The TETRAD Project:
Constraint Based Aids to Causal Model Specification

Richard Scheines, Peter Spirtes, Clark Glymour,
Christopher Meek and Thomas Richardson
Department of Philosophy, Carnegie Mellon University[1]

The statistical community has brought logical rigor and mathematical precision to the problem of using data to make inferences about a model's parameter values. The TETRAD project, and related work in computer science and statistics, aims to apply those standards to the problem of using data and background knowledge to make inferences about a model's specification. We begin by drawing the analogy between parameter estimation and model specification search. We then describe how the specification of a structural equation model entails familiar constraints on the covariance matrix for all admissible values of its parameters; we survey results on the equivalence of structural equation models, and we discuss search strategies for model specification. We end by presenting several algorithms that are implemented in the TETRAD II program.

1. Motivation

A principal aim of many sciences is to model causal systems well enough to provide sound insight into their structures and mechanisms, and to provide reliable predictions about the effects of policy interventions. In order to succeed in that aim, a model must be specified at least approximately correctly. Unfortunately, this is not an easy problem. When some of the causes are unknown and/or unobserved, there are an infinity of possible causal models and it is not obvious how to go about constructing plausible ones. To make matters worse, there may be many models that are compatible with background knowledge and the data, but which lead to entirely different causal conclusions.

The process of statistical modeling is typically divided into at least two distinct phases: a model specification phase in which a model (with free parameters) is specified, and a parameter estimation and statistical testing phase in which the free parameters of the specified model are estimated and various hypotheses are put to a statistical test. Both

[1] Research for this paper was supported by the National Science Foundation through grant 9102169 and the Navy Personnel Research and Development Center and the Office of Naval Research through grants N0014-93-0568 #N00014-95-1-0684. Correspondence should be directed to Richard Scheines, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: RS2L@andrew.cmu.edu.

model specification and parameter estimation can fruitfully be thought of as search problems. Parameter estimation can be thought of as a search through a large space for a particular vector of values that satisfies a given set of constraints. For example, it is a search problem to find a vector of parameter values that maximize the likelihood of the data given the model. Statisticians have fruitfully investigated a number of parameter estimation problems under a variety of background assumptions, e.g., normal and non-normal distributional assumptions, recursive or non-recursive models, etc. Even though model specification affects parameter estimation (Kaplan, 1988) and predictions about the effects of adopting different policies (Strotz & Wold, 1960; Spirtes, et al., 1993), the great bulk of theoretical attention on statistical causal models has been devoted to estimating their parameters, to developing statistics for testing individual parameters and overall model fit (Bollen and Long, 1993), and to techniques for making minor respecifications of models that fit data poorly (Kaplan, 1990; Saris, et al., 1987; Sorbom, 1989; Bentler, 1986; Jöreskog and Sörbom, 1993).

We believe the problem of model specification is in many respects analogous to the problem of parameter estimation, and that the same kind of rigor brought to the development and evaluation of algorithms for parameter estimation can and should be applied to algorithms for model specification. In a model specification problem a class of models is searched, and various models assessed. Our concern is with structuring such searches so as to have guarantees of reliability analogous to those available for parameter estimators.

This is *not* to say that it is our intention to *replace* well-founded theoretical sources of model specification with automatic procedures. Where theory and domain knowledge provide justified constraints on model specification, those constraints should be used, and one of the important desiderata for model search procedures is that they make use of whatever domain knowledge is available. But rarely, if ever, is social scientific theory and background knowledge sufficiently well confirmed and sufficiently strong to entail a unique model specification. There are typically many theoretically plausible alternatives to a given model, some of which support wholly different causal conclusions and thus lead to different conclusions about which policies should be adopted.

Just as in the case of parameter estimation, the results we will present here do not free one from having to make assumptions; instead, they make rigorous and explicit what can and cannot be learned about the world if one is willing to make certain assumptions and not others. If, for example, one is willing to assume that causal relations are approximately linear and additive, that there is no feedback, that error terms are i.i.d and uncorrelated, and that the Causal Independence and Faithfulness assumptions (explained in detail in sections 3.1 and 3.2) are satisfied, then quite a lot can be learned about the causal structure underlying one's data. If one is only willing to make weaker assumptions, then less can be learned, although what can be learned may still be useful. Our aim is to make precise exactly what can and cannot be learned in each context. As with proofs of properties of estimators, the results are mathematical, not moral: they do not say what assumptions ought to be made.

1.1 Structural Equation Models and Directed Graphs

Many of the results and procedures we will describe are very general and apply to models of categorical as well as continuous data, but for the sake of concreteness we will illustrate them with linear structural equation models (hereafter, SEMs) (Bollen, 1989; James, Mulaik, and Brett, 1982). SEMs include linear regression models (Weisberg, 1985), path analytic models (Wright, 1934), factor analytic models (Lawley & Maxwell, 1971), panel models (Blalock, 1985; Wheaton, et al., 1977), simultaneous equation models (Goldberger & Duncan, 1973), MIMIC models (Bye, et al., 1985), and multiple indicator models (Sullivan, et. al., 1979). This section introduces the terminology we will use throughout the rest of the paper.

The variables in a SEM can be divided into two sets, the "error variables" or "error terms," and the substantive variables. Corresponding to each substantive variable X_i is a linear equation with X_i on the left hand side of the equation, and the direct causes of X_i plus the error term ε_i on the right hand side of the equation. Since we have no interest in first moments, without loss of generality each variable can be expressed as a deviation from its mean.

Consider, for example, two SEMs S_1 and S_2 over $\mathbf{X} = \{X_1, X_2, X_3\}$, where in both SEMs X_1 is a direct cause of X_2 and X_2 is a direct cause of X_3. The structural equations[2] in Figure 1 are common to both S_1 and S_2.

$$
\begin{aligned}
X_1 &= \varepsilon_1 \\
X_2 &= \beta_1 X_1 + \varepsilon_2 \\
X_3 &= \beta_2 X_2 + \varepsilon_3
\end{aligned}
$$

Figure 1: Structural Equations for SEMs S_1 and S_2

In these equations, β_1 and β_2 are free parameters ranging over real values, and ε_1, ε_2 and ε_3 are unmeasured random variables called error terms. Suppose that ε_1, ε_2 and ε_3 are distributed as multivariate normal. In S_1 we will assume that the correlation between each pair of distinct error terms is fixed at zero. The free parameters of S_1 are $\theta = \langle \beta, \mathbf{P} \rangle$, where β is the set of linear coefficients $\{\beta_1, \beta_2\}$ and $\mathbf{P}$ is the set of variances of the error terms. We will use $\Sigma_{\mathbf{S1}}(\theta_1)$ to denote the covariance matrix parameterized by the vector θ_1 for model S_1, and occasionally leave out the model subscript if the context makes it clear

[2] We realize that it is slightly unconventional to write the trivial equation for the exogenous variable X_1 in terms of its error, but this serves to give the error terms a unified and special status as providing all the exogenous source of stochastic variation for the system.

which model is being referred to. If all the pairs of error terms in a SEM S are uncorrelated, we say S is a SEM with **uncorrelated errors**.

Let S_2 contain the same structural equations as S_1, but in S_2 allow the errors between X_2 and X_3 to be correlated, i.e., make the correlation between the errors of X_2 and X_3 a free parameter, instead of fixing it at zero, as in S_1. In S_2 the free parameters are $\theta = <\beta$, **P'**$>$, where β is the set of linear coefficients $\{\beta_1,\beta_2\}$ and **P'** is the set of variances of the error terms and the correlation between ε_2 and ε_3. If the correlations between any of the error terms in a SEM are not fixed at zero, we will call it a SEM with **correlated errors**.[3]

It is possible to associate with each SEM with uncorrelated errors a directed graph that represents the causal structure of the model and the form of the linear equations. For example, the directed graph associated with the substantive variables in S_1 is $X_1 \rightarrow X_2 \rightarrow X_3$, because X_1 is the only substantive variable that occurs on the right hand side of the equation for X_2, and X_2 is the only substantive variable that appears on the right hand side of the equation for X_3. We generally do not include error terms in our path diagrams of SEMs unless the errors are correlated. We enclose measured variables in boxes, latent variables in circles, and leave error variables unenclosed.

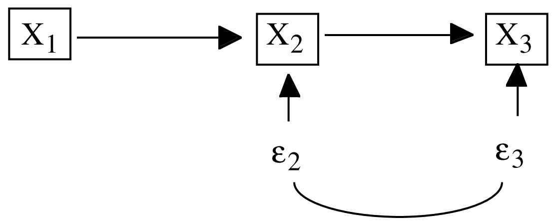


Figure 2. SEM S_2 with correlated errors

The typical path diagram that would be given for S_2 is shown in Figure 2. This is not strictly a directed graph because of the curved line between error terms ε_2 and ε_3, which indicates that ε_2 and ε_3 are correlated. It is generally accepted that correlation is to be explained by some form of causal connection. Accordingly, if ε_2 and ε_3 are correlated we will assume that either ε_2 causes ε_3, ε_3 causes ε_2, some latent variable causes both ε_2 and ε_3, or some combination of these. In other words, curved lines are an ambiguous representation of a causal connection. In section 3.1, for each SEM S with correlated errors we will show how to construct a directed acyclic graph G with latent variables that represents important causal and statistical features of S.

Finally, a directed graph is **acyclic** if it contains no directed path from a variable back to itself. A SEM is said to be **recursive** (an RSEM) if its directed graph is acyclic.

---

[3]We do not consider SEMs with other sorts of constraints on the parameters, e.g., equality constraints.

1.2 Causal Structure: Predicting the Effects of Manipulations

In this section we will consider the causal interpretation of RSEMs (Strotz & Wold, 1960). The causal interpretation of non-recursive SEMs is not well understood and we will not discuss it here. Consider the following hypothetical situation. For each member of a population it is recorded how many cigarettes they have smoked in the last month (S) and how yellow their fingers are (Yf). Let us suppose that the correct causal description of this system is given by RSEM M in Figure 3.

$$Yf = a\,S + \varepsilon_{Yf}$$
$$S = \varepsilon_S$$
$$\rho(\varepsilon_{Yf}, \varepsilon_S) = 0,\ \text{mean}(\varepsilon_Y) = 0, \text{mean}(\varepsilon_{Yf}) = 0, \text{var}(\varepsilon_S) = 1, \text{var}(\varepsilon_{Yf}) = 1 - a^2$$

$$\boxed{S} \longrightarrow \boxed{Yf}$$

Figure 3. RSEM M

The graph of this RSEM is asymmetric because it contains an arrow from S to Yf, but not from Yf to S. What consequences does the causal asymmetry have?

The causal asymmetry is reflected in the predictions that M makes about the effects of interventions on the values of the random variables S and Yf respectively.[4] The prediction of the effect of an intervention on a system is a counterfactual prediction, that is, it is a prediction not about the existing population, but about a population that does not exist and might never exist. Of course, how an actual intervention would affect other variables would depend upon *how* we intervened in the system. We will consider theories that predict the effects of a kind of ideal intervention in which the only variables in the system affected *directly* are those that we manipulate by setting their value. For example, suppose we intervene ideally in the population originally described by M to eliminate smoking, i.e., we set S = 0. Then the new causal system that is the result of this ideal intervention would be described by model M_S (Figure 4), in which the only change is that the equation for S in M is replaced by a new equation in which all of the coefficients relating S to other variables (in this case just ε_S) are set to 0, and S is set equal to a constant.[5]

[4] The causal asymmetry is also reflected in the quite different statistical relationships between ε_s and Yf on the one hand, and ε_{Yf} and S on the other hand. From the Causal Independence assumption introduced in section 3, it follows that in M, ε_{Yf} and S are uncorrelated, but it does not entail that ε_S and Yf are uncorrelated.

[5] In this example, the constant is zero, but we could of course just as easily have an ideal intervention which set S to any other value.

$$Yf = a\,S + \varepsilon_{Yf}$$
$$S = 0$$
$$\rho(\varepsilon_{Yf}, \varepsilon_S) = 0,\ \ \text{mean}(\varepsilon_S) = 0, \text{mean}(\ \varepsilon_{Yf}) = 0, \text{var}(\varepsilon_S) = 0, \text{var}(\varepsilon_{Yf}) = 1 - a^2$$

$$\boxed{S} \longrightarrow \boxed{Yf}$$

Figure 4. RSEM M_S

If, for an arbitrary parameterization of M_S, f_S is the density function according to M_S (and similarly for M and f_M) then the effect on Yf of ideally intervening to set S to 0, i.e., $f_S(Yf \mid S = 0)$, is equal to the density, in RSEM M, of Yf conditional on S = 0, i.e., $f_M(Yf \mid S = 0)$.

M_S is a correct theory of the results of the kind of ideal intervention in which the only variable in the system that is directly affected is S. Of course, whether some particular course of action is an ideal intervention of this kind is an empirical question that is outside the scope of M or M_S. For example, we could try and force people not to smoke (in such a way that the only variable directly affected is smoking) by passing a law against smoking. If the law is effective, and it does not directly affect the values of other variables, then M_S is the correct description of the new system; otherwise it is not.

Suppose now that we were to perform an ideal intervention on yellowed fingers (Yf) in the system described by M, i.e. we were to intervene in such a way that the only change to the system is that the equation for Yf in M is replaced by a new equation in which all of the coefficients relating Yf to other variables are set to 0, and Yf is set equal to a constant. Call this new theory M_{Yf}. Note that we have removed the edge from S to Yf in the graph of M_{Yf}, because variations in S no longer cause variations in Yf and hence the coefficient of S in the equation for Yf is set to zero.

$$Yf = 0$$
$$S = \varepsilon_S$$
$$\rho(\varepsilon_{Yf}, \varepsilon_S) = 0,\ \ \text{mean}(\varepsilon_S) = 0, \text{mean}(\ \varepsilon_{Yf}) = 0, \text{var}(\varepsilon_S) = 0, \text{var}(\varepsilon_{Yf}) = 1 - a^2$$

$$\boxed{S} \qquad\qquad \boxed{Yf}$$

Figure 5. RSEM M_{Yf}

One consequence of the causal asymmetry in M is that if we perform an ideal intervention on S in M, then $f_M(Yf)$ changes, but when we perform an ideal intervention on Yf in M, $f_M(S)$ does not change (because Yf is not a cause of S). It is also important to note that the distribution of smoking *conditional* on Yf = 0, i.e., $f_M(S \mid Yf = 0)$, is *not* the same as the distribution of S when an ideal intervention is performed on Yf, i.e., $f_{Yf}(S \mid Yf = 0)$. In the case of any ideal intervention on Yf, $f_{Yf}(S \mid Yf = c) = f_{Yf}(S)$.

Given an RSEM M in which all parameters are identified, and information about how an intervention affects a given variable in the system, predicting the effects of ideal interventions is easy; one can simply make the suitable changes to M in the manner described above. The problem is much more difficult if M is only partially specified (e.g. the directions of only some of the arrows in the graph of M are known), or if M contains latent variables and not all of the parameters are identifiable. A general theory of representing interventions in causal systems (not limited to RSEMs) in a graphical framework, and of predicting the effects of interventions for a partially specified model is presented in Spirtes, et al. 1993, chapter 7. Examples of making predictions from a partially specified model will be presented in section 5. Robins (1986) made important advances on the problem of predicting the effects of interventions in models with latent variables. Pearl (1995) gives a more general solution of the problem of predicting the effects of interventions in models with latent variables using the graphical representations of interventions presented in Spirtes, et al. 1993.

1.3 Parameter Estimation and SEM Specification

Having clarified the objects under discussion, we now return to our analogy between parameter estimation and model specification search. Ideally, there are four properties that an estimation procedure should have:

- *Identification*. An estimation procedure should be able to determine whether or not a parameter is identifiable, i.e. determine whether or not there is a unique estimate that satisfies the given constraints.
- *Consistency.* If a parameter is identified, it should be the case that as the sample size grows without limit, the probability approaches one that the difference between the true value and the estimated value approaches zero.
- *Error Probabilities*. The sampling distribution of the estimator should be known.
- *Practical Reliability*. The estimation procedure should be reliable on samples of realistic size, and relatively robust against small violations of the operative assumptions.

We will examine each of these desiderata in more detail, and point out analogies (and some disanalogies) between parameter estimation and model specification procedures. To see how the analogy extends to model specification procedures, we will consider for the sake of concreteness properties of the PC algorithm, which is a model specification search procedure implemented in the Build module of TETRAD II (described in sections 5.1.1 and 8.2). For the points we make in this section, it is not necessary to know the details of the PC algorithm. The only features relevant to this section are that it takes as input: 1) a sample covariance matrix (under the assumption of multivariate normality, and 2) background knowledge, and it outputs a graphical object called a *pattern* that represents a class of RSEMs without latent variables or correlated errors that are statistically equivalent (in a sense we make precise in section 4 below). Again, for concreteness, we will use maximum likelihood (ML) estimation and the algorithms that implement it as the example of a parameter estimator.

1.3.1 Identifiability

If there is a unique ML estimate of a parameter in a SEM, then the parameter is said to be identifiable. When a parameter is not identifiable, it has more than one value for which the likelihood of the data is maximal given the model. Although many special cases have been solved (e.g., see Becker, et al., 1994), necessary and sufficient conditions for SEM parameter identifiability are not known.

In the case of SEM specification procedures there is a problem analogous to parameter non-identifiability. There are many pairs of RSEMs R_1 and R_2 that have the same set of measured variables, and no latent variables or correlated errors, that are **covariance equivalent** in the following sense: for every parameterization $\theta_{\mathbf{i}}$ of R_1 there is a parameterization $\theta_{\mathbf{j}}$ of R_2 such that $\Sigma_{\mathbf{R1}}(\theta_{\mathbf{i}}) = \Sigma_{\mathbf{R2}}(\theta_{\mathbf{j}})$, and vice versa. When R_1 and R_2 have no latent variables or correlated errors, then covariance equivalence has the following consequence: for any covariance matrix over the measured variables, if R_1 and R_2 are both parameterized by the respective ML estimates of their free parameters $\Sigma_{\mathbf{R1}}(\theta_{\mathbf{ML}})$ and $\Sigma_{\mathbf{R2}}(\theta_{\mathbf{ML}})$, then the p-values of the χ^2 likelihood ratio test for R_1 and R_2 will be identical. Thus the data cannot help us distinguish between R_1 and R_2. This is a kind of "causal underidentification."

The slogan that "correlation is not causation" expresses the idea that from data including only the existence of a single significant correlation between variables A and B, the causal structure governing A and B is underidentified. That is, a correlation between two variables A and B could be produced by A causing B, B causing A, a latent variable that causes both A and B, or some combination of these. But just as a single example of an underidentified SEM does not show that parameters are always underidentified, or that parameter estimation is always impossible or useless, the existence of a single example of covariance equivalent SEMs does not show that specification search for SEMs is always impossible or useless.

For some SEMs, certain parameters may be identifiable while others are not. Similarly, certain features of an RSEM R might be common to every R' that is covariance equivalent to R. We will show examples in which a covariance equivalence class of RSEMs *all* share the feature that some variable A is a (possibly indirect) cause of B; we will show other examples in which *none* of the members of a covariance equivalence class is A (even indirectly) a cause of B. As explained in detail in section 4, for various special cases, necessary and sufficient, or necessary conditions for various kinds of statistical equivalence are known. Because of the problem of covariance equivalence, the output of our algorithms will generally not be a single RSEM. Instead the output will be an object that represents a *class* of RSEMs consistent with the assumptions made and which marks those features shared by all of the members of the RSEMs output.

By outputting a representation of covariance equivalence class of RSEMs, rather than a single SEM, the PC algorithm addresses the problem that there may be many different structural equation models that are compatible with background knowledge and fit the data equally well (as measured by a p-value, for example). However, it may be the case that there are SEMs which are not covariance equivalent, but nonetheless fit the data *almost* equally well; ideally an algorithm should output all such models, rather than simply choose the best. This problem could be addressed by outputting multiple patterns, rather than a single pattern. Devising an algorithm (or modifying the PC algorithm) to output representations of all models that fit the data well and are compatible with background knowledge is an important area of future research.

1.3.2 Consistency and Correctness

A SEM parameter estimation algorithm takes as input a sample covariance matrix **S** and distributional assumptions about the population from which **S** was drawn, and produces as output an estimate θ_{est} of the population parameters θ_{pop}. If the measured variables are indeed multivariate normal, and the specified model holds in the population, then asymptotically, as the sample size goes to infinity, the sampling distribution of θ_{ML} goes to $N(\theta_{pop}, J^{-1}(\theta))$, where $J(\theta)$ is the Fisher information matrix (cf. Tanner, 1993, p. 16). So θ_{ML} is a **consistent** estimator in the sense that as the sample size grows without bound the difference between θ_{pop} and θ_{ML} will, with probability 1, converge to zero.

The PC algorithm takes as input a sample covariance matrix **S**, the assumption that **S** is drawn from a multivariate normal population described by an RSEM R_{pop} with no latent variables or correlated errors, and produces as output a **pattern** which represents a class of RSEMs that are covariance equivalent to R_{pop} (see section 4).[6] Let $\mathbf{M_{PC}}$ be the pattern output by the PC algorithm, and $\mathbf{M_{pop}}$ be the pattern that represents the class of RSEMs covariance equivalent to R_{pop}. Since there is no obvious metric to express the difference between $\mathbf{M_{PC}}$ and $\mathbf{M_{pop}}$, we will not follow the analogy with parameter estimation and say that the PC algorithm is consistent. We can, however, state and prove a closely related property which we call correctness. The PC algorithm is **correct** in the following sense: if the Causal Independence and Faithfulness assumptions are satisfied, then, as the sample grows without bound, the probability that $\mathbf{M_{PC}} = \mathbf{M_{pop}}$ converges to one.[7]

[6] In some cases the input to the PC algorithm is not consistent with the assumptions made. In these cases it is possible that the output of the PC algorithm is not strictly a pattern.

[7] The PC algorithm performs a series of statistical tests of zero partial correlations; the asymptotic

1.3.3 Sampling Distribution

In a SEM, to estimate the sampling distribution of θ_{ML} on finite samples, we have two choices. First, if the sample size is reasonably large we can use θ_{ML} as an estimate of θ_{pop}, and then use the asymptotic theory described above ($\theta_{ML} \sim N(\theta_{pop}, J^{-1}(\theta))$) as an estimate of the sampling distribution of θ_{ML}. Second, we can approximate the sampling distribution of θ_{ML} empirically by Monte Carlo techniques (Boomsma, 1982). We can do this by assuming $\Sigma(\theta_{ML}) = \Sigma(\theta_{pop})$. We can then repeatedly sample from $\Sigma(\theta_{pop})$, and calculate the ML estimate for each sample (Figure 6). Although for small N the sampling distribution of θ_{ML} is not multivariate normal (Boomsma, 1982), it can be still be usefully summarized by the standard deviation (standard errors) and the mean.

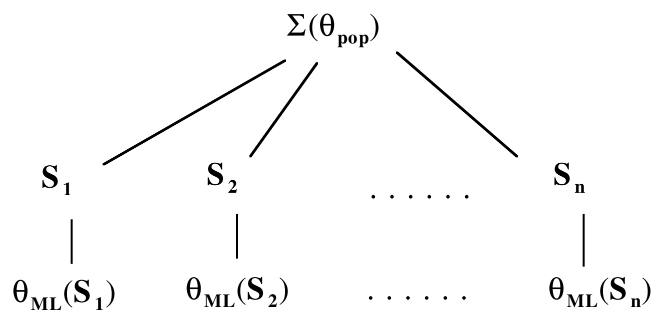


Figure 6. Monte Carlo approximation of the sampling distribution for θ_{ML}

On samples from a given model with specified parameters, the sampling distribution of $\mathbf{M}_{PC}$ is well defined. However, $\mathbf{M}_{PC}$ is not a vector of real valued parameters as θ_{ML} is, but rather a graphical object (see section 5.1.1) that represents an equivalence class of RSEMs. Hence $\mathbf{M}_{PC}$ is a categorical variable with no meaningful ordering of the categories. Thus the variance and mean are not very useful summaries of features of the distribution. We do not know how to calculate an analytic approximation of the sampling distribution for $\mathbf{M}_{PC}$ on finite samples. But we can apply empirical techniques parallel to those mentioned above for ML parameter estimation. To approximate the sampling distribution for $\mathbf{M}_{PC}$ on finite samples, consider Figure 7, which is analogous to Figure 6.

results assume that we systematically lower the significance level as the sample size increases, in order to decrease the probabilities of both type I and type II errors to zero.

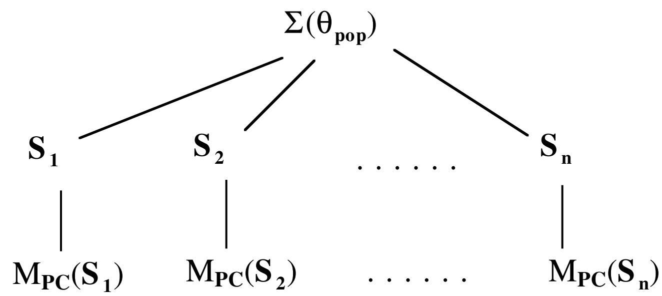


Figure 7. Monte Carlo approximation to sampling distribution for $\mathbf{M_{PC}}$

A slight disanalogy occurs in estimating $\Sigma(\theta_{pop})$. In the maximum likelihood setting, $\Sigma(\theta_{ML})$ is used as an estimate of $\Sigma(\theta_{pop})$. To obtain $\Sigma(\theta_{ML})$ from our sample **S** and $\mathbf{M_{PC}}$, we can pick an arbitrary member M_i of the equivalence class of RSEMs represented by $\mathbf{M_{PC}}$ and then calculate θ_{ML} for M_i and **S**. (The resulting covariance matrix $\Sigma_{\mathbf{Mi}}(\theta_{ML})$ is the same regardless of which member M_i of $\mathbf{M_{PC}}$ we choose.) We can then use $\Sigma_{\mathbf{Mi}}(\theta_{ML})$ as an estimate of $\Sigma(\theta_{pop})$.

1.3.4 Practical Reliability

Finally, we want to know if the estimation procedure is reliable in practice. θ_{ML} has, by definition, the property that there is no $\theta_i \neq \theta_{ML}$ s.t. $p(\mathbf{S}|\theta_i) > p(\mathbf{S}|\theta_{ML})$. On samples of realistic size, however, iterative procedures that search the parameter space such as those implemented in LISREL (Jöreskog, 1993) and EQS (Bentler, 1995) cannot guarantee that they will find θ_{ML}. They must begin, for example, from some starting point in the parameter space and hill climb, and the likelihood surface might have local maxima (Scheines, Hoijtink, & Boomsma, 1995). We can investigate the practical reliability of an ML estimation procedure at a given sample size by 1) drawing an RSEM R from a distribution over RSEMs, 2) drawing a parameterization $\theta_{\mathbf{i}}$ from a distribution over the parameters of R to give a population $\Sigma_{\mathbf{R}}(\theta_i)$, and 3) drawing a sample **S** from $\Sigma_{\mathbf{R}}(\theta_i)$. We can then use **S** as the input to the implemented estimator E, finally comparing $\Sigma_{\mathbf{R}}(\theta_E)$ to $\Sigma_{\mathbf{R}}(\theta_i)$ or just θ_E to θ_i.

We can investigate the reliability of model specification algorithms in an analogous way. The PC algorithm performs statistical tests of vanishing partial correlations, and if it cannot reject the null hypothesis at a significance level set by the user, then the procedure accepts the null hypothesis. If the null hypothesis is wrongly accepted or rejected, the output of the procedure can be incorrect. On finite samples, the reliability of the model specification algorithms depends upon the power of the statistical tests, the significance level used in the tests, the distribution over the models, and the parameters of the models.

In (Spirtes, et al., 1993), and in (Scheines, et al., 1994), we report on systematic Monte Carlo simulation studies to approximate features of the sampling distribution over the PC algorithm (and a variety of our other RSEM specification algorithm) by 1) drawing an RSEM R from a distribution over RSEMs, 2) drawing a parameterization $\theta_{\mathbf{i}}$ from a distribution over the parameters for R, 3) drawing a sample **S** from the multivariate normal population $\Sigma_{\mathbf{R}}(\theta_i)$, and then using **S** as input to the PC algorithm. Finally, we compare $\mathbf{M_{PC}}$ to $\mathbf{M_{pop}}$.

These tests indicate that the PC algorithm is reliable with respect to determining which variables are adjacent in the population causal graph as long as the sample sizes are on the order of 500 and the population RSEM is not highly interconnected (i.e. that not everything is either a cause or an effect of everything else). For example, at sample size 500 for sparsely connected RSEMs with 50 variables, the PC algorithm incorrectly hypothesized an adjacency less than once in 1,000 times such a mistake was possible, and incorrectly omitted an adjacency approximately 10% of the time, with the accuracy improving as the sample size grows (see page 155 of Spirtes, et. al., 1993). We should note, however, that these simulation tests satisfied all the distributional assumptions underlying the algorithm and did not allow parameter values close to 0. We have not yet systematically explored the effect of small violations of these or other assumptions.

1.4 Difficulty of Search

In practice, model specification problems are very difficult for (at least) the following reasons:

- Data sets may fail to record variables (confounders) that produce associations among recorded variables.
- When no limitation is placed on the number of "latent variables," the number of alternative SEMs may be literally infinite.
- Many distinct SEMs may produce the same, or nearly the same distributions of recorded variables.
- Natural and social populations may be mixtures of SEMs with different causal graphs.
- Values of quantities recorded for some units in a data set may be missing for other units.
- There may be “selection bias”--that is, a measured variable may be causally connected to whether an individual is or is not included in the sample.
- The causal structure may involve feedback loops.
- The functional relations between causes and their effects may be non-linear.
- Actual distributions may not be closely approximated by any well known probability distributions.

In the last fifteen years a movement in computer science and statistics has made theoretical progress on a number of these issues,[8] progress that has led to computer based methods to aid in model specification. These results and the methods that implement them appear to be little known and rarely used in the social science communities. This paper is an introductory description of some of the more important theoretical ideas and of some of the computational procedures that have arisen out of these discoveries, offered in the hope that social and behavioral scientists will make more use of these methods and help to improve them. (Mathematical details and related results can be found in Pearl, 1988; Spirtes, et al., 1993; Scheines, et al., 1994).

1.5 Search Procedures

Two approaches to RSEM specification have been pursued in the statistics and computer science literature. The first focuses on searching for the RSEM or RSEMs that maximize some score.[9] The second approach focuses on searching for the RSEMs that satisfy a set of constraints judged to hold in the population (e.g., Spirtes, et al., 1993). (See Richardson (1996) for a correct algorithm for searching for non-recursive SEMs without latent variables.) Searches based on maximizing a score have been developed for RSEMs with no latent variables (e.g., Geiger & Heckerman, 1994; Cooper & Herskovits, 1992); typically they are either stepwise forward (they add edges), stepwise backward (they take away edges), or some combination of stepwise forward and backward. Most regression searches are of this type, although they are restricted to searching a very restricted class of RSEMs. The "modification index" searches based on the Lagrange Multiplier statistic (Bentler, 1986; Kaplan, 1989, 1990; Jöreskog & Sörbom, 1993; Sörbom, 1989) in LISREL and EQS are restricted versions of this strategy. They typically begin with a given SEM M and perform a stepwise forward search (EQS can also perform a stepwise backward search). One difficulty with searches that maximize a score is that no proofs of correctness are yet available. A more difficult problem is that there are as of now no feasible score-maximization searches that include SEMs with latent variables. The modification index searches cannot suggest adding or removing a latent variable, for example. Also, these searches output a single SEM, rather than an equivalence class of SEMs. Another search strategy based upon maximizing a score is to search not RSEMs themselves, but covariance equivalence classes of RSEMs (Spirtes and Meek, 1995).

[8] Some of this literature is published in the annual proceedings of the conferences on Uncertainty in Artificial Intelligence, Knowledge Discovery in Data Bases, and the bi-annual conference on Artificial Intelligence and Statistics. Examples of important papers in this tradition include: (Buntine 1991; Cooper & Herskovits, 1992; Geiger, 1990; Geiger & Heckerman, 1991, 1994; Geiger, Verma, and Pearl, 1990, Hand, 1993; Lauritzen, et al., 1990; Lauritzen & Wermuth, 1984; Pearl, 1988; Pearl & Dechter, 1989; Pearl and Verma, 1991; Robins, 1986; Spiegelhalter, 1986).

[9] For example, the Bayesian Information Criterion (Raftery, 1993), or the posterior probability, (Geiger & Heckerman, 1994).

In contrast to a score maximization search, a constraint search uses some testing procedure for conditional independence, vanishing partial correlations, vanishing tetrad differences, or other constraints on the covariance matrix. One advantage of this kind of search is that there are provably correct search algorithms for certain classes of RSEMs. For example, we will later discuss correct algorithms for multivariate normal RSEMs even when the population RSEM may contain latent variables (Spirtes, et al. 1993).

In order to understand model specification search procedures based on constraints, one must first understand how SEMs entail constraints on the covariance matrix. Various equivalence relations between SEMs also need to be explained. We turn to those topics in the next sections.

2. *Constraints Entailed by SEMs*

We use two kinds of correlation constraints in our searches: zero partial correlation constraints, and vanishing tetrad constraints.

2.1 Zero Partial Correlation Constraints

In a SEM some partial correlations may be equal to zero for *all* values of the model's free parameters (for which the partial correlation is defined). (See Blalock 1962; Kiiveri & Speed, 1982). In this case we will say that the SEM **entails** that the partial correlation is zero.[10] For example, in SEM S_1 (Figure 1), where all of the error terms are uncorrelated, $\rho_{X1,X3.X2} = 0$ for all values of the free parameters of S_1.

Judea Pearl (1988) discovered a fast procedure that can be used to decide, for any partial correlation $\rho_{A,B.\mathbf{C}}$ and any RSEM with uncorrelated errors, whether the RSEM entails that $\rho_{A,B.\mathbf{C}}$ is zero. Pearl defined a relation called **d-separation** that can hold between three disjoint sets of vertices in a directed acyclic graph. A simple consequence of theorems proved by Pearl, Geiger, and Verma shows that in an RSEM R with uncorrelated errors a partial correlation $\rho_{A,B.\mathbf{C}}$ is entailed to be zero if and only if {A} and {B} are d-separated by **C** in the directed graph associated with R (Pearl 1988). More details about their discovery, which is considerably more general than the description given here, are given in section 8.1. Spirtes (1995) showed that these connections between graphical structure and vanishing partial correlations hold as well for non-recursive SEMs, i.e. in a SEM with uncorrelated errors a partial correlation $\rho_{A,B.\mathbf{C}}$ is entailed to be zero if and only if {A} and {B} are d-separated given **C**. (The if part of the theorem was shown independently in Koster (forthcoming)).

[10] Correlations and partial correlations are zero exactly when the corresponding covariances and partial covariances are zero. While there may be important different statistical properties of partial correlations and partial covariances, they are not germane to the discussion of the constraints entailed by a SEM.

There is also a way to decide which partial correlations are entailed to be zero by a SEM with correlated errors, such as S_2 (Figure 2). This is done by first creating a directed graph G with latent variables and then applying d-separation to G to determine if a zero partial correlation is entailed. The directed graph G (with latent variables but without correlated errors) that we associate with a SEM S with correlated errors is created in the following way. Start with the usual graphical representation of S, that contains undirected lines connecting correlated errors (e.g. SEM S_2 in Figure 2). For each pair of correlated error terms ε_i and ε_j, introduce a new latent variable T_{ij}, and edges from T_{ij} to X_i and X_j. Finally replace ε_i and ε_j with uncorrelated errors ε_i' and ε_j'. When this process is applied to SEM S_2, the result is shown in Figure 8.

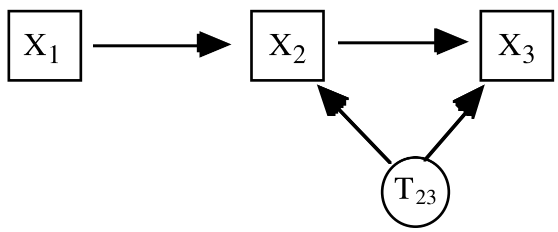


Figure 8. SEM S_2': Correlated Errors in S_2 Replaced by Latent Common Cause

In a SEM like S_2, with correlated errors, one can decide whether $\rho_{X1,X3.X2}$ is entailed to be zero by determining whether $\{X_1\}$ and $\{X_3\}$ are d-separated given $\{X_2\}$ in the graph in Figure 8. In this way the problem of determining whether a SEM with correlated errors entails a zero partial correlation is reduced to the already solved problem of determining whether a SEM without correlated errors entails a zero partial correlation. (In general if S is a SEM with correlated errors, and G is the latent variable graph with uncorrelated erorrs associated with S, it is *not* the case that for every linear parameterization θ_1 of S there is a linear parameterization θ_2 of G such that $\Sigma_{\mathbf{S}}(\theta_1) = \Sigma_{\mathbf{G}}(\theta_2)$. We are making the weaker claim that d-separation applied to G correctly describes which zero partial correlations are entailed by S. For the proof, see Spirtes, et. al, 1996.

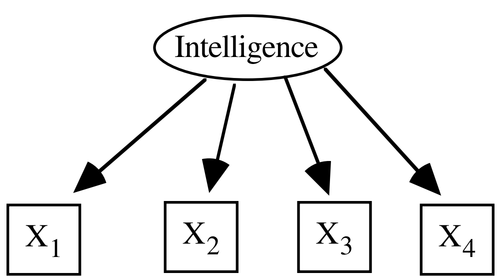


Figure 9. Factor Model of Intelligence

2.2 Vanishing Tetrad Constraints

In SEMs containing latent variables, zero partial correlation constraints among the measured covariances Σ are often uninformative. For example, consider Figure 9 in which Intelligence is a latent variable. The only correlations entailed to be zero by this SEM are those that are partialed on at least Intelligence. Since Intelligence is unmeasured, however, our data will only include partial correlations among the measured variables $\mathbf{X} = \{X_1, X_2, X_3, X_4\}$, and there is no partial correlation involving only variables in $\mathbf{X}$ that is entailed to be zero by this SEM.

The vanishing tetrad difference (Spearman, 1904, Glymour, et al., 1987), however, can provide extra information about the specification of this model. A tetrad difference involves two products of correlations, each of which involve the same four variables but in different permutations. In the SEM of Figure 9 there are three tetrad differences among the measured correlations that are entailed to vanish for all values of the free parameters (for which the correlations are defined):

$$\rho_{X1,X2}\ \rho_{X3,X4} - \rho_{X1,X3}\ \rho_{X2,X4}$$
$$\rho_{X1,X2}\ \rho_{X3,X4} - \rho_{X1,X4}\ \rho_{X2,X3}$$
$$\rho_{X1,X3}\ \rho_{X2,X4} - \rho_{X1,X4}\ \rho_{X2,X3}$$

If a SEM S entails that $\rho_{X1,X2}\ \rho_{X3,X4} - \rho_{X1,X3}\ \rho_{X2,X4} = 0$ for all values of its free parameters we say that S **entails** the vanishing tetrad difference. The tetrad differences that are entailed to vanish by a SEM without correlated errors are also completely determined by the directed graph associated with the SEM. The graphical characterization is given by the Tetrad Representation Theorem (Spirtes, 1989; Spirtes, et al. 1993; Shafer et al., 1993), which leads to a general procedure for computing the vanishing tetrad differences entailed by a SEM, implemented in the Tetrads module of the TETRAD II program (Scheines, Spirtes, Glymour and Meek, 1994). Bollen and Ting (1993) discuss the advantages of using vanishing tetrad differences in SEM analysis, e.g. they can be used to compare underidentified SEMs.

3. Assumptions Relating Probability to Causal Relations

3.1 The Causal Independence Assumption

The most fundamental assumption relating causality and probability that we will make is the following:

> **Causal Independence Assumption:** If A does not cause B, and B does not cause A, and there is no third variable which causes both A and B, then A and B are uncorrelated.

This assumption provides a bridge between statistical facts and causal features of the process that underlies the data. In certain cases the assumption allows us to draw a *causal* conclusion from *statistical* data and lies at the foundation of the theory of randomized experiments. If the value of A is randomized, the experimenter knows that the randomizing device is the sole cause of A. Hence the experimenter knows B did not cause A, and that there is no other variable which causes both A and B. This leaves only two alternatives: either A causes B or it does not. If A and B are correlated in the experimental population, the experimenter concludes that A does cause B, which is an application of the Causal Independence assumption.

The Causal Independence assumption entails that if two error terms are correlated, such as ε_2 and ε_3 in S_2 (see Figure 2), then there is at least one latent common cause of the explicitly modeled variables associated with these errors, i.e., X_2 and X_3.

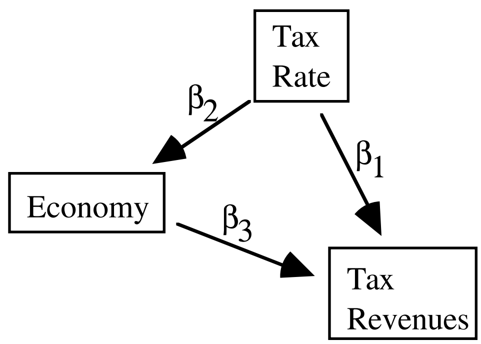


Figure 10. Distribution is Unfaithful to SEM when $\beta_1 = -(\beta_2\beta_3)$

3.2 The Faithfulness Assumption

In addition to the zero partial correlations and vanishing tetrad differences that are entailed for *all* values of the free parameters of a SEM, there may be zero partial correlations or vanishing tetrad differences that hold only for *particular* values of the free parameters of a SEM. For example, suppose Figure 10 is the directed graph of a SEM that describes the relations among the Tax Rate, the Economy, and Tax Revenues.

In this case there are no vanishing partial correlation constraints entailed for all values of the free parameters. But if $\beta_1 = -(\beta_2\beta_3)$, then Tax Rate and Tax Revenues are uncorrelated. The SEM postulates a direct effect of Tax Rate on Revenue (β_1), and an indirect effect through the Economy ($\beta_2\beta_3$). The parameter constraint indicates that these effects *exactly* offset each other, leaving no total effect whatsoever. In such a case we say that the population is **Unfaithful** to the SEM that generated it. A distribution is **Faithful** to SEM M (or its corresponding directed graph) if each partial correlation that is zero in the distribution is entailed to be zero by M, and each tetrad difference that is zero in the distribution is entailed to be zero by M.

> **Faithfulness Assumption:** If the directed graph associated with a SEM M correctly describes the causal structure in the population, then each partial correlation and each tetrad difference that is zero in $\Sigma_M(\theta_{\mathbf{pop}})$ is entailed to be zero by M.

The Faithfulness assumption is a kind of simplicity assumption. If a distribution P is faithful to an RSEM R_1 without latent variables or correlated errors, and P also results from a parameterization of another RSEM R_2 to which P is not faithful, then R_1 has fewer free parameters than R_2.

The Faithfulness assumption limits the SEMs considered to those in which population constraints are entailed by structure, not by particular values of the parameters. If one assumes Faithfulness, then if A and B are *not* d-separated given **C**, then $\rho_{A,B.\mathbf{C}} \neq 0$, (because it is not entailed to equal zero for all values of the free parameters.) Faithfulness should not be assumed when there are deterministic relationships among the substantive variables, or equality constraints upon free parameters, since either of these can lead to violations of the assumption. Some form of the assumption of Faithfulness is used in every science, and amounts to no more that the belief that an improbable and unstable cancellation of parameters does not hide real causal influences. When a theory cannot explain an empirical regularity save by invoking a special parameterization, most scientists are uneasy with the theory and look for an alternative.

It is also possible to give a personalist Bayesian argument for assuming Faithfulness. For any SEM with free parameters, the set of parameterizations of the SEM that lead to violations of Faithfulness are Lebesgue measure zero. Hence any Bayesian whose prior over the parameters is absolutely continuous with Lebesgue measure assigns a zero prior probability to violations of Faithfulness. Of course, this argument is not relevant to those Bayesians who place a prior over the parameters that is not absolutely continuous with Lebesgue measure and assign a non-zero probability to violations of Faithfulness. All of the algorithms we have developed assume Faithfulness, and from here on we use it as a working assumption.

The Faithfulness assumption is necessary to guarantee the correctness of the model specification algorithms used in TETRAD II. It does *not* guarantee that on samples of finite size the model specification algorithms are reliable.

4. SEM Equivalence

Two SEMs S_1 and S_2 with the same substantive variables (or their respective directed graphs) are **covariance equivalent** if for every parameterization θ_i of S_1 with covariance matrix $\Sigma_{S1}(\theta_i)$ there is a parameterization θ_j of S_2 with covariance matrix $\Sigma_{S2}(\theta_j)$ such that $\Sigma_{S1}(\theta_i) = \Sigma_{S2}(\theta_j)$, and vice versa. Two SEMs with the same substantive variables (or their respective directed graphs) are **partial correlation equivalent** if they entail the same set of zero partial correlations among the substantive variables.

If two SEMs contain latent variables, and the same set of measured variables **V**, we may be interested if they are equivalent on the measured variables. Two SEMs S_1 and S_2 (or their respective directed graphs) are **covariance equivalent over a set of measured variables V** if for every parameterization θ_i of S_1 with covariance matrix $\Sigma_{S1}(\theta_i)$ there is a parameterization θ_j of S_2 with covariance matrix $\Sigma_{S2}(\theta_j)$ such that the margin of $\Sigma_{S1}(\theta_i)$ over **V** = the margin of $\Sigma_{S2}(\theta_j)$ over **V**, and vice versa. Two SEMs are **partial correlation equivalent over a set of measured vertices V** if they entail the same set of zero partial correlations among variables in **V**.

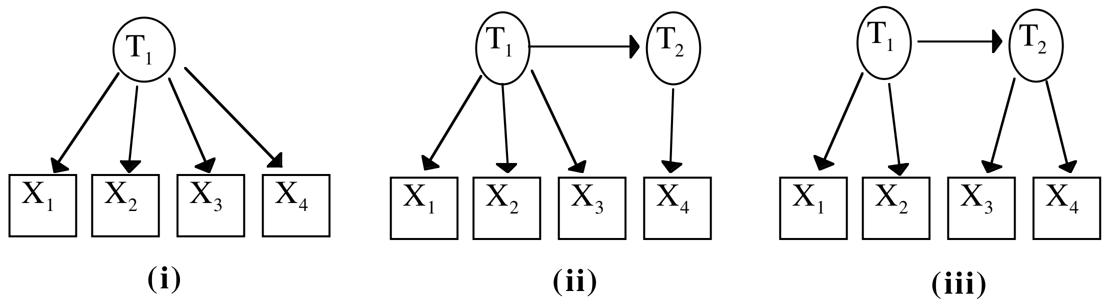


Figure 11. Three SEMs

We illustrate the difference between equivalence and equivalence over a set **V** with the models in Figure 11. Models i and ii do not share the same set of substantive variables, so they are not covariance or partial correlation equivalent. Models ii and iii share the same substantive variables, but are not covariance equivalent or partial correlation equivalent because, for example, model iii entails $\rho_{X2,X3.T2} = 0$ while model ii does not. For $\mathbf{V} = \{X_1, X_2, X_3, X_4\}$, however, the situation is quite different. All three models are partial correlation equivalent over **V**, and models i and ii are covariance equivalent over **V**. Models ii and iii are not covariance equivalent over **V** because, for example, model ii entails that $\rho_{X1,X2}\,\rho_{X3,X4} = \rho_{X1,X3}\,\rho_{X2,X4}$ while model iii does not. The next four subsections will outline what is known about these various kinds of equivalence in both recursive and non-recursive SEMs.

4.1 Covariance and Partial Correlation Equivalence in Recursive SEMs

In this section we consider equivalence over RSEMs with no correlated errors. For two such RSEMs, covariance equivalence holds if and only if zero partial correlation equivalence holds (Spirtes et al. 1993). In RSEMs, only two concepts need to be defined to graphically characterize covariance (or partial correlation) equivalence: **adjacency** and **unshielded collider**. Two variables X and Y are adjacent in a directed graph G just in case X $\rightarrow$ Y is in G, or Y$\rightarrow$ X is in G.

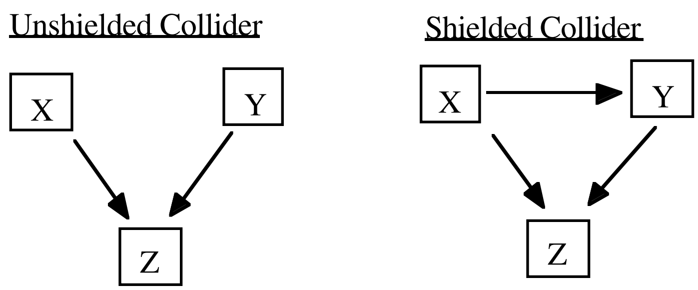


Figure 12.

A triple of variables <X,Z,Y> is a **collider** in G just in case X $\rightarrow$ Z $\leftarrow$ Y is in G, and Z is an **unshielded collider** between X and Y just in case <X,Z,Y> is a collider and X and Y are not adjacent (Figure 12). The first theorem stated below is a simple consequence of a theorem proved in Verma and Pearl (1990), and in Frydenberg (1990).

> **RSEM Partial Correlation Equivalence Theorem**: Two RSEMs with the same variables and no correlated errors are partial correlation equivalent if and only if their respective directed graphs have the same adjacencies and the same unshielded colliders.

> **RSEM Covariance Equivalence Theorem**: Two RSEMs with the same variables and no correlated errors are covariance equivalent if and only if their respective directed graphs have the same adjacencies and the same unshielded colliders.

By the first theorem, if two RSEMs with the same variables and no corelated errors have the same adjacencies and unshielded colliders, then they are partial correlation equivalent. The latter theorem can be proved by induction on the number of edges taken out of two exactly identified SEMs. Take an exactly identified RSEM R with no correlated errors and no latent variables. Then R can be parameterized to fit any covariance matrix. Take an edge out of R to form R'. R' now entails a single vanishing partial correlation $\rho_{XY\cdot Z}$, so it can be parameterized to fit any covariance matrix in which $\rho_{XY\cdot Z}$, = 0. If M is partial correlation equivalent to R, and also has no correlated errors or latent variables, then it too is exactly identified and thus can also can parameterize any covariance matrix. If we form M' by taking an edge away from M that corresponds to the same adjacency we took out of R to form R', and M' and R' are still partial correlation equivalent, then M' can parameterize any covariance matrix in which $\rho_{XY.Z'} = 0$.

4.2 Covariance and Partial Correlation Equivalence Over the Measured Variables in RSEMs

We now consider the case where there may be latent variables and/or correlated errors, and the question is whether two SEMs are covariance equivalent or partial correlation equivalent over a set of measured variables **V**. Since an RSEM with correlated errors is partial correlation equivalent to another RSEM with a latent variable but no correlated errors, the problem of deciding partial correlation equivalence over the measured variables when there are correlated errors reduces to the problem of deciding partial correlation equivalence over the measured variables when there are no correlated errors.

Covariance equivalence over the measured variables entails partial correlation equivalence over the measured variables, but the converse does not hold. Consider the directed graphs i and ii in Figure 13, where the set of measured variables $\mathbf{V} = \{X_1, X_2, X_3, X_4\}$ and the errors are uncorrelated. Although these graphs are partial correlation equivalent over **V** (neither entails any partial correlations among the measured variables), they are not covariance equivalent over **V**, since model i but not model ii entails that

$$\rho_{X1,X2}\ \rho_{X3,X4} = \rho_{X1,X3}\ \rho_{X2,X4} = \rho_{X1,X4}\ \rho_{X2,X3}$$

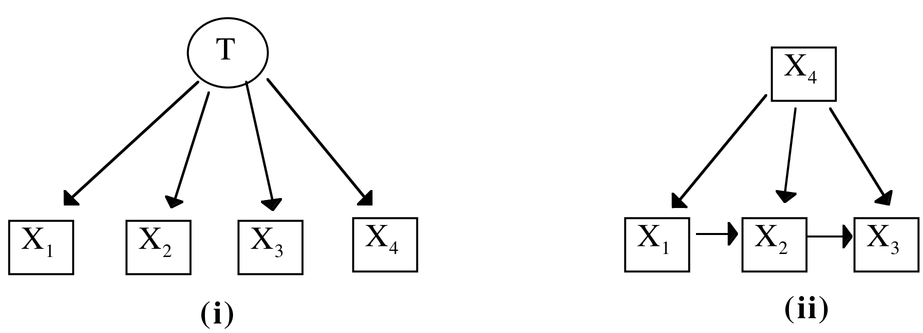


Figure 13: Two graphs that are partial correlation equivalent over $\{X_1, X_2, X_3, X_4\}$, but not covariance equivalent over $\{X_1, X_2, X_3, X_4\}$.

Spirtes, Meek, and Richardson (1995) have given a polynomial (in the number of variables in the two RSEMs) time algorithm for deciding when two RSEMs with uncorrelated errors are partial correlation equivalent over the measured variables. The algorithm is too complex to present here, but some examples of partial correlation equivalence are given in section 5.1. A feasible algorithm for deciding covariance equivalence over a set of measured variables is not known.

4.3 Covariance and Partial Correlation Equivalence in Non-recursive SEMs

Assuming uncorrelated errors, Richardson (1994, 1995) has given an algorithm for deciding when two non-recursive SEMs are partial correlation equivalent that is polynomial in the number of variables in the two SEMs. It is not known whether partial correlation equivalence entails covariance equivalence in this case.

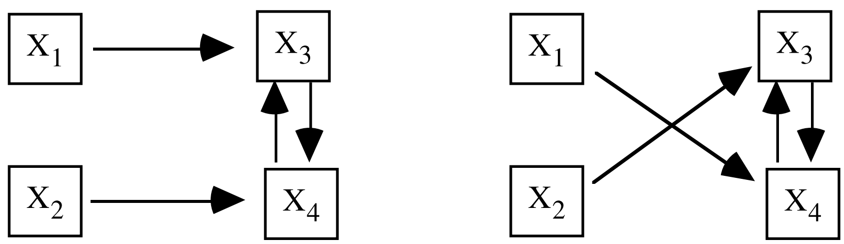


Figure 14. Partial Correlation Equivalent Cyclic SEMs

One noteworthy corollary of Richardson's theorem is that for every SEM with a directed cycle, there is another partial correlation equivalent SEM with a cycle reversed in direction. And while partial correlation equivalent RSEMs without correlated errors always have the same adjacencies, partial correlation equivalent SEMs without correlated errors can have directed cyclic graphs with different adjacencies. For example, the two SEMs in Figure 14 are partial correlation equivalent but do not have the same adjacencies.

4.4 Covariance and Partial Correlation Equivalence over the Measured Variables in Non-recursive SEMs

No feasible general algorithm for deciding either partial correlation or covariance equivalence over a set of measured variables is known for non-recursive SEMs when the measured variables are a proper subset of the substantive variables in the SEM.

5. Search Algorithms in TETRAD II

In this section we describe some of the constraint based, provably correct search procedures that we have implemented in TETRAD II. Our approach is to design algorithms that search for all RSEMs consistent with background knowledge that entail constraints on the covariance matrix that are judged to hold in the population. Depending on the type of background knowledge, and what kind of RSEM is sought, we use either vanishing partial correlation constraints or vanishing tetrad constraints. Because in many cases the number of possible constraints is too large to examine exhaustively, some of the algorithms we describe make sequential decisions about constraints and thus test only a subset of the possible constraints during the search process. These sequential procedures are still correct in the sense we defined in section 1.3.2, but might not be optimal on realistic samples because mistakes about constraints made early in the sequence can ramify into mistakes made later.

5.1 The Build Algorithm

The Build module[11] of TETRAD II takes as input:

1. sample data (either raw, or as a covariance matrix) and
2. background knowledge that constrains RSEM specification,

and gives as output:

1. a representation of the partial correlation equivalence class of RSEMs that is consistent with the background knowledge, and
2. a set of features that this class of RSEMs has in common.

Build performs statistical tests of hypotheses that specific partial correlations vanish in the population, and if it cannot reject the null hypothesis at a significance level set by the user, then the procedure accepts the null hypothesis (see the appendix in Scheines, et al., 1994). Because Build uses only information about which partial correlations are zero, it cannot distinguish between any members of a partial correlation equivalence class; hence its output is a representation of a partial correlation equivalence class of RSEMs consistent with background knowledge. In order to achieve enough efficiency to be practical for large numbers of variables (up to 100), the algorithms in Build use the results of tests of lower order partial correlation (i.e., correlations conditional on small sets of variables) to restrict the tests it needs to perform on partial correlations of higher order.

The algorithms are correct in the sense of section 1.3.2. That is, the background knowledge a user enters may include assumptions about:

1. whether the population RSEM contains correlated errors, or latent common causes;
2. time order among the variables;
3. known causal relationships among the variables;
4. causal relationships among the variables known not to hold.

[11] The Build module is documented in (Scheines, et al., 1994), and its algorithms described in detail in (Spirtes, et al, 1993).

5.1.1 Build for RSEMs Without Correlated Errors or Latent Common Causes

If you assume that the generating RSEM contains no latent common causes, then Build runs the PC algorithm, which is documented and traced in the appendix, and in (Spirtes, et. al, 1993; Scheines, et. al, 1994). The output of the PC algorithm is a **pattern**, (Verma and Pearl, 1990) which is a compact representation of a partial correlation (and covariance) equivalence class of RSEMs without correlated errors or latent common causes. A pattern contains a mixture of directed and undirected edges. If a pattern contains an edge A $\rightarrow$ B, then the directed graph of *every* RSEM represented by the pattern contains the edge A $\rightarrow$ B. If a pattern contains an edge A — B then A and B are adjacent in the directed graph of *every* RSEM represented by the pattern, but the graphs of some RSEMs represented by the pattern may contain the edge A $\rightarrow$ B, and others may contain the edge A $\leftarrow$ B. If a pattern contains no adjacency between A and B, then in every RSEM represented by the pattern A and B are not adjacent.

Suppose we measure only two variables A and B and find that they are significantly correlated. There are two RSEMs without correlated errors or latent variables containing just A and B that are compatible with A and B being correlated in the population: A $\rightarrow$ B, and A $\leftarrow$ B. The output of Build in this case is the pattern A — B, which represents the two RSEMs in this equivalence class. This illustrates the slogan "correlation is not causation", because the statistical information is not sufficient to predict the results of an ideal intervention on A or B.

In this example the output of Build is not useful for predicting the effects of ideal interventions. The next example shows how the output of Build can in some cases provide more useful causal knowledge. Suppose that for four measured variables, A, B, C, and D, from sample data we conclude that in the population, $\rho_{A,B} = 0$, $\rho_{A,D.C} = 0$, and $\rho_{B,D.C} = 0$, but that no other partial correlations (other than those entailed by those listed) vanish. In that case the output of Build is the pattern in Figure 15, which represents an equivalence class of RSEMs with only one member, also shown in Figure 15.[12]

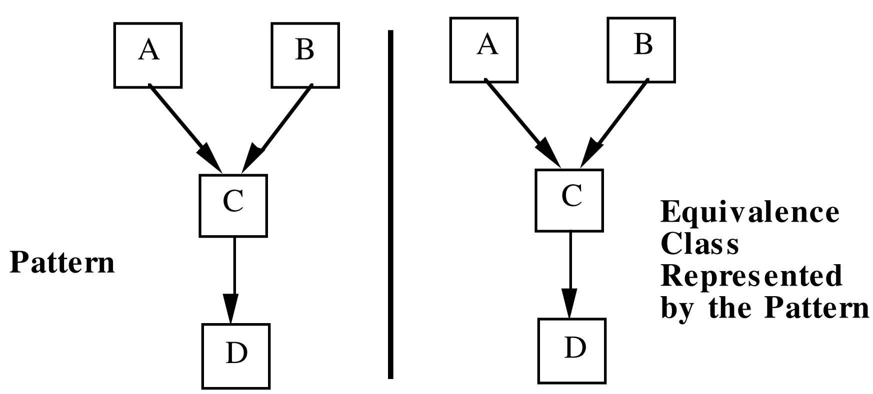


Figure 15

[12] See the appendix for an example in which the equivalence class is larger.

In this case, the output of Build is sufficient to predict the results of ideally intervening on A, B, C, or D. Of course, the assumption of no correlated errors or latent variables is a very strong one, and in the next section we consider what happens when it is abandoned.

5.1.2 Build for RSEMs with Correlated Errors

If you allow that the RSEM that generated the data might have correlated errors or latent common causes, then Build runs the FCI algorithm, which is documented in (Spirtes, et. al., 1993, chapter 6). The output of the FCI algorithm is a **partial ancestor graph** (PAG).[13] A is an **ancestor** of B in a directed graph when there is a directed path from A to B. Just as patterns represent features common to a partial correlation equivalence class of RSEMs without latent variables, PAGs represent features common to a set of RSEMs that are partial correlation equivalent over the measured variables. (In this section, for the sake of brevity, we will refer to the PAG simply as an equivalence class.) We will illustrate with the two examples from the previous section.

Again, suppose we measure two variables A and B, and find that they have a significant correlation and conclude that they are correlated in the population. The output of Build in this case is the partial ancestor graph shown in Figure 16. Because we have placed no limit on the number of distinct latent variables, the equivalence class represented by the output is actually infinite, and we have shown only a few members of the equivalence class in Figure 16. The presence of a "o" at both ends of an edge in a PAG makes no claim about the ancestor relationship common to every member of the equivalence class. Note that in some of the RSEMs represented by the PAG (e.g. (i) and (iv) of Figure 16), A is an ancestor of B, and in others (e.g. (ii) and (iii) of Figure 16) it is not. Similarly, in some of the members of the equivalence class (e.g. (ii) of Figure 16) B is an ancestor of A, and in others (e.g. (i), (iii) and (iv) of Figure 16) it is not. Thus this PAG shows us that we cannot predict the results of ideally intervening to change either A or B from this data without further background knowledge.

[13] In fact the output is described in (Spirtes, et. al., 1993, and Scheines, et. al., 1994) as a POIPG, or partially oriented inducing path graph. POIPGs can, without loss of generality, be interpreted much more naturally as PAGs.

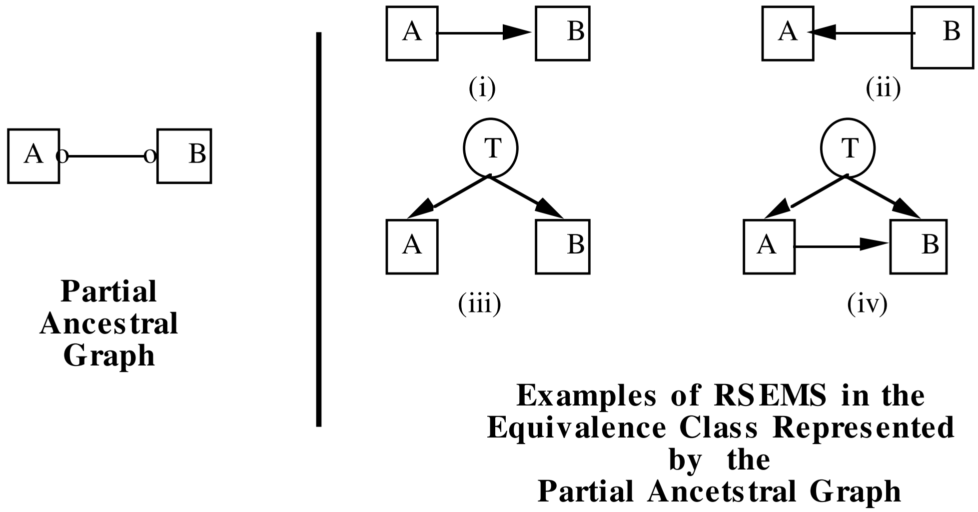


Figure 16

Whereas the pattern A — B informed us that either A is a cause of B or B is a cause of A, the PAG A o—o B informs us that either A is a cause of B, B is a cause of A, there is a latent common cause, or there is some combination of these causal connections responsible for the correlation. The next example shows how PAGs output by Build can be used to predict the effects of some ideal interventions. Consider the example from Figure 15 again, where there are four measured variables A, B, C, and D, and we conclude from the data that in the population $\rho_{A,B} = 0$, $\rho_{A,D.C} = 0$, and $\rho_{B,D.C} = 0$, but that no other partial correlations vanish. Assuming that correlated errors might exist in the generating RSEM, the output of Build is the PAG in the left hand side of Figure 17.

A and C are adjacent in the PAG because the correlation of A and C conditional on every subset of the measured variables does not vanish (i.e. $\rho_{A,C}$, $\rho_{A,C.B}$, $\rho_{A,C.D}$, $\rho_{A,C.BD}$ do not vanish.) The "o" at the A end of the edge between A and C entails neither that A is an ancestor of B in every member of the equivalence class nor that A is not an ancestor of B in every member of the equivalence class. The ">" at the C end of the edge between A and C in the PAG means that C is not an ancestor of A in any RSEM in the partial correlation equivalence class. Similarly, C is not an ancestor of B, and D is not an ancestor of C in any RSEM in the partial correlation equivalence class. Finally, a "—" at the C end of the edge between C and D means that C is an ancestor of D in every RSEM in the equivalence class.

From this PAG we can make predictions about the effects of some ideal interventions, but not others. For example, it is not possible to determine if an ideal intervention on A will affect C, because in some members of the equivalence class A is a cause of C, and in others it is not. On the other hand, it is possible to determine that an ideal intervention on C will affect D, because C is a cause of D in every RSEM in the equivalence class. (And given the distributional assumption, it is also possible to determine the size of the effect that an ideal intervention on C will have on D. See Spirtes, et al. 1993, chapter 7).

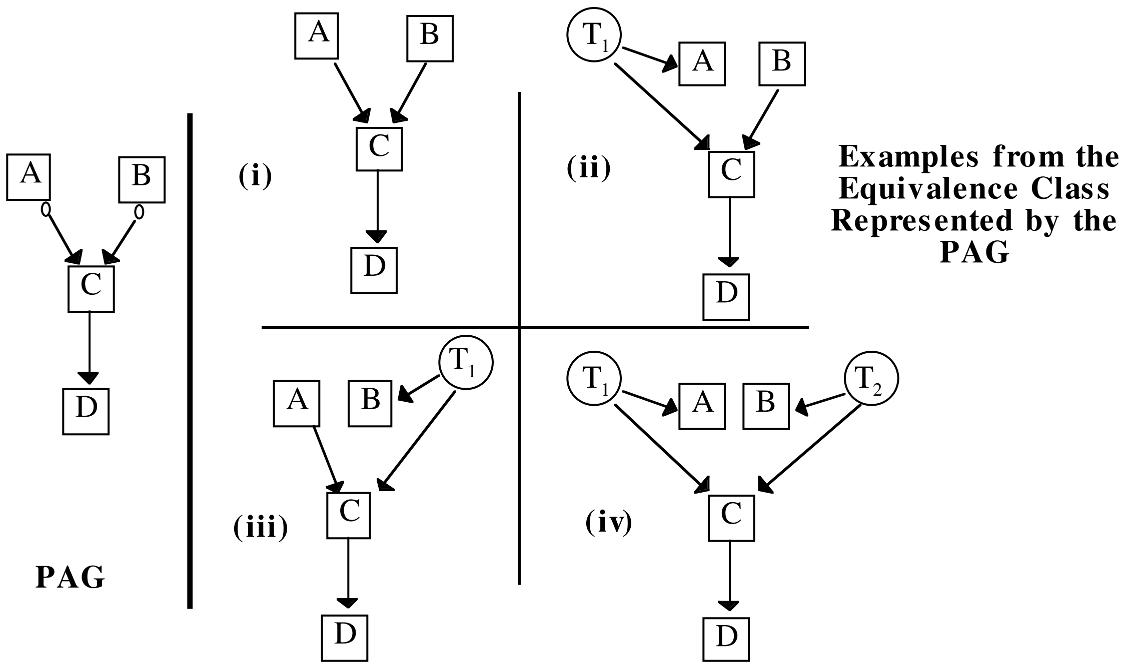


Figure 17

The partial correlation equivalence class in Figure 17, which includes RSEMs with latent variables, is much larger (in fact it is infinite) than the partial correlation equivalence class in Figure 15, which does not include models with latent variables. This in turn means that the conclusions that we can draw are weaker than if we assume that the generating RSEM has no correlated errors or latent common causes. For example, with this assumption we can conclude that A is a cause of C; without it we cannot. With the assumption we can estimate the size of the effect that an ideal intervention on A will have on C; without it we cannot. While the conclusions that can be drawn even without the assumption of no latent variables are weaker than when the assumption is made, they are not trivial. Asymptotically, we can reliably conclude that C is a cause of D, and we can estimate the size of the effect an ideal intervention on C will have on D.

It should also be noted that even though in general not all of the members of the partial correlation equivalence class are covariance equivalent, this does not affect the reliability of the conclusions. It simply means that there may be stronger conclusions that could be drawn if we used more information than simply which partial correlations vanish.

Finally, we note that there are examples in which there is no RSEM without latent variables that is compatible with a correlation matrix, but there are RSEMs with latent variables that are. Suppose that we measure A, B, C, and D and from the data conclude that in the population, $\rho_{A,C} = 0$, $\rho_{A,D} = 0$, and $\rho_{B,D} = 0$, but that no other partial correlations (other than those entailed by these three) vanish. In that case the output of Build is A o$\rightarrow$ B $\leftrightarrow$ C $\leftarrow$o D. The double headed arrow between B and C means that in every member of the equivalence class represented by the PAG B is not an ancestor of C and C is not an ancestor of B. This is only possible in an RSEM with a latent variable causing both B and C, so every member of the equivalence class contains a latent variable.

5.1.3 What Can Go Wrong

In general, the correctness of Build's output depends upon several factors:

1. The correctness of the background knowledge input to the algorithm.
2. Whether the recursiveness condition holds, i.e., that there are no feedback loops.
3. Whether the Causal Independence assumption holds.
4. Whether the Faithfulness assumption holds.
5. Whether the distributional assumptions made by the statistical tests hold.
6. The power of the statistical tests against alternatives.
7. The significance level used in the statistical tests.

In the case of Build under the assumption of no latent variables, it is not difficult to take the output pattern which represents a partial correlation equivalence class (and a covariance equivalence class) of RSEMs, and use it to find a single RSEM in the equivalence class. A sketch of this process is described in the TETRAD II manual and can be automated (Meek, 1995). Once this is done, the user can estimate and test the selected RSEM using such programs as EQS, or LISREL. (All RSEMs in the partial correlation equivalence class parameterized by their respective ML parameter estimates have the same p-value.) In addition, the user can approximate the sampling distribution using the method described in section 1.3.3. The user should keep in mind however, that the sampling distribution of the output may show that even when the RSEMs suggested by TETRAD II fit the data very well, it is possible that there are other RSEMs that will also fit the data well and are equally compatible with background knowledge, particularly when the sample size is small. This suggests that further research on the search is needed, and that Build might be improved by outputting multiple patterns--something which can be done in a limited way in the present implementation by varying the significance level used in the procedure. Also, at large sample sizes, even slight deviations from normality or linearity can lead to the rejection of an otherwise correct RSEM. Finally, if a model produced by search is tested on the data used to find the model specification, the p-value of the test is *not* a measure of the error probability of the model specification procedure. For a discussion of the meaning of such p-values, see Glymour, et al. (1987). Where possible, models generated from one sample should be cross-validated on others.

In the case of Build under the assumption of latent variables, more research is needed to find out how to construct (efficiently) from the PAG which represents the entire partial correlation equivalence class a single, representative RSEM. In this case the output partial correlation equivalence class is not a covariance equivalence class, so that different RSEMs represented by the output can have different p-values when parameterized by their ML parameter estimates. More research is needed on estimating and testing the output of Build under the assumption of latent variables. Spirtes, et al. 1993 describes some algorithms that can be used for predicting the effects of some policy interventions from a given PAG.

5.2 Specification Search for Latent Variable RSEMs: Purify and MIMbuild

In many applications of structural equation modeling, the focus of interest is the causal relationships among latent variables. In many such cases the latent variables are measured with multiple indicators, and the output of Build on data for these indicators is correct but uninformative; the correct RSEM entails no zero partial correlation constraints on the indicators alone and the output of Build on the indicators is completely connected and completely undirected, whether it is a pattern or a PAG. In these cases the Purify and MIMbuild modules of TETRAD II can help in RSEM specification. Purify helps locate unidimensional measurement models (Anderson, Gerbing, & Hunter, 1987; Anderson & Gerbing, 1988; Scheines, 1993). The basic idea of unidimensionality is that each indicator measures exactly one latent and all error terms are uncorrelated (the exact definition is more complicated, and presented in section 8.3 of the appendix). Finding a unidimensional measurement model is one way in which the correlations among the latent variables may be estimated consistently. Also, given a unidimensional measurement model, the MIMbuild module uses vanishing tetrad constraints to search the space of structural models, i.e., RSEM models containing only the latent variables.

5.2.1 Purify

We make our explanation of both Purify and MIMbuild concrete by accompanying it with an example taken from the user's manual to TETRAD II (Scheines, et al, 1994, chapter 9). The example shows how Purify can aid in finding a unidimensional measurement model and why it is important to do so. The population RSEM is shown in Figure 18. Our data for the example consist of the correlations among the X variables in a pseudo-random multivariate normal sample drawn from a random parameterization of this RSEM (N=2,000).

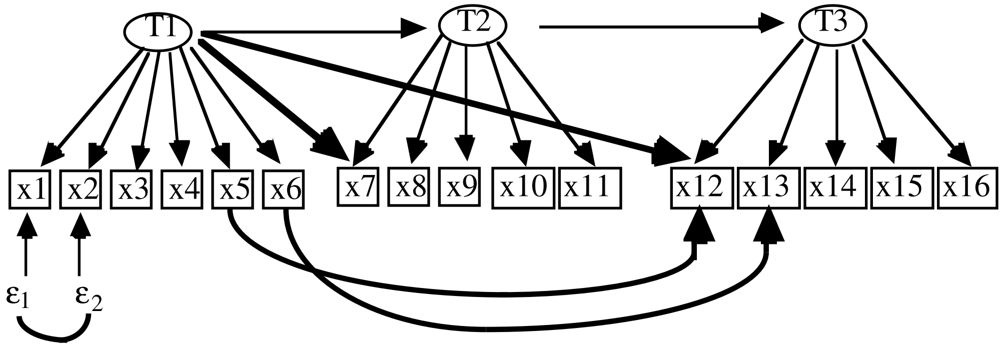


Figure 18: Population RSEM

Suppose our interest is in the causal relationships between the three latent variables T_1, T_2, and T_3. The part of the RSEM specifying the relationships between the latent variables is called the structural model; the rest is called the measurement model. In this case the population structural model is shown in Figure 19, and the population measurement model is shown in Figure 20.

Figure 19: Population Structural Model

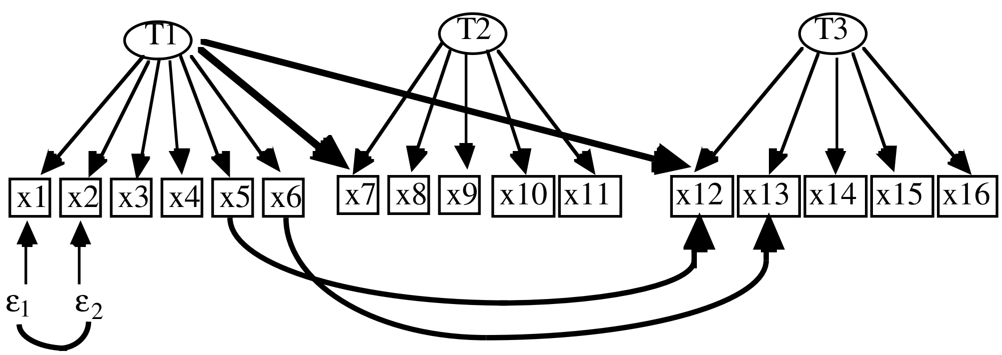


Figure 20: Population Measurement Model

One approach to this problem is to use background knowledge to build a measurement model for each latent variable, and then perform a specification search for the structural model constrained by background knowledge and aided by a computer, e.g., the Search module of TETRAD II, or the modification indices of LISREL, or the Lagrange Multiplier statistic of EQS. There are several problems with this approach.

First, while background knowledge may often be sufficient to construct part of the population measurement model (i.e. we may know which of the latent variables each indicator variable is a measure of), background knowledge is seldom detailed enough to completely specify the full population measurement model (e.g. an indicator may be a measure of several latent variables, or indicator variables may have correlated errors). This means that the specification search must also seek to correct the hypothesized measurement model, as well as discover the structural model. Because there are often a large number of indicator variables, this search space is astronomically large. Moreover, a search that at each step chooses to add the edge (free the parameter) that will most improve the fit can easily go wrong for several reasons. First, it may be that freeing a number of different parameters improves the fit to the same degree, so there is no way to choose which parameter to free at that point in the search. In addition, there may be pairs of parameters which if freed will greatly improve the overall fit, even though freeing either parameter by itself does not improve the fit much. Also, when the initial RSEM to be modified is far from the population RSEM, the parameter estimates may be far from their population values, which can affect the estimates of the Lagrange multipliers, or may prevent the estimation algorithms from converging at all.

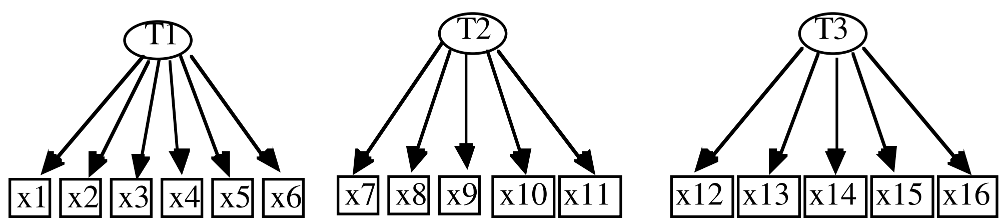


Figure 21: Hypothesized Measurement Model

The Purify module represents a different approach to the problem that is a provably correct[14] algorithm for finding unidimensional measurement models (Scheines, 1993). Instead of searching for parameters to free, i.e., edges to add, Purify searches for a submodel of the originally specified measurement model that contains a subset of the indicators originally specified, but that is correctly specified as unidimensional. Such a submodel can be used to find consistent estimates of the correlations between the latent variables, and thus aid in the search for structural models.[15]

For example, the measurement model in Figure 20 is not unidimensional because of the edges and correlated errors that are in boldface. But note that the population model

[14] Purify is a correct search procedure in the following sense. Given that there are correctly specified unidimensional submodels of the initially specified measurement model with at least three indicators for each latent, then as the sample grows without bound the probability that Purify will find one of the unidimensional submodels converges to one.

[15] This two stage search process was also suggested by Anderson and Gerbing, (1988).

does contain a unidimensional submodel, shown in Figure 22, which is obtained by simply removing X_1, X_7, X_{12}, and X_{13} from the model.

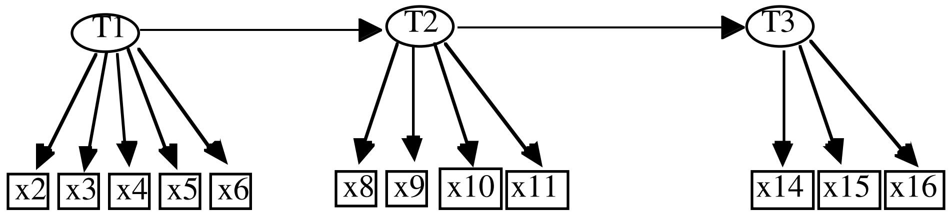


Figure 22. Model with Correctly Specified Unidimensional Measurement Model

Purify searches for unidimensional submodels in the following way. First we suppose that we are given as input a hypothetical measurement model which is unidimensional, for example, the measurement model shown in Figure 21. We assume the input measurement model is a submodel of the population measurement model, that is, every edge specified in the input measurement model exists in the population measurement model. However, we do not assume that the input measurement model is complete; the population measurement model may be non-unidimensional because a single indicator may be caused by multiple latents, cause other indicators, or have correlated errors with other variables.

Given this input, if the population measurement model is unidimensional, it entails a characteristic set of vanishing tetrad differences, *regardless of the population structural model* (Scheines, 1993). For example, if the population measurement model is unidimensional, and X_1, X_2, and X_3 measure a single latent, then $\rho_{X1,X2}\, \rho_{X3,X4} - \rho_{X1,X3}\, \rho_{X2,X4}$ is entailed to be zero regardless of the population structural model. This means that Purify can test whether the specified measurement model is truly unidimensional without knowing the structural model. If the characteristic set of vanishing tetrad differences entailed by a unidimensional measurement model is judged to hold in the population, Purify concludes that the measurement model specified is truly unidimensional, and halts. If the population measurement model is the one in Figure 20, some of the tetrad differences entailed by the initially specified model in Figure 21, e.g., $\rho_{X1,X2}\, \rho_{X3,X4} - \rho_{X1,X3}\, \rho_{X2,X4}$, are not entailed to vanish, and with a representative sample Purify will conclude that the population measurement model among the given set of indicator variables is not unidimensional. Purify then begins to search for a submodel that is unidimensional by sequentially eliminating indicators. In general, searching all subsets of the given measured indicators for a set of indicators that form a unidimensional measurement model would take too long, due to the enormous number of subsets. But by examining *which* vanishing tetrad differences do not hold in the population, the algorithm can greatly narrow the search, making it feasible to handle initially specified measurement models with more than 50 measured variables in minutes. In this case on simulated data it correctly removes X_1, X_7, X_{12}, and X_{13} from the measurement model, leaving a set of indicators that have a measurement model correctly specified as unidimensional.

5.2.2 MIMbuild

In many studies the theoretical question addressed cannot be reduced to the significance of a single parameter in an otherwise reliably specified model. There might be many latent variables, and the problem of finding a reasonable structural model is then difficult. With just four latent variables there are well over 700 structural models with no correlated errors. Even with substantial background knowledge, this is a large space to search. With eight latent variables the space is astronomical. Several strategies for automatic structural model search are possible. One might begin with a null structural model and do a Lagrange Multiplier search limited to structural parameters. To the best of our knowledge no one has studied the behavior of this strategy. One might estimate the correlations among the latent variables and then apply Build to the latents as if they were measured. In our experience, this works well in simulation studies at moderate to large sample sizes, but we do not know how to properly adjust the sample size when testing for vanishing partial correlations among latents that are being treated "as if" they are measured. A third alternative takes further advantage of the vanishing tetrad difference constraint.

We have already seen that if the population measurement model is unidimensional, a SEM entails a characteristic set of vanishing tetrad differences, regardless of what the population structural model may be. But if the measurement model is unidimensional, there are other tetrad differences which are entailed to vanish for some structural models, but are not entailed to vanish for other structural models. These constraints are extremely easy to compute and test, and the tests are not susceptible to specification error in other parts of the structural model (Scheines, 1993). For example, in the model in Figure 22, (where the population measurement model is unidimensional) all three tetrad constraints involving one indicator from T_1, two from T_2, and one from T_3 are entailed by the model if and only if there is no edge between T_1 and T_3. The MIMbuild algorithm uses tests of vanishing tetrad differences to construct a set of structural models that entail vanishing partial correlations among latent variables judged to hold in the population.

The set of structural models that MIMbuild outputs entail the same set of unconditional correlations and partial correlations with only one variable in the conditioning set. Because it can output models which are not fully partial correlation equivalent or covariance equivalent, MIMbuild represents only a partial solution to the RSEM structural model specification problem. A "?" is attached to those parts of MIMbuild's output that might change if second order or higher partial correlations among latent variables could be tested. Section 8.4 of the appendix details the sense in which MIMbuild is a correct estimator of structural models.

6. Applications

The TETRAD II procedures have been used to study job satisfaction among military personnel (Callahan & Sorensen, 1992), to develop psychiatric measures (Prigerson, et. al., 1995), to predict survival and death in pneumonia patients (Cooper, et. al., 1995), to study mechanisms in plant biology (Shipley, forthcoming), to study dropout rates in American universities (Druzdzel & Glymour, 1994), even to recalibrate instruments on orbiting satellites (Waldemark & Norqvist, 1995). In this section we will illustrate the application of the search procedures to three data sets, two published and one simulated. In the case of the empirical examples we do not mean to endorse the assumptions made by the researchers who used the data sets, or the scales they constructed. In the first case our intent is to show how the search procedures implemented in TETRAD II can be used to find plausible alternatives to a published model. The existence of these alternatives weakens the evidential support for conclusions published, but it is not our intent to claim that the alternatives found by TETRAD II are correct. In the second case we ran the Purify and Search procedures on a published data set with the same results as those published, and in the third case we show how the procedures perform on a very large search space. Other applications can be found in (Scheines, et. al., 1994), and in (Spirtes, et. al., 1993).

6.1 Finding Alternative Models

Before giving a proposed hypothesis any great credence, good scientific practice ought to try to articulate and investigate every serious alternative. A frequent objection to causal models in any discipline is that they are arbitrarily selected without any sound arguments that would exclude alternative explanations of data. That some cherished causal model cannot be rejected statistically is little reason to believe its causal claims: There might be alternatives that also cannot be rejected statistically, but that make contrary causal claims. Published studies may be defective in their general distributional assumptions, in their data collection procedures, or in their assumptions about what is influencing what. Here is an illustration of how TETRAD II is meant to be used to help search for and articulate alternative causal explanations under varying background assumptions. In a study published in the American Sociological Review, Timberlake and Williams (1984) claimed that foreign investment in Third World or "peripheral" nations causes the exclusion of various groups from the political process. In other words, foreign investment inhibits democracy. Their empirical case for this claim rests on fitting a linear regression.

PO	FI	EN	CV
1.000			
-0.175	1.000		
-0.480	0.330	1.000	
0.868	-0.391	-0.430	1.000

Table 1. Political Repression Data (N = 72)

They develop measures of political exclusion (PO), foreign investment penetration (FI), energy development (EN), civil liberties (CV) (measured on an ordered scale from 1 to 7, with lower values indicating *greater* civil liberties.) We show the correlations given by Timberlake and Williams for these variables on 72 "non-core" countries in Table 1.

An apparent embarrassment to their claim is that political exclusion is negatively correlated with foreign investment; further, foreign investment is negatively correlated with the civil liberties scale (and hence because of their reverse ordering of the civil liberties scale, *positively* correlated with civil liberties). To defeat this objection, Timberlake and Williams regress PO on the other variables on the assumption that the coefficient relating FI to PO is a superior measure of FI's causal influence on PO than is their simple correlation. A regression on the correlations above yields:

$$\begin{array}{lcccccc} PO & = & .227*\mathbf{FI} & - & .176*\mathbf{EN} & + & .880*\mathbf{CV} + \varepsilon \\ & & (.058) & & (.059) & & (.060) \\ & & 3.941 & & -2.985 & & 14.604 \end{array}$$

You can see that the crucial coefficient is positive and highly significant. Timberlake and Williams took this as evidence to support the claim that foreign investment causes more political exclusion. They do not explicitly consider any alternative models.

But a regression model is only one among many that might describe the relations among these four variables. To search for alternatives, we again use TETRAD II's Build module, again without considering whether the linearity and normality assumptions are warranted.[16] Without assuming that all common causes are included in the variables measured, and using a significance level (α) of .05 for its statistical hypothesis tests, Build's output is the PAG in Figure 23.

[16]Because our aim is to illustrate the use of TETRAD II in finding alterantives to a given model, the correctness of the distribution and linearity assumptions made by Timberlake and Williams is not at issue. We note, however, that we were unable to reproduce their correlation matrix from the sources they cite.

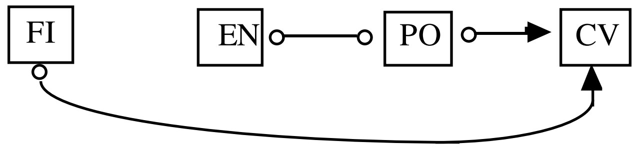


Figure 23. Build output at α = .05

Since this structure entails that foreign investment (FI) and political exclusion (PO) are uncorrelated, we increase α to .15, at which point Build produces the PAG in Figure 24.

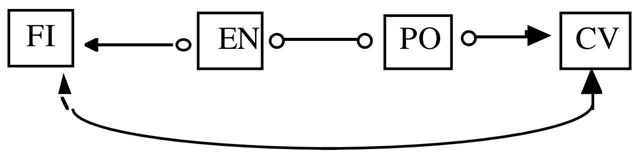


Figure 24. Build output at α = .15

Because all of the connections in the PAG involving FI have arrowheads directed into FI, these data indicate that foreign investment is not a cause of any of the other variables. A large number of causal models are members of the equivalence class represented by the output in Figure 24. The model in Figure 25 is one of the simplest in this class, and is plausible besides.

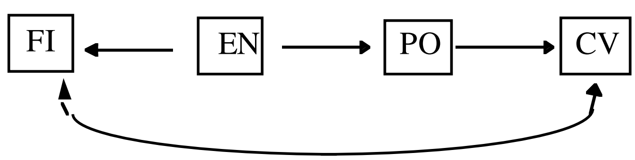


Figure 25. An Alternative to Timberlake and William's Model

This model asserts that EN (a measure of economic development) causes both the level of foreign investment and the level of political exclusion. Political exclusion causes the lack of civil liberties, and there is some unmeasured common cause connecting foreign investment and civil liberties (or in other terms, that their errors are correlated). Estimating this model with EQS yields a χ^2 = .136 with 2 degrees of freedom, with p(χ^2) = .934. We give the coefficients with their standard errors and t-statistics in Figure 26 below.

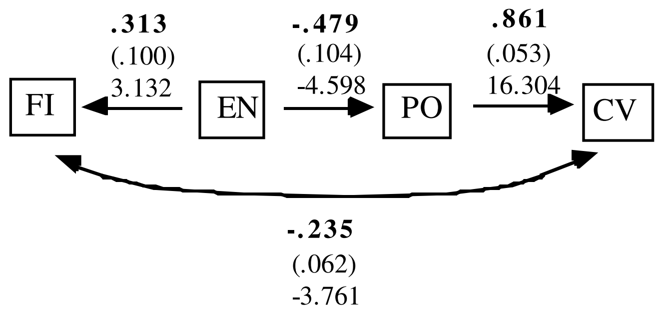


Figure 26. Estimated Alternative Model

The signs of the coefficients suggest that the relation between FI and PO is negative and mediated by a common cause, contrary in two ways to Timberlake and Williams' hypothesis. We do not mean to suggest that this analysis shows our alternative to be correct. At this small a sample size statistical tests have little power against alternatives, so it is difficult to statistically distinguish between two models even when they are not statistically equivalent. Our point is to show how the Build module can be used to search for plausible alternatives to a given model.

6.2 Specifying Measurement Models of Political Democracy

Bollen (1980) studied whether a number of measures of political democracy were unidimensional indicators of a common feature of societies. Bollen used the following measures:

PF	press freedom
FG	freedom of group opposition
GS	government sanctions
FE	fairness of elections
ES	executive selection
LS	legislature selection

He considered the unidimensional factor model specified in Figure 27, where it is understood that for each of the measured variables there is an error term.

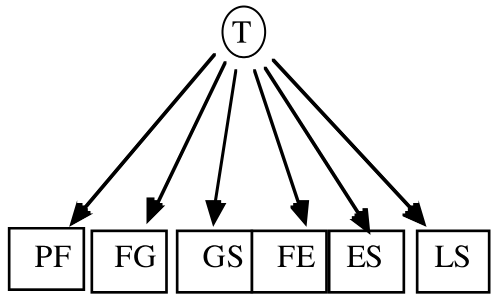


Figure 27. Initial Measurement Model of Political Democracy

Bollen estimated this model with LISREL and found that the data reject it.[17] Instead of attempting to locate and discard the impure indicators, Bollen elaborated his original model by correlating error terms (Figure 28). When estimated with EQS, this model has a χ^2 of 6.009 based on 6 degrees of freedom, with $p(\chi^2) = 0.42218$. The Search module of TETRAD II, which uses vanishing tetrad differences to search for elaborations of an initial model, arrives at a set of factor models which contains Bollen's model and others.

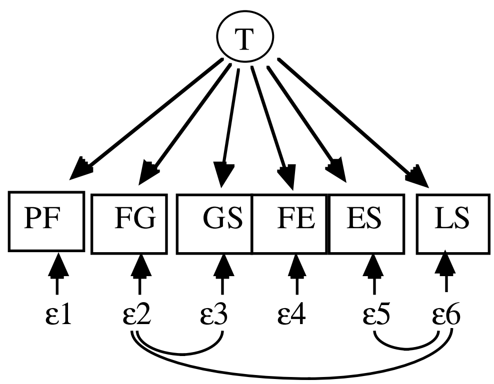


Figure 28. Bollen's Respecification of the Measurement Model

Although Bollen's final measurement model of democracy fits the data well, it is not unidimensional. To find a unidimensional submodel, we can run Purify on Bollen's original model (Figure 27) and data. Giving the initial model and the measured covariances to Purify, FG and LS are identified as impure indicators and discarded, resulting in the measurement model we picture in Figure 29.

[17]EQS yields a $\chi^2 = 42.076$ based on 9 degrees of freedom, with $p(\chi^2) < 0.001$.

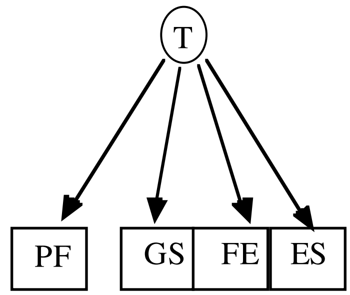


Figure 29. Sub-model found by Purify

Estimating the resulting unidimensional measurement model (Figure 29) with EQS yields a $\chi^2 = 1.687$ based on 2 degrees of freedom, with $p(\chi^2) = 0.43013$.

6.3 A Large Search Space: The Alarm Network

By interviewing several medical experts, Beinlich, et. al., (1989) developed a large causal model of the probabilistic relations in emergency medicine (Figure 30).[18] Using the directed graph associated with this model (Figure 30), called the ALARM network, linear coefficients with values between .1 and .9 were randomly assigned to each directed edge in the graph. Using a standard joint normal distribution (mean 0, variance 1) on the exogenous variables, three sets of simulated data were generated, each with a sample size of 2,000. The covariance matrix and sample size were given to the TETRAD II program. No information about the orientation of the variables was given to the program. With 37 variables, the space of possible models is astronomical,[19] yet the program required less than fifteen seconds to return a pattern on a Decstation 3100. In each trial the output pattern omitted two edges in the ALARM network; in one of the cases it also added one edge that was not present in the ALARM network.

[18] Beinlich's network was over discrete variables, and we have run Build on a discrete version of this network with results similar to those we report here for a SEM interpretation of the structure.

[19] With 37 variables there are 666 pairs of variables. Assuming that each pair <X,Y> has either 1) no edge between therm, or 2) an edge from X to Y, or 3) an edge from Y to X, the number of possible models is 3^{666}. The actual number to search is smaller, because some of these models will contain cycles, but the space remaining is still far too big to search by evaluating each member.

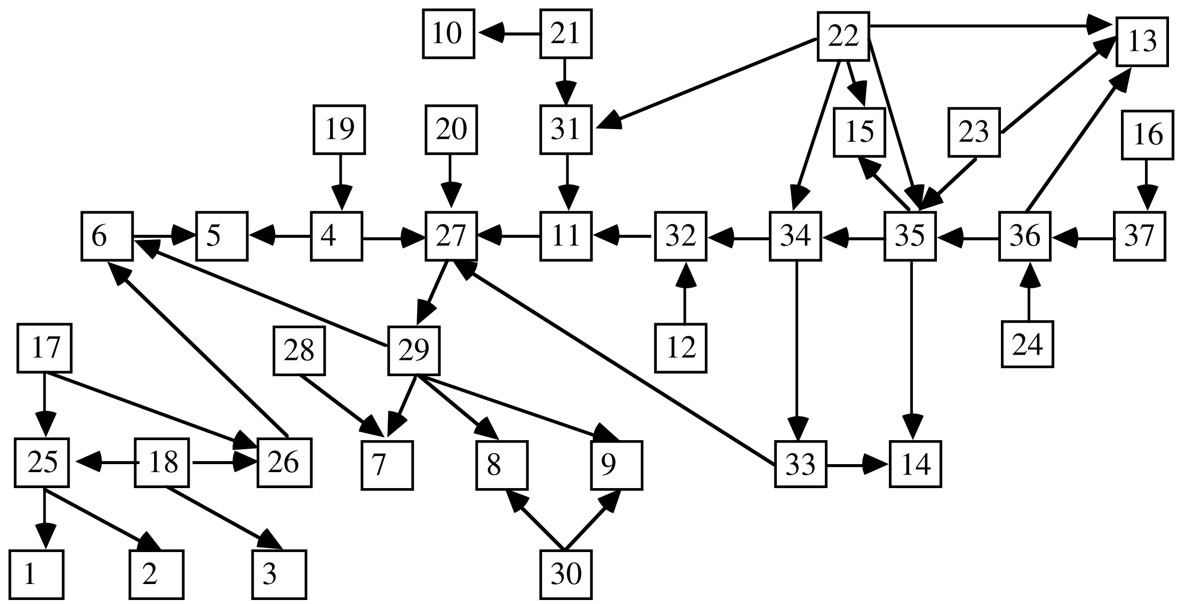


Figure 30. The ALARM Network.

7. *Conclusion*

The work we have described is based on assumptions that are implicit, and sometimes explicit, throughout scientific practice. The Causal Independence assumption, for example, posits a relation between the absence of causal connection and statistical independence that is fundamental to experimental design; the Faithfulness assumption states a preference for explanation by structure over explanation by coincidence that, in various forms, is used in every science. Consequences of these assumptions were worked out for special cases by many social scientists, for example by Simon (1954), by Blalock (1962) and by Costner (1971).

The methods we have described are incomplete, and there is a great deal of research that remains to be done and that should lead to improved modeling. Important outstanding problems include improving the reliability of model search through better statistical and algorithmic procedures, deriving computationally tractable algorithms for testing covariance equivalence of linear, latent variable models, finding correct methods for clustering measured variables that share a latent common cause, completing and implementing known algorithms for predicting the outcomes of interventions from partial causal and distributional specifications, implementing searches for non-recursive models, and much more.

We hope the TETRAD II procedures will become a useful part of the methodological toolkit used by quantitative social scientists, that the statistical and social scientific communities will investigate the questions we have raised, and improve the techniques we have suggested.

8. *Appendix*

8.1 D-Separation

An **undirected path** between X_1 and X_n in a graph G is a sequence of vertices $<X_1, ..., X_n>$ such that for each pair of vertices X_i and X_{i+1} ($1 \leq i < n$) that are adjacent in the sequence, either there is an edge $X_i \rightarrow X_{i+1}$ or an edge $X_{i+1} \rightarrow X_i$ in G. A **directed path** between X_1 and X_n in a graph G is a sequence of vertices $<X_1, ..., X_n>$ such that for each pair of vertices X_i and X_{i+1} ($1 \leq i < n$) that are adjacent in the sequence, there is an edge $X_i \rightarrow X_{i+1}$ in G. X is a **descendant** of Y in directed graph G if and only if there is a directed path from Y to X or Y = X. In graph G a vertex X_i is a **collider on undirected path** U if and only if U contains a subpath $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$. Otherwise if X_i if on U, X_i is a **noncollider on** U. Following Pearl (1988), in a directed acyclic graph G, for disjoint sets of vertices **X**, **Y**, and **W**, **X** and **Y** are **d-separated** given **W** in G if and only if there exists no undirected path U between a member of **X** and a member of **Y**, such that (i) every collider on U has a descendent in **W** and (ii) no other vertex on U is in **W**. An illustration of d-separation is given in the directed acyclic graph shown in Figure 31.

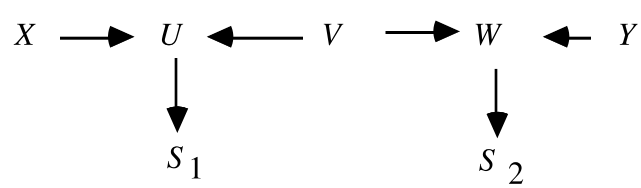


Figure 31

{X} and {Y} are d-separated given the empty set
{X} and {Y} are not d-separated given set $\{S_1, S_2\}$
{X} and {Y} are d-separated given the set $\{S_1, S_2, V\}$

8.2 The Build Algorithm with the Assumption of No Correlated Errors

The Build module uses the PC algorithm (Spirtes, et al., 1993) when it is assumed that the RSEM that generated the data has no correlated errors or latent common causes. Let **Adjacencies**(C,A) be the set of vertices adjacent to A in a graph C.[20] In the algorithm, the graph C is continually updated, so **Adjacencies**(C,A) changes as the algorithm

[20]Note that C is not defined to be a directed graph, so that edges can be either directed or undirected.

progresses. **A** \ **B** is the set of members of **A** that are not elements of **B**. We adopt the convention that if **S** is the empty set, then $\rho_{X,Y.S}$ is $\rho_{X,Y}$.[21]

PC Algorithm:

A) Form the complete undirected graph C on the vertex set **V**.

B) n = 0.

repeat

  repeat

    select an ordered pair of variables X and Y that are adjacent in C such that the number of vertices in **Adjacencies**(C,X)\{Y} is greater than or equal to n;

    repeat

      select a subset **S** of **Adjacencies**(C,X)\{Y} with n vertices;

      if the statistical test fails to reject $\rho_{X,Y.S} = 0$, then delete edge X — Y from C and set **Sepset**(X,Y) = **S** and **Sepset**(Y,X) = **S**;

    until every subset **S** of **Adjacencies**(C,X)\{Y} with n vertices has been selected or some subset **S** has been found for which $\rho_{X,Y.S} = 0$;

  until all ordered pairs of adjacent vertices X and Y such that **Adjacencies**(C,X)\{Y} has greater than or equal to n vertices have been selected;

  n = n + 1;

until for each ordered pair of adjacent vertices X, Y, **Adjacencies**(C,X)\{Y} has less than n vertices.

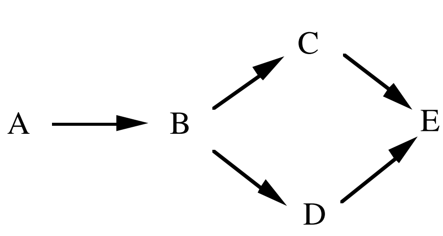
C) For each triple of vertices X, Y, Z such that the pair X, Y and the pair Y, Z are each adjacent in C but the pair X, Z are not adjacent in C, orient X — Y — Z as X → Y ← Z if and only if Y is not in **Sepset**(X,Z).

D) **repeat**

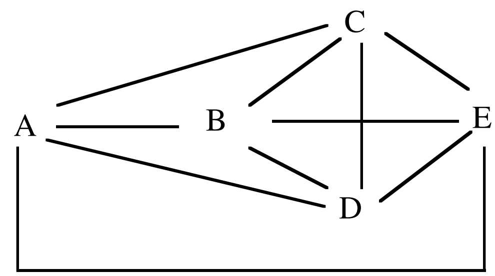
  If X → Y — Z in C, and X and Z are not adjacent in C, then orient as Y → Z.

until no more edges can be oriented.

[21] The simplified version of the algorithm presented here does not make all of the orientations that are theoretically possible, because we have found in practice that additional orientation rules, while theoretically correct, are in practice unreliable until the sample size is very large.



True Graph



Complete Undirected Graph

n = 0

No zero correlations.

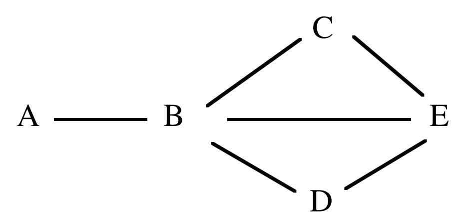
n = 1

$\rho_{AC.B} = 0$ Sepset(A,C) = {B}

$\rho_{AE.B} = 0$ Sepset(A,E) = {B}

$\rho_{AD.B} = 0$ Sepset(A,D) = {B}

$\rho_{CD.B} = 0$ Sepset(C,D) = {B}



n = 2

$\rho_{BE.CD} = 0$ Sepset(B,E) = {C,D}

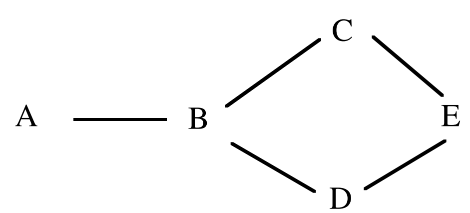


Figure 32: A trace of the adjacency stage of the PC algorithm

Figure 32 traces the operation of the first two parts of the PC algorithm for input Faithful to the true graph in Figure 32.

Although it does not in this case, stage B) of the algorithm may continue testing for some steps after the correct undirected graph has been identified. After stage B) has been completed, the undirected graph at the bottom of Figure 32 is partially oriented in step C of the PC algorithm. The triples of variables with only two adjacencies among them are:

A — B — C; A — B — D;
C — B — D; B — C — E;
B — D — E; C — E — D;

E is not in **Sepset**(C,D) so C — E and E — D collide at E. None of the other triples form colliders. The final pattern produced by the algorithm is shown in Figure 33.

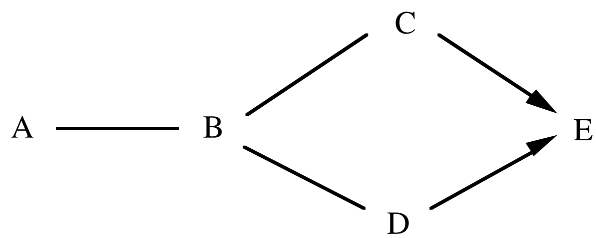


Figure 33. Final Pattern Output by PC.

The pattern in Figure 33 represents the partial correlation (and covariance) equivalence class of RSEMs we show in Figure 34.

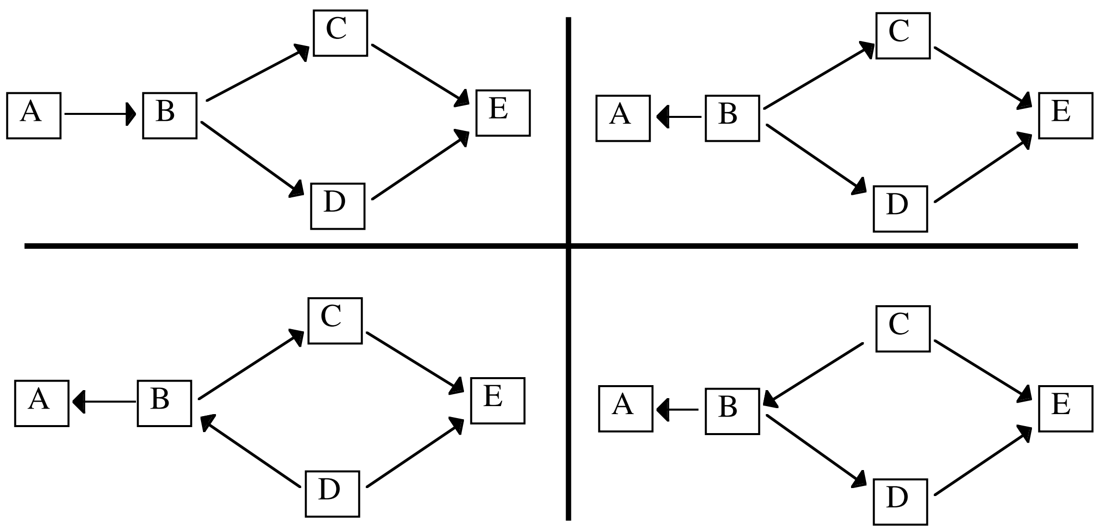


Figure 34. Equivalence Class of RSEMs represented by the Pattern in Figure 33.

8.3 Purify

We say that an indicator X **measures** T in an RSEM if there is a directed edge from T to X in the directed graph associated with the RSEM. If G is the directed graph of an RSEM with latent variables **T** and a measurement model with indicators **X** such that every $X_i \in$ **X** measures some latent T in **T**, then X_i is a **pure indicator** in G if and only if

X_i is d-separated from every other indicator by T. A measurement model is pure, or **unidimensional**, if and only if all of its indicators are pure.

8.4 MIMbuild

MIMbuild takes as input a unidimensional measurement model and covariance data over the indicators in this model, and outputs a modified pattern Π, where the adjacencies can be either: $\rightarrow$, —, ?$\rightarrow$?, or ?—?. The edges labeled with a "?" indicate that the MIMbuild algorithm cannot determine if there is an edge in the population graph or not. Suppose that G is the directed graph of an RSEM with no correlated errors that has latent variables **T** and indicators **X'**. Then if Π is the output of MIMbuild on a correctly specified unidimensional measurement model for **T** and $\mathbf{X} \subseteq \mathbf{X'}$, and the statistical decisions about vanishing tetrad differences among **X** made on a sample from a Faithful parameterization of G are correct, then

1. If T_i and T_j are not adjacent in Π, then they are not adjacent in G.
2. If T_i and T_j are adjacent in Π and the edge is not labeled with a "?", then T_i and T_j are adjacent in G.
3. If $T_i \rightarrow T_j$ is in Π, then T_j is not an ancestor of T_i in G.
4. If $T_i \rightarrow T_j$ is in Π and the edge between T_i and T_j is not labeled with a "?", then $T_i \rightarrow T_j$ is in G.

9. References

Anderson, J., & Gerbing, D. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.

Anderson, J., and Gerbing, D., & Hunter, J. (1987). On the assessment of unidimensional measurement: internal and external consistency and overall consistency criteria. *Journal of Marketing Research*, 24, 432-437.

Becker, P., Merckens, A., and Wansbeek, T. (1994). *Identification, equivalent models, and Computer Algebra*. Academic Press, San Diego, CA.

Beinlich, I., Suermondt, H., Chavez, R., and Cooper, G. (1989). The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. *Proc. Second European Conference on Artificial Intelligence in Medicine*, London, England. 247-256.

Bentler, P. (1986). Lagrange multiplier and Wald tests for EQS and EQS/PC. BMDP Statistical Software, LA

Bentler, P. (1995). *EQS: Structural Equations Program Manual*. Multivariate Software, Inc. Encino, CA.

Blalock, H. (1962). Four-variable causal models and partial correlations. *American Journal of Sociology*, 68, 182-194.

Blalock, H. (ed.) (1985). *Causal models in panel and experimental designs*. Aldine, NY.

Bollen, K. (1980). Issues in the comparative measurement of political democracy. American Sociological Review, 45, 370-390.

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

Bollen, K. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review*, 45, 370-390.

Bollen, K. (1990). Outlier screening and a distribution-free test for vanishing tetrads. *Sociological Methods and Research*, 19, 80-92.

Bollen, K. & Long, J. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Bollen, K., & Ting, K. (1993). Confirmatory tetrad analysis. *Sociological Methodology, 23*, 147-175. Oxford: Blackwell Publishers.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K.G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Part I, pp. 149-173). Amsterdam: North-Holland.

Buntine, W. (1991). Theory refinement on Bayesian networks. G. Piatetski-Shapiro (Ed.), Workshop Notes from the Ninth National Conference on Artificial Intelligence, Knowledge Discovery in Data Bases. Anaheim, CA.

Bye, B., Gallicchio, S., & Dykacz, J. (1985). Multiple-indicator, multiple-cause models for a single latent variable with ordinal indicators. *Sociological Methods & Research*, 13, 487-509.

Callahan, J., & Sorensen, S. (1992). Using TETRAD II as an automated exploratory tool. *Social Science Computer Review*. Vol. 10, No. 3, Pages 329-336.

Cheeseman, P. and Oldford, R. (1994). *Selecting Models from Data.* Lecture Notes in Statistics, v. 89, Springer-Verlag.

Cooper, G., Aliferis, C., Ambrosino R., Aronis, J., Buchanan, B., Caruana, R., Fine, M. Glymour, C., Gordon G., Hanusa, B. Janosky, J., Meek, C., Mitchell, T., Richardson, T., Spirtes, P. (1995). An Evaluation of Machine-Learning Methods for Predicting Pneumonia Mortality., *Technical Report CMU-PHIL-66.* Carnegie Mellon University, Pittsburgh, PA.

Cooper, G., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, 308-347.

Costner, H. (1971). Theory, deduction and rules of correspondence. *Causal Models in the Social Sciences*, Blalock, H. (ed.). Aldine, Chicago.

Druzdzel, M., & Glymour, C. (1994). Application of the TETRAD II program to the study of student retention in U.S. colleges. *Proceedings of the AAAI-94 Workshop on Knowldege Discovery in Databases (KDD-94)*, Seattle, WA, 419-430.

Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics,* 17, 333-353.

Geiger, D. (1990). *Graphoids: A Qualitative Framework for Probabilistic Inference*. Ph.D. Thesis, University of California, Los Angeles.

Geiger, D., & Heckerman, D. (1994). Learning Gaussian Networks. Microsoft *Technical Report, MSR-TR-94-10*. Redmond, WA.

Geiger, D., and Heckerman, D. (1991) Advances in probabilistic reasoning. *Proc. Seventh Conference on Uncertainty in AI*, B. D'Ambrosio et al. (ed.). Morgan Kaufman, Los Angeles, California.

Geiger, D., Verma, T., and Pearl, J. (1990) Identifying independence in Bayesian networks. *Networks* 20, 507-533.

Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (1987). *Discovering Causal Structure*. Academic Press, San Diego, CA.

Goldberger, A., Duncan, O. (eds.) (1973). *Structural Equation Models in the Social Sciences*. Seminar Press, New York.

Hand, D. (1993). *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*. Chapman and Hall.

Hausman, D. (1984). Causal priority. Nous 18, 261-279.

James, L., Mulaik, S., and Brett, J. (1982). *Causal analysis: assumptions, models and data*. Sage Publications, Beverly Hills, CA.

Jöreskog, K., & Sörbom, J. (1993) *LISREL 8: User's reference guide*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, 23, 69-86.

Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, 24, 285-305.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25, 137-55.

Kiiveri, H. and Speed, T. (1982). Structural analysis of multivariate data: A review. *Sociological Methodology*, Leinhardt, S. (ed.). Jossey-Bass, San Francisco.

Koster, J.T.A. (forthcoming). Markov Properties of Non-Recursive Causal Models. *Annals of Statistic*.

Lauritzen, S., Dawid, P., Larsen, B., Leimer, H. (1990) Independence Properties of Directed Markov Fields. *Networks*, 20, 491-505.

Lauritzen, S. and Wermuth, N. (1984). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Stat*. 17, 31-57.

Lawley, D., and Maxwell, A. (1971). *Factor Analysis as a Statistical Method.* Butterworth, London.

Lee, S. (1987). *Model Equivalence in Covariance Structure Modeling*. *Ph.D. Thesis*, Department of Psychology, Ohio State University.

Luijben, T., Boomsma, A., and Molenaar, I. (1986). Modification of factor analysis models in covariance structure analysis. A Monte Carlo study. *On Model Uncertainty and its Statistical Implications*. Lecture Notes in Economics and Mathematical Systems 307, Dijkstra, T. (ed.). Springer-Verlag, Berlin.

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin* 100, 107-120.

Meek, C. (1995) Causal inference and causal explanation with background knowledge. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Philippe Besnard and Steve Hanks (Eds.), Morgan Kaufmann Publishers, Inc., San Mateo, CA, 403-410.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems***.** Morgan Kaufmann, San Mateo.

Pearl, J. (1995). Causal Diagrams for Empirical Research, Biometrika. 82(4), December. 669--709,

Pearl, J. and Dechter, R. (1989). Learning structure from data: A survey. *Proceedings COLT '89*, 30-244.

Pearl, J. and Verma, T. (1991). A theory of inferred causation. Principles of Knowledge Representation and Reasoning: *Proceedings of the Second International Conference*, Morgan Kaufmann, San Mateo, CA.

Prigerson, H., Maciejewski, K., Reynolds, C., Bierhals, A., Newsom, J., Frank, E., Miller, M., Doman, J., & Fasiczka, A. (1995). The Inventory of Complicated Grief: A Scale to Measure Certain Maladaptive Symptoms of Loss. Submitted to *Psychiatry Research*.

Raftery, A.E. (1993). Bayesian model selection in structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 163-180). Newbury Park, CA: Sage.

Richardson, T. (1994). *Properties of Cyclic Graphical Models.* Master's Thesis, Dept. of Philosophy, Carnegie Mellon University, Pgh, PA, 15213.

Richardson, T. (1995). *A Polynomial-Time Algorithm for Deciding Markov Equivalence of Directed Cyclic Graphical Model*s, Technical Report CMU-PHIL-63, Philosophy Department, Carnegie Mellon University, Pgh, PA, 15213.

Richardson, T. (1996). Discovering Cyclic Causal Structure, Technical Report CMU-PHIL-68, Philosophy Department, Carnegie Mellon University, Pgh, PA, 15213.

Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect. *Mathematical Modeling*, 7, 1393-1512.

Saris, W., Satorra, A., & Sorbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. Clogg (Ed.), *Sociological methodology* (pp. 105-129). San Francisco, CA: Jossey-Bass.

Scheines, R. (1993). Unidimensional linear latent variable models. *Technical Report CMU-PHIL-39*. Carnegie Mellon University, Pittsburgh, PA.

Scheines, R. (1994). "Inferring Causal Structure Among Unmeasured Variables," in *Selecting models from data: AI and Statistics IV,* P. Chessman and R. W. Oldford (Eds.), Springer-Verlag, pp. 197-204.

Scheines, R., Spirtes, P., Glymour, C., and Meek, C. (1994). *TETRAD II: Users Manual*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Scheines, R., Hoijtink, H., & Boomsma, A. (1995). *Bayesian Estimation and Testing of Structural Equation Models*, Technical Report CMU-PHIL-66, Dept. of Philosophy, Carnegie Mellon Univ., Pgh, PA, 15213.

Shafer, G., Kogan, A., and Spirtes, P. (1993). Generalization of the Tetrad Representation Theorem. GSM Working Paper #93-36, Rutgers Graduate School of Management.

Shipley, B. (Forthcoming). Structured interspecific determinants of specific leaf area in 34 species of herbaceous angiosperms. *Functional Ecology.*

Simon, H. (1953). Causal ordering and identifiability. *Studies in Econometric Methods*. Hood and Koopmans (eds). 49-74.Wiley, NY.

Simon, H. (1954). Spurious Correlation: A Causal Interpretation. *Journal of the American Statistical Association,* 49, 467-79.

Sörbom, D. (1989) Model modification. *Psychometrika*, 54, 371-384.

Spearman, C. (1904). General intelligence objectively determined and measured. American *Journal of Psychology* 15, 201-293.

Spiegelhalter, D. (1986). Probabilistic reasoning in predictive expert systems. *Uncertainty in Artificial Intelligence*, Kanal, K. and Lemmer, J. (eds.). North-Holland, Amsterdam.

Spirtes, P. (1989). A necessary and sufficient condition for conditional independencies to imply a vanishing tetrad difference. *Technical Report CMU-LCL-89-3*, Laboratory for Computational Linguistics, Carnegie Mellon University, Pittsburgh, PA.

Spirtes, P. (1995) Directed cyclic graphical representation of feedback models. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, ed. by Philippe Besnard and Steve Hanks, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 491-498.

Spirtes, P., Meek, C., and Richardson, T. (1995) Causal inference in the presence of latent variables and selection bias. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, ed. by Philippe Besnard and Steve Hanks, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 499-506.

Spirtes, P., Meek, C. (1995). Learning Bayesian networks with discrete variables from data. *Proceedings of The First International Conference on Knowledge Discovery and Data Mining*, ed. by Usama M. Fayyad and Ramasamy Uthurusamy, AAI Press, 294-299.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., and Glymour, C. (1996). Using d-separation to calculate zero partial correlations in linear models with correlated errors. *Technical Report CMU-PHIL-72*, Dept. of Philosophy, Carnegie Mellon University, Pittsburgh, PA, 15213.

Spirtes, P. (1994a). "Conditional Independence in Directed Cyclic Graphical Models for Feedback." *Technical Report CMU-PHIL-54*, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA.

Spirtes, P. (1994b). "Discovering Causal Relations Among Latent Variables in Recursive Structural Equation Models." *Technical Report CMU-PHIL-69*, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA.

Spirtes, P., Scheines, R., & Glymour, C. (1990). Simulation studies of the reliability of computer aided specification using the TETRAD II, EQS, and LISREL programs. *Sociological Methods and Research*, 19, 3-66.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. Springer-Verlag Lecture Notes in Statistics, 81. Springer-Verlag.

Strotz, R., and Wold, H. (1960) Recursive versus nonrecursive systems: an attempt at synthesis. *Econometrica*, 28, 417-427.

Sullivan, J., & Ferldman, S. (1979). *Multiple indicators: an introduction*. Sage Publications, Beverly Hills, CA.

Tanner, M.A. (1993). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (2nd ed.). New York: Springer.

Timberlake, M. and Williams, K. (1984). Dependence, political exclusion, and government repression: Some cross-national evidence. American Sociological Review 49, 141-146.

Verma, T. and Pearl, J. (1990). Equivalence and synthesis of causal models. *Proc. Sixth Conference on Uncertainty in AI*. Association for Uncertainty in AI, Inc., Mountain View, CA, 220-227.

Waldemark, J., & Norqvist, P. (1995). In-Flight Calibration of Satellite Ion Composition Data Using Artificial Intelligence Methods. Manuscript, Dept. of Applied Physics and Electronics, Umea University, S-90187 Umea, Sweden.

Weisberg, S. (1985). *Applied Linear Regression*, 2nd ed. Wiley, New York.

Wheaton, B., Muthen, B., Alwin, D., and Summers, G. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology 1977*, Heise, D. (ed.). Jossey-Bass, San Francisco.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Stat*. 5, 161-215.