Searching for the Causal Structure of a Vector Autoregression*

Selva Demiralp† and Kevin D. Hoover‡

†Division of Monetary Affairs, Monetary and Reserve Analysis, Board of Governors of the Federal Reserve System, Washington, DC, USA
(e-mail: selva.demiralp@frb.gov)
‡Department of Economics, University of California, Davis, CA, USA
(e-mail: kdhoover@ucdavis.edu)

Abstract

We provide an accessible introduction to graph-theoretic methods for causal analysis. Building on the work of Swanson and Granger (*Journal of the American Statistical Association*, Vol. 92, pp. 357–367, 1997), and generalizing to a larger class of models, we show how to apply graph-theoretic methods to selecting the causal order for a structural vector autoregression (SVAR). We evaluate the PC (causal search) algorithm in a Monte Carlo study. The PC algorithm uses tests of conditional independence to select among the possible causal orders – or at least to reduce the admissible causal orders to a narrow equivalence class. Our findings suggest that graph-theoretic methods may prove to be a useful tool in the analysis of SVARs.

I. 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 'Macroeconomics and reality' by Sims (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

$$\mathbf{B}(L)\mathbf{Y}_t = \mathbf{U}_t, \quad \text{where } \mathbf{B}(0) = \mathbf{I}. \tag{1}$$

The subscript t indexes time; $\mathbf{Y}_t$ is an $n \times 1$ column vector of the contemporaneous values of the variables Y_{it}, $i = 1, 2, \ldots, n$; $\mathbf{B}(L)$ is a conformable square matrix whose terms are polynomials in the lag operator; and $\mathbf{U}_t$ is a column vector of 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 $\mathbf{u}_i = [u_{i1}, u_{i2}, \ldots, u_{iT}]$ be the time series for u_{it} and $\mathbf{U}$ without a time subscript be the $n \times T$ matrix whose rows are the $\mathbf{u}_i$. (Equivalently, here and in analogous cases, omitting time subscripts indicates the matrix whose columns are the $\mathbf{U}_t$.) The contemporaneous covariance matrix is $\Sigma = \mathrm{E}(\mathbf{U}\mathbf{U}')$, where E is the expectations operator. In general, Σ 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 $\mathbf{C}\mathbf{C}' = \Sigma$. Premultiplying both sides of equation (1) by $\mathbf{C}^{-1}$ yields

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

The covariance matrix of equation (2) is $\mathrm{E}(\mathbf{C}^{-1}\mathbf{U}(\mathbf{C}^{-1}\mathbf{U})') = \Psi$, where Ψ is diagonal (Hamilton 1994, p. 320). 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 $Y_2, Y_3, \ldots, Y_n$, while the shock to Y_2 feeds contemporaneously into $Y_3, Y_4, \ldots, Y_n$, but into Y_1 only with a lag, and so on. The order of the Y_{it} in the vector $\mathbf{Y}_t$ is arbitrary; but, for each such ordering, the Choleski decomposition is unique.[1] What is more, orthogonal-

[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.

izing transformations are not restricted to Choleski decompositions, but may involve non-triangular matrices **P**, such that $E(\mathbf{P}^{-1}\mathbf{U}(\mathbf{P}^{-1}\mathbf{U}')) = \Psi$, 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 (i.e. orthogonalizing transformations), so that the economist must appeal to *a priori* knowledge. As 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 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, but it is also 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.

All just-identified SVARs derivable from the same VAR are in fact observationally equivalent. As we shall see presently, overidentified SVARs belong to equivalence classes – not only distinct from the class of just-identified SVARs, but also from each other. It would be useful if these classes could be determined from the data. Some of these classes may have only one member.

The underlying premise of SVAR analysis is that some SVAR corresponds to the data-generating process. Let the data-generating process be

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

where $\mathbf{E}_t = [\varepsilon_{ijt}]$ is a column vector of error terms at time t; $\mathbf{A}(L)$ is a conformable matrix whose terms are polynomials in the lag operator, such that $\mathbf{A}(0)$ is a lower triangular with ones on the main diagonal and, possibly, zeroes for some elements in the lower triangle; and the covariance matrix $\Omega = E(\mathbf{EE}')$ is diagonal. The true SVAR is distinguished from the

---

[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.

orthogonalized VAR by the fact that its error terms are *independent* rather than merely uncorrelated by construction. It is the independence of the error terms that marks the VAR as *structural*.

Let $\mathbf{A}_0 \equiv \mathbf{A}(0)$ be the matrix of the zeroth-order terms of the matrix $\mathbf{A}$ – i.e. the typical element of $\mathbf{A}_0$ is α_{ij0}. The reduced form of equation (3), is then

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

Equating terms with equation (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

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

The independence of the ε_{it} and the structure of the SVAR embodied in $\mathbf{A}_0$ implies the relationships of interdependence, independence, and conditional independence among the u_{it}, 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 equation (1) into equation (3) is, therefore, a privileged one in that it is the only one that recovers the independent errors, the ε_{it}.

The reduced form [equation (4)] may be transformed into a variety of (pseudo) structural VARs – each appearing to possess a contemporaneous causal order different from the data-generating process and each with uncorrelated error terms. But although the error terms of the pseudo SVARs are mutually uncorrelated, they are *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), 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.

The central question is, then, whether, starting from the equation (1), empirical evidence can help us to recover the privileged transformation that corresponds to the true equation (3)? In principle, the answer is yes.

Over the past 20 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

---

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

have been used in a variety of social sciences other than economics, but are unfamiliar to most economists.[4]

Most of this research on graph-theoretic methods assumes that the causally ordered data are cross-sectional. Time series data may show statistical dependencies unrelated to causal relationships because of failures of stationarity, temporal aggregation, or omitted simultaneity.[5] Furthermore, it is unclear whether past values of a variable should be regarded as causing future values or that autocorrelation would be evidence for causal connectedness. 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 fully in section III below.) Swanson and Granger (1997) restrict the admissible structures to causal chains (for example, *A* causes *B* causes *C*; but not *A* causes *B*, and *A* and *B* cause *C*). In contrast, we allow every possible ordering. In this paper we contribute three things. First, we provide an accessible account of the underlying rationale for the graph-theoretic approach to causal order in general. Second, we use the PC algorithm embedded in Spirtes *et al.*'s (1996) Tetrad 3 software to extend Swanson and Granger's methodology to the entire class of acyclic (recursive) graphs, rather than to causal chains only. Third, using a simulation study, we investigate the efficacy of our approach to select the causal order of SVARs.

II.   The graph-theoretic analysis of causal structure

Start with a structure defined by equation (3) with the added stipulation that the matrix $\mathbf{A} = \mathbf{A}_0$, i.e. 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

[4]Hoover (2001, ch. 7) gives a critical description of these methods, and LeRoy (2002) has recently discussed them in a review of Pearl (2000). Some earlier applications to economics include Sheffrin and Triest (1998) and Akleman, Bessler and Burton (1999). Swanson and Granger (1997) and Demiralp (2000, ch. 4) are particularly concerned with the causal order of the VAR. Reale and Tunnicliffe Wilson (2000, 2002), Tunnicliffe Wilson and Reale (2002), and Tunnicliffe Wilson, Reale and Morton (2001) apply substantially different, but still graph-theoretically based methods based on the work of Lauritzen and Spiegelhalter (1988) (see also Lauritzen and Richardson 2002) to VARs using economic data. A comparison of these methods with the PC algorithm would be interesting but beyond the scope of this article. Many of the authors who have used economic data are not economists and virtually all of the work using economic data has been published in statistics journals or edited volumes or circulates only in working paper format (Bessler and Lee, 2002 is an exception to the rule).

[5]Spirtes *et al.* (2000, p. 296). See also Hoover (2003) and Tunnicliffe Wilson and Reale (2002) on non-stationary variables and causality and Breitung and Swanson (2002) and Hoover (2001, ch. 6, section 2) on temporal aggregation.

$$\mathbf{Y} = \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \alpha_{31} & 0 & 1 & 0 & 0 \\ \alpha_{41} & \alpha_{41} & 0 & 1 & 0 \\ 0 & \alpha_{52} & 0 & \alpha_{54} & 1 \end{bmatrix},$$

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, *CAD* is a path from *C* to *D*. A *directed path* follows the direction of causation. For example, *BDE* is a directed path from *B* to *E*; while *CAD* is a path, but not a directed path, from *C* to *D*. 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 acylical graphs, we restrict our simulations to causally sufficient, acylical models.

Returning to the initial model, as the ε_i are independent random shocks, the matrix **A** and its corresponding graph (Figure 1), represent a causal structure that defines the patterns of dependence or independence among the variables.
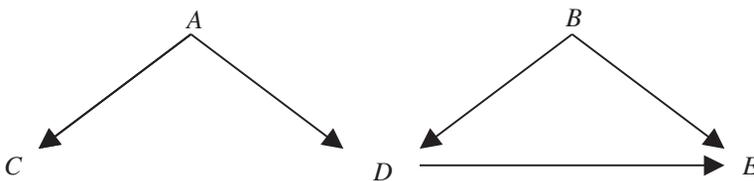
Figure 1. A directed graph

In this case, it is easy to see that C and D are not independent because both depend on A. A is said to be their *common cause*. It is also intuitive that C is independent of D *conditional* on A.

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 ($A \leftarrow B$) or they have a common cause ($A \leftarrow C \rightarrow B$). 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 **V** with a probability distribution P. Let **W** be a subset of **V**. G and P satisfy the *causal Markov condition* if, and only if, for every **W** in **V**, **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).[6]

A few further examples illustrate how to apply these ideas. Consider the two causal graphs $A \rightarrow C \rightarrow B$ and $A \leftarrow C \leftarrow B$. 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. In Figure 1, A and B are unconditionally uncorrelated. They are however correlated conditional on D. The classic example is $A =$ the car's battery being charged; $B =$ the car's starter switch being on; and $D =$ the car's starting. A and B may be completely independent. Yet, *if we know that the car does not start*, then knowing that the switch is on raises the probability that the battery is dead. Node D in Figure 1 is called an *unshielded collider on the path ADB* (or *BDA*). It is a 'collider' because the arrowheads come together at D, and is 'unshielded' because there is no direct causal connection between A and B. Node E is a *shielded collider on the path DEB*. The link $B \rightarrow D$ acts as a shield in that B and D are correlated even without conditioning on the common effect.

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* (Spirtes *et al.*, 2000, p. 31) or *stable* (Pearl, 2000, p. 38) if and only if, there is a one-to-one mapping between the relationships of conditional independence implied by the causal Markov condition applied to G and those found in P (Spirtes *et al.*, 2000, p. 31). Consider a graph like Figure 1 except that the link $A \rightarrow D$ is missing. The parameters of the causal structure might happen to take specific values such that the correlation

[6]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.

between *B* and *E*, present in the graph, is not present in the probability distribution. In that case, the graph and the probability distribution are not faithful. If the direct link $D \rightarrow E$ was also missing, however, the parameter-specific loss of correlation could not occur, and the graph would be faithful.[7]

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 or many elements. According to the *observational equivalence theorem* (Pearl, 2000, p. 19, Theorem 1.2.8) any probability distribution that can be faithfully represented in a causally sufficient, acyclical graph can be equally well represented by any other acyclical graph that has the same skeleton and the same unshielded colliders (cf. 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 leaves the reversible links undirected. This partial causal ordering defines an equivalence class whose members correspond to the permutations of the orientations of the undirected links. A just-identified model has no unshielded colliders. It follows immediately that all just identified models are observationally equivalent.

Causal search algorithms start with the empirical probability distribution of the variables (typically characterized by the covariance, or (equivalently) the correlation, matrix). Although there are other algorithms, we investigate the PC algorithm of Spirtes *et al*. (2000), because it is the most commonly used and is readily implemented using easily available software. Appendix A presents a detailed description of the PC algorithm. It is easily understood. The causal Markov condition applied to any graph indicates the conditional independence relations that should, or should not, be present in the data. The algorithm tests for the implied independence relations and works backward to the graph as far as possible.[8] The algorithm begins by assuming that each variable is linked to every other variable through an undirected link. It then proceeds in two major stages: elimination and orientation.

---

[7]Hoover (2001, pp. 45–49, 151–153, 168–169) considers cases in which failures of faithfulness arise naturally in economic contexts and cannot be dismissed as suggested by Spirtes *et al*. (2000, p. 41) as of 'Lebesque measure zero' (see also Pearl, 2000, p. 63).

[8]Tests of conditional independence are implemented using conditional correlations. The unconditional correlation coefficient between *A* and *B* is denoted r_{AB}. The correlation of *A* and *B* conditional on *C* is then defined as $r_{AB|C} = (r_{AB} - r_{AC}r_{BC})/(\sqrt{1 - r_{AC}^2}\sqrt{1 - r_{BC}^2})$. See Johnston (1972, p. 135) for the generalization to conditioning on more than one variable. The statistical significance of the conditional correlation can be computed using Fisher's *z*-statistic. As we observed in section I, 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 I. Hoover (2001, ch. 2, section 4, and ch. 7, section 1) discusses cases in which these vanishing correlations arise from economically meaningful optimal control.

At the *elimination stage*, it first removes connections between any variables that are not unconditionally correlated. Next it tests for the correlation of each pair of variables conditional on a third variable, eliminating the link between any pair that is conditionally uncorrelated. Continuing in the same vein, it tests for absence of correlation conditional on pairs of variables, then on sets of three, four, or more and eliminates links whenever there is no conditional correlation.

Once the possibilities for elimination of links are exhausted, the PC algorithm proceeds to the *orientation stage*. It considers every pair of variables that is conditionally uncorrelated and causally connected along an undirected path through a third variable. If the pair is correlated conditional on the third variable, then the members of the pair are unshielded colliders on that path, and the arrows from the pair are oriented towards the third variable. Finally, some unoriented links may be oriented based on screening relationships: if two variables (A and B) are not directly connected, but are connected through a third variable (C), so that one link points to the third variable (say, $A \rightarrow C$) and the other link is undirected ($C — B$), then the undirected link is oriented to point away from the third variable ($C \rightarrow B$). This follows because, given the previous steps, the intervening variable is a screen and not an unshielded collider, and so the link cannot point towards it.

The PC algorithm may fail to orient some links. This will happen, even when small sample issues are not in play, whenever the graph of the true data-generating process contains links whose orientations can be reversed without affecting the number and location of the unshielded colliders.

III.   The effectiveness of the PC (causal search) algorithm

The simulation methodology

Spirtes *et al.* (1996, user's manual, ch. 13), Spirtes *et al.* (2000, pp. 113–122), and Cooper (1999, section 10) present some simulation evidence of the effectiveness of the PC algorithm. 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 $\mathbf{A} = \mathbf{A}_0 + \bar{\mathbf{A}}$, where the elements of $\bar{\mathbf{A}}$ are $\bar{A}_{ij} = a_{ij1}L + a_{ij2}L^2 + a_{ij3}L^3 + a_{ij4}L^4 + \cdots + a_{ijK}L^K$. Each equation in the SVAR has an identical lag structure, i.e. for each $j = 1, 2, \ldots, N$ and each $k = 1, 2, \ldots, K$, and for all $i = 1, 2, \ldots, N$ and $h = 1, 2, \ldots, N$: $\alpha_{ijk} = \alpha_{hjk}$.[9] For

---

[9]Parameter values are chosen to ensure stationarity. The parameter values for all own lags ($i = j$, $k = 1, 2, 3, 4$) is (a_{ijk}) = (0.0403, 0.162, 0.0654, 0.0264), and the values for all cross lags ($i \neq j$) is (a_{ijk}) = (0.054, 2.92×10^{-3}, 5.17×10^{-4}, 8.50×10^{-6}).

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 non-zero terms in $\mathbf{A}_0$. Given the causal structure and a particular choice of values for the non-zero a_{ij0}, the data are generated recursively drawing the error terms from a random-number generator. The $\varepsilon_{ijt} \sim N(0, 1)$. To eliminate problems with initial values, 1,500 realizations are generated and only the last 500 retained for analysis.[10]

2. In order to evaluate a range of signal-to-noise ratios, we generate 50,000 realizations for each model with the non-zero 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 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 $\hat{\Sigma} = \mathbf{U}\mathbf{U}'/(T - K)$ 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 indexes the equivalence class to which the true graph belongs. It 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. It is the graph that PC algorithm would select under the best circumstances (i.e. with an infinite amount of data).

   Every possible link is evaluated. The possible outcomes are:

    (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.

---

[10]Many standard U.S. macroeconomic time series begin in 1959. Using monthly data, there are 528 observations between 1959 and 2003. Since many data are quarterly, and many VARs are estimated for shorter periods, it would be of interest in subsequent work to investigate the sensitivity of our results to sample size.

(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) to (vi) can occur only if a link is present in the reference graph.

Three 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 a more complex model.

Model 1

The graph of model 1 is depicted in Figure 2. Corresponding to the graph is

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha_{410} & \alpha_{420} & \alpha_{430} & 1 \end{bmatrix},$$

where $\alpha_{ij0} \neq 0$. Model 1 is symmetrical around y_4 (that is, swapping the positions of y_1, y_2, and y_3 produces isomorphic graphs). The variable y_4 is an unshielded collider on three separate paths: $y_1 y_4 y_2$, $y_1 y_4 y_3$, and $y_2 y_4 y_3$. As with
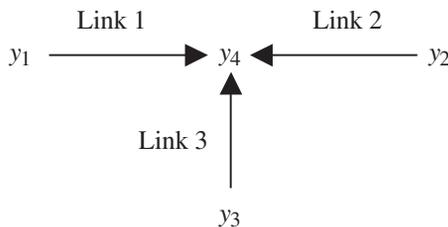
Figure 2. Model 1

TABLE 1

Model 1 simulation outcomes by signal strength

| Signal strength† | | Outcomes as fraction of possible outcomes | | | | | | |
| | | | Skeleton | | Link errors | | | |
By link‡	Average§	Correct	correct	Committed	Omitted	Reversed	Unresolved	Total
LLL	1.00	0.46	0.60	0.10	0.71	0.00	0.28	0.99
MLL	1.83	0.47	0.71	0.08	0.50	0.00	0.47	0.98
HLL	3.00	0.48	0.71	0.07	0.52	0.01	0.45	0.97
MLM	2.67	0.50	0.81	0.07	0.31	0.02	0.60	0.93
MMM	3.50	0.54	0.91	0.06	0.11	0.07	0.67	0.84
HML	3.83	0.55	0.82	0.07	0.29	0.03	0.53	0.85
HMM	4.67	0.63	0.93	0.07	0.08	0.11	0.50	0.69
HHL	5.00	0.65	0.83	0.05	0.27	0.03	0.32	0.62
HHM	5.83	0.73	0.95	0.06	0.04	0.15	0.29	0.49
HHH	7.00	0.83	0.98	0.05	0.00	0.15	0.13	0.29

Notes:

†Signal strength measured as the expected z-statistic.

‡L = low ($0 \leq z^* < 2$), M = medium ($2 \leq z^* < 5$), H = high ($5 \leq z^* \leq 9$); each letter indicates one link (e.g. *HML* means link 1 is high, link 2 medium, link 3 low).

§Average expected z-statistic over three links.

all identified SVARs, the $\mathbf{A}_0$ matrix can be cast into a lower triangular form with the overidentifying restrictions indicated by zeroes below the main diagonal. 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.

The main question to be addressed is how does the effectiveness of the PC algorithm vary with the signal-to-noise ratios of the causal links? We begin by classifying 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 $\mathbf{A}_0$ that corresponds to the link (e.g. α_{410} for link 1):[11] $0 < z^* < 2$ 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. As 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 2, then *HMH* should have very similar results to *HHM* and *MHH*. We, therefore, record only the 10 non-redundant patterns.

The results for model 1 are reported in Table 1. The nominal size for the conditional correlation tests used in the PC algorithm is set at 10%. Each outcome is expressed as a proportion of the number of times it might have

[11]Expected values ($z^*_{\alpha_{ij0}}$) are determined using predicted values from the regression $z_{aij0} = \gamma\alpha_{ij0} + \varepsilon_{ij}$.

occurred. 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 ('skeleton correct') reports the proportion of links correctly identified as present even if their direction is reversed or unresolved.

The different combinations of signal strength are indicated in the first column of Table 1. The number in the second column is the mean population value of z^*-statistic for that combination. The data are ordered in ascending order of the proportion correct link identifications. The data are nearly, but not quite perfectly, lexicographically ordered by number of L, M, and H links. They are also nearly, but not quite perfectly, monotonically ordered by average expected z-statistic.

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 (contrary 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 is the same as the nominal size of the z-statistics and for HHH is half the nominal size.[12] 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 nearly lexicographically between these extremes.

The PC algorithm recovers the correct graph moderately well (improving with signal strength). It recovers the skeleton of the graph (ignoring the orientation of links) at a high rate, even when the signal strengths are low. The columns grouped as 'Link errors' report the various ways in which the PC algorithm fails to correctly identify a true link in model 1. 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

---

[12]In simulations not reported in the table, the same pattern is repeated for nominal test sizes of 5% and 20%: *LLL* shows commission rates near nominal test size; error rates fall as signal strength rises; *HHH* shows rates about half of nominal test size. What is more, rates of omission rise somewhat as nominal text size falls, confirming the usual tradeoff between size and power. These results are consistent with findings of well-behaved size in non-causal search algorithms (see Hendry and Krolzig, 1999; Hoover and Perez, 1999, 2002; 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).

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*) 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* and *HHL*).

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%. The total error rate for true links bottoms out at 29%. 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 of commission is very small. It almost always identifies the skeleton of the graph.

Model 2

The graph of model 2 is depicted in Figure 3. Corresponding to the graph is

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \alpha_{340} \\ \alpha_{410} & \alpha_{420} & 0 & 1 \end{bmatrix},$$

where $\alpha_{ij0} \neq 0$. Model 2 has the same skeleton as model 1. The $y_3 y_4$ link is reversed. Model 2 has only one unshielded collider: B on the path $y_1 y_4 y_2$. By rearranging the order of the variables in the vector $\mathbf{Y}_t$, we could recast $\mathbf{A}_0$ as a lower triangular matrix, emphasizing its econometric identifiability. In doing so, however, we would obscure the close relationship (only a single reversed link) between the graphical representations of models 1 and 2.

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). Table 2 reports rates of errors of commission similar to that for model 1: at a 10% nominal test size, the maximum number of commissions occurs at *LLM* (10%) and the minimum at *HHM* (5%).
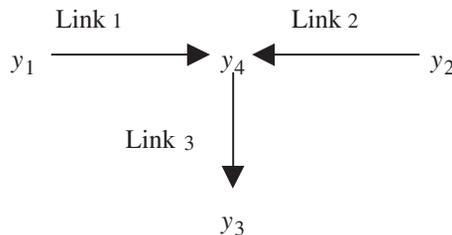
Figure 3. Model 2

TABLE 2

Model 2 simulation outcomes by signal strength

| Signal strength† | | Outcomes as fraction of possible outcomes§ | | | | | | | | | | |
| | | | | | Link errors for links 1 and 2 | | | | Link errors for link 3 | | | |
(1) By link‡	(2) Average§	(3) Correct	(4) Skeleton correct	(5) Committed	(6) Omitted	(7) Reversed	(8) Unresolved	(9) Total	(10) Omitted	(11) Reversed	(12) Unresolved	(13) Total
LLL	1.00	0.46	0.59	0.09	0.72	0.00	0.28	1.00	0.76	0.00	0.23	1.00
LLM	1.83	0.46	0.68	0.10	0.43	0.00	0.55	0.98	0.78	0.01	0.21	1.00
MLL	1.83	0.46	0.68	0.09	0.42	0.01	0.56	0.99	0.77	0.01	0.21	1.00
LLH	3.00	0.47	0.70	0.09	0.37	0.01	0.59	0.97	0.78	0.01	0.20	0.99
HLL	3.00	0.47	0.70	0.08	0.39	0.01	0.58	0.98	0.79	0.01	0.20	1.00
MML	2.67	0.47	0.79	0.08	0.44	0.01	0.51	0.97	0.12	0.04	0.83	0.99
HML	3.83	0.48	0.81	0.07	0.40	0.01	0.55	0.96	0.12	0.05	0.81	0.98
HHL	5.00	0.48	0.83	0.06	0.41	0.01	0.54	0.96	0.00	0.05	0.93	0.98
MLM	2.67	0.49	0.79	0.08	0.11	0.01	0.80	0.93	0.79	0.01	0.19	0.99
MMM	3.50	0.51	0.90	0.08	0.13	0.03	0.75	0.90	0.12	0.06	0.76	0.94
HLM	3.83	0.53	0.81	0.08	0.07	0.01	0.72	0.80	0.80	0.02	0.15	0.96
MLH	3.83	0.54	0.81	0.07	0.06	0.01	0.73	0.80	0.79	0.01	0.16	0.96
MMH	4.67	0.57	0.93	0.06	0.06	0.03	0.70	0.79	0.13	0.06	0.66	0.85
HMM	4.67	0.57	0.92	0.06	0.08	0.04	0.66	0.78	0.11	0.08	0.65	0.84
HHM	5.83	0.57	0.94	0.05	0.09	0.04	0.66	0.79	0.00	0.08	0.74	0.82
HLH	5.00	0.66	0.83	0.06	0.00	0.01	0.46	0.47	0.81	0.01	0.09	0.92
HMH	5.83	0.73	0.95	0.06	0.00	0.02	0.43	0.45	0.13	0.07	0.37	0.57
HHH	7.00	0.74	0.97	0.06	0.00	0.04	0.42	0.45	0.00	0.06	0.42	0.48

Notes:

†Signal strength measured as the expected *z*-statistic.

‡L = low ($0 \leq z^* < 2$), M = medium ($2 \leq z^* < 5$), H = high ($5 \leq z^* \leq 9$); each letter indicates one link (e.g. *HML* means link 1 is high, link 2 medium, link 3 low).

§Average expected *z*-statistic over three links.

Because of the asymmetry of model 2, Table 2 reports link errors for links 1 and 2 in one group and link 3 in a separate group. For links 1 and 2, there are three clearly defined sets. First, if link 3 or both links 1 and 2 have a low signal strength (see column 1), then total link errors for links 1 and 2 are nearly 100%, with reversals and omissions accounting for about half each. Secondly, 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%. Omissions account for about 10 percentage points of the total, and unresolved links for most of the rest. Thirdly, 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%. Omissions fall to almost zero, and unresolved links account for almost all of the total. Reversals remain very low for all combinations.

For link 3 the total error rate is high if any of the three links has a low signal. Omissions about 80% if either link 1 or 2 has low signal strength. If both links have medium or high strength, then omissions stand at 11–13%, 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%.

Table 2 shows that at any reasonable signal strength, the algorithm performs well at recovering the skeleton. Column 4 indicates its unqualified success both at recovering the skeleton and properly orienting the causal arrows. The difference between columns 3 and 4 is a measure of the total number of link errors over all links. 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, as model 2 has only one unshielded collider, whereas model 1 has three; it is the presence of unshielded colliders that makes orientation of links possible.

Model 3

Model 3 can be seen as an elaboration of model 1. The graph of model 3 is depicted in Figure 4. The core graph is the same as model 1. The link added between y_1 and y_2 acts as a shield, so that y_4 on path $y_1 y_4 y_2$ 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: y_4 on paths $y_1 y_4 y_3$ and $y_2 y_4 y_3$; and y_2 on $y_1 y_2 y_5$.

As link 6 can be reversed to run $y_3 \rightarrow y_6$ 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 (PC-true graph) is to recover a graph with
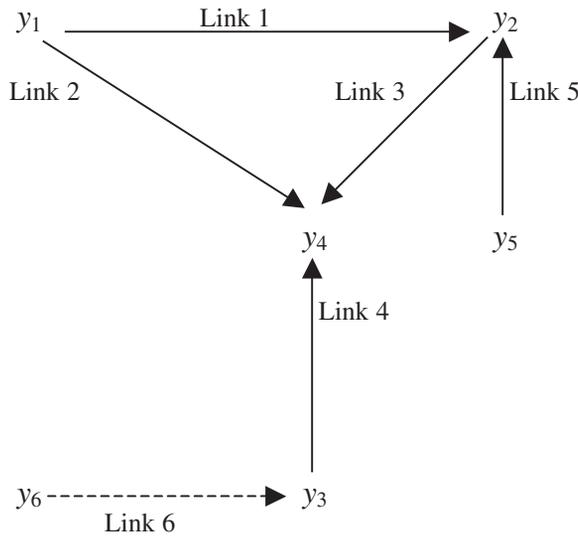
Figure 4. Model 3

link 6 undirected (indicated by a dashed arrow in Figure 4). The true graph is used to generate the simulated data, while the PC-true graph is used to score the outcomes.

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. As before, if we restrict ourselves to three levels of signal strength, 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 5 we simplify by reporting results for the average signal strengths across all six lengths.

Figure 5 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% 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
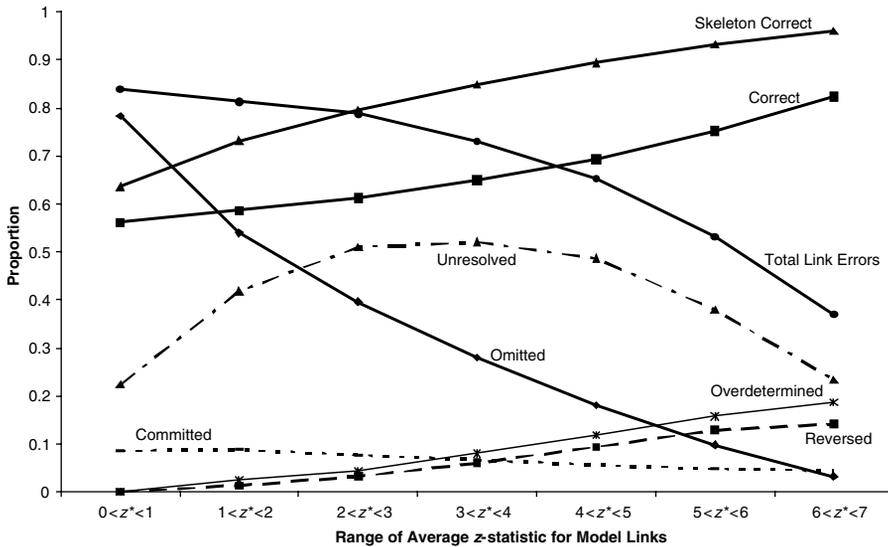
Figure 5. Outcomes by average signal strength for model 3

96%, while unqualified success at recovering the correct graph rises in parallel, toping out at just over 82%.

    In earlier work, we also investigated a model with a graph similar to that in Figure 4, except that link 4 was reversed. The results (available on request from the authors) are broadly similar to those reported for model 3.

IV. An application to US macro data

Using updated quarterly data previously investigated by King *et al.* (1991), Swanson and Granger (1997, pp. 362–363) apply the filtering methods described in section III to a four-variable VAR for the period January 1949 to February 1990 using eight lags. We now investigate a model using similar data for the period January 1949 to April 2002. The time series are the logarithms of real per capita consumption expenditures (C), real per capita gross private domestic fixed investment (I), per capita real balances (M), and real per capita private gross domestic product (Y). (Sources and transformations are described in Appendix B.)

    Because they restrict the class of admissible models to linear chains, Swanson and Granger need test only for first-order independence, using tests of correlation between two variables conditional on a third. They do not use the PC algorithm nor do they consider unshielded colliders. So they must appeal to extra-statistical information to orient the links between variables. Swanson and Granger identify the following graph: $M — C — I — Y$, which they orient as
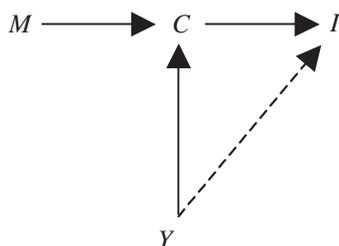
Figure 6. US macroeconomic data

$M \to C \to I \to Y$ on the grounds that at least one of M, C, I must causally precede Y in the current quarter. Using Swanson and Granger's methods, we are able to reproduce their results on the updated data both for their period and for the most recent available data (January 1949 to April 2000).

   For the most recent data, the PC algorithm identifies the graph in Figure 6. The link between Y and I is shown as dashed because the algorithm could not orient it. However, as the algorithm presupposes acylical graphs, only the orientation with the link $Y \to I$ is consistent with that assumption. Notice that Y is ordered recursively ahead of C and I and is independent of M, contradicting Swanson and Granger's identifying assumption that one of the other variables must be causally prior to Y in the current period. The example illustrates the advantages of the PC algorithm: first, that it admits a wider class of models than either linear chains or the frequently encountered Choleski orderings and, second, that it may at least partially orient the skeleton without appeal to debatable non-statistical considerations.

V.   Conclusions and directions for future work

We are now in a position to draw some preliminary conclusions from our simulation studies.

   (i) Graph-theoretic methods can be an effective data-based tool in selecting (or at least in narrowing the equivalence class of) the contemporaneous causal order of SVARs. In the systems examined, the PC algorithm was 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. 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.

  (ii) Contrary to the fears often expressed in relation to search methodol-
       ogies, the PC algorithm appears to have well behaved statistical
       properties. In the worst case, commission rates are approximately the
       nominal size of the test statistic used to assess conditional independ-
       ence and, in the best cases, at about half that rate. There is a clear
       tradeoff between increasing the rate of commission (type I error) and
       decreasing the rate of omission (type II error).
 (iii) Errors in causal ordering have multiple dimensions. 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.
 (iv) 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.

The last point suggests an important direction of future research. The
Monte Carlo studies here refer to specific models, and the degree to which
they apply generically is unclear. We are developing and validating a
bootstrap-type procedure to provide some measure of the reliability of causal
identifications when, as it real-world applications, the true data-generating
process is unknown.

Our study assumed a fixed sample size of 500 that, while reasonable
in some applications, is too long for others (especially ones using
quarterly data). We plan to extend the study to include a variety of realistic
sample sizes.

We also assumed that the true lag length of the VAR was known. But in
practice selecting the lag length is also a statistical problem – one that may
interact with selecting the causal ordering. We plan to investigate the joint
problem in future simulations.

Finally, we have restricted our SVAR to be stationary. We plan to also
plan to investigate the efficacy of the search algorithm in nonstationary
settings.

Final Manuscript Received: October 2003

References

Akleman, D. G., Bessler, D. A. and Burton, D. M. (1999). 'Modeling corn exports and
    exchange rates with directed graphs and statistical loss functions', in Glymour, C., Cooper,
    G. F. (eds), *Computation, Causation, and Discovery*, American Association for Artificial
    Intelligence, Menlo Park, CA and MIT Press, Cambridge, MA, pp. 497–520.

Board of Governors of the Federal Reserve System (1976). *Banking and Monetary Statistics: 1941–1970*, Board of Governors of the Federal Reserve System, Washington, DC.

Bessler, D. A. and Lee, S. (2002). 'Money and prices: US data 1869–1914 (a study with directed graphs)', *Empirical Economics*, Vol. 27, pp. 427–446.

Breitung, J. and Swanson, N. R. (2002). 'Temporal aggregation and spurious instantaneous causality in multiple time series models', *Journal of Time Series Analysis*, Vol. 23, pp. 651–656.

Cooley, T. F. and LeRoy, S. F. (1985). 'Atheoretical macroeconometrics: a critique', *Journal of Monetary Economics*, Vol. 16, pp. 283–308.

Cooper, G. F. (1999). 'An overview of the representation and discovery of causal relationships using Bayesian networks', in Glymour C. and Cooper G. F. (eds), *Computation, Causation, and Discovery*, American Association for Artificial Intelligence, Menlo Park, CA and MIT Press, Cambridge, MA, pp. 3–64.

Demiralp, S. (2000). 'The structure of monetary policy and transmission mechanism', PhD dissertation, Department of Economics, University of California, Davis, CA.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Hendry, D. F. and Krolzig, H.-M. (1999). 'Improving on "Data Mining Reconsidered" by K. D. Hoover and S. J. Perez', *Econometrics Journal*, Vol. 2, pp. 202–218.

Hoover, K. D. (1995). 'In defense of data mining: some preliminary thoughts', in Hoover K. D. and Sheffrin S. M. (eds), *Monetarism and the Methodology of Economics: Essays in Honour of Thomas Mayer*, Edward Elgar, Aldershot.

Hoover, K. D. (2001). *Causality in Macroeconomics*, Cambridge University Press, Cambridge.

Hoover, K. D. (2003). 'Nonstationary time series, cointegration, and the principle of the common cause'. *British Journal for the Philosophy of Science*, Vol. 54, pp. 527–551.

Hoover, K. D. and Perez, S. J. (1999). 'Data mining reconsidered: encompassing and the general-to-specific approach to specification search', *Econometrics Journal*, Vol. 2, pp. 167–191.

Hoover, K. D. and Perez, S. J. (2000). 'Three attitudes towards data-mining', *Journal of Economic Methodology*, Vol. 7, pp. 195–210.

Hoover, K. D. and Perez, S. J. (2002). *Truth and Robustness in Cross-country Growth Regressions*, Unpublished working paper, University of California, Davis, CA.

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

King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W. (1991). 'Stochastic trends and economic fluctuations', *American Economic Review*, Vol. 81, pp. 819–840.

Krolzig, H.-M. and Hendry, D. F. (2001). 'Computer automation of general-to-specific model selection procedures', *Journal of Economic Dynamics and Control*, Vol. 25, pp. 831–866.

Lauritzen, S. L. and Richardson, T. S. (2002). 'Chain graphs and their causal interpretations', *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 64, pp. 321–361.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). 'Local computations with probabilities on graphical structures and their applications to expert systems' [with discussion], *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 50, pp. 157–224.

Leamer, E. (1985). 'Vector autoregressions for causal inference?' in Brunner K. and Meltzer A. H. (eds), *Understanding Monetary Regimes*, Carnegie-Rochester Conference Series on Public Policy, Vol. 22, North Holland, Amsterdam, 225–304.

LeRoy, S. F. (2002). 'Causality: models, reasoning, and inference: a review of Judea Pearl's *Causality*', *Journal of Economic Methodology*, Vol. 9, pp. 100–103.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge.

Reale, M. and Tunnicliffe Wilson, G. (2000). *Identification of Vector AR Models with Recursive Structural Errors Using Conditional Independence Graphs*, Unpublished typescript, Department of Mathematics and Statistics, University of Canterbury (New Zealand) and Department of Mathematics and Statistics, Lancaster University (UK).

Reale, M. and Tunnicliffe Wilson, G. (2002). 'The sampling properties of conditional independence graphs for structural vector autoregressions', *Biometrica*, Vol. 89, pp. 457–461.

Reichenbach, H. (1956). *The Direction of Time*, University of California Press, Berkeley and Los Angeles.

Tunnicliffe Wilson, G., Reale, M. and Morton, A. S. (2001). *Developments in Multivariate Time Series Modeling*, Unpublished typescript, Department of Mathematics and Statistics, Lancaster University (UK) and Department of Mathematics and Statistics, University of Canterbury (New Zealand).

Tunnicliffe Wilson, G. and Reale, M. (2002). *Causal Diagrams for I(1) Structural VAR Models*, Unpublished typescript, Department of Mathematics and Statistics, University of Canterbury (New Zealand) and Department of Mathematics and Statistics, Lancaster University (UK).

Sheffrin, S. M. and Triest, R. K. (1998). *A New Approach to Causality and Economic Growth*, Unpublished typescript, University of California, Davis, CA.

Sims, C. A. (1980). 'Macroeconomics and reality', *Econometrica*, Vol. 48, pp. 1–48.

Sims, C. A. (1986). 'Are forecasting models usable for policy analysis?' *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 10, pp. 2–15.

Spirtes, P., Scheines, R., Meek, C., Richardson, T., Glymour, C., Hoijtink, H. and Boomsma, A. (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.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd edn, MIT Press, Cambridge, MA.

Swanson, N. R. and Granger, C. W. J. (1997). 'Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions', *Journal of the American Statistical Association*, Vol. 92, pp. 357–367.

Appendix A: 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 nth-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: If $X \rightarrow Z$ and $Z — Y$ and X and Y are not directly connected, then orient $Z — Y$ as $Z \rightarrow Y$.[13]

Appendix B: The data

All the raw series, except M1, are from the US National Income and Product Accounts and were downloaded from the Haver Analytics *United States Economic Statistics* database. Except where noted, they are seasonally adjusted, stated in billions of constant 1996 dollars, and cover January 1947 to April 2002. Haver codes are in bold type. Personal Consumption Expenditure = **CH**; Gross Private Domestic Fixed Investment = **FH**; M1 monetary aggregate (*MN*) for January 1947 to April 1958 = M1 monetary aggregate from Board of Governors of the Federal Reserve System (1976), Table 1.1, pp. 17–18, column 2 ('Money Stock: Total') × 0.97966 and for January 1959 to April 2002 = **FM1** (data are billions of current dollars; quarterly values are averages of monthly values); Government Consumption Expenditure and Gross Investment (current dollars) = **G**; Government Consumption Expenditure and Gross Investment (constant dollars) = **GH**; Gross National Product (current dollars) = **GNP**; Gross National Product (constant dollars) = **GNPH**; Civilian Noninstitutional Population over 16 Years Old = **LNN**.

Data used in paper are constructed as follows: $C = \log(\textbf{CH}/\textbf{LNN})$; $I = \log(\textbf{FH}/\textbf{LNN})$; $M = \log[\textbf{MN}/(P \times \textbf{LNN})]$, where $P = (\textbf{GNP} - \textbf{G})/(\textbf{GNPH} - \textbf{GH})$; $Y = \log[(\textbf{GNPH} - \textbf{GH})/\textbf{LNN}]$.

---

[13]Cooper (1999), Pearl (2000) and Spirtes *et al.* (2000) all describes a final step to the algorithm that is omitted in Tetrad 3 as well as in our simulations: If there is a directed path from X to Y and if there is an undirected edge between X and Y, orient $X — Y$ as $X \rightarrow Y$. We appeal to this step in orienting the link between I and Y in section IV.