Hoover, Kevin D.; Demiralp, Selva; Perez, Stephen J.

Working Paper

A bootstrap method for identifying and evaluating a structural vector autoregression

Working Paper, No. 06-14

Provided in Cooperation with:
University of California Davis, Department of Economics

Suggested Citation: Hoover, Kevin D.; Demiralp, Selva; Perez, Stephen J. (2006) : A bootstrap method for identifying and evaluating a structural vector autoregression, Working Paper, No. 06-14, University of California, Department of Economics, Davis, CA

This Version is available at:
https://hdl.handle.net/10419/31328

Department of Economics
Working Paper Series

A Bootstrap Method for Identifying and Evaluating a Structural Vector Autoregression

Kevin Hoover
University of California, Davis

Selva Demiralp
Koç University

Stephen J. Perez
California State University, Sacramento

March 28, 2006

Paper # 06-14

Graph-theoretic methods of causal search based in the ideas of Pearl (2000), Spirtes, Glymour, and Scheines (2000), and others have been applied by a number of researchers to economic data, particularly by Swanson and Granger (1997) to the problem of finding a data-based contemporaneous causal order for the structural autoregression (SVAR), rather than, as is typically done, assuming a weakly justified Choleski order. Demiralp and Hoover (2003) provided Monte Carlo evidence that such methods were effective, provided that signal strengths were sufficiently high. Unfortunately, in applications to actual data, such Monte Carlo simulations are of limited value, since the causal structure of the true data-generating process is necessarily unknown. In this paper, we present a bootstrap procedure that can be applied to actual data (i.e., without knowledge of the true causal structure). We show with an applied example and a simulation study that the procedure is an effective tool for assessing our confidence in causal orders identified by graph-theoretic search procedures.

UCDAVIS

Department of Economics
One Shields Avenue
Davis, CA 95616
(530)752-0741

http://www.econ.ucdavis.edu/working_search.cfm

A Bootstrap Method for Identifying and Evaluating a Structural Vector Autoregression*

Selva Demiralp[†]

Kevin D. Hoover[‡]

Stephen J. Perez[††]

Revised, 28 March 2006

† Department of Economics, Koç University, Sariyer, Istanbul, 34450 Turkey. Tel. +212-338 1842. E-mail sdemiralp@ku.edu.tr

‡Department of Economics, University of California, One Shields Avenue, Davis, California 95616-8578, U.S.A. Tel. +530-752-2129. E-mail kdhoover@ucdavis.edu Web page www.econ.ucdavis.edu/faculty/kdhoover/

††Department of Economics, California State University, Sacramento, 6000 J. Street, Sacramento, California 95819-6082, U.S.A. Tel. +916-278-6919. E-mail sjperez@csus.edu Web page http://www.csus.edu/indiv/p/perezs/

*We thank Katarina Juselius, Søren Johansen, David Hendry, and the participants in seminars at the University of Copenhagen and Oxford University for comments on an earlier draft.

Abstract

of

A Bootstrap Method for Identifying and Evaluating a Structural Vector Autoregression

Graph-theoretic methods of causal search based in the ideas of Pearl (2000), Spirtes, Glymour, and Scheines (2000), and others have been applied by a number of researchers to economic data, particularly by Swanson and Granger (1997) to the problem of finding a data-based contemporaneous causal order for the structural autoregression (SVAR), rather than, as is typically done, assuming a weakly justified Choleski order. Demiralp and Hoover (2003) provided Monte Carlo evidence that such methods were effective, provided that signal strengths were sufficiently high. Unfortunately, in applications to actual data, such Monte Carlo simulations are of limited value, since the causal structure of the true data-generating process is necessarily unknown. In this paper, we present a bootstrap procedure that can be applied to actual data (i.e., without knowledge of the true causal structure). We show with an applied example and a simulation study that the procedure is an effective tool for assessing our confidence in causal orders identified by graph-theoretic search procedures.

Keywords: vector autoregression (VAR), structural vector autoregression (SVAR), causality, causal order, Choleski order, causal search algorithms, graph-theoretic methods

JEL Codes: C30, C32, C51

A Bootstrap Method for Identifying and Evaluating a Structural Vector Autoregression

1. Identification of the SVAR

The *structural vector autoregression* (*SVAR*) has become a workhorse of empirical macroeconomics. The main hurdle to using the SVAR is to identify the system so that unique random shocks can be associated with particular variables. While other identification schemes are available (e.g., Blanchard and Quah (1989) identify the SVAR from assumptions about long-run properties), the most common scheme – and the only one that we address here – achieves identification through imposing zero restrictions on the matrix of contemporaneous correlations among the variables. It is widely believed that identification depends on *a priori* assumptions. On the contrary, starting with Swanson and Granger (1997), a number of authors have demonstrated how *graph-theoretic* techniques can, in some cases, use the data themselves to identify the SVAR. These methods exploit patterns of conditional independence in the data. In cases in which unique identification is not possible, they may nonetheless reduce the class of admissible identifying assumptions considerably.[1] Demiralp and Hoover (2003) provide Monte Carlo evidence that, provided signal-to-noise ratios are high enough, these graph-theoretic methods can recover effectively the true contemporaneous structure of SVARs. Monte Carlo results are too often specific to particular simulations and do not necessarily provide generic guidance. To remedy this problem, we develop a bootstrap method that

[1] Graph-theoretic methods are relatively new in economics, although increasingly used in other fields from sociology and psychology to medicine and biology. Hoover (2001, ch. 7) provides a critical description (see also Hoover 2003, 2005). LeRoy (2002) has recently discussed them in a review of Pearl (2000). There have been relatively few applications of graph-theoretic search algorithms to economic data. Some examples include Sheffrin and Triest (1998), Akleman, Bessler, and Burton (1999), Bessler and Loper (2001), Yang and Bessler (2004), Haigh, Nomikos, and Bessler (2004). Swanson and Granger (1997) and Demiralp (2000, ch. 4), Bessler and Lee (2002), Yang and Bessler (2003), Haigh, Nomikos, and Bessler (2004), Moneta (2003), and Awokuse (2005) are particularly concerned with the causal order of the VAR.

allows one to assess the reliability of a data-based identification scheme for arbitrary structures. We show that this technique allows us, in cases in which we do not know the true underlying structure, to present simulation evidence that closely mimics Monte Carlo simulations for which we know the underlying structure *ex hypothesi*. It therefore allows us to give reasonable assessments of the reliability of data-based, graph-theoretic identifications of SVARs.

The SVAR can be written as:

$$\mathbf{A}_0\mathbf{Y}_t = \mathbf{A}(L)\mathbf{Y}_{t-1} + \mathbf{E}_t\,, \tag{1}$$

where $\mathbf{Y}_t$ is an $N \times 1$ vector of contemporaneous variables, $\mathbf{A}_0$ is an $N \times N$ matrix with ones on the main diagonal and possibly non-zero off-diagonal elements; $\mathbf{A}(L)$ is a polynomial in the lag operator, L; and $\mathbf{E}_t$ is an $N \times 1$ vector of error terms with $\mathbf{E} = [\mathbf{E}_t]$, $t = 1, 2, \ldots, T$ and the covariance matrix $\mathbf{\Sigma} = \mathrm{E}(\mathbf{E}\mathbf{E}')$ diagonal. The individual error terms (shocks) can be assigned unequivocally to particular equations because $\mathbf{\Sigma}$ is diagonal. The matrix $\mathbf{A}_0$ defines the causal interrelationships among the contemporaneous variables.

Premultiplying equation (1) by $\mathbf{A}_0^{-1}$ yields the reduced-form or *vector autoregression* (*VAR*):

$$\mathbf{Y}_t = \mathbf{A}_0^{-1}\mathbf{A}(L)\mathbf{Y}_{t-1} + \mathbf{A}_0^{-1}\mathbf{E}_t = \mathbf{B}(L)\mathbf{Y}_{t-1} + \mathbf{U}_t\,, \tag{2}$$

with $\mathbf{U} = [\mathbf{U}_t]$, $t = 1, 2, \ldots, T$. While equation (2) is easily estimated, the covariance matrix, $\mathbf{\Lambda} = E(\mathbf{U}\mathbf{U}')$ in general will not be diagonal, so that it will be impossible to evaluate the effects of shocks to particular variables. The identification problem reduces to this: if we know $\mathbf{A}_0$, then it is easy to recover equation (1) from our estimates of (2); but how do we know $\mathbf{A}_0$?

Identification schemes typically start with the property that $\mathbf{\Sigma}$, the covariance matrix of $\mathbf{E}_t$, is diagonal. True identification would permit us to transform equation (2) into (1) and recover the diagonal $\mathbf{\Sigma}$. There are a large number of $N \times N$ matrices, $\mathbf{P}_i$ such that the covariance matrix $\mathbf{\Omega} = E(\mathbf{P}_i^{-1}\mathbf{U}(\mathbf{P}_i^{-1}\mathbf{U})')$ is diagonal. Let $\mathbf{P} = \{\mathbf{P}_i\}$ be the set of all such orthogonalizing transformations. Each element of $\mathbf{P}$ can be thought of as a potential candidate for $\mathbf{A}_0$. Most commonly economists have based their choices of $\mathbf{P}_i$ on highly informal arguments appealing to general plausibility or weak theoretical considerations. Typically, although not uniformly (see, for instance, Bernanke and Mihov 1998), they have restricted themselves to just-identified, lower-triangular matrices. These are based on the *Choleski decompositions* of the covariance matrix of $\mathbf{U}$. There are, in general, $n!$ such matrices, each corresponding to one *Wold-* or *recursive* causal ordering of the variables in $\mathbf{Y}$.

We too restrict ourselves to recursive orderings, but widen the class of possible matrices to include the over-identified – that is, to matrices that may have zeroes among the off-diagonal elements of $\mathbf{P}_i$. Graph-theoretic search algorithms work according to the following general plan: the true $\mathbf{A}_0$ induces a set of conditional independence relations among the elements of $\mathbf{U}_t$. The algorithm thoroughly tests for conditional independence

relations among the estimated $\hat{\mathbf{U}}_t$. It then selects the class of $\mathbf{P}_i$ – perhaps unique – that is consistent with those independence relations.

The Monte Carlo simulations of Demiralp and Hoover (2003) generate data from a variety of *known* specifications of SVARs like equation (1) and then address the question of how successfully $\mathbf{A}_0$ can be recovered from estimates of VARs like equation (2). The problem for empirical analysts is to evaluate the reliability of such identifications when $\mathbf{A}_0$ and, indeed, the entire specification of the SVAR is *unknown*. We employ a bootstrap strategy. Starting with the original data, we estimate the VAR (equation (2)) and retain the residuals $\hat{\mathbf{U}} = [\hat{\mathbf{U}}_t]$, $t = 1, 2, \ldots, T$. In order to maintain the contemporaneous correlations among the variables, we resample the residuals *by columns* from $\hat{\mathbf{U}}$. The resampled residuals are used in conjunction with the coefficient estimates of equation (2) to generate simulated data. A large number of simulated data sets are created. For each one, we run the search algorithm, record the results, and compute summary statistics.

Graph-theoretic, causal search algorithms are increasingly used in biological sciences, physical sciences, and social sciences, other than economics. And while there are by now a number of applications within economics, the ideas behind graph-theoretic search algorithms are not well known among economists. Consequently, we begin with a review of the basic ideas behind the algorithms. Both the search algorithm and the bootstrap procedure that we propose to evaluate our confidence in the outcome of the search are best understood in a concrete example. We, therefore, illustrate both with the same data set used in Swanson and Granger (1997). The central question of this study is,

how reliable is this bootstrap procedure? We conclude with a simulation study that addresses that question.

2. The Causal Search Algorithm

To keep things simple, assume that the variables are a cross-section with an interesting causal structure, but no time-series structure. The most fundamental notion is that causally related variables are connected by *edges*. On the one hand, if the direction of causal influence is ambiguous the edge is shown as a line or *undirected edge*. On the other hand, if it is determinate, then an arrow head indicates the causal direction, and the edge is referred to as *directed.* Sprites, Glymour, Scheines (2000) and Pearl (2000) show how the mathematics of graph theory relates a graphical representation of causal relations to implications for the probability distributions of the variables.

Suppose that $A \rightarrow B \rightarrow C$ (that is, *A* causes *B* causes *C*). *A* and *C* would be probabilistically dependent; but, conditional on *B*, they would be independent.[2] Similarly for $A \leftarrow B \leftarrow C$. In each case, *B* is said to *screen A* from *C*. Suppose that $A \leftarrow B \rightarrow C$. Then, once again *A* and *C* would be dependent; but, conditional on *B*, they would be independent. *B* is said to be the *common cause* of *A* and *C*. Now suppose that *A* and *B* are independent conditional on sets of variables that exclude *C* and its descendants (i.e., variables that have *C* as a direct or indirect cause, and none of the variables that cause *A* or *B* directly causes *C*. Then, if *A*, *B*, and *C* are related as $A \rightarrow C \leftarrow B$, then, conditional on *C*, *A* and *B* are dependent. *C* is called an *unshielded collider* on the path *ACB*. (A *shielded* collider would have a direct edge between *A* and *B* and would, therefore, be dependent whether or not one conditioned on *C*.)

[2] This and the next five paragraphs are based closely on Hoover (2005), pp. 71-74.

Causal search algorithms use a statistical measure of independence, commonly a measure of conditional correlation, to check systematically the patterns of conditional independence and dependence and to work backwards to the class of admissible causal structures.[3] We employ the SGS algorithm of Spirtes *et al.* (2000, pp. 82-83). The SGS algorithm is a close relative of the PC algorithm (Spirtes *et al.* 2000, pp. 84-85). Demiralp and Hoover (2003) investigated the PC algorithm using Monte Carlo methods. The SGS algorithm assumes that graphs are *acylical* (what VAR analysts typically refer to as "recursive") – that is, there are no loops in causal chains such that an effect feeds back onto a direct or indirect cause. Acyclicality rules out simultaneous equations. The SGS algorithm proceeds in three stages:

Stage 1. Elimination of Edges

1. Start with a graph in which each variable is assumed to be connected by an undirected causal edge.
2. Test for the unconditional correlation of each pair of variables, eliminating the edge in the graph whenever the absence of correlation cannot be rejected.
3. Test for the correlation of each pair of variables conditional on a third variable, again eliminating the edge if correlation is absent. Continue testing pairs conditional on pairs, triples, quadruples, and so on until the graph is pared down as far as the data permit.[4]

[3] Absence of conditional correlation is a necessary, but not sufficient, condition for statistical independence.

[4] This step distinguishes the SGS algorithm (Spirtes *et al.* 2000, pp. 82-83) from the more commonly used PC algorithm (Spirtes *et al*. 2000, pp. 84-85, Pearl 2000, pp. 49-51, Cooper 1999, p. 45, figure 22, Demiralp and Hoover 2003, pp. 766-767). The PC algorithm tests independence between two variables that had been connected by an edge in the previous round conditional only on variables that were themselves adjacent to one of the two variables in the previous round. The SGS algorithm tests the pair of

Stage 2. Statistical Orientation of Edges

4. For each conditionally uncorrelated pair of variables (i.e., ones without a direct edge) that are connected through a third variable, test whether they become correlated conditional on that third variable. If so, the third variable is an unshielded collider. Orient the edges as pointing into the unshielded collider.

Stage 3. Logical Orientation of Edges

5. If there are any pairs *A* and *C* that are not directly connected but are indirectly connected $A \rightarrow B$ — *C*, then orient the second edge toward *C*, so that the triple is $A \rightarrow B \rightarrow C$.
6. If there is a pair of variables, *A* and *B* connected both by an undirected edge and a directed path, starting at *A*, through one or more other variables to *B* (i.e., a path in which the arrows all orient in a chain), then orient the undirected edge as $A \rightarrow B$.

Steps 1-4 are statistical; the next two steps are logical. Step 5 follows logically, because orienting the undirected edge in the other direction would turn the pattern into an unshielded collider, which would have already been identified in Step 4. Step 6 follows because orienting the undirected edge in the other direction would, contrary to assumption, render the graph cyclical.

---

variables against every set of variables of the right number for that round, whether or not they were connected to the test pair in the previous round. The SGS algorithm may detect conditional independence better in the presence of nonlinearities, but its computational complexity rises exponentially even when the true underlying graph is sparsely connected. Comparison of our results in section 4, especially Figures 6 and 7, with those of Hoover and Demiralp (2003) confirm that the two algorithms generate similar results for small numbers of variables.

The SGS algorithm can be illustrated with the example in Figure 1. Panel A shows the graph of the data-generating process. It determines just what the tests should find, small-sample problems to one side. The graph corresponds to a particular matrix $\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ a_{YW} & a_{YX} & 1 & 0 \\ a_{ZW} & 0 & a_{ZY} & 1 \end{bmatrix}$, where the variables are ordered *WXYZ*, the rows correspond to effects and the columns to causes, and the a_{ij} to the non-zero elements.

Step 1 starts with panel B in which each variable is connected to every other variable by an undirected edge. Step 2 eliminates edge 1, because *W* and *X* are unconditionally uncorrelated in the true graph. Step 3 eliminates edge 5 (*X* and *Z* are uncorrelated conditional on *Y*). Step 4 orients edges 2 and 3 toward *Y* (*W* and *X* are correlated conditional on *Y* – i.e., *Y* is an unshielded collider on *WYX*). Step 5 orients edge 6 towards *Z*. Step 6 orients edge 4 toward *Z*. The algorithm is able to recover the true graph.

Not every true graph can be recovered uniquely. A graph and a probability distribution are *faithful* when the independence relationships in the graph stand in one-to-one correspondence with those implied by the probability distribution. The *skeleton* of a graph is the pattern of its causal linkages ignoring their direction. The *observational equivalence theorem* (Pearl 2000, p. 19, Theorem 1.2.8) states that any probability distribution that can be faithfully represented by an acyclical graph, can equally well be represented by another acyclical graph with the same skeleton and the same unshielded colliders. A graph identical to panel A of Figure 1 except that edge 6 was reversed would not be observationally equivalent to panel A because it would add an unshielded collider

(*Y* on *XYZ*). A graph that reversed edge 4 would be observationally equivalent to the graph in panel A because it would have the same skeleton and neither add nor subtract unshielded colliders. While the graph with edge 4 reversed illustrates the observational equivalence theorem, Step 6 of the algorithm rules it out, since it possesses a cycle ($W \rightarrow Y \rightarrow Z \rightarrow W$), which violates the antecedent of the observational equivalence theorem.

3. An Application of the Bootstrap Procedure

Swanson and Granger's (1997) essential contribution was to realize that identification of the SVAR depends only on the contemporaneous correlations among variables and that the VAR itself filters out the lagged dynamics so that the residuals of a well specified VAR (the $\hat{\mathbf{U}}_t$ in an estimated version of equation (2)) capture the relevant contemporaneous characteristics of the original variables. Each element of $\hat{\mathbf{U}}_t$ can be thought of as a filtered variable purged of its time-series properties. A graph-theoretic search algorithm can then be applied to attempt to recover the causal order among these filtered variables.

In one case, Swanson and Granger (1997, pp. 362-363) applied this filtering method and a causal search algorithm to updated quarterly data previously investigated by King *et al.* (1991). Swanson and Granger's VAR used 8 lags of 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*) for the period 1949:1-1990:2. Swanson and Granger's algorithm restricted attention to linear chains in which each variable was allowed at most one direct cause

(e.g., $A \rightarrow B \rightarrow C \rightarrow D$ would be admissible, but $A \rightarrow B \leftarrow C \rightarrow D$ would not). With this restriction their algorithm considered only measures of first-order (i.e., unconditional) independence, which they implemented using tests of correlation between two variables conditional on a third, ignoring unshielded colliders. They identified the skeleton of the graph as $M — C — I — Y$. Appealing to the extra-statistical assumption that at least one of *M*, *C*, or *I* should cause *Y* in the current period, they oriented the edges as $M \rightarrow C \rightarrow I \rightarrow Y$. Demiralp and Hoover (2003, pp. 762-763) applied the PC algorithm to similar data for the period 1949:1-2002:4 and identified the structure shown as the heavy black lines in Figure 2. (Sources and construction of these data are described in Appendix A.)

But what confidence should anyone place in the identification selected by the algorithm? We use the bootstrap procedure to evaluate the selected identification. Appendix B reports the details of the bootstrap procedure. Here we provide only a sketch.

We start with the VAR that is used to identify the causal order. A *bootstrap realization* is created by using the estimated coefficients from this VAR as well as its matrix of residuals ($\hat{\mathbf{U}}$) resampled by columns with replacement,. For each bootstrap realization, a new VAR is estimated and the matrix of its residuals ($\tilde{\mathbf{U}}$) is retained as the time series of the filtered variables. These are then fed into the SGS search algorithm and the results are recorded. We construct a large number of bootstrap realizations and record the outcomes of the search.

The bootstrap results are displayed in two ways. Table 1 gives the complete distribution of outcomes for each possible edge. The column headed *No Edge* displays the percentage of bootstrap realization for which an edge was not found. The other

columns each refer to possible orientations of the edge and report each as a percentage of the total number of realizations in which some edge was found.

An undirected edge (—) implies that the algorithm cannot select a unique causal direction, so that *ceteris paribus* the selected graph identifies an equivalence class with one member with the edge directed one way and the other with the edge reversed.

A bidirectional edge (↔) appears to violate the assumption of acylicality. The algorithm sometimes may select a bidirectional edge nonetheless, because it orients edges based on triples of variables, rather than on the whole graph. Inconsistencies can arise for three reasons. First, they may be artifacts of small samples. Second and more interestingly, practitioners of graph-theoretic methods have seen them as evidence of omitted latent variables in otherwise acyclical graphs (see Scheines et al. 1994, pp. 90). Third, economists naturally may interpret them as evidence of simultaneous causality. This is an especially attractive option if we are willing to maintain the assumption of *causal sufficiency* – that is, that there are no important omitted variables.

The results give strong support for the skeleton identified in Figure 2. That skeleton omits two edges and the bootstrap typically omits those edges as well: between *I* and *M* in 94 percent of the realizations and between *M* and *Y* in 100 percent of the realizations. The highest number of omissions for any of the edges in the selected skeleton is 16 percent, and two edges are never omitted.

The data are repackaged in Figure 2 to provide easy-to-grasp summary statistics. Each edge is associated with a triple *Exists*/*Directed*/*Net Direction.* The first number, *Exists*, is the percentage of realizations in which an edge is found (i.e., it is the complement of the *No Edge* value). The second, *Directed*, indicates the percentage of

existing edges for which there is a definite direction – that is, the sum of the three columns headed "→" and "←" and "↔" divided by (1–*No Edge*). Finally, *Net Direction* indicates the difference between unidirectional edges going from earlier to later ordered variables and those going from later to earlier as a percentage of the directed edges – that is, the difference between the columns headed "→" and "←" divided by the sum of the same two columns. The order of variables is that in the left-hand column of Table 1, shown in Figure 2 as the clockwise arrangement of the variables starting from the upper left corner (*C*). A negative number, therefore indicates an edge directed from a higher to a lower ordered variable – for example, from *Y* to *I*.

To illustrate, consider the edge between *C* and *M*, which the SGS algorithm identified it as $C \leftarrow M$ (we adopt the convention of always writing the lower ordered variable on the left). Its statistics, 87/75/-100, indicate that it is selected by the bootstrap procedure in 87 percent of the realizations, found to have a definite direction in 75 percent of the realizations in which it is selected, and is 100 percentage points more frequently oriented $C \leftarrow M$ than $C \rightarrow M$.

Overall, the bootstrap provides support for the orientation of the edges, as well as for the skeleton. Both the *Directed* and *Net Direction* statistics correspond to the orientation of the identified edges. Two edges require further comment. First, the edge between *I* and *Y* is directed in the bootstrap only 40 percent of the times that it is selected, although when it is directed it is oriented as in the identified graph ($I \leftarrow Y$) by 95 points more than the reverse. While this orientation would appear to be only weakly supported, notice that reversing the orientation, holding all other edges fixed, would violate the maintained hypothesis that the graph is acyclical by introducing a loop: $I \rightarrow Y \rightarrow C \rightarrow I$.

Second, the search algorithm directs the edge between *C* and *I* in 67 percent of the cases in which it is selected, but with only a 19 point advantage for the selected direction ($C \rightarrow I$) over the reverse. This is confirmed in Table 1 which shows a fairly even division between the edge being undirected and it taking each of the unidirectional orientations. However, given the high level of confidence in the orientation of $C \leftarrow Y$ and $C \leftarrow M$ edges, directing the edge as $C \leftarrow I$ would not be attractive. It would imply that *C* was an unshielded collider that had not been identified. *C* is found to be an unshielded collider in only 20 percent of the realizations. Thus, despite the two edges the directions of which are individually more weakly identified edges, the bootstrap procedure, viewed as a whole, provides moderately strong support for the graph identified in Figure 2.

Figure 3 demonstrates why it matters which order is selected. In the four-variable system, there are sixteen impulse-response functions. We display three that illustrate the range of variation. Each panel displays one of these impulse-response functions for each of the three identification schemes:

1. A Choleski ordering *M*, *C*, *I*, *Y*, corresponding to $\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{CM} & 1 & 0 & 0 \\ a_{IM} & a_{IC} & 1 & 0 \\ a_{YM} & a_{YC} & a_{YI} & 1 \end{bmatrix}$;

2. Swanson and Granger's (1997) causal chain $M \rightarrow C \rightarrow I \rightarrow Y$, corresponding to

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{CM} & 1 & 0 & 0 \\ 0 & a_{IC} & 1 & 0 \\ 0 & 0 & a_{YI} & 1 \end{bmatrix};$$

3. the order selected by the SGS algorithm (Figure 2), corresponding to

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{CM} & 1 & 0 & a_{CY} \\ 0 & a_{IC} & 1 & a_{IY} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In each panel, the impulse-response functions for the Choleski ordering and Swanson and Granger's causal chain are similar, while those for the SGS-selected order are quite distinct. In panel A, the impulse-response function of *M* for a shock to C for the SGS-selected order is positive at all horizons and globally increasing, while those for the other two orders are negative. While the impulse-responses in the other two panels do not display such a qualitative contrast, they are nonetheless quantitatively distinct. All three impulse-responses of *Y* to *M* in panel *B* show a similar pattern, but the impulse response corresponding to the order chosen by the SGS algorithm lies well below the other two, especially for the first half of the forecast period. The quantitative gap is even larger and shows less tendency to converge over time with the impulse responses of *Y* to a shock to *Y*. And where the impulse responses for both Swanson and Granger's and the Choleski order start positive and then turn negative, the impulse response for the order chosen by the SGS algorithm never becomes negative at all.

The close similarity of impulse-response functions for the Choleski order and Swanson and Granger's causal chain is not too surprising as the causal chain (in contrast to the SGS-selected order) is an overidentifying restriction on that particular Choleski order. The SGS-selected order is, however, an overidentifying restriction of two other Choleski orders: *M, Y, C, I* and *Y, M, C, I*. The impulse-response functions for these two Choleski orders are in fact very close to those for the SGS-selected order. This similarity

points to the usefulness of causal search algorithms in choosing among just-identified causal orders that could not be distinguished on, for example, a likelihood criterion alone.

4. How Well Does the Bootstrap Indicate the Uncertainty of Causal Search?

Now that we have seen how to apply the bootstrap procedure, we need to ask whether its guidance is reliable. We know from Monte Carlo simulations that the performance of the causal search algorithm depends on the signal-to-noise ratios of the causal linkages in the true data-generating process. Monte Carlo simulations are based on the specification of an artificial "true" data-generating process. In practice, of course, we do not know the true data-generating process. Our assessment of reliability is guided by the following idea: the bootstrap procedure is reliable if it gives us guidance that would mimic that of a Monte Carlo simulation, if we somehow had access to the truth.

We begin our assessment of the bootstrap procedure with a *known SVAR*. We evaluate the performance of SGS search algorithm in a Monte Carlo simulation. This simulation then becomes the standard against which we judge the success of the bootstrap procedure. Since the bootstrap is a solution to an essentially statistical problem, we evaluate its success with respect to the four statistical steps of the SGS algorithm – namely, how well it does at identifying the skeleton and unshielded colliders of the graph of the true SVAR. The bootstrap procedure would work ideally if it reproduced the results of the Monte Carlo simulation.

Appendix C reports the details of the design of the simulation experiments. Here we provide only a sketch.

1. We start with two graphs and the $\mathbf{A}_0$ matrix of their associated SVARs. Model 1 (Figure 4) represents three unshielded colliders: y_4 on three different paths –

$y_1\, y_4\, y_2$, $y_1\, y_4\, y_3$, and $y_2\, y_4\, y_3$. Model 2 (Figure 5) is an elaboration of Model 1: the edge added between y_1 and y_2 acts as a shield, so that y_4 on path $y_1y_4y_2$ is no longer an unshielded collider. The additional edge 5 adds another unshielded collider, while the additional edge 6 does not. Model 2, therefore, also has three unshielded colliders: y_4 on paths $y_1y_4y_3$ and $y_2y_4y_3$; and y_2 on $y_1y_2y_5$. And since edge 6 can be reversed without changing the skeleton or the unshielded colliders, the best that the SGS algorithm could do is to find an undirected edge between y_3 and y_6.

We also consider two further models. Model 3 (not shown) is identical to Model 1 except that it reverses edge 3. This reduces the number of unshielded colliders to just one: y_4 on the path $y_1y_4y_3$. Similarly, Model 4 (not shown) is identical to Model 2 except that it reverses edge 4. There are now only two unshielded colliders: y_2 on $y_1y_2y_5$ and y_3 on $y_4y_3y_6$. The algorithm cannot identify the direction of edge 2.

2. The SVARs corresponding to Models 1-4 are used to create simulated sets of *pseudo-real-world data*. Each realization constructs a time-series of 500 observations based on the assignment of a different set of random values to the non-zero off-diagonal coefficients (the α_{ij}) of the $\mathbf{A}_0$ matrix. The coefficients are drawn uniformly from an interval that ensures that the *ex ante t*-statistics, which provide a measure of the signal-to-noise ratio, are bound between 0 and about 10.
3. We then estimate a VAR like equation (2) for each realization of the pseudo-real-world data, retain the residuals, and treat them as filtered variables that form the input to the SGS search algorithm. The output of the search is then recorded and

compared to the known graph of the model. These comparisons are reported as the outcomes of the Monte Carlo simulations. These simulations are the standard against which the bootstrap procedure is judged.

4. We apply the bootstrap procedure, constructing a large number of bootstrap realizations, for each realization of the pseudo-real-world data and record the outcomes.

The key statistical decision in the SGS algorithm starts from the null hypothesis that a conditional correlation equals zero. Rejection of the null implies, in steps 1-3, that an edge is not removed from the graph and, in step 4, that an unshielded collider is identified. An error of commission (falsely including an edge or falsely identifying an unshielded collider) is, therefore, an example of type I error, and the rate of type I error corresponds to the size of the procedure. Analogously, an error of omission (falsely omitting an edge or failing to identify an unshielded collider) is an example of type II error, and the complement of the rate of type II error corresponds to the power the procedure.

Figure 6 reports the size and power for identification of the skeleton of Model 1 for both the Monte Carlo and the bootstrap procedure. The horizontal axis reports the signal-to-noise ratio of the parameterization of the $\mathbf{A}_0$ matrix measured as the mean *ex-ante t*-statistic for the edges in Figure 4.

The critical values of the tests of conditional correlation for the Monte Carlo simulations are set at 10 percent. Notice that, when signal-to-noise ratios are very low, the size of the procedure is also approximately 10 percent; but, when signal-to-noise ratios are very high, the size of the procedure falls by about half to 6 percent. This is

consistent with the results of Monte Carlo experiments on the PC algorithm reported by Demiralp and Hoover (2003). Figure 7 reports results for size and power for the unshielded colliders in Model 1. The pattern is similar to that in Figure 6, though at high signal strengths, type I error falls to nearly zero.

Preliminary experimentation on Model 1 showed that using a 10-percent critical value for the tests of conditional correlation in the SGS algorithm applied to the bootstrap resulted in a size of the procedure systematically higher than that for the Monte Carlo simulations. Experimentation demonstrated that by reducing the critical value to 2.5 percent it was possible to match the size of the Monte Carlo procedure extremely well as shown in Figure 6. The comparative size reported in Figure 6 does not, therefore, represent an independent test of the quality of the bootstrap procedure. However, the same adjustment to the critical value is applied to Model 2-4 with no further model-specific calibration. Table 2 shows that the match between the size of the Monte Carlo and bootstrap procedures is very close for both the skeletons and the unshielded colliders (look at the errors of commission, ignoring Model 2A, which is discussed below).

Turning to errors of omission, Figure 6 shows that the power of the bootstrap procedure to identify the skeleton of Model 1 is uniformly lower than that of the Monte Carlo. The power of both the Monte Carlo and the bootstrap increase as signal-to-noise ratios increase. The mean difference between the two curves is 8.6 percentage points.

Figure 7 shows that for Model 1, when signal-to-noise ratios are low, both the Monte Carlo and the bootstrap procedures falsely identify unshielded colliders with a size of 10 percent (again approximately the critical value of the conditional correlation tests in the Monte Carlo), and almost never falsely identify them when signal strengths are high.

The power of the search algorithm to identify unshielded colliders correctly as indicated by the Monte Carlo is very low when signal strengths are low, but rises to 85 percent when they are high. The bootstrap procedure also shows this rising pattern, but remains systematically less powerful than the Monte Carlo with a mean difference between the two curves of 11.3 percentage points.

All four models show similar patterns of type I and type II errors. The main question that we want to answer is how well the bootstrap algorithm mimics the results that we would find in a Monte Carlo study, were we lucky enough to know the true structure. We have already seen that the bootstrap algorithm with the adjustment to the critical value mimics type I error closely, but differs on type II error to a non-trivial degree. Table 2 presents the differences between the bootstrap algorithm and the Monte Carlo for all four models. The pattern of differences for Models 2, 3, and 4 are similar to those for Model 1.

To investigate the effect of the critical value adjustment somewhat further, we also compare the Monte Carlo to the bootstrap algorithm for Model 2 with the critical value in the bootstrap set to 10 percent (rather than the adjusted 2.5 percent). The results (shown in Table 2 as Model 2A) nearly reverse the pattern with the adjusted critical value. Now the bootstrap matches the Monte Carlo on errors of omission for the skeleton very closely, but differs by 9.2 percentage points (on average over different signal strengths) on errors of commission. The pattern is much the same with respect to unshielded colliders. These patterns are reflective of the usual tradeoff between size and power. They show that we can calibrate the bootstrap to either extreme, and possibly through an intermediate setting of the critical value in the bootstrap might choose some

more desirable compromise in matching the Monte Carlo properties on one or other type of error.

We conclude from these simulation experiments that the bootstrap procedure provides a reasonable method for assessing the reliability of causal structures identified by the SGS algorithm. With an adjustment to the critical value in the tests of conditional correlation, size can be controlled with reasonable precision. Power is systematically lower than Monte Carlo simulations indicate. This implies that the bootstrap procedure will suggest the omission of a causal linkage or fail to direct it too frequently – although the mean errors appear moderate. To look at this result more positively, if the bootstrap confirms a causal linkage, that confirmation is unlikely to be spurious: the bootstrap method should lead to few false positives.

5. The Utility of Graph-theoretic Algorithms

Demiralp and Hoover (2003) showed that the graph-theoretic PC algorithm was very successful in finding the skeleton and somewhat successful in directing the edges of the causal graph corresponding to the contemporaneous causal order of structural vector autoregressions. These findings held out the promise that data-based methods might replace the *ad hoc* and largely unconvincing *a priori* arguments that VAR analysts have typically used to choose the causal order of the SVAR. Unfortunately, while it was possible to show good results for particular structures, it was not clear how one could assess the reliability of a causal ordering chosen by a search algorithm in the typical cases in which, unlike Monte Carlo experiments, we do not have the true order to hand to serve as a standard of comparison.

In this paper, we have proposed a bootstrap procedure that allows the investigator to assess the performance of the SGS algorithm (a close relative of the PC algorithm) using only information readily to hand – that is, the data themselves. Using data from Swanson and Granger (1997), we showed that this bootstrap procedure was an effective tool. And, by comparing the results of the bootstrap procedure to those of Monte Carlo experiments on known models, we were able to show that the bootstrap sufficiently well mimics the Monte Carlo that it can provide useful guidance on the reliability of inference using graph-theoretic algorithms to determine the causal order of SVARs.

Appendix A. The Data[5]

All the raw series, except M1, are from the U.S. 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 1947:1-2002:4. Haver codes are in bold type. Personal Consumption Expenditure = ***CH***; Gross Private Domestic Fixed Investment = ***FH***; M1 monetary aggregate (*MN*) for 1947:1-1958:4 = M1 monetary aggregate from Board of Governors (1976), Table 1.1, pp. 17-18, column 2 ("Money Stock: Total") × 0.97966 and for 1959:1-2002:4 = ***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: $C = \log(\boldsymbol{CH}/\boldsymbol{LNN})$; $I = \log(\boldsymbol{FH}/\boldsymbol{LNN})$; $M = \log[MN/(P \times \boldsymbol{LNN})]$, where $P = (\boldsymbol{GNP} - \boldsymbol{G})/(\boldsymbol{GNPH} - \boldsymbol{GH})$; $Y = \log[(\boldsymbol{GNPH} - \boldsymbol{GH})/\boldsymbol{LNN}]$.

Appendix B. The Bootstrap Procedure

1. Start with an estimate of the reduced-form using the notation of section 1 with the addition that a "hat" (^) indicates an estimated value:

   (B.1) $$\mathbf{Y}_t = \hat{\mathbf{B}}(L)\mathbf{Y}_{t-1} + \hat{\mathbf{U}}_t .$$

   Let $\hat{\mathbf{U}} = [\hat{\mathbf{U}}_t]$, where $\hat{\mathbf{U}}$ is an $N \times T$ matrix with columns $\hat{\mathbf{U}}_t$.

2. Let $k = 1, 2, \ldots K$ be the bootstrap realization, and $j = 1, 2, \ldots J$, where $J >> T$, be the pseudo-time period. (In the simulations in this paper, $T = 500$, and $J = 1{,}500$.) Construct a set of bootstrap pseudo-observations recursively, starting with the initial condition $\tilde{\mathbf{Y}}_{k0} = \mathbf{0}$, with

   (B.2) $$\mathbf{Y}_{kj}^{BS} = \hat{\mathbf{B}}(L)\mathbf{Y}_{k(j-1)}^{BS} + \mathbf{U}_{kj}^{BS},$$

   where each $\mathbf{U}_{kj}^{BS}$ is a scaling factor, $\sqrt{\frac{T-L}{T-L-V}}$ times one of the columns of $\hat{\mathbf{U}}$ drawn randomly with replacement, where L is the order of the lags of the VAR and $V = N + 1$ (i.e., the number of variables in each equation of the VAR including the constant). The scaling factor corrects for losses in degrees of freedom in estimating the residuals from the VAR. Define the pseudo-data set as

[5] This appendix is drawn verbatim from Demiralp and Hoover (2003), Appendix B.

$\tilde{\mathbf{Y}}_k = [\tilde{\mathbf{Y}}_{kt}] = [\mathbf{Y}_{ki}^{BS}]$ to be the last T of the J columns of recursive pseudo-observations. That is, the observations for $t = 1, 2, \ldots T$ correspond to those for $i = J - T + 1, J - T + 2, \ldots J - 1, J$. (The long "startup period" is meant to eliminate dependence on the initial condition.)

3. For each k, run the casual search algorithm on the residuals ($\tilde{\mathbf{U}}_k$) from a reduced form VAR of $\tilde{\mathbf{Y}}_k$ and record the graph. (In this paper, we use the SGS algorithm, although the PC algorithm or other algorithms could be used.)
4. For each possible edge between variables (elements of $\mathbf{Y}$) record the number of times it appears a) missing (*no edge*); b) undirected (—); c) directed from lower to a higher ordered variable(←); d) directed from a higher to a lower order variable (→); and e) bidirectional (↔).

Appendix C. Design of the Simulation Experiments

1. We evaluate four SVAR models with contemporaneous causal orders described in section 4 of the main text. Evaluations are made with respect to the reference graph as described in point 5 below.
2. We start with a set of artificially generated data constructed by Monte Carlo methods as follows:
   a) A set of contemporaneous parameterizations is indexed $p = 1, 2, \ldots P$; and, for each parameterization, a set of Monte Carlo realizations is indexed $m = 1, 2, \ldots M$, so that there are PM total realizations. (For all models, P is set to 200; and, for Models 1 and 3, M is set to 200, yielding 40,000 realizations; while, for Models 2 and 4, M is set to 100, yielding 20,000 realizations.
   b) The data for each Monte Carlo realization are generated recursively. For each realization, let $j = 1, 2, \ldots J$, indicate time, start with the initial condition $\mathbf{Y}_0^{MC} = \mathbf{0}$, and construct subsequent observations according to

   (B.2) $$\mathbf{Y}_{pmj}^{MC} = \mathbf{A}_{p0}{}^{-1}\mathbf{A}\mathbf{Y}_{pm(j-1)}^{MC} + \mathbf{A}_{p0}{}^{-1}\tilde{\mathbf{E}}_{pmj}\,.$$

   where:

   i) Each model includes 1 lag of each variable, and $\mathbf{A}$ is a fixed $N \times N$ matrix of coefficients on the lagged variables, for which all diagonal elements are set to 0.25 and all off-diagonal elements are set to 0.05.
   ii) $\mathbf{A}_{pm0} = [a_{pmgh}]$ is an $N \times N$ ($g = 1, 2, \ldots, N$; $h = 1, 2, \ldots, N$) matrix of contemporaneous coefficients with ones on the main diagonal, and zeroes placed to correspond to the causal orders for each model in point 1 above. The non-zero elements of $\mathbf{A}_{pm0}$ are drawn randomly from a range that yields *ex ante t*-statistics between 0 and 10 (averaging over all edges), with

oversampling in the range that produces average *ex ante t*-statistics between 0 and 1. (See subpoint v below on the evaluation of *ex ante t*-statistics.)

iii) $\tilde{\tilde{\mathbf{E}}}_{pmj} = [E_{pmnj}]$ is an N-element column vector with each element drawn randomly from an independent normal distributions with mean 0 and variance 1.

iv) To determine the *ex ante t*-statistics: 1. each model is simulated 1000 times according to subpoint i with randomly drawn contemporaneous coefficients, estimated, and the coefficient values (β) and the corresponding *t*-statistics recorded; 2. a regression equation is estimated $t = \varphi + \mu\beta + error$ is estimated; 3. in the simulation experiments, the *ex ante t*-statistic for any coefficient is given as $\hat{t} = \hat{\varphi} + \hat{\mu}\beta$, where the hats ("^") indicate estimated or predicted values.

Define the pseudo-data set as $\breve{\mathbf{Y}}_{pm} = [\breve{\mathbf{Y}}_{pmt}] = [\mathbf{Y}^{MC}_{pmi}]$, t = 1, 2, . . . 500, to be the last 500 of the J columns of recursive pseudo-observations, where $J >> 500$. That is, the observations for t = 1, 2, . . . 500 correspond to those for $i = J - 499, J - 498, \ldots J - 1, J$. (As with the bootstrap method in Appendix B, the long "startup period" is meant to eliminate dependence on the initial condition.)

3. For each Monte Carlo realization, run the SGS algorithm on the residuals from a reduced form VAR of $\breve{\mathbf{Y}}_{pm}$ and record: a. the *ex ante t*-statistics for each edge and b. whether edges are present or absent and, if present, how they are oriented. Results are assessed by comparison to the known reference graph, as described in point 5 below. The resulting *PM* realizations in the current study, correspond to the Monte Carlo simulations reported in Demiralp and Hoover (2003) for the PC algorithm. The results provide a practical measure of how well the SGS algorithm does in discovering the true causal order underlying the data.

4. For each Monte Carlo realization run the bootstrap procedure described in Appendix B using $\breve{\mathbf{Y}}_{m}$ as the input; and, for each bootstrap realization (k), record: a. the *ex ante t*-statistics for each edge; b. whether edges are present or absent; c. if edges are present, how they are oriented; and d. the unshielded colliders. Results are assessed by comparison to the known graph (see point 1 above), as described in point 5 below. (The number of bootstrap realizations for each Monte Carlo realization is set to K = 100, so that, for Models 1 and 3, the total number of bootstrap realizations is 4,000,000 (=PMK = 200 × 200 × 100) and, for Models 2 and 4, 2,000,000 (=PMK = 200 × 100 × 100). The relatively small number of bootstrap realizations per Monte Carlo realization is needed to keep the computing time manageable.) The results provide a measure of how closely the bootstrap conforms to the Monte Carlo in particular cases.

5. To assess outcomes in points 3 and 4 above, the graphs identified by the SGS algorithm applied to simulated data are compared edge by edge to the graphs that are

used to generate the data (see point 1 and point 2 above). Because of Pearl's observational equivalence theorem (see section 2), an ideally functioning search algorithm working with infinite data will not be able to orient some edges. Comparisons are not made to the true graph that generates the data, but to a *reference graph* that represents the equivalence class (that is, to the graph that leaves edges unoriented if, in the true graph, they can be reversed without changing the number or location of the unshielded colliders). Results are assessed according to success at the statistical identification of the skeleton and the unshielded colliders. There are three possible outcomes: i. *correct*: an edge or unshielded collider is identified as absent when it is absent in the reference graph or identified as present when it is present in the reference graph; ii. *omitted*: an edge or unshielded collider that is present in the reference graph is identified as absent; iii. *committed*: an edge or unshielded collider that is absent in the reference graph is identified as present.

Results are reported as rates taking the number of possible realizations of a particular outcome as the base. For example, in Model 2 (Figure 5), there are six edges out of a total of fifteen possible between six variables. The base for computing the rate of omissions is, therefore, six per search. In contrast, the base for the rate of commissions is nine – that is, the number of possible edges that are in fact missing in the reference graph.

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*, American Association for Artificial Intelligence, Menlo Park, CA and MIT Press, Cambridge, MA, pp. 497-520.

Awokuse, T. O. (2005) "Export-led Growth and the Japanese Economy: Evidence from VAR and Directed Acyclical Graphs," *Applied Economics Letters* 12(14), 849-858.

Bernanke, Ben S. and Illian Mihov. (1998) "Measuring Monetary Policy," *Quarterly Journal of Economics* 113(3), 869-902

Bessler, David A. and N. Loper. (2001) "Economic Development: Evidence from Directed Acyclical Graphs" *Manchester School* 69(4), 457-476.

Bessler, David A. and Seongpyo Lee. (2002). 'Money and prices: U.S. data 1869-1914 (a study with directed graphs)', *Empirical Economics*, Vol. 27, pp. 427-46.

Blanchard, Olivier J. and Danny Quah. (1989) "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* 79(4), 655-73.

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

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*, American Association for Artificial Intelligence, Menlo Park, CA and MIT Press, Cambridge, MA, 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.

Demiralp, Selva and Kevin D. Hoover. (2003) "Searching for the Causal Structure of a Vector Autoregression," *Oxford Bulletin of Economics and Statistics* 65(supplement), pp. 745-767.

Haigh, M.S., N.K. Nomikos, and D.A. Bessler (2004) "Integration and Causality in International Freight Markets: Modeling with Error Correction and Directed Acyclical Graphs," *Southern Economic Journal* 71(1), 145-162.

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

Hoover, Kevin D. (2003) "Some Causal Lessons from Macroeconomics," *Journal of Econometrics* 112(1), 121-125.

Hoover, Kevin D. (2005) "Automatic Inference of the Contemporaneous Causal Order of a System of Equations," *Econometric Theory* 21(1), 69-77.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. (1991) "Stochastic Trends and Economic Fluctuations," *American Economic Review* 81(4), 819-40.

LeRoy, Stephen F. (2002). 'Causality: models, reasoning, and inference: a review of Judea Pearl's *Causality*', *Journal of Economic Methodology*, Vol. 9, pp. 100-3.

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

Scheines, Richard, Peter Spirtes, Clark Glymour, and Christopher Meek. (1994) *Tetrad II: Tools for Causal Modeling: User's Manual*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sheffrin, Steven M. and Robert K. Triest. (1998). 'A new approach to causality and economic growth', unpublished typescript, University of California, Davis.

Spirtes, Peter, Clark Glymour, and Richard Scheines. (2000) *Causation, Prediction, and* Search, 2nd edition. Cambridge, MA: MIT Press.

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*, Vol. 92, pp. 357-67.

Yang, J. and D.A. Bessler. (2004) "The International Price Transmission in Stock Index Futures Markets," *Economic Inquiry* 42(3), 370-386.

Table 1
Bootstrap Evaluation of the SGS-selected Causal Order for the King *et al.* Data Set

SGS-selected Causal Order			Edge Identification (percent of bootstrap realizations)				
Edge direction			—	←	no edge	→	↔
C	→	*I*	28	23	16	**33**	0
C	←	*M*	22	**65**	13	0	0
C	←	*Y*	32	**67**	0	1	0
I	no edge	*M*	0	6	**94**	0	0
I	←	*Y*	61	**39**	0	1	0
M	no edge	*Y*	0	0	**100**	0	0

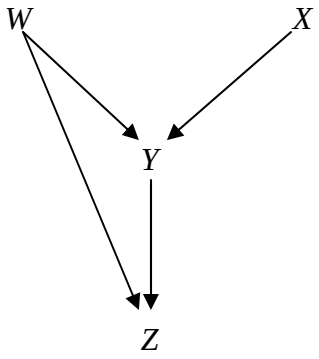
Notes: Entries in bold type correspond to the SGS-selected direction.
See text (section 2) for a description of the SGS search algorithm and Appendix B for a description of the bootstrap procedure.
See Appendix A for data definitions and sources.

Table 2
How Well Does the Bootstrap Mimic the Monte Carlo Simulations?

		Difference between Bootstrap and Monte Carlo Simulations (percentage points)								
		Signal-to-Noise Ratios (measured by *ex ante t*-statistics)								
Model	**Errors of:**	***t* < 1**	**1 < *t* < 2**	**2 < *t* < 3**	**3 < *t*< 4**	**4 < *t* < 5**	**5 < *t* < 6**	**6 < *t* < 7**	***t* > 7**	**mean**
					Skeleton					
Model 1	Commission	0.9	1.4	1.1	1.6	0.9	0.5	0.7	0.2	0.9
Model 2	Commission	1.5	1.1	0.9	1.2	1.3	0.9	0.5	0.2	0.9
Model 2A	Commission	12.4	11.8	11.0	9.7	8.5	7.5	6.7	5.8	9.2
Model 3	Commission	1.1	1.1	1.0	1.0	0.6	0.7	0.5	0.6	0.8
Model 4	Commission	1.1	1.0	1.0	0.8	0.8	0.6	0.6	0.3	0.8
Model 1	Omission	4.3	11.8	14.1	13.1	8.8	7.1	6.6	2.7	8.6
Model 2	Omission	4.3	9.6	13.0	13.9	8.7	6.5	3.3	2.2	7.7
Model 2A	Omission	-7.6	-1.6	2.0	2.6	0.4	-0.1	0.2	0.6	-0.4
Model 3	Omission	7.3	11.7	14.0	10.1	6.8	6.1	5.0	3.0	8.0
Model 4	Omission	6.1	10.9	14.0	10.0	6.7	6.2	4.3	2.7	7.6
					Unshielded Colliders					
Model 1	Commission	0.3	1.7	2.6	2.1	1.6	0.9	0.7	0.3	1.3
Model 2	Commission	-0.1	0.2	-0.5	-1.5	-0.9	-0.4	1.2	2.6	0.1
Model 2A	Commission	-7.6	-1.6	2.0	2.6	0.4	-0.1	0.2	0.6	-0.4
Model 3	Commission	-0.3	-0.5	-0.7	-0.4	0.0	0.3	0.5	0.8	0.0
Model 4	Commission	-0.2	-0.6	-0.9	-0.4	-0.1	0.1	0.7	0.7	-0.1
Model 1	Omission	1.2	7.4	14.8	19.8	13.5	13.1	13.5	7.3	11.3
Model 2	Omission	1.8	5.5	11.9	17.6	16.1	12.5	7.9	8.1	10.2
Model 2A	Omission	-3.0	0.8	6.8	10.0	7.5	7.3	9.2	10.4	6.1
Model 3	Omission	2.7	9.0	15.2	13.5	10.8	11.8	8.9	8.1	10.0
Model 4	Omission	1.2	6.7	15.5	14.3	11.0	11.3	9.4	4.9	9.3

Figure 1

Panel A
The Causal Data-Generating Process



Panel B
Densely Connected Skeleton

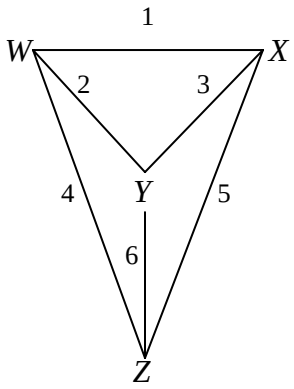
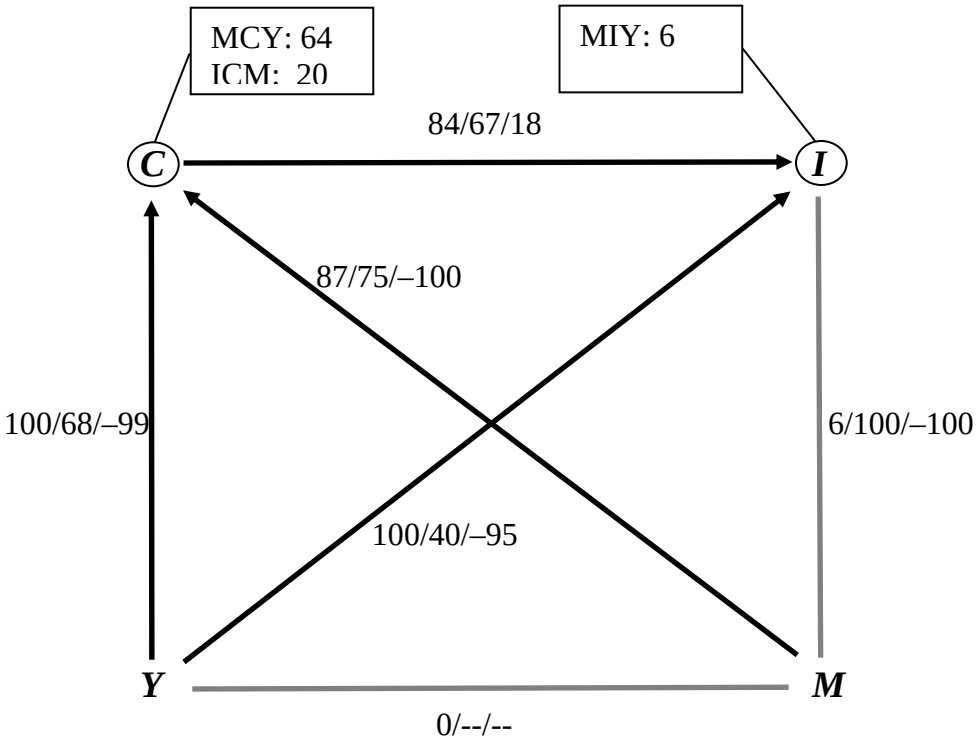


Figure 2
Causal Order for King *et al.* Data



Notes: See test (section 2) for a description of the SGS search algorithm and Appendix B for a description of the bootstrap procedure.
See Appendix A for data definitions and sources.
Causal connections selected and oriented by the SGS search algorithm shown as black arrows; those omitted as gray lines.
Triples associated with each edge are *Exist*/*Directed*/*Net Direction* where
Exist = the percentage of bootstrap replications in which an edge is selected;
Directed = edges directed as a percentage of edges selected;
Net Direction = difference between edges directed low-to-high (→) and high-to-low (←), where higher variables are the more clockwise starting in the upper left corner (at *C*), as a percentage of the directed edges.
Circled variables are found as unshielded colliders on paths, and with rates (percentage of replications) as indicated in boxed notes. Any unshielded collider found in fewer than 1 percent of replications is omitted.
Data may differ from values computed from Table 1 due to rounding error.

Figure 3
Impulse-response Functions for Alternative Causal Orderings of the Swanson-Granger Data

Figure 3.A
Impulse Responses of Money to a Consumption Shock for Three Contemporaneous Causal Orderings

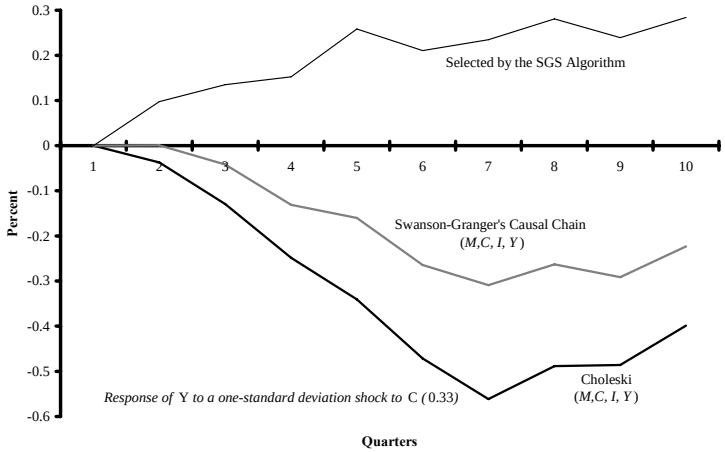


Figure 3.B
Impulse Responses of GDP to a Money Shock for Three Contemporaneous Causal Orderings

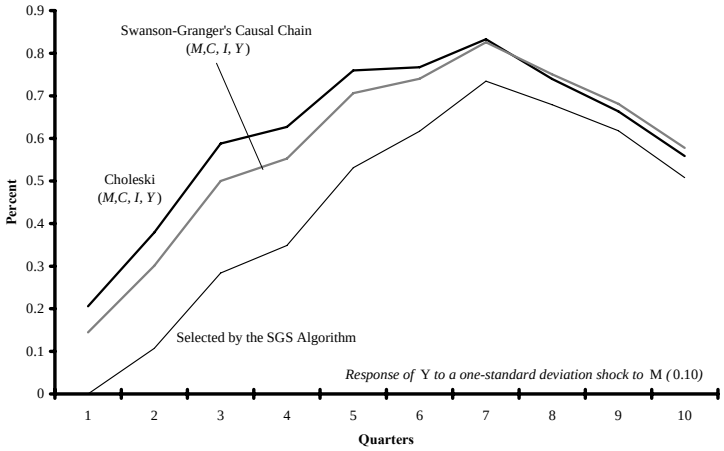


Figure 3.C
Impulse Responses of GDP to a GDP Shock for Three Contemporaneous Causal Orderings

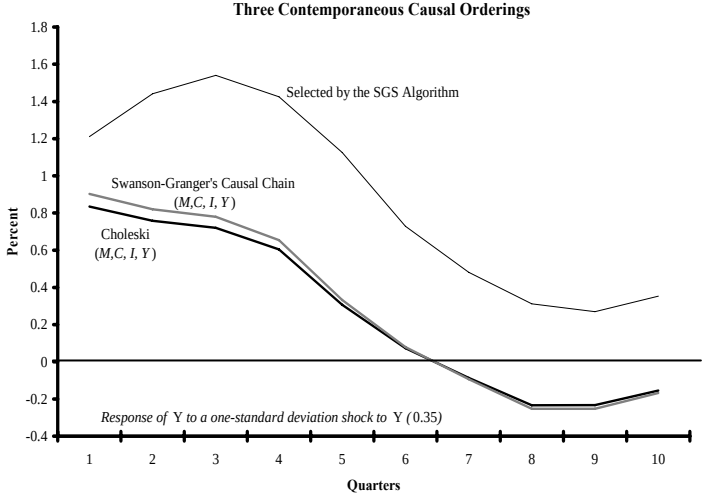
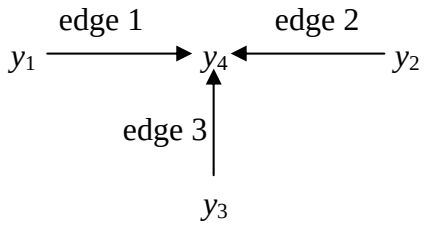
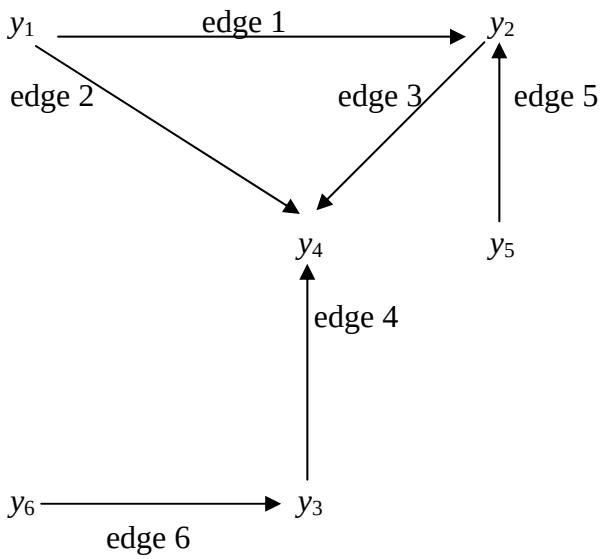


Figure 4
Model 1



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

Figure 5
Model 2



$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ a_{210} & 1 & 0 & 0 & a_{250} & 0 \\ 0 & 0 & 1 & 0 & 0 & a_{360} \\ a_{410} & a_{420} & a_{430} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 6
Monte Carlo versus Bootstrap for the Model 1 Skeleton

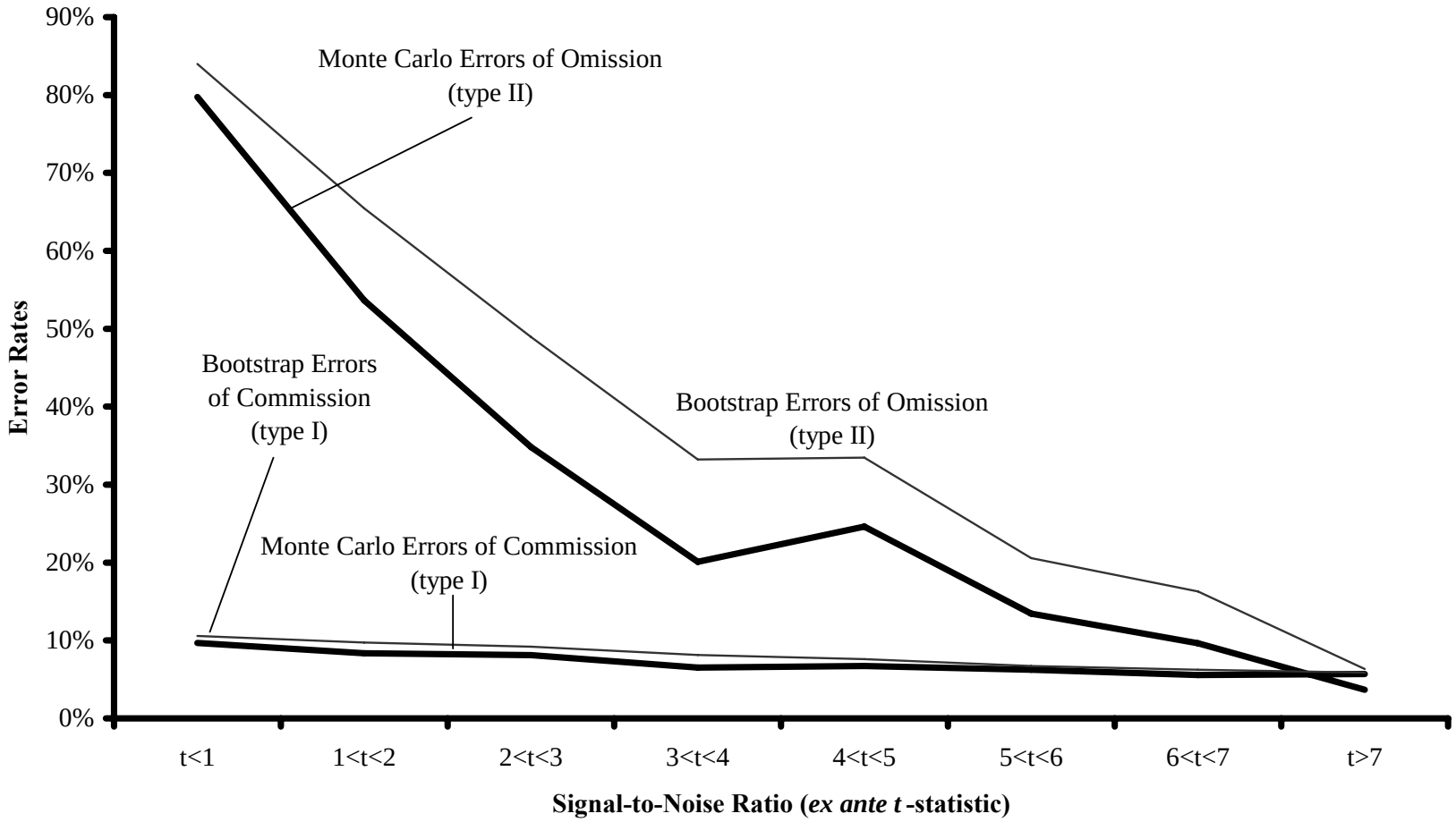


Figure 7
Monte Carlo versus Bootstrap for the Model 1 Unshielded Colliders