) and (B) are real constraints on the measured variables and should, if possible, be explained.

These two equations are interesting exactly *because they are the kind of relationship among correlations that can be explained by causal structure*. The first equation 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, the first equation will be explained provided the causal connections between political exclusion and foreign investment are illustrated in fig. 11. In the same way, the

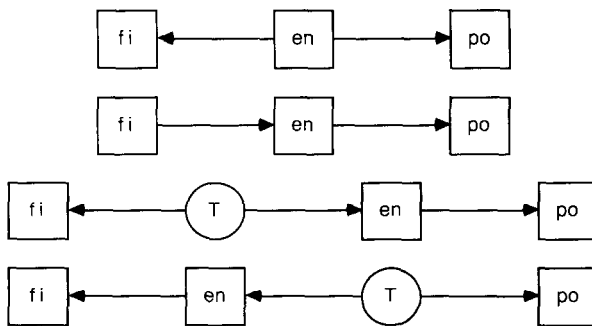


Fig. 11. Causal explanations of eq. (A).

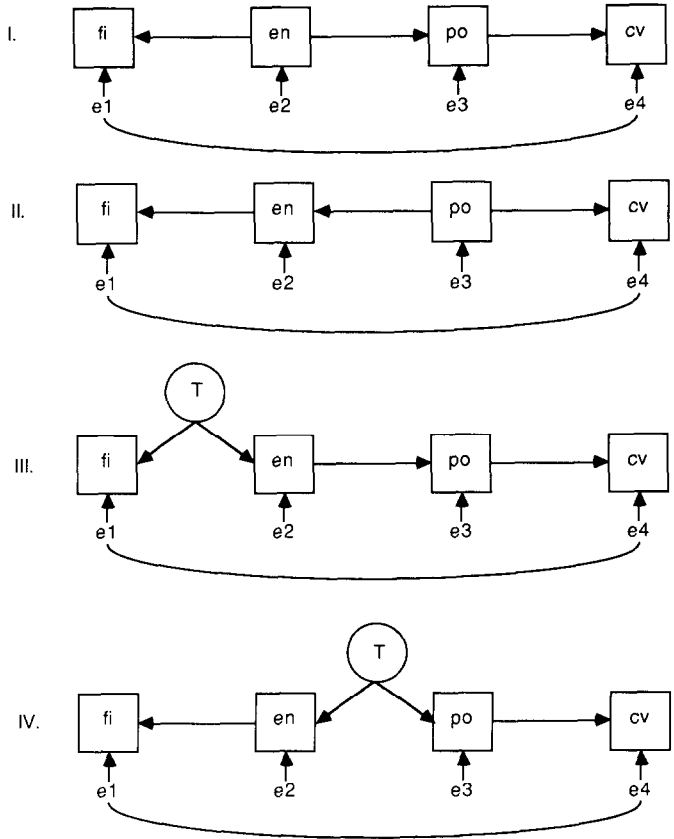


Fig. 12. Alternatives to the regression model.

second equation (B) 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' model does not provide any causal explanation of relations (A) and (B), but it is easy to find assumptions that do explain these patterns, and explain them rather neatly. We exhibit some alternative explanations pictorially in fig. 12. Here *T* signifies a latent common cause. The causal hypotheses in all figures, under the assumption of linearity, imply that both (A) and (B) hold in the population, *no matter what the values of the linear coefficients may be.*

We used the EQS program [Bentler (1985)] to estimate and test model (II). All linear coefficients are very significant. The coefficient giving the depen-

dence of  $fi$  on  $en$  is positive; the coefficient giving the dependence of  $po$  on  $en$  is negative; and the coefficient giving the dependence of  $cv$  on  $po$  is positive. The  $p$ -value for the chi-square statistic with two degrees of freedom is 0.94.

If one accepts model (II), then the conclusion is that foreign investment in 'peripheral' nations neither promotes nor inhibits the development of democracy and civil liberties, but raising the energy consumption per capita promotes both foreign investment and more representative government, and through representative government increases respect for civil liberties. We would not on this data, and given the alternatives, argue that model (II) should be accepted, but it, and very likely the other alternatives suggested here, are preferable to Timberlake and Williams' regression model.

## 11. Time series and latent transfer models

The class of models for which the TETRAD program was designed does not include time-series models such as ARMA or transfer-function models. These classes of models play an important role in econometrics, where forecasting the values of variables is of major interest. It is natural to ask whether the techniques we have described for distinguishing direct causation from common causation, and for determining when latent variables should be introduced, can be extended to time-series models.

Assume that all of the models considered are stationary; that is  $p(y_t, \dots, y_{t+k}) = p(y_{t+m}, \dots, y_{t+k+m})$  for any  $t, k, m$ . The most general class of models that we will consider are transfer function models. To characterize them we require some additional notation:

*The backward shift operator  $B$*  when applied to a variable shifts the variable one time period backwards; i.e.,  $B^n y_t = y_{t-n}$ .  $\tau(B)$  represents a polynomial function of the operator  $B$ , i.e.,  $\tau_1 B + \tau_2 B^2 + \dots + \tau_n B^n$ .

*A transfer function model* relates the value of a dependent variable to lagged values of itself, current and lagged values of independent variables, and an error term. So a univariate transfer function model can be written as

$$y_t = v^{-1}(B)\omega(B)x_t + \pi^{-1}(B)\tau(B)v_t,$$

where  $v_t$  is a normally distributed error term. Moving-average (MA), autoregressive (AR), and moving-average autoregressive (ARMA) models are all special cases of transfer function models.

There are obvious difficulties in extending the techniques that TETRAD uses to time-series models. For example, TETRAD was designed to calculate constraints on correlation matrices, not autocorrelation matrices. More fundamentally, transfer function models contain infinite graphs, whereas TETRAD

represents and analyzes only finite graphs. It would be theoretically possible to treat time-series models within the current framework by regarding a variable at different times as separate variables and by approximating an infinite graph by some finite segment of it, but these are not practical procedures. The number of data points for a time-series model can easily extend into the hundreds, which is far beyond the number of variables that can be analyzed by the TETRAD program. TETRAD would ignore the repeated *patterns* of edges occurring in time-series models, leading to very large numbers of redundant calculations.

In order to extend TETRAD's techniques to time-series models, the following four questions would have to be answered affirmatively:

- Are there overidentifying constraints on autocorrelation matrices that can be explained by some models no matter what values are given to their respective free parameters?
- Are these overidentifying constraints determined by the graphs alone, i.e., are they independent of the variances of the variables appearing in the model?
- Are there representations of infinite graphs, and algorithms that can be performed upon such representations, that can quickly calculate which overidentifying constraints are implied by a graph?
- Are there patterns of constraints on autocorrelations that can only be robustly explained by the introduction of latent variables?

The following examples below will show that the first two questions can be answered affirmatively; the last two questions are still open.

### 11.1. Existence of analogous constraints determined by graphs

The following example shows that there are overidentifying constraints upon autocovariances in time-series models that are analogous to the tetrad constraints found among covariances, and that these constraints are determined by the graph alone.

Consider the AR(1) model with 0 mean (see fig. 13), which has the equation

$$y_t = \pi_1 y_{t-1} + \epsilon_t.$$

It is well known that the autocovariance for a  $k$ -lag displacement (which we denote by  $\gamma_k$ ) is equal to  $\pi^k \gamma_0$ . Hence,

$$\gamma_k \gamma_j = \gamma_n \gamma_m,$$

as long as  $k + j = n + m$ .

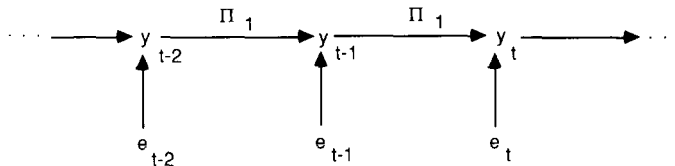


Fig. 13. AR(1) model.

This example shows that there are constraints on autocovariances, analogous to the tetrad constraints upon covariances, that are implied for all values of the equation coefficients. Furthermore, it is clear that this constraint is determined by the graph alone, since the constraint is implied regardless of the value of the variances of the variables.

However, it does not seem likely that models more complex than the AR(1) model imply constraints among autocorrelations analogous to tetrad or partial correlation constraints. For example, it is easy to show that no MA model of any order implies a ‘tetrad’ constraint among four distinct autocorrelations. The general formula for  $\rho_k$  in a moving-average process of order  $q$  is

$$\frac{-\tau_k + \tau_1\tau_{k+1} + \dots + \tau_{q-k}\tau_q}{1 + \tau_1^2 + \tau_2^2 + \dots + \tau_q^2},$$

for  $k = 1, \dots, q$  and 0 for  $k > q$ . Hence the numerator of  $\rho_u\rho_v$  contains a term equal to  $\tau_q^2\tau_{q-u}\tau_{q-v}$ . Similarly, the numerator of  $\rho_w\rho_x$  contains a term equal to  $\tau_q^2\tau_{q-w}\tau_{q-x}$ . If  $\rho_u\rho_x = \rho_w\rho_v$  for all values of  $\tau_{q-u}$ ,  $\tau_{q-v}$ ,  $\tau_{q-w}$  and  $\tau_{q-x}$ , it follows that  $\tau_q^2\tau_{q-u}\tau_{q-v} = \tau_q^2\tau_{q-w}\tau_{q-x}$ . This implies that  $\{u, v\} = \{w, x\}$ .

We strongly suspect that, if there are robust constraints on the autocorrelation matrix of a graph, then the constraints will be implied by the graph alone. But whether or not there are ARMA or AR models [other than AR(1)] that robustly imply constraints is an open question.

### 11.2. Calculation of implied constraints

There is no problem in giving a finite representation an infinite graph as long as there is sufficient periodicity in the graph. In a transfer-function model, if the longest direct causal connection is from  $t - k$  to  $t$ , then all of the information in the graph is contained in the finite segment from  $t - k$  to  $t$ . Infinite graphs can, however, present problems for calculating covariances. The method that TETRAD uses to calculate covariances is based upon the following theorem.

*Theorem 1. In a finite acyclic graph the covariance of any two distinct variables  $x$  and  $y$  is equal to the sum of the product of the labels of edges in the treks between  $x$  and  $y$  times the variance of the source of the trek.*

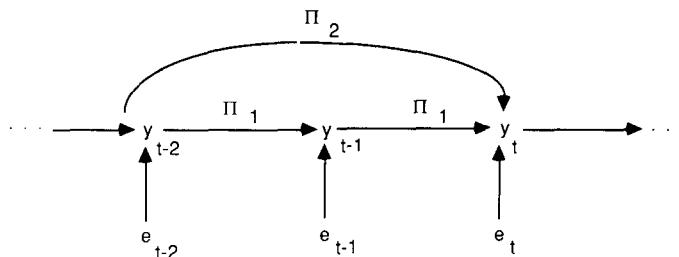


Fig. 14. AR(2) model.

TETRAD calculates the covariances between variables  $x$  and  $y$  by taking the sum of the product of the labels of edges in the treks between  $x$  and  $y$ .

A number of examples suggest that the method of calculating covariances based upon the preceding theorem is correct for infinite as well as finite graphs, although we have no proof of this as yet. For example, in the model discussed above, for any  $k$ , the only trek between  $y_{t-k}$  and  $y_t$  is a path of length  $k$  from  $t_{t-k}$  to  $y_t$ . The product of the edge labels in such a path is clearly  $\pi^k$ . Thus, the method that TETRAD uses to calculate covariances could be carried over virtually unchanged in this case. But this method of calculating covariances is not feasible in an AR(2) model, because there are an infinite number of treks between  $y_t$  and  $y_{t-1}$ . The equation for an AR(2) model is

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \epsilon_t.$$

See fig. 14.

In this case,

$$\gamma_1 = \frac{\pi_1 \gamma_0}{1 - \pi_2}.$$

Once again, the formula for calculating covariances remains true in this particular case since the sum of the product of the treks between  $y_{t-1}$  and  $y_t$  is

$$\gamma_1 = \gamma_0 (\pi_1 + \pi_1 \pi_2 + \pi_1 \pi_2^2 + \dots) = \frac{\pi_1 \gamma_0}{1 - \pi_2}.$$

So, while the formula upon which TETRAD's method of calculating covariances is based remains applicable in this case, the actual method that TETRAD uses to calculate covariances could not be applied, since it would lead to an infinite sum. It is an open question whether there is an efficient algorithm applicable to such a graph that can correctly calculate the auto-

variances and the constraints upon covariances implied by the graph. Nor do we know whether there are patterns of constraints in time series models that can only be explained by the introduction of latent variables. The resolution of these questions will not only help to clarify the proper role of latent variables in econometric time series; it may also lead to the development of effective aids in model specification.

## References

- Aigner, D. and A. Goldberger, 1977, *Latent variables in socio-economic models* (North-Holland, Amsterdam).
- Ancot, J.P. and G. Duru, 1984, *Structural analysis, hierarchical functions, and weak structures*, in: J.P. Auray, ed., *Analysing the structure of econometric models* (Martinus Nijhoff, The Hague).
- Bentler, P., 1985, *Theory and implementation of EQS: A structural equations program* (BMDP Statistical Software, Los Angeles, CA).
- Boutillier, M., 1984, *Reading macroeconomic models and building causal structures*, in: J.P. Auray, ed., *Analysing the structure of econometric models* (Martinus Nijhoff, The Hague).
- Fontela, E. and A. Gabus, 1974, *Events and economic forecasting models*, *Futures* 6, 329–333.
- Fontela, E. and M. Gilli, 1977, *The causal structure of economic models*, *Futures* 9, 510–518.
- Gaberly, M. and M. Gilli, 1984, *Two approaches in reading model interdependencies*, in: J.P. Auray, ed., *Analysing the structure of econometric models* (Martinus Nijhoff, The Hague).
- Gilli, M. and R. Rossier, 1981, *Understanding complex systems*, *Automatica* 17, 647–652.
- Glymour, C., R. Scheines, P. Spirtes and K. Kelly, 1987, *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling* (Academic Press, Orlando, FL).
- Granger, C., 1969, *Investigating causal relations by econometric models and cross-spectral methods*, *Econometrica* 37, 424–438.
- Holland, P., 1985, *Statistics and causal inference*, Lecture to the American Statistical Association.
- Klein, L.R., 1961, *An econometric model of the United Kingdom* (Oxford University, Institute of Statistics, Oxford).
- Lewis, D., 1973, *Counterfactuals* (Harvard University Press, Cambridge, MA).
- Magidson, J., 1979, *Advances in factor analysis and structural equation models* (Abt Books, Cambridge, MA).
- 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, 509–521.
- Suppes, P., 1970, *A probabilistic theory of causality* (North-Holland, Amsterdam).
- Timberlake, M. and K.R. Williams, 1984, *Dependence, political exclusion, and government repression: Some cross-national evidence*, *American Sociological Review* 49, 141–146.
- Warfield, J.N., 1976, *Social systems: Planning, policy, and complexity* (Wiley, New York).
- Wishart, J., 1928/9, *Sampling errors in the theory of two factors*, *British Journal of Psychology* 19, 180–187.