

Searching for the Causal Structure

of a

Vector Autoregression

Selva Demiralp
Division of Monetary Affairs
Monetary and Reserve Analysis
Board of Governors of the Federal
Reserve System
Washington, D.C. 20551

Tel. (202) 452-6483
E-mail selva.demiralp@frb.gov

Kevin D. Hoover
Department of Economics
University of California
1 Shields Avenue
Davis, California 95616-8578
Tel. (530) 752-2129
Fax (530) 752-9382
E-mail kdhoover@ucdavis.edu

First Draft, 16 October 2002
Revision, 6 March 2003

We thank Marcus Cuda for his help with programming and computational design and Derek Stimel for able research assistance, Oscar Jorda and the participants in the European Community Econometrics Conference (EC)², University of Bologna, Italy, 13-14 December 2002, for comments.

The views expressed do not necessarily reflect the views of the Federal Reserve System.

Abstract of

Searching for the Causal Structure of a Vector Autoregression

Vector autoregressions (VARs) are economically interpretable only when identified by being transformed into a structural form (the SVAR) in which the contemporaneous variables stand in a well-defined causal order. These identifying transformations are not unique. It is widely believed that practitioners must choose among them using *a priori* theory or other criteria not rooted in the data under analysis. We show how to apply graph-theoretic methods of searching for causal structure based on relations of conditional independence to select among the possible causal orders – or at least to reduce the admissible causal orders to a narrow equivalence class. The graph-theoretic approaches were developed by computer scientists and philosophers (Pearl, Glymour, Spirtes among others) and applied to cross-sectional data. We provide an accessible introduction to this work. Then building on the work of Swanson and Granger (1997), we show how to apply it to searching for the causal order of an SVAR. We present simulation results to show how the efficacy of the search method algorithm varies with signal strength for realistic sample lengths. Our findings suggest that graph-theoretic methods may prove to be a useful tool in the analysis of SVARs.

Keywords: search, causality, structural vector autoregression, graph theory, common cause, causal Markov condition, Wold causal order, identification; PC algorithm

JEL Classification: C15, C32, C49, C51

Searching for the Causal Structure of a Vector Autoregression

1. The Problem of Causal Order

Drawing on recent work on the graph-theoretic analysis of causality, we propose and evaluate a statistical procedure for identifying the contemporaneous causal order of a structural vector autoregression.

Since the publication of Christopher Sims's "Macroeconomics and Reality" (1980), the vector autoregression (VAR) has become the dominant tool of empirical macroeconomics in the United States – if somewhat less so in Europe. Dissatisfied with the "incredible identifying restrictions" imposed on structural macroeconometric models, Sims proposed the use of the VAR – an unrestricted reduced form.

A VAR can be written as

$$(1) \quad \mathbf{B}\mathbf{Y}_t = \mathbf{U}_t,$$

where t indexes time; \mathbf{Y}_t is an $n \times 1$ column vector of the contemporaneous values of the variables Y_{it} , $i = 1, 2, \dots, n$; \mathbf{B} is a conformable square matrix whose terms are polynomials in the lag operator – e.g., $B_{ij}(L)$; and \mathbf{U}_t is a column vector of structural residuals with elements u_{it} .

Although the VAR is easily estimated, difficulties begin when we turn to policy analysis. A typical problem would be to work out the effects of a shock to one of the variables on all the other variables of the system. Let $\mathbf{u}_i = [u_{i1}, u_{i2}, \dots, u_{iT}]$ be the time series for u_{it} and \mathbf{U} without a time subscript be the $n \times T$ matrix whose elements are the \mathbf{u}_i . The contemporaneous covariance matrix is $\mathbf{\Sigma} = E(\mathbf{U}\mathbf{U}')$, where E is the expectations operator. In general, $\mathbf{\Sigma}$ is not diagonal. The non-zero off-diagonal elements imply that one variable, say Y_{1t} , cannot be shocked through its corresponding random error term, u_{1t} , without having simultaneously to deliver correlated shocks to other variables. Without independence it makes little sense to think of shocks, say, to the money supply or to employment.

Sims (1980) advocated orthogonalizing the shocks using a Choleski decomposition. There is a unique lower triangular matrix \mathbf{C} , such that $\mathbf{C}\mathbf{C}' = \mathbf{\Sigma}$. Premultiplying both sides of (1) by \mathbf{C}^{-1} yields

$$(2) \quad \mathbf{C}^{-1}\mathbf{B}\mathbf{Y}_t = \mathbf{C}^{-1}\mathbf{U}_t.$$

The covariance matrix of (2) is $\mathbf{E}(\mathbf{C}^{-1}\mathbf{U}(\mathbf{C}^{-1}\mathbf{U})') = \mathbf{I}$ (Hamilton 1994, p. 320). Unit shocks can be delivered to any of the variables of the system, and their effects traced out. The Choleski decomposition imposes a Wold causal order on the variables so that the shock to Y_1 feeds contemporaneously into Y_2, Y_3, \dots, Y_n , while the shock to Y_2 feeds contemporaneously into Y_3, Y_4, \dots, Y_n , but into Y_1 only with a lag, and so on. While the Choleski decomposition is unique, it differs under differing orderings of the Y_i in the \mathbf{Y} vector, and these orderings are arbitrary.¹ What is more, orthogonalizing transformations are not restricted to Choleski decompositions, but may involve non-triangular matrices \mathbf{P} , such that $\mathbf{E}(\mathbf{P}^{-1}\mathbf{U}) = \mathbf{I}$, providing that at least $n(n-1)/2$ restrictions are imposed for identification.² A VAR identified through restrictions on contemporaneous variables is known as a *structural vector autoregression (SVAR)*.

It is widely believed that there is no empirical or statistical basis for the choice of the contemporaneous causal orderings (that is, orthogonalizing transformations), so that the economist must appeal to *a priori* knowledge. Since there are transformations that impose every possible order, there is a family of SVARs for which the original VAR (equation (1)) is the common reduced form. Each member of the family has the same reduced form and, therefore, the same likelihood function. Practitioners typically regard the members of the family as observationally equivalent. Only outside knowledge would allow the researcher to choose among them.

But where is such knowledge to come from? Only rarely does economic theory imply particular contemporaneous causal orderings. Generally, practitioners of SVAR

¹ Sims (1980) initially underplayed the interpretive ambiguity implied by the different orderings. Under criticism from Cooley and LeRoy (1985), Leamer (1985) and others, Sims (1986) conceded that useful interpretation of VARs required choosing among the possible orthogonalizing transformations.

² Identification may be achieved in other ways, although in this paper we shall be concerned only with zero restrictions on the contemporaneous coefficient matrix.

methods appeal to plausible stories about which variables could or could not affect which other variables in the course of a month or quarter, depending on the periodicity of the data. The problem with this approach is that sometimes equally plausible stories can be told for competing causal orderings. Not only does such story-telling not inspire much confidence, it is ironic that a method that originated as a way of getting away from incredible identifying restrictions relies so heavily on hardly more credible stories to identify contemporaneous causal ordering.

Contrary to the widespread belief, all the SVARs derivable from a VAR are *not* observationally equivalent. The underlying premise of SVAR analysis is that some SVAR corresponds to the data-generating process. Let the data-generating process be

$$(3) \quad \mathbf{A}\mathbf{Y}_t = \mathbf{E}_t,$$

where $\mathbf{E}_t = [\varepsilon_{ijt}]$ is a column vector of error terms at time t , and the covariance matrix $\mathbf{\Omega} = \mathbf{E}(\mathbf{E}\mathbf{E}')$ is diagonal. (Omitting time subscripts indicates the matrix whose columns are the \mathbf{E}_t , and analogously in other cases.) The error terms of this SVAR are not merely uncorrelated; they are, in fact, *independent*. The diagonal covariance matrix, which indicates the independence of the error terms from each other, marks the VAR *structural*.

Each equation in such an SVAR can be shocked independently of the others. Let \mathbf{A}_0 be the matrix of the zeroth-order terms of the matrix \mathbf{A} – i.e., the typical element of \mathbf{A}_0 is $A_{ij}(0)$. The reduced form of (3), then is

$$(4) \quad \mathbf{A}_0^{-1} \mathbf{A} \mathbf{Y}_t = \mathbf{A}_0^{-1} \mathbf{E}_t.$$

Equating terms with (1), we see that \mathbf{A}_0 connects the reduced-form errors from the ordinary VAR (u_{it}) with the structural errors (ε_{it}), so that

$$(5) \quad \mathbf{U}_t = \mathbf{A}_0^{-1} \mathbf{E}_t.$$

The independence of the ε_{ij} and the structure of the SVAR embodied in \mathbf{A}_0 implies the relationships of interdependence, independence, and conditional independence among

the u_{ij} , the elements of \mathbf{U} . These are robust relationships in the sense that they are invariant to different values of the α_{ij0} , the elements of \mathbf{A}_0 .³

The transformation that converts VAR (1) into SVAR (3) is a privileged one in that it is the only one that recovers the independent errors, the ε_{ij} . The well known fact that statistical independence implies an absence of correlation, but that an absence of correlation does not imply statistical independence comes into play here. The reduced form (4) may be transformed into (pseudo) structural VARs – each appearing to possess a contemporaneous causal order different from the data-generating process. But appearances are deceiving. The error terms of the pseudo SVARs are mutually uncorrelated but *not* independent. Each pseudo structure carries with it the constraints implied by \mathbf{A}_0 . And unlike the causal order of the true structure (3), which is robust to alternative values for the non-zero elements of \mathbf{A}_0 , the causal orders of the transformed structures are well-defined only for a particular set of values implied by \mathbf{A}_0 . If the α_{ij0} change, the error terms of the pseudo SVAR will no longer appear to be orthogonal.

Since one transformation is privileged, the central question then becomes whether starting from the VAR (1), empirical evidence can help us select the transformation that corresponds to the true SVAR (3)? In principle, the answer is yes.

Over the past twenty years, a group of philosophers and computer scientists have developed a graph-theoretic analysis of causal structure and demonstrated the relationship between particular causal orders and relationships of conditional independence embedded in the likelihood function. Pearl (2000) and Spirtes, Glymour, and Scheines (2000) provide detailed accounts of this approach, as well as search algorithms for implementing it. These methods have been used in a variety of social sciences other than economics, but are virtually unknown to economists.⁴

³ Hoover (2001, chs. 2-4) provides a detailed discussion of the role of independence and invariance under parameter change as hallmarks of the true causal structure.

⁴ Some earlier applications to economics include Swanson and Granger (1997), Sheffrin and Triest (1998), Akleman, Bessler, and Burton (1999), Demiralp (2000, ch. 4), and Bessler and Lee (2002). Hoover (2001, ch. 7) gives a critical description of these methods, and LeRoy (2002) has recently discussed them in a review of Pearl (2000).

Most of this research on graph-theoretic methods assumes that the causally ordered data are cross-sectional, and many of the main theoretical results do not apply directly to time series. We follow Swanson and Granger's (1997) suggestion of how to adapt graph-theoretic methods to the problem of finding the causal order of the SVAR. (The method is described more fully in Section 3 below.) Unlike Swanson and Granger, who restrict the admissible structures to the class of Choleski orderings, we allow every possible ordering. We use the PC algorithm embedded in Spirtes *et al.*'s *Tetrad 3* (1996) software. Implementation is straightforward, but the nagging question that macroeconometricians are entitled to ask is: just how well does this method work in practice? In this paper we contribute two things. First, we provide an accessible account of the underlying rationale for the graph-theoretic approach to causal order in general. Second, using a simulation study, we address the efficacy of the most common algorithm for implementing this approach to selecting the causal order of SVARs.

2. The Graph-theoretic Analysis of Causal Structure

Start with a structure defined by equation (3) with the added stipulation that the matrix $\mathbf{A} = \mathbf{A}_0$ – that is, there are only contemporaneous variables. Each row of \mathbf{A} represents the equation for the corresponding element of \mathbf{Y} , and the non-zero off-diagonal elements determine which are the explanatory variables of the equation represented by each row. A causal structure can be represented by a graph in which arrows run from causes to the caused variable, and the graph corresponds to the pattern of non-zero elements of \mathbf{A} . For

example, if $\mathbf{Y} = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$ and $\mathbf{A} = \begin{bmatrix} 1 & 0 & \alpha_{13} \\ 0 & 1 & \alpha_{23} \\ 0 & 0 & 1 \end{bmatrix}$, where the α_{ij} designate non-zero elements,

then the causal structure can be represented by Figure 1, where the arrows represent one-way causal influence.

It is helpful to define some terms used in graph theory. Causal connections between variables are indicated by lines (known as *edges* or *links*) that may or may not have arrowheads indicating the direction of causation. The map of a set of variables showing the causal connections and their directions is a *graph* such as that depicted in Figure 1. The map showing just the variables and their connections but ignoring the

directions is the *skeleton* of the graph. A *path* is a chain of causal connections between two variables. For example, in Figure 1, ABC is a path from A to B . A *directed path* follows the direction of causation (the direction of the arrowheads). For example, in the graph $A \rightarrow B \rightarrow C$, ABC is a directed path from A to C , but in Figure 1, there is no directed path from A to B . If a variable A is connected to another variable B by an arrow originating at A and running into B , then A is the *parent* of B , and B is the *child* of A . If there is a directed path between A and B , then A is an *ancestor* of B , and B is a *descendant* of A . If there are no directed paths from a descendant to its own ancestor, then the graph is *acyclic*. If each cause of every variable in a graph is also a variable in that graph, then the graph is *causally sufficient*.

Errors terms in each equation could be treated as causes of deterministic variables. When error terms are independent and, therefore, affect one variable each, it is conventional to omit them from a graph and to treat the variables as stochastic. When they are not independent, it is conventional to show them explicitly as latent, unobservable variables or to indicate bidirectional causal linkages between the variables. Graphs with latent variables are not causally sufficient. Because the graph-theoretic account is best developed for acyclical graphs, we restrict our simulations to causally sufficient, acyclical models.

Returning to the initial model, since the ε_i are independent random shocks, the matrix \mathbf{A} and its corresponding graph (Figure 1), represent a causal structure that defines the patterns of dependence or independence among the variables. In this case, it is easy to see that A and B are not independent because both depend on C . C is said to be their *common cause*. It is also intuitive that A is independent of B *conditional on* C .

Causal search algorithms are based on patterns of conditional independence, invoking Reichenbach's (1956, p. 156) *principle of the common cause*: if any two variables, A and B , are truly correlated, then either A causes B ($A \rightarrow B$) or B causes A or ($A \leftarrow B$) or they have a common cause (as in Figure 1). The common cause, C , may be a complex of parent variables.

The principle of the common cause can be generalized as the *Causal Markov Condition*:

Definition. Let G be a causal graph relating a set of variables \mathbf{V} with a probability distribution P . Let \mathbf{W} be a subset of \mathbf{V} . G and P satisfy the *causal Markov condition* if, and only if, for every \mathbf{W} in \mathbf{V} , \mathbf{W} is independent of every set of variables that does not contain its descendants, conditional on its parents. (Spirtes *et al.* 2000, p. 29; see also Pearl 2000, p. 30).⁵

Essentially, the causal Markov condition holds when a graph corresponds to the conditional independence relationships in the associated probability distribution. A graph is said to be *faithful* if, and only if, there is a one-to-one mapping between the relationships of conditional independence relation implied by the causal Markov condition applied to G and those found in P (Spirtes *et al.* 2000, p. 31).

A few further examples illustrate how to apply these ideas. Consider the two causal graphs in Figure 2. In each case, A and B are dependent, but are independent conditional on C . C is said to *screen-off* A from B .

Causal structure can induce conditional dependence as well as eliminate unconditional dependence. Consider the graph in Figure 3. A and B are unconditionally uncorrelated. They are however correlated conditional on C . The classic example is A = the car's battery being charged; B = the car's starter switch being on; and C = the car's starting. A and B may be completely independent. Yet, *if we know that the car does not start*, then knowing that switch is on raises the probability that the battery is dead. The configuration in Figure 3 is called an *unshielded collider on the path ACB* (or *BCA*). It is a “collider” because the arrowheads come together at C , and is “unshielded” because there is no direct causal connection between A and B . The graph in Figure 4 is a *shielded collider*. Because they are directly causally connected in Figure 4, A and B are correlated even without conditioning on the common effect.

Causal search algorithms start with the empirical probability distribution of a set of variables represented by the covariance matrix or its normalized form, the unconditional correlation matrix. Tests of conditional independence are implemented using conditional correlations. The unconditional correlation coefficient between A and B is denoted r_{AB} .

⁵ The graph-theoretic account uses a dauntingly complex and unfamiliar terminology. Here, as elsewhere, we follow closely the version of Spirtes *et al.* (2000), but translate it into a more accessible language.

The correlation of A and B conditional on C is then defined as $r_{AB|C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{1 - r_{AC}^2} \sqrt{1 - r_{BC}^2}}$.

The statistical significance of the conditional correlation can be computed using Fisher's z -statistic.⁶

Each causal graph implies a set of independence relationships in the associated probability distribution. Unfortunately, different graphs may imply the same set, so that a probability distribution defines a class of observationally equivalent causal structures. This class may have only one element or it may have many. An important theorem says that any probability distribution that can be faithfully represented in a causally sufficient, acyclical graph can equally well be represented by any other acyclical graph that has the same skeleton and the same unshielded colliders (Pearl 2000, p. 19, and Spirtes *et al.* 2000, ch. 4). As a result, there may be observationally equivalent causal structures in which some causal links are reversed but all unshielded colliders preserved. In those cases, the algorithm yields only partial causal orderings.

It should be recognized that the equivalence class in this case has a different membership condition than that of the class of SVARs the observational equivalence of which we rejected in Section 1.

There are several search algorithms available. In this paper we use the PC algorithm of Spirtes *et al.* (2000). An illustration shows how search algorithms work. Consider the true causal graph given in Figure 5, panel (i). (Notice that there are two unshielded colliders: C on the path ACE and D on the path BDF . B is a shielded collider on ABC .) Starting from the correlation matrix, how would the PC algorithm proceed? It begins with a graph (panel (ii)) in which every variable is connected to every other, but the links are not oriented. It then eliminates connections between any variables that are not unconditionally correlated (panel (iii)). Next it tests for the correlation of each pair of variables conditional on a third variable. It eliminates the link between any pair that is conditionally uncorrelated (panel (iv)). Continuing in the same vein, it tests for absence

⁶ As we observed in Section 1, independence implies an absence of correlation, but not the converse. There may be highly specific parameter values for which correlations vanish, even though the variables are not independent. These correspond to the non-robust transformations of the true SVAR mentioned in Section 1. Hoover (2001, ch. 2, section 4, and ch. 7, section 1) discusses cases in which these vanishing correlations arise from economically meaningful optimal control.

of correlation conditional on pairs of variables and eliminates links whenever there is no conditional correlation (panel (v)). In principle, it would test for a lack of correlation conditional on triples, sets of four, five, and more variables. In this case, however, it has exhausted the possibilities at pairs. It then considers every pair of variables that is conditionally uncorrelated and causally connected along an undirected path through a third variable. If conditioning on the third variable renders them conditionally correlated, then there is an unshielded collider and the arrows are oriented accordingly (panel (vi)). Finally, some unoriented links may be oriented based on screening relationships. We know from panel (v) that C screens the correlation between B and E . Usually, this would mean either that $B \rightarrow C \rightarrow E$ or that $E \rightarrow C \rightarrow B$. Since we already know that $E \rightarrow C$, only $E \rightarrow C \rightarrow B$ is consistent as shown in panel (vii). Notice that the link between A and B remains unoriented. This is because the graph in panel (viii) has the same skeleton and unshielded colliders as the true structure in panel (i), even though it reverses the link between A and B . The two graphs are observationally equivalent, and the search algorithm cannot choose between them. A precise description of the PC algorithm is given in Appendix.

3. The Effectiveness of the Causal Search Algorithm

3.1 THE SIMULATION METHODOLOGY

The PC algorithm can be implemented with *Tetrad 3* (Spirtes *et al.* 1996, program). Spirtes *et al.* (1996, user's manual, ch. 13), Spirtes *et al.* (2000, pp. 113-122), and Cooper (1997, section 10) present some simulation evidence of its effectiveness. However, no previous studies have investigated its effectiveness in the context of ordering the contemporaneous variables in an SVAR. We proceed in the following steps:

1. Each SVAR takes the form of equation (3). We can write $\mathbf{A} = \mathbf{A}_0 + \overline{\mathbf{A}}$, where the elements of $\overline{\mathbf{A}}$ are $\overline{A}_{ij} = a_{ij1}L + a_{ij2}L^2 + a_{ij3}L^3 + a_{ij4}L^4 + \dots + a_{ijK}L^K$. Each equation in the SVAR has an identical lag structure – i.e., for each $j = 1, 2, \dots, N$ and each $k = 1,$

- 2, . . . , K, and for all $i = 1, 2, . . . , N$ and $h = 1, 2, . . . , N$: $\alpha_{ijk} = \alpha_{hjk}$.⁷ For concreteness $K = 4$. Models to be evaluated differ in the number of variables and the causal structure of the contemporaneous terms defined by the placement of nonzero terms in \mathbf{A}_0 . Given the causal structure and a particular choice of values for the nonzero a_{ij0} , the data are generated recursively drawing the error terms from a random-number generator. The $\varepsilon_{ijt} \sim \mathbf{N}(0, 1)$. To eliminate problems with initial values, 1500 realizations are generated and only the last 500 retained for analysis.
2. In order to evaluate a range of signal-to-noise ratios, we generate 50,000 realizations for each model with the nonzero a_{ij0} chosen at each realization using a random number generator with the range calibrated to generate Fisher's z-statistics for these parameters in the maximum likelihood estimates of the SVAR covering a range of roughly 0 to 9. The distribution is weighted to oversample the 0 to 2 range.
 3. A VAR of the form of equation (1) with a lag length of four ($K = 4$) is estimated for each realization. The estimated residuals $\hat{\mathbf{U}}_t$ are retained as the *filtered* \mathbf{Y}_t . The sample covariance of the filtered \mathbf{Y} is $\hat{\Sigma} = \frac{\mathbf{U}\mathbf{U}'}{T - K}$ and serves as input to *Tetrad 3* from which it calculates all the needed conditional correlations.
 4. *Tetrad 3* is run for each realization using the PC algorithm and assuming causal sufficiency. To evaluate the success of the algorithm, the graph of the model selected by Tetrad (the *selected graph*) is compared to a *reference graph* (the *PC-true graph*). The PC-true graph is not the graph of the model that generated the data (i.e., it is not the *true graph*). It is, instead, the graph that the PC algorithm would select under the best circumstances (i.e., with an infinite amount of data). This is the graph that has the same skeleton as the true graph, but leaves undirected links wherever a link can be reversed without altering the identities of the unshielded colliders.

Every possible link is evaluated. The possible outcomes are:

⁷ The vector of values for all own lags ($i = j$) is $[a_{ijk}] = [0.0403, 0.162409, 0.065450827, 0.026376683281]$ and the vector of values for all cross lags ($i \neq j$) is $[a_{ijk}] = [0.054, 0.002916, 0.000157464, 0.000008503056]$.

- (i) *Correct*: the link is present and oriented the same way in both graphs or it is absent in both graphs;
- (ii) *Committed*: the link is absent in the reference graph but present in the selected graph.
- (iii) *Omitted*: the link is present in the reference graph but absent in the selected graph.
- (iv) *Reversed*: the link is present in both graphs, but points in opposite directions.
- (v) *Unresolved*: the link is oriented in the reference graph and, although present, cannot be oriented in the selected graph.
- (vi) *Overdetermined*: the link cannot be oriented in the reference graph, but is oriented in the selected graph.

Errors fall into two groups. *Errors of commission*: outcome (ii) can occur only if a link is missing in the true (and, therefore, reference) graph. *Link errors*: outcomes (iii) through (vi) can occur only if a link is present in the reference graph.

3.2 FOUR MODELS

The strategy of causal identification used in the PC algorithm makes use of the whole structure. It is likely to work best when there are a relatively large number of unshielded colliders and a relatively low density of causally connected variables. We begin with two very simple models. Although these should be difficult for Tetrad to identify, they are easily grasped by the analyst and can be used to identify some salient issues. We then consider two more complex models.

Model 1

The graph of Model 1 is depicted in Figure 6. Corresponding to the graph is

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_{210} & 1 & \alpha_{230} & \alpha_{240} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $\alpha_{ij0} \neq 0$. Model 1 is symmetrical around B (that is, switching the positions of A , C , and D produces isomorphic graphs). The variable B is an unshielded collider on three separate paths: ABC , ABD , and CBD . In principle, the PC algorithm can identify Model 1 (i.e., there is only one graph in the equivalence class), so the true graph and the PC-true graph are identical. There cannot, therefore, be any errors of overdetermination.

We address two questions: First, how does the effectiveness of the PC algorithm depend on the nominal size of the z -statistics that *Tetrad 3* uses to assess conditional correlations? Second, how does the effectiveness of the PC algorithm vary with the signal-to-noise ratios of the causal links?

To answer these questions, we classify signal-to-noise ratios into categories according to the expected value of the z -statistic ($z_{\alpha_{ij0}}^* = E(z_{\alpha_{ij0}})$) for the coefficient in A_0 that corresponds to the link (e.g., α_{210} for link 1).⁸ $0 < z^* < 2$ is classified as L (low); $2 \leq z^* < 5$ as M (medium); and $5 \leq z^* \leq 9$ as H (high). There are in principle $3^3 = 27$ different combinations of signal strengths for Model 1 using these classifications. Since Model 1 is fully symmetrical, combinations with the same number of links in a particular category should yield nearly the same results. For instance, if we label a particular draw by the order of its links as numbered in Figure 6, then HMM should have very similar results to HHM and MHH. We, therefore, record only the ten nonredundant patterns.

Figure 7 compares three nominal sizes for the test statistics in *Tetrad 3*: 5, 10, and 20 percent. For each size it shows the omitted and committed outcomes. Each is expressed as a proportion of the number of times it might have occurred. A link can be omitted only if it is actually included in Model 1. A link can be committed only if it is actually excluded in Model 1. While the statistics reported here are not classic test statistics, the proportion of outcomes committed is analogous to the size of a classical test statistic (i.e., type I error), while the proportion omitted is analogous to the complement of the power (i.e., type II error). But there are degrees of errors of omission. The worst would be to omit a link altogether, but even if a link is included it may not be correctly directed. A third summary statistic – the total correct links irrespective of direction – is

⁸ Expected values ($z_{\alpha_{ij0}}^*$) are determined using predicted values from the regression $z_{\alpha_{ij0}}^* = \gamma\alpha_{ij0} + \varepsilon_{ij}$.

also reported. This statistic counts a success any time a true link is identified even if its direction is reversed or unresolved. It shows the success of the algorithm at identifying the skeleton of the model.

The different combinations of signal strength are indicated along the horizontal axis of Figure 7. The number associated with each combination in the labels on the x-axis is the mean population value of z^* -statistic for that combination. The data are ordered in descending order of the proportion of omissions at the 10-percent test size. The figure clearly shows the usual tradeoff between size and power. As we move from nominal sizes of 5 percent to 10 and 20 percent, the proportion of commissions rises and the proportion of omissions falls.

The PC algorithm is able to recover the skeleton of the graph at a high rate, even when the signal strengths are low. Its ability to do so hardly varies with the different nominal test sizes. It turns out that the pattern of differences among different test sizes is robust across all the models we examine. We therefore shall report only the results for 10 percent size in the rest of the paper as it represents an intermediate case with a good balance between errors of omission and commission.

The usual criticism of specification searches involving repeated testing is that the true size rises substantially above the nominal size of the test statistic. Although the PC algorithm tests repeatedly, only tests involved in orienting links (as opposed to tests establishing the existence of a link) involve conditional decisions, which are the usual targets of opponents of data mining. The cost of search appears low in this case: the proportion of commissions for *LLL 1.00* is very close to the nominal size of the z -statistics and for *HHH 7.00* is about half the nominal size.⁹ The multiple testing used in the PC algorithm appears to be well-behaved on that front. In this simple, symmetrical model, the ability of the algorithm depends on the relative number of weak links. It is least effective when there are three low-strength lengths and most effective when there are three high strength links with the remaining combinations ordered lexicographically between these extremes.

⁹ These results are consistent with finds of well-behaved size in non-causal search algorithms (see Hoover and Perez (1999), Hendry and Krolzig (1999), Krolzig and Hendry (2001). For a general defense of well-regulated search as a respectable econometric practice, see Hoover (1995) and Hoover and Perez (2000).

Figure 8 reports the various ways in which the PC algorithm fails to correctly identify a true link in Model 1. The data are ordered by the proportion of total link errors, which again turns out to be lexicographically from *LLL* to *HHH*. Omissions are high if two or more links have low strength. When it does omit a link, the other errors (reversed or unresolved links) cannot occur. Typically, as the proportion omitted falls rapidly as signal strength rises, the proportion unresolved rises rapidly to fill the gap. Failures to resolve peak when all links have a medium signal strength (*MMM 3.50*) and falls as the number of high signal strengths increases. Irrespective of the average signal strength, even a single low-strength link noticeably increases the omission rate (compare, for example, omissions for *MMM 3.50* and *HHL 5.00*).

There is a clear hierarchy of error: omissions yield to failures to resolve yield to reversals. Reversals occur only when signal strengths are high. They peak at about 15 percent. The total error rate for true links bottoms out at 29 percent. This understates the success of the algorithm, first because once signal strengths are even moderately high it almost always never omits a link and because its error rate on true omitted links is very small. It almost always identifies the skeleton of the graph.

Model 2

The graph of Model 2 is depicted in Figure 9. Corresponding to the graph is

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_{210} & 1 & \alpha_{230} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \alpha_{420} & 0 & 1 \end{bmatrix},$$

where $\alpha_{ij0} \neq 0$. Model 2 has the same skeleton as Model 1. The *BD* link is reversed.

Model 2 has only one unshielded collider: *B* on the path *ABC*.

Model 2 is symmetrical only with respect to links 1 and 2. There are, therefore, more distinct combinations of signal strengths than was the case with Model 1 (18 in all). The rate of errors of commission is similar to that for Model 1: with a 10 percent nominal size, a maximum of 9.5 percent at *LLM 1.83* and a minimum of 5.5 percent at *HHM 5.83*.

Because of the asymmetry of Model 2 we investigate link errors for links 1 and 2 as a pair in Figure 10 and link 3 separately in Figure 11. Figure 10 is arranged in descending order of total link errors. There are three clearly defined sets. First, if link 3 or both links 1 and 2 have a low signal strength, then total link errors (for links 1 and 2) are nearly 100 percent with reversals and omissions accounting for about half each. Second, if link 3 has a medium or high signal strength and at most one of links 1 and 2 has a low signal strength, then the total error rate falls to around 80 percent. Omissions account for about 10 percentage points of the total, and unresolved links for most of the rest. Third, if link 3 and at least one of links 1 and 2 have a high signal strength, then the total error rate falls to 50 percent. Omissions fall to almost zero, and unresolved links account for almost all of the total. Reversals remain very low for all combinations.

Link errors for link 3 are decomposed in Figure 11, which is arranged in descending order of the total link error rate. The total error rate is high if any of the three links has a low signal. Omissions are about 80 percent if either link 1 or 2 has low signal strength. If both links have medium or high strength, then omissions stand at 11 to 13 percent, whatever value link 3 takes. If both are high, then omissions fall to zero. Failure to resolve are inversely related to omission rates and fall only when the total error rate itself falls when all signal strengths are medium or high. The maximum reversal rate is just over 10 percent.

Figure 12 provides an overall summary of the success of the algorithm at recovering the structure of Model 2. Data are ordered by increasing success at recovering the skeleton of the graph (that is, they are ordered so that the upper line, “correct irrespective of direction,” is monotonically increasing). At any reasonable signal strength, the algorithm performs well at recovering the skeleton. The lower line indicates its unqualified success both at recovering the skeleton and properly orienting the causal arrows. The difference between the two lines is a measure of the total number of link errors reported in Figures 10 and 11. The algorithm performs well at recovering the skeleton, so long as no more than one link has a low signal. The PC algorithm is less good at recovering either the skeleton or the true graph of Model 2 than it was of Model 1. This is not surprising, since Model 1 has only one unshielded collider, whereas Model 2

has three; it is the presence of unshielded colliders that makes orientation of links possible.

Model 3

Models 3 and 4 can be seen as elaborations of Models 1 and 2. The graph of Model 3 is depicted in Figure 13. The core graph is the same as Model 1. The link added between A and C acts as a shield, so that B on path ABC is no longer an unshielded collider. The additional link 5 adds another unshielded collider, while the additional link 6 does not. Model 3 has three unshielded colliders: B on paths ABD and CBD ; and C on ACE .

Since link 6 can be reversed to run $D \rightarrow F$ without changing the skeleton or the number of unshielded colliders, Model 3 is one member of a two-member equivalence class. The PC-algorithm cannot recover the true graph in principle. The best that it can do, the PC-true graph is shown in Figure 14. The simulated data is generated using the true graph, while the scoring uses the PC-true graph.

Recall that success at recovering particular links depends not only on those links directly, but on all of the links in the graph. Even the simple Model 2 presented some complexity. If we restrict ourselves to three levels of signal strength as before, there are $729 (= 3^6)$ combinations to be considered for each of the six links, yielding 4,374 evaluations. This is too complex to grasp easily, so some simplifications are necessary. In Figure 15 we simplify by reporting results for the average signal strengths across all six lengths. (It is important to recall in reading the graph that proportions are expressed as the number of times links are classified in a particular category as a fraction of the number of times that they could have been truly in that category. There are nine possibilities per realization to make an error of commission, six to make an error of omission, five to make reversals or to fail to resolve direction, and only one to overdetermine the direction. The last error was not possible for Models 1 and 2, which unlike the PC-true graph of Model 3 did not possess undirected links.)

Figure 15 shows that the PC-algorithm performs quite similarly with the more complex model as it did with Models 1 and 2. Errors of commission are approximately 10 percent when average signal strengths are below $z^* = 1$, and fall monotonically as average signal strength rises. Errors of omission start very high and fall rapidly as signal strength

rises. As omissions fall, unresolved links rise peaking at the moderate average signal strength of $3 < z^* \leq 4$, and falling thereafter. Reversals and overdeterminations appear only at higher signal strengths when failures to resolve direction at all become fewer. (Reversals and overdetermination occur more frequently at low average signal strengths because the omission of some links also interferes with the correct identification of unshielded colliders, which are essential to the correct orientation of the remaining links.) Overall success as measured by the recovery of the skeleton rises rapidly with signal strength and tops out at about 96 percent, while unqualified success at recovering the correct graph rises in parallel, toping out at just over 82 percent.

Model 4

Model 4 bears the same relationship to Model 2 as Model 3 does to Model 1. The true graph of Model 4 is depicted in Figure 16. It has two unshielded colliders: *C* on the path *ACE* and *D* on the path *BDF*. Since link 2 can be reversed without altering the skeleton or the unshielded colliders, the PC algorithm cannot direct it. The PC-true graph for Model 4 is depicted in Figure 17:¹⁰

The success of the PC algorithm in recovering Model 4 is shown in Figure 18. It is clearly similar both qualitatively and quantitatively to Figure 15, which refers to Model 3. Comparison of the uppermost increasing lines in each figure shows that the algorithm is equally good at recovering the skeleton of Model 4 as of Model 3. The difference is that it is slightly better at identifying the true links – direction as well as connection – for Model 3. Total link errors and unresolved links are higher for Model 4 at every average signal strength. This is consistent with the fact that Model 4 has fewer unshielded colliders than Model 3. Unshielded colliders are needed to discover the direction of causation but not the fact that variables are causally connected irrespective of direction. The zig-zag pattern of omissions and failures to resolve direction seen in Model 2 (Figure 11) are not recapitulated in the related Model 4, partly because grouping by average signal

¹⁰ Except for layout on the page it is the same graph as Figure 5 (vii).

strength mixes a variety of patterns of signal strengths in the same cell and partly because the additional unshielded collider increases the ability of the algorithm to orient links.

4. Preliminary Conclusions

Although the current paper reports work in progress, we believe that we are in a position to draw some preliminary conclusions from our simulation studies.

1. Causal search using graph-theoretic methods sometimes allows us to choose contemporaneous causal orders for SVARs based on relations of conditional independence in the data without appealing to *a priori* theory or other nonstatistical criteria. Even when these methods cannot select a unique causal order, they can narrow the equivalence class.
2. Contrary to the fears often expressed in relation to search methodologies, the PC algorithm appears to have well behaved statistical properties. In particular, the rate of falsely identifying a causal link when none exists in the true model occurs in the worst cases at a rate approximately equal to the size of the test statistic used to assess conditional independence and, in the best cases, at about half that rate. There is a clear tradeoff between increasing the rate of commission (type II error) and decreasing the rate of omission (type I error).
3. Error in causal ordering is more subtle than a type I/type II classification would suggest. A causal link can be omitted, but it can also be included but reversed, overdetermined or unresolved. In general errors of omission fall rapidly with signal strength. Other types of error can occur only if errors of omission do not. Reversals and overdetermination are generally relatively low, but increase with average signal strength, while failures to resolve the direction of links peaks in the mid-range of signal strengths and then falls.
4. All types of link errors are sensitive to the fine details of the causal structure. These can be well understood in very simple systems, but are hard to characterize as complexity grows even moderately.
5. In the systems examined, the PC algorithm was completely successful in identifying the correct causal structures with reasonable reliability only when

signal strengths were relatively high. It was, however, substantially better at identifying the skeletons of causal structures, even when it failed to resolve the directions of causal influence. Extra-statistical information may suggest the direction of particular causal links and when combined with knowledge of the skeleton may provide a firmer basis for a complete causal ordering than either the PC algorithm or the extra-statistical information could provide separately.

Appendix. The PC Algorithm

Descriptions of the PC algorithm are found in Spirtes *et al.* (2000), pp. 84 and 85 and Pearl (2000), pp. 49-51. This description is based on Cooper (1999, p. 45, Figure 22).

1. Start with a graph C in which each variable is connected by an edge to every other variable.
2. Set $n = 0$. Test for n^{th} -order conditional correlation between every pair of variables conditioning on every subset of variables size n . (For $n = 0$, the conditioning set is the null set, so that conditional correlation is equivalent to unconditional correlation). If a pair of variables is conditionally uncorrelated, eliminate the edge between them.
3. Set $n = n + 1$ and repeat step 2 until all possible conditionings have been exhausted. Call the resulting graph F .
4. Consider each pair of variables (X and Y) in F that are unconnected by a direct edge but are connected through an undirected path through a third variable (Z). Orient $X - Z - Y$ as $X \rightarrow Z \leftarrow Y$ if, and only if, X and Y are dependent when conditioned on every subset of variables, excluding X and Y , that includes Z . Call the resulting graph F' .
5. Repeat until no more edges in F' can be oriented:
 - a. If $X \rightarrow Z$ and $Z - Y$ and X and Y are not directly connected, then orient $Z - Y$ as $Z \rightarrow Y$.
 - b. If there is a directed path between X and Y (i.e., a path in which all edges have a consistent orientation with the arrowhead of the initial edge at X and all intermediate edges pointing to the tail of the next edge along the path until reaching Y) and if there is an undirected edge between X and Y , orient $X - Y$ as $X \rightarrow Y$.

References

- Akleman, Derya G., David A. Bessler, and Diana M. Burton. (1999) "Modeling Corn Exports and Exchange Rates with Directed Graphs and Statistical Loss Functions," in Clark Glymour and Gregory F. Cooper (eds.) *Computation, Causation, and Discovery*. Menlo Park, CA: American Association for Artificial Intelligence and Cambridge, MA: MIT Press, 1999, pp. 497-520.
- Bessler, David A. and Seongpyo Lee. (2002) "Money and Prices: U.S. Data 1869-1914 (A Study with Directed Graphs)," *Empirical Economics* 27, 427-446.
- Cooley, Thomas F. and Stephen F. LeRoy. (1985) "Atheoretical Macroeconometrics: A Critique," *Journal of Monetary Economics* 16(3), pp. 283-308.
- Cooper, Gregory F. (1999) "An Overview of the Representation and Discovery of Causal Relationships Using Bayesian Networks," in Clark Glymour and Gregory F. Cooper (eds.) *Computation, Causation, and Discovery*. Menlo Park, CA: American Association for Artificial Intelligence and Cambridge, MA: MIT Press, 1999, pp. 3-64.
- Demiralp, Selva. (2000) "The Structure of Monetary Policy and Transmission Mechanism," unpublished Ph.D dissertation, Department of Economics, University of California, Davis.
- Hendry, David F. and Hans-Martin Krolzig. (1999) "Improving on 'Data Mining Reconsidered' by K.D. Hoover and S.J. Perez," *Econometrics Journal*, 2(2), 202-218.
- Hoover, Kevin D. (1995) "In Defense of Data Mining: Some Preliminary Thoughts," in Kevin D. Hoover and Steven M. Sheffrin (eds.) *Monetarism and the Methodology of Economics: Essays in Honour of Thomas Mayer*. Aldershot: Edward Elgar, 1995.
- Hoover, Kevin D. (2001) *Causality in Macroeconomics*. Cambridge: Cambridge University Press.
- Hoover, Kevin D. and Stephen J. Perez. (1999) "Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search," *Econometrics Journal* 2(2), 167-191.
- Hoover, Kevin D. and Stephen J. Perez. (2000) "Three Attitudes Towards Data-mining," *Journal of Economic Methodology* 7(2), 195-210.
- Hoover, Kevin D. and Stephen J. Perez. (2002) "Truth and Robustness in Cross-country Growth Regressions," unpublished working paper, University of California, Davis.
- Krolzig, Hans-Martin and David F. Hendry. (2001) "Computer Automation of General-to-Specific Model Selection Procedures," *Journal of Economic Dynamics and Control*. 25(6-7), June-July, 831-66.

- Leamer, Edward. (1985) "Vector Autoregressions for Causal Inference?" in Karl Brunner and Alan H. Meltzer (eds.) *Understanding Monetary Regimes*, Carnegie-Rochester Conference Series on Public Policy, vol. 22 (Spring). Amsterdam: North Holland, pp. 225-304.
- LeRoy, Stephen F. (2002) "Causality: Models, Reasoning, and Inference: A Review of Judea Pearl's *Causality*," *Journal of Economic Methodology* 9(1), 100-103.
- Pearl, Judea. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Sheffrin, Steven M. and Robert K. Triest. (1998) "A New Approach to Causality and Economic Growth," unpublished typescript, University of California, Davis.
- Sims, Christopher A. (1980) "Macroeconomics and Reality," *Econometrica* 48(6), 1-48.
- Sims, Christopher A. (1986) "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review* 10(1), 2-15.
- Spirtes, Peter, Clark Glymour, and Richard Scheines. (2000) *Causation, Prediction, and Search*, 2nd edition. Cambridge, MA: MIT Press.
- Spirtes, Peter, Richard Scheines, Christopher Meek, Thomas Richardson, Clark Glymour, Herbert Hoijtink and Anne Boomsma. (1996) *TETRAD 3: Tools for Causal Modeling*, program (beta version, October) and user's manual on the worldwide web at <http://www.phil.cmu.edu/tetrad/tet3/master.htm>.
- Swanson, Norman R. and Clive W.J. Granger. (1997) "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association* 92(437), 357-367.

Figure 1

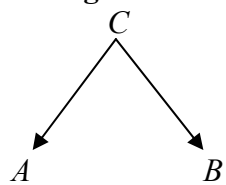


Figure 2

(i)
 $A \rightarrow C \rightarrow B$

(ii)
 $A \leftarrow C \leftarrow B$

Figure 3

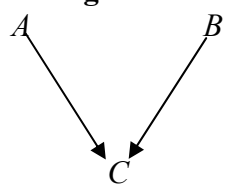


Figure 4

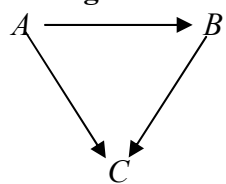
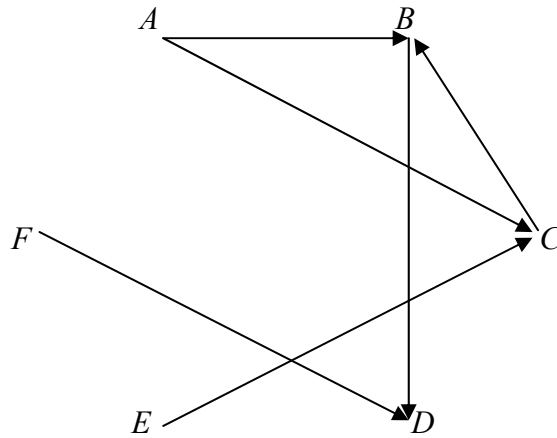


Figure 5. How the PC Algorithm Works

- (i) The true structure determines the correlations that will be found in the data, and which can be eliminated or oriented in the various steps of algorithm (shown in subsequent panels).**



- (ii) Start with a graph in which every variable is connected by an undirected edge to every other.**

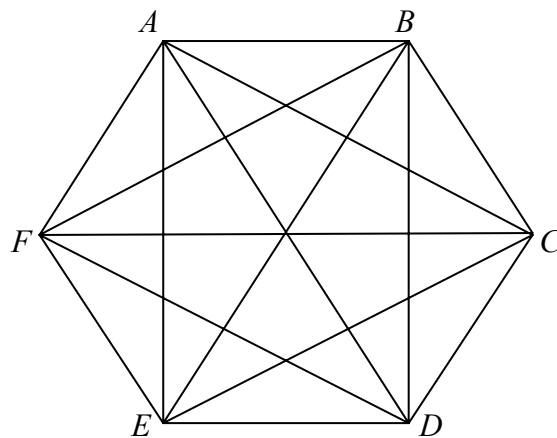


Figure 5 (continued)

(iii) Edges between variables that are not unconditionally correlated are eliminated: *AE* is broken by the collider at *C*; *AF*, *BF*, *CF*, and *EF* are broken by the collider at *D*. (Dashed lines or arrows indicate the links affected by the current step.)

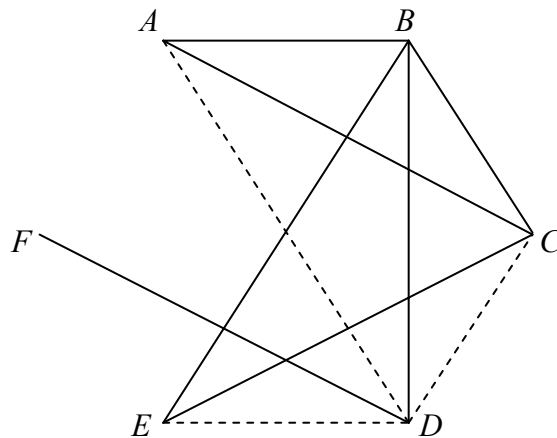
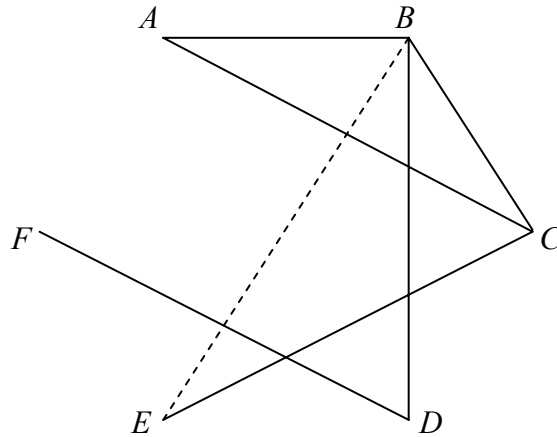


Figure 5 (continued)

- (v) Edges between variables that are uncorrelated conditional on a two variables are eliminated: C and A screen B from E .



- (vi) Conditioning on more than two variable at a time will not remove any further edges. Check every pair of variables that are not directly connected (and are therefore uncorrelated conditional on some subset of the variables) but are connected through a third variable. If conditioning on the third variable renders them correlated, then orient the edges as arrows pointing into the third variable: C is identified as a collider on ACE and D as a collider on BDF .

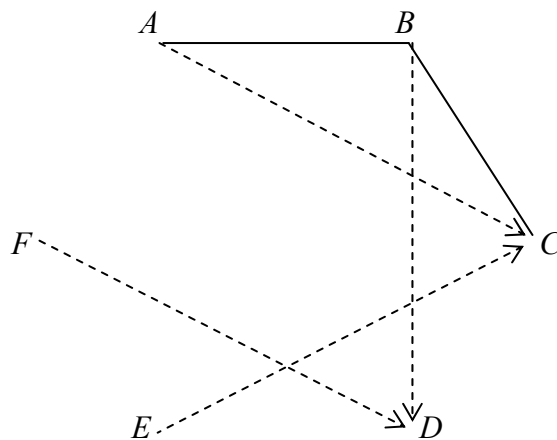
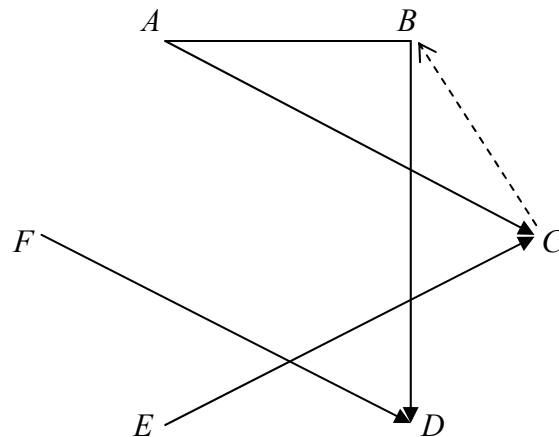


Figure 5 (continued)

(vii) Since C screens B from E and an arrow runs from E into C , the edge from C to B must be oriented toward B . No further edges can be oriented, so (vii) is the final graph. It is identical to (i) in all but the AB edge, which in (vii) is missing the arrowhead into B that is present in (i). The algorithm cannot orient that edge.



(viii) The unoriented edge on AB in (vii) shows that the true graph in (i) is not the only member of the equivalence class. The following graph has the same skeleton and unshielded colliders as (i), differing only in the orientation of the arrow between A and B.

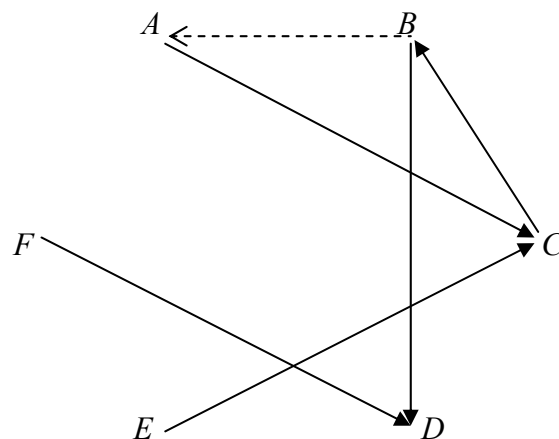


Figure 6

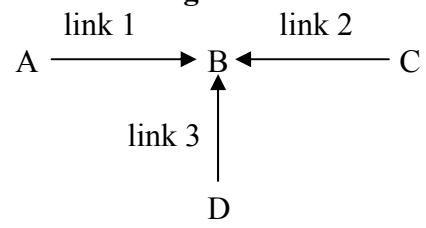


Figure 7. Performance of the PC Algorithm for Model 1 at Different Nominal Test Sizes

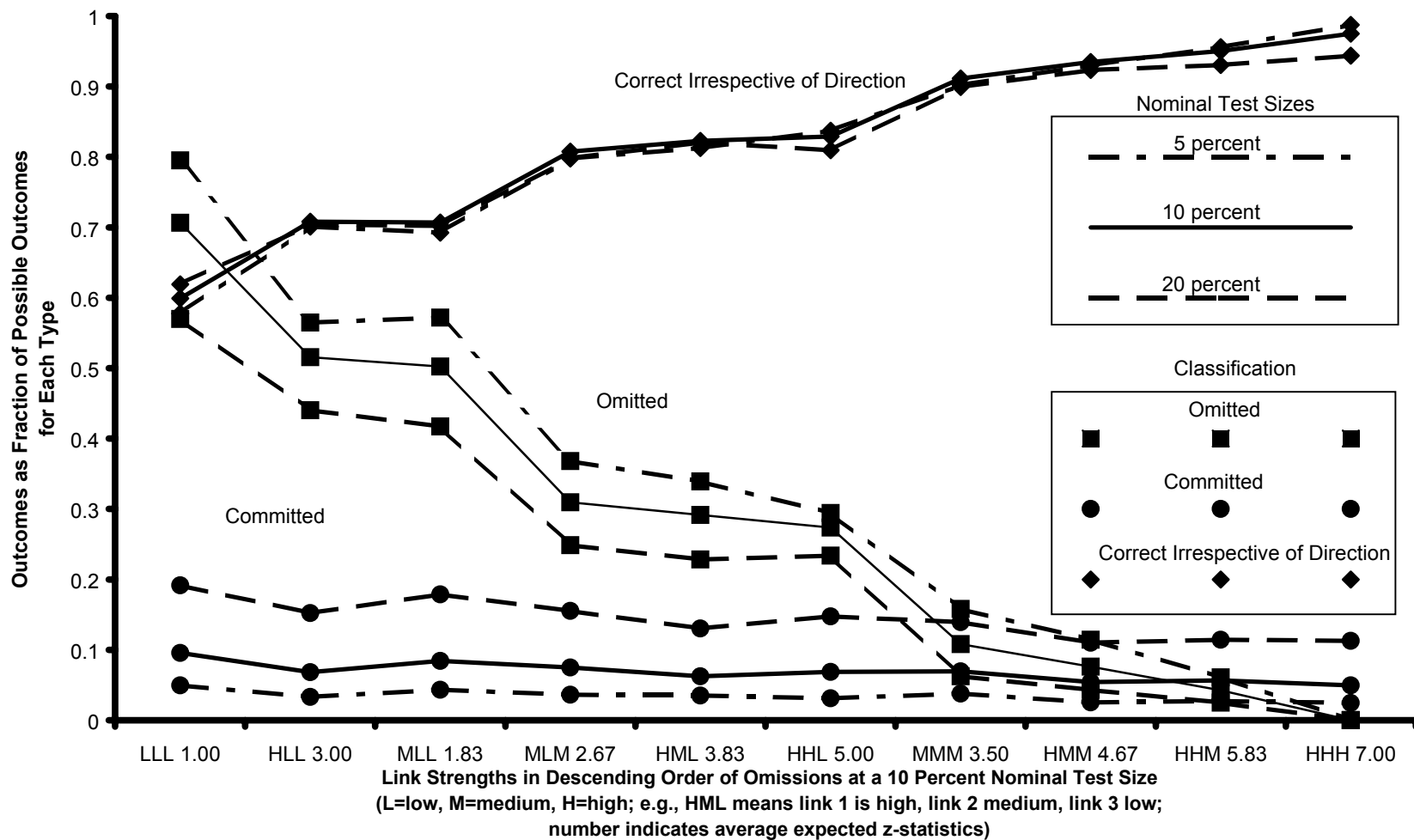


Figure 8. Model 1 Link Errors

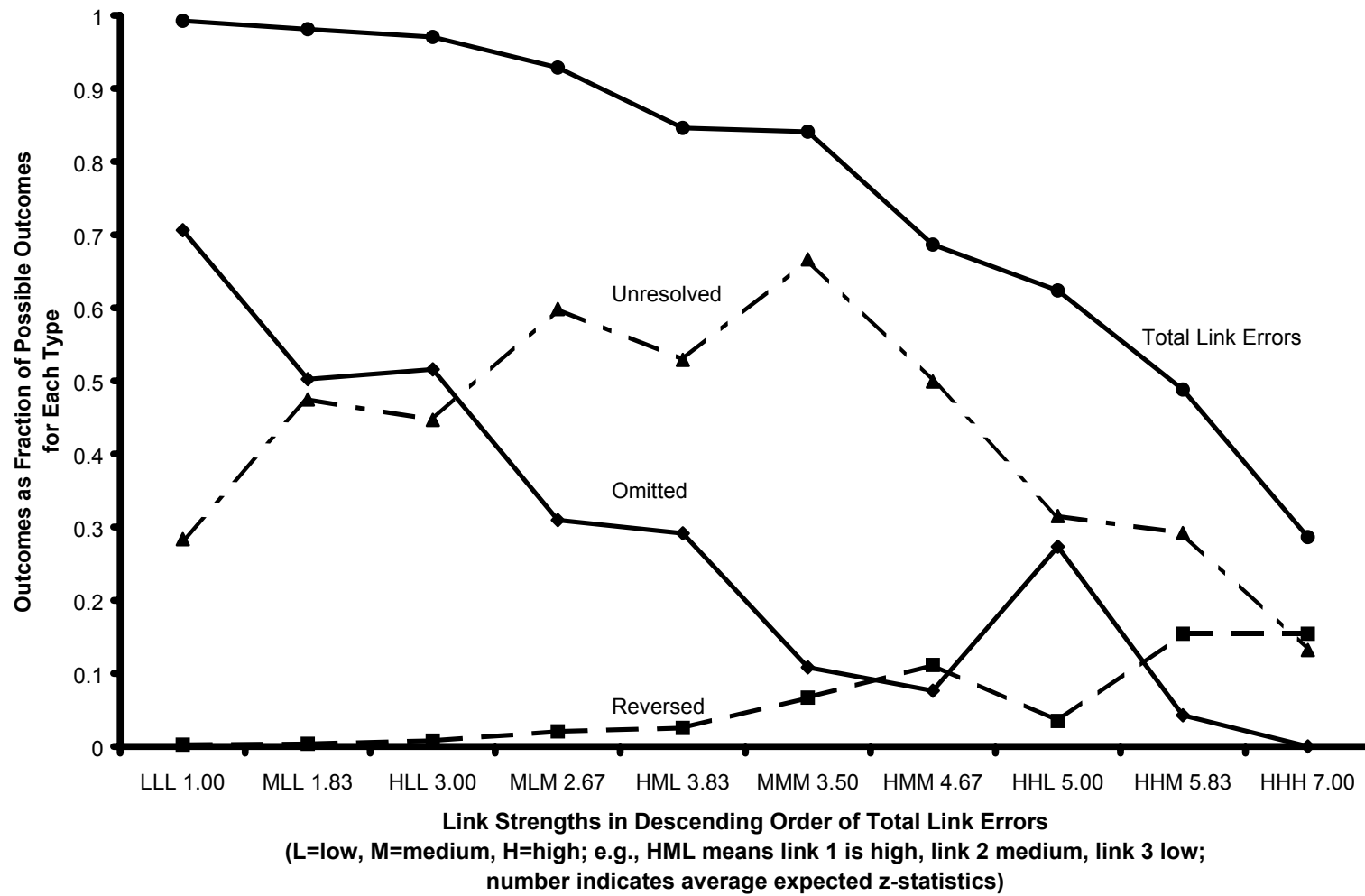


Figure 9

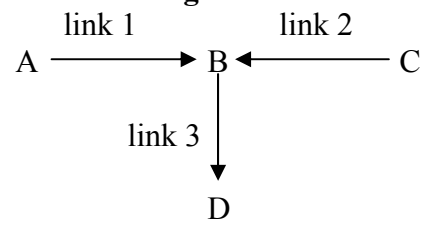


Figure 10. Model 2 Link Errors for Links 1 and 2

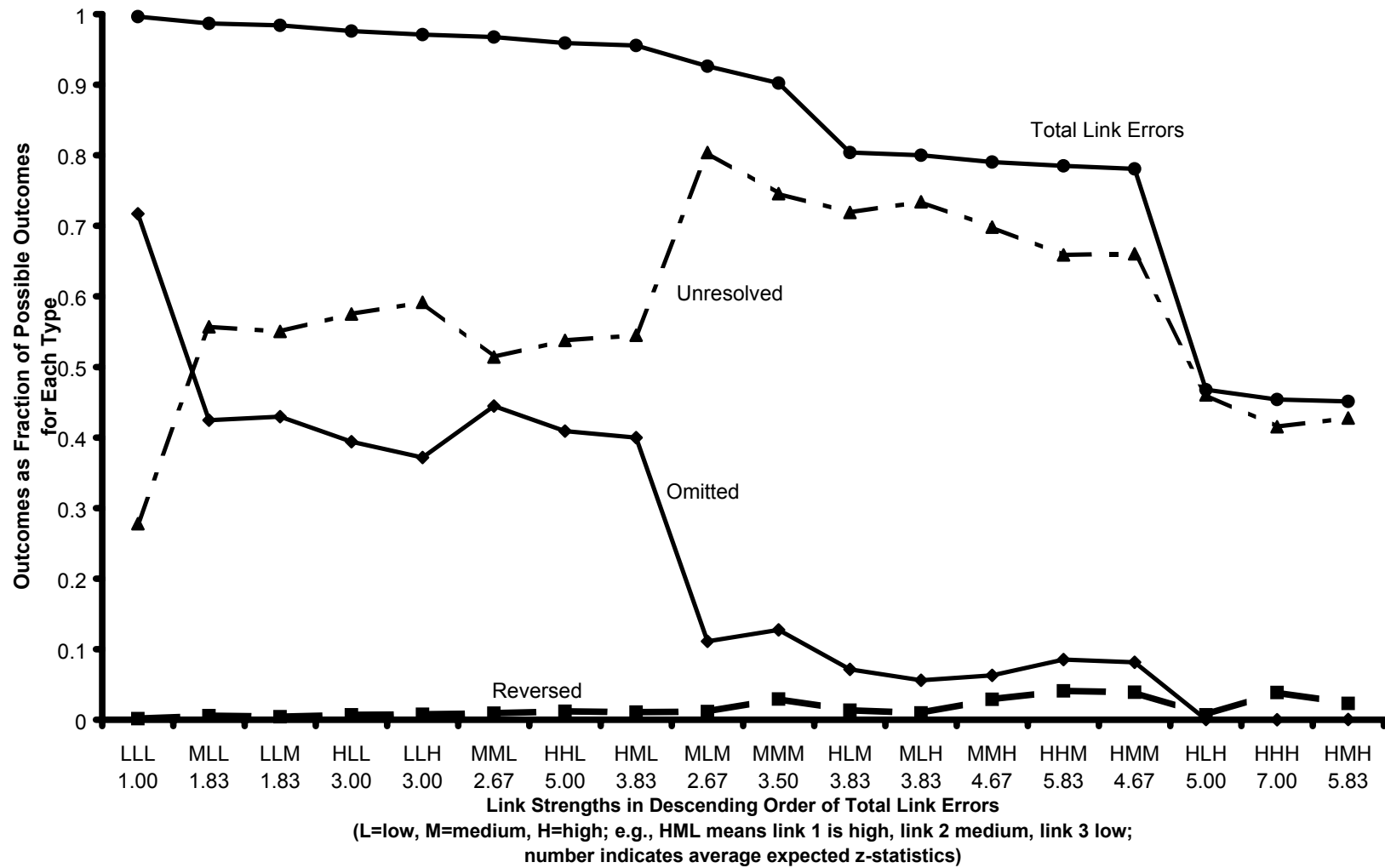


Figure 11. Model 2 Link Errors for Link 3

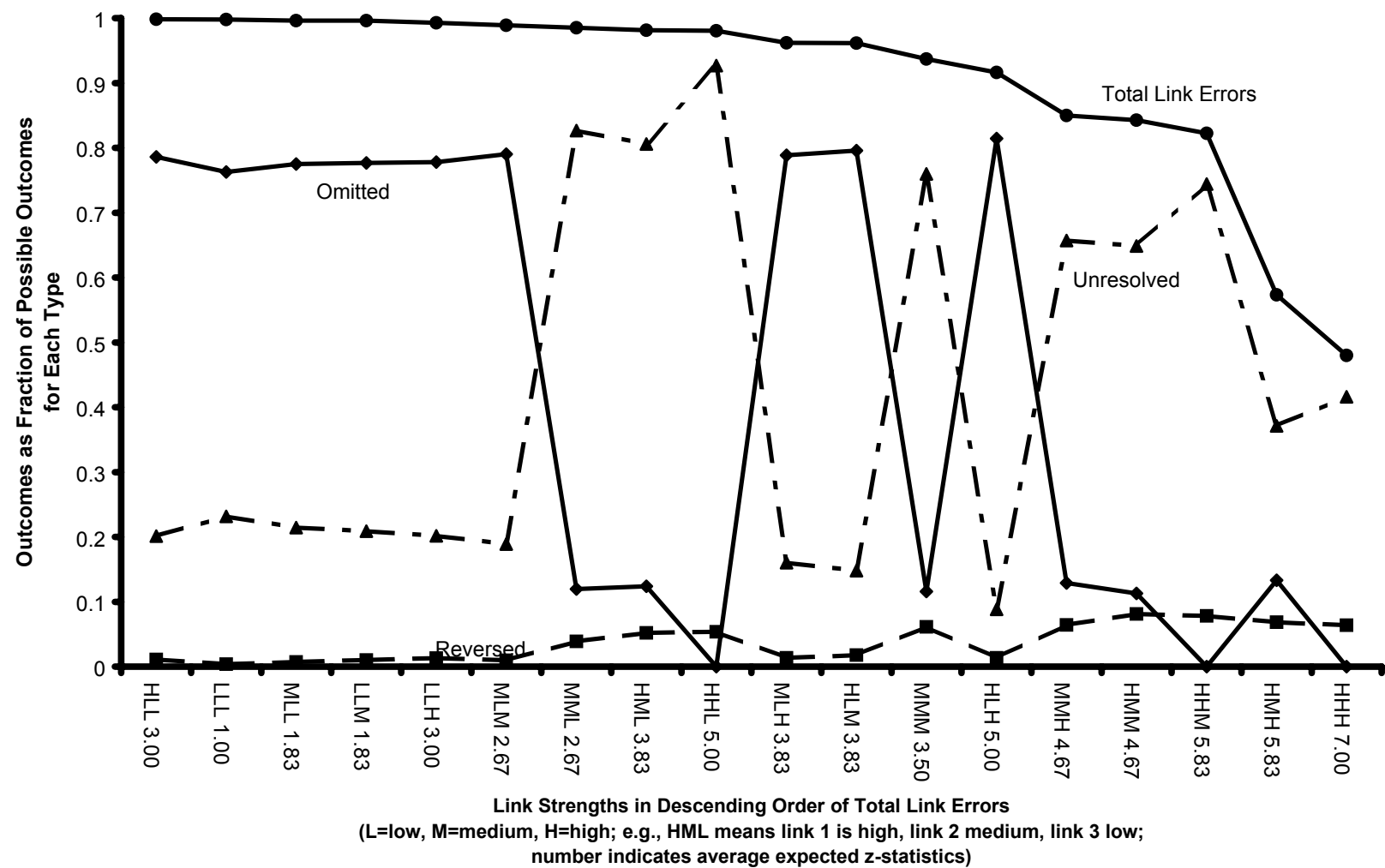


Figure 12. Overall Success of Model 2

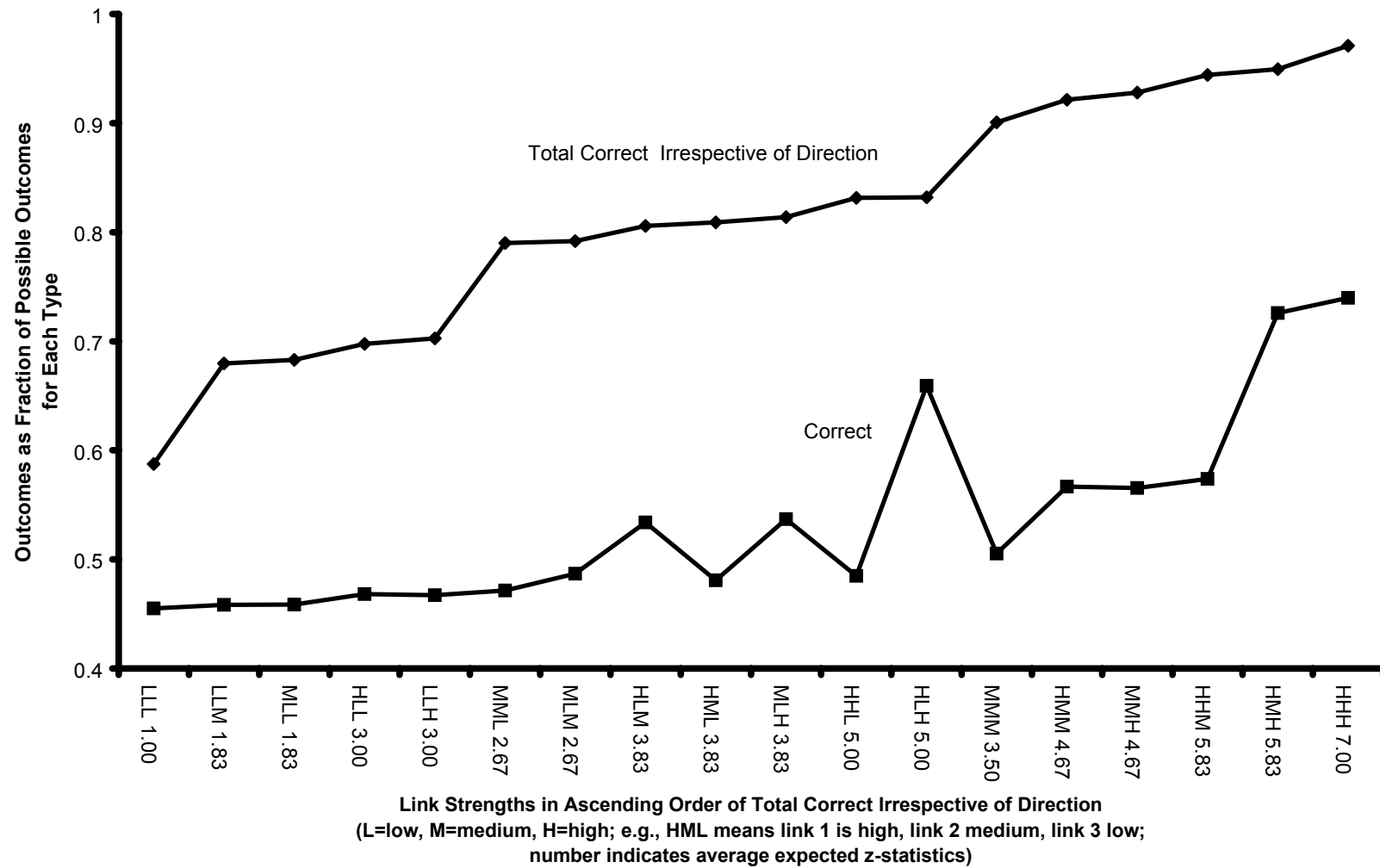


Figure 13. The True Graph of Model 3

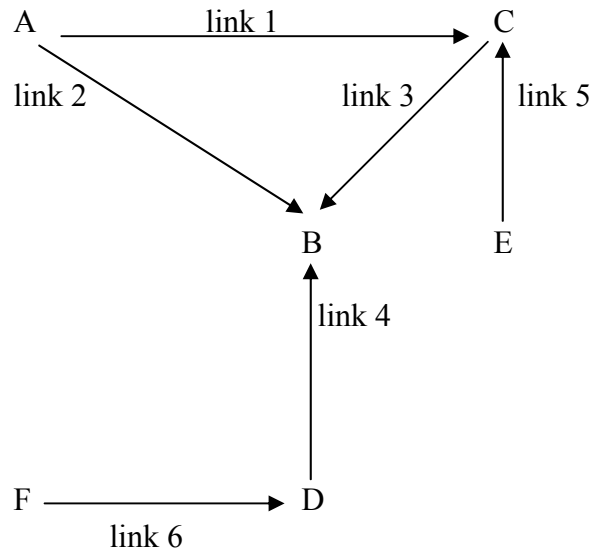


Figure 14. The PC-true Graph of Model 3

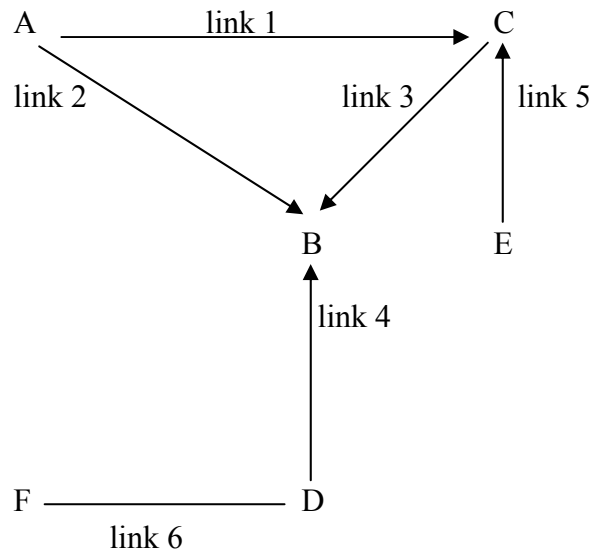


Figure 15. Outcomes by Average Signal Strength for Model 3

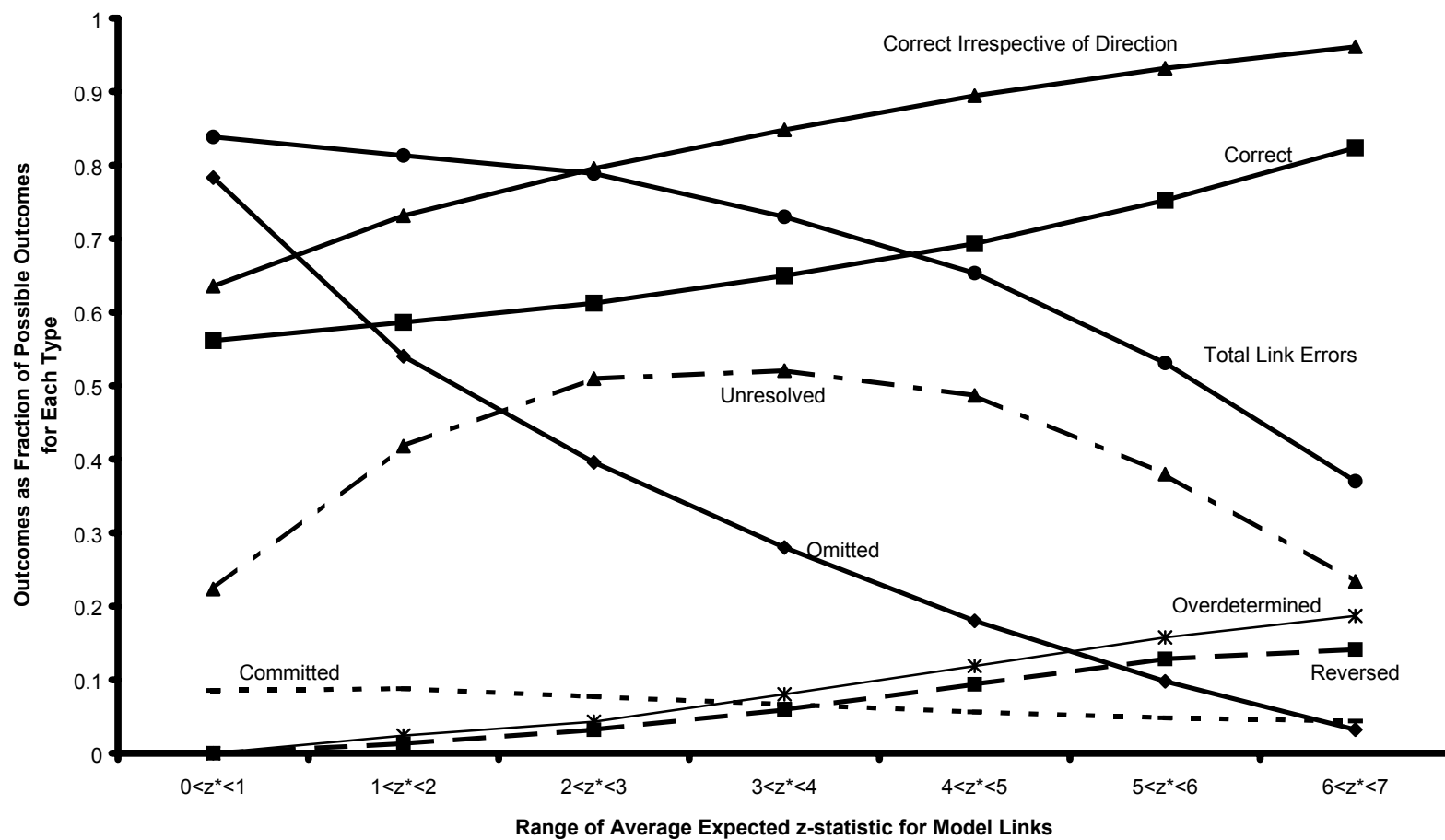


Figure 16. The True Graph of Model 4

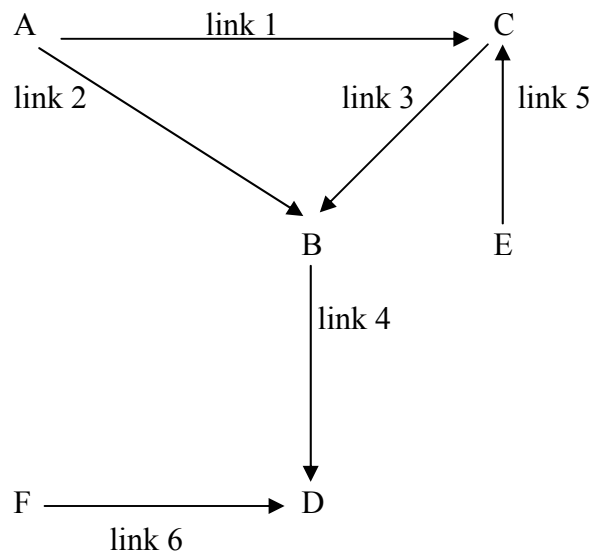


Figure 17. The PC True Graph of Model 4

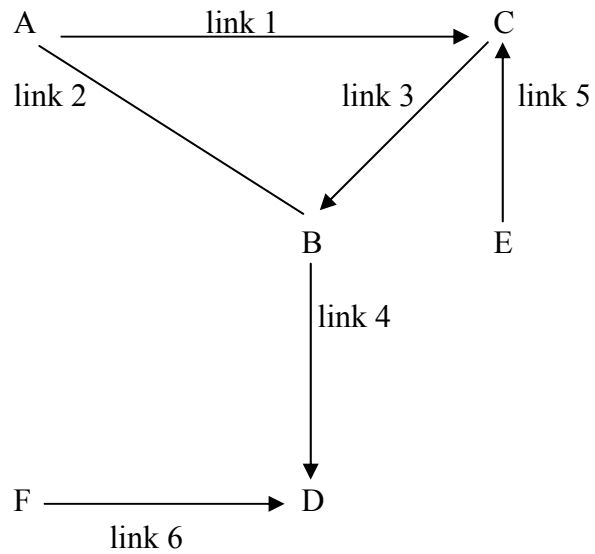


Figure 18. Outcomes by Average Signal Strength for Model 4

