Econometric Theory, **21**, 2005, 69–77. Printed in the United States of America.
DOI: 10.1017/S026646660505005X

AUTOMATIC INFERENCE OF THE CONTEMPORANEOUS CAUSAL ORDER OF A SYSTEM OF EQUATIONS

KEVIN D. HOOVER
University of California, Davis

1. THE PROBLEM OF CONTEMPORANEOUS CAUSAL ORDER

When Stephen Perez and I first began our Monte Carlo studies of the efficacy of general-to-specific search methodologies in 1995, we were keenly aware of our limited ability to capture the tacit knowledge of the skilled time-series econometrician operating in the London School of Economics (LSE) tradition (Hoover and Perez, 1999a, 1999b). Econometrics, we believed, was an art, and our algorithm was not intended to replace the artist. David Hendry and Hans-Martin Krolzig's subsequent development of PcGets did not, in fact, eliminate the art of econometrics. Power tools did not eliminate the art of the cabinetmaker but changed where his value added lay and—importantly—made new things possible. PcGets is likewise a new, powerful tool, useful in the hands of a skilled craftsman.

But no tool solves every problem. One open problem is briefly touched on in Hendry's answer to question 16:

> When the reduced-form VAR has a diagonal covariance matrix, then all possible reductions of the system can be efficiently estimated by OLS, and model-selection procedures can operate equation-by-equation without any loss in efficiency. For a structural VAR (SVAR), with a recursive specification as in Wold (1949), a similar result holds for OLS being efficient.

The suggestion is that, if a recursive (or Wold causal) order is known for the contemporaneous variables in the SVAR, then PcGets can be applied equation by equation to find a parsimonious lag structure. But where is the knowledge of the causal order to come from?

The SVAR can be written as

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

I thank Peter Phillips, Selva Demiralp, Stephen Perez, and an anonymous referee for helpful comments on an earlier draft. Address correspondence to Kevin D. Hoover, Department of Economics, University of California, One Shields Avenue, Davis, CA 95616-8578, USA; e-mail: kdhoover@ucdavis.edu.

 0266-4666/05 $12.00

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 nonzero 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 $\mathbf{\Sigma} = \mathrm{E}(\mathbf{EE}')$ 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. The system is identified provided that there are $n(n-1)/2$ zero restrictions on $\mathbf{A}_0$.[1] For any just-identified system, $\mathbf{A}_0$ can be rendered lower triangular by selecting the appropriate order of the variables $\mathbf{Y}$ along with the conformable order of the rows of $\mathbf{A}_0$. This is the *recursive* (or *Wold causal*) *order*.

Starting with the SVAR as the data-generating process (DGP), premultiplying by $\mathbf{A}_0^{-1}$ yields the reduced-form or VAR:

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

The VAR program in macroeconometrics began with the assertion of Christopher Sims (1980) that the belief that we could know $\mathbf{A}_0$ and $\mathbf{A}$ a priori was "incredible." He advocated starting with the VAR (equation (2)), in which $\mathbf{B}$ is easily estimated. He realized that because the covariance matrix $E(\mathbf{UU}')$ is not diagonal, it is not possible to shock the individual equations of the system independently to trace out impulse responses or to calculate variance decompositions. He suggested working with orthogonalizing transformations of the VAR to secure a one-to-one correspondence between shocks and equations, but he was cavalier about the problem of choosing the appropriate transformation. Critics such as Cooley and LeRoy (1985) and Leamer (1985) convinced Sims (1986) that sensible economic interpretation required the identification of $\mathbf{A}_0$—if not of $\mathbf{A}$. This is the central problem of VAR analysis and one that, so far, has not been effectively resolved.[2] Neither PCGets nor related search algorithms improves in this regard on the common practices of the VAR literature.

2. IDENTIFICATION

If we *knew* $\mathbf{A}_0$, then recovery of the SVAR (equation (1)) from the easily estimated VAR (equation (2)) would be straightforward. There are, however, a large number of $n \times n$ matrices $\mathbf{P}_i$ that may be used to premultiply equation (2) 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.

For each ordering of the variables in $\mathbf{Y}$, there is a unique lower triangular $\mathbf{P}_i \in \mathbf{P}$ such that $\mathbf{P}_i\mathbf{P}_i' = \mathbf{\Omega}$. This is the Choleski decomposition of the covariance matrix and corresponds to a Wold causal ordering of the variables. Because the ordering of the variables in $\mathbf{Y}$ is arbitrary, there are as many such orderings as there are permutations of the elements of $\mathbf{Y}$. Each such ordering is just-identified and, therefore, observationally equivalent. There are also other overidentified causal orderings—that is, $\mathbf{P}_i$ for which there are more than $n(n-1)/2$ zero restrictions.

The central identification problem for SVARs is to choose the one member of $\mathbf{P}$ that corresponds to the DGP: that is, to choose $\mathbf{P}_i = \mathbf{A}_0$ when $\mathbf{A}_0$ is unknown. The other elements can be thought of as defining pseudo-SVARs. But on what basis should we choose? There are at least two options. First, we can appeal to economic theory to tell us what the causal order should be. This is, in fact, what almost all practitioners of VAR methodologies profess to do. Unfortunately, formal economic theory is rarely decisive about causal order. In reality, VAR practitioners follow one of two strategies: They choose the order arbitrarily, sometimes with an accompanying claim that their results are robust to alternative causal orderings—apparently unaware that such robustness really amounts to a claim that the contemporaneous terms do no real work at all, so that causal order is irrelevant. Sometimes they appeal not so much to theory as to "just so" stories. Intuition or common sense tells them that, say, financial markets adjust more quickly than goods markets, so that interest rates, for instance, ought to be causally ordered ahead of real GDP. It is usually easy, however, to tell a "just so" story to justify most any order—the time order of variables that are contemporaneously related at the given frequency of observation being especially unreliable. There is a special irony that this strategy should be so commonly accepted among VAR practitioners. After all, the motivation of Sims (1980) in initiating the VAR program was to avoid the need to appeal to "incredible identifying restrictions."

3. GRAPH-THEORETIC METHODS

A second method of choosing $\mathbf{P}_i$ is to try to extract more information out of the data. Graph-theoretic causal search is an approach (really a family of approaches) to this problem very much in the spirit of general-to-specific model selection. (Spirtes, Glymour, and Scheines, 2000, and Pearl, 2000, provide the most developed accounts of the approach.)[3] In a causal graph, arrows connecting causal variables to their effects represent causal relationships. The mathematics of graph theory can be used to analyze the causal structures. It is important that it can be shown that there are isomorphisms between graphs and the probability distributions of variables. In particular, certain graphical patterns imply conditional independence and dependence relationships among the variables. The graph of the DGP can also be represented through the restrictions on $\mathbf{A}_0$. Working backward from statistical measures of conditional independence and dependence, it is possible to infer the class of graphs compatible with the data. Sometimes that class has only a single member, and then $\mathbf{A}_0$ can be identified statistically.

The key ideas of the graph-theoretic approach are simple. Suppose that $A \rightarrow B \rightarrow C$ (i.e, A causes B causes C). Here A and C would be dependent, but conditional on B, they would be independent. 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 or its descendants, and $A \rightarrow C \leftarrow B$, and none of the variables that cause A or B directly causes C. Then, conditional on C, A and B are dependent. Here C is called an *unshielded collider* on the path ACB. (A *shielded* collider would have a direct link between A and B.)

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 backward to the class of admissible causal structures.[4] The PC algorithm is the most commonly used in the literature (Spirtes et al., 2000, pp. 84–85; Pearl, 2000, pp. 49–51; Cooper, 1999, p. 45, fig. 22).[5] It assumes that graphs are *acyclical*—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. There are six steps.

(1) Start with a graph in which each variable is assumed to be connected by an undirected causal link.
(2) Test for the unconditional correlation of each pair of variables, eliminating the link 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 link 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) For each conditionally uncorrelated pair of variables (i.e., ones without a direct link) 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 links as pointing into the unshielded collider.
(5) If there are any pairs A and C that are not directly connected but are linked $A \rightarrow B - C$, then orient the second link 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 link 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 link as $A \rightarrow B$.

Steps 1–4 are based in statistical inference. Step 5 follows logically, because orienting the undirected link 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 link in the other direction would, contrary to assumption, render the graph cyclical.

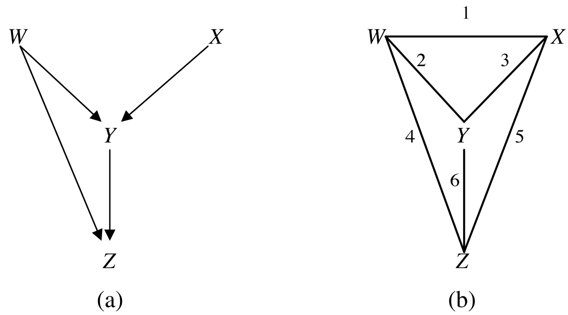

FIGURE 1. Illustration of the PC algorithm.

The PC algorithm can be illustrated with the example in Figure 1. Figure 1a shows the graph of the DGP. 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 nonzero elements.

Step 1 starts with Figure 1b, in which all the variables are connected. It is analogous to the general unrestricted model (GUM) of PcGets. Step 2 eliminates link 1, because *W* and *X* are unconditionally uncorrelated in the DGP. Step 3 eliminates link 5 (*X* and *Z* are uncorrelated conditional on *Y*). Step 4 orients links 3 and 4 toward *C* (*W* and *X* are correlated conditional on *Y*—i.e., *Y* is an unshielded collider on *WYX*). Step 5 orients link 6 toward *Z*. Step 6 orients link 2 toward *Z*. The algorithm is able to recover the DGP.

Not every DGP 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, Theorem 1.2.8, p. 19) 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 Figure 1a except that link 6 was reversed would not be observationally equivalent to Figure 1a because it would add an unshielded collider (*Y* on *XYZ*). A graph that reversed link 2 would be observationally equivalent to the graph in Figure 1a because it would have the same skeleton and neither add nor subtract unshielded colliders. Although the graph with link 2 reversed illustrates the observational equivalence theorem, step 6 of the algorithm rules it out, because it possesses a cycle ($W \rightarrow Y \rightarrow Z \rightarrow W$), which violates the antecedent of the observational equivalence theorem.

Graph-theoretic search, like LSE general-to-specific methods, is sometimes taxed with the "heroic" assumption that DGP lies within the search universe or even that there exists a true DGP. The reply is the same for the one method as the other. First, if there is any regularity in the world—and surely this is not in question—then there is a mechanism that generates it. It may be exceedingly complex and unstable, so that there are levels of detail at which it defies characterization by the benighted econometrician. Nevertheless, the conviction—based perhaps in faith and hope—that sustains all science is that characterizations are possible that, if they do not recover the DGP, at least bear a systematically useful relationship to it.

Second, although a search may be conducted on too narrow a basis, serious criticism—that is, criticism that suggests concretely what might have been omitted—provides the basis for new, more expansive searches, so that understanding advances dialectically. More particularly, with respect to causal search, Sprites et al. (2000) have developed methods for inferring the existence of omitted (latent) variables and working out their causal consequences.

4. APPLICATIONS TO VECTOR AUTOREGRESSIONS

Swanson and Granger (1997) were the first to introduce graph-theoretic search into the analysis of the contemporaneous causal order of VARs. Swanson and Granger restrict the class of orderings to causal chains—that is, to orderings in which each element of $\mathbf{Y}$ has at most one direct cause and one direct effect (e.g., $W \rightarrow X \rightarrow Y \rightarrow Z$). Graph-theoretic methods were generally not conceived with time-series data in mind. Swanson and Granger realized that the relevant information for the contemporaneous causal ordering of the SVAR is actually contained in the covariance matrix of the VAR error terms in equation (2). They estimate (2) and calculate $\hat{\Omega}$, from which all the conditional correlations needed by the search algorithm can be calculated.[6] Demiralp (2000), Bessler and Lee (2002), Moneta (2003), and Demiralp and Hoover (2004) have extended their strategy to the less restricted class of structures compatible with the PC algorithm. Demiralp and Hoover (2004) provide Monte Carlo evidence that shows that the PC algorithm is highly effective at recovering the skeleton of the DGP graph and moderately effective at recovering the directions of individual links provided that signal-to-noise ratios are high enough.

The observational equivalence theorem implies that some structures cannot be recovered in principle. For example, if the DGP really displays a Wold causal order ($\mathbf{A}_0$ is lower triangular), then there are no unshielded colliders, so no links can be directed, and all possible Wold causal orders are observationally equivalent. Even when the DGP cannot be recovered, the class of data-admissible models will generally be narrowed. Theory may in some instances permit some links to be oriented, which may, according to steps 5 and 6, imply other orderings. Undirected links might also be ordered by exploiting information about regime changes.[7]

Acyclical graphs are not fully adequate to economics, as so much of economics is represented in the form of simultaneous systems. Some economists, including Wold (1949) and Granger (1969, 1988), argue that there is no true simultaneity. For Granger, causality is a temporal notion—causes must precede effects—so that simultaneity or instantaneous causality cannot be fundamental. Simultaneity or instantaneous causality may appear because of omitted causal variables or because of temporal aggregation—most macroeconomic data are monthly or quarterly.[8] Granger argues that such simultaneity would disappear if fine enough cuts of the data were taken and the models took account of the relevant causal variables.

Hoover (2001, Ch. 6) argues that an adequate account of causality must permit simultaneity and instantaneous causality. As for Pearl and Spirtes et al., Hoover's notion of causality is structural. The potential controllability of one variable by another is a hallmark of the true causal relationship. On the one hand, such controllability may characterize equilibrium configurations independent of temporal processes. On the other hand, some concepts may lose economic meaning long before fine enough temporal disaggregation has left them with a strict temporal ordering. Surely, hourly GDP, for example, is not an economically meaningful quantity. On this view, a causal analysis of *cyclical* graphs is a vital element of future research (see Pearl, 2000, pp. 95–96, 142–143; and Richardson, 1996).

In the meantime, a natural extension of general-to-specific single-equation modeling would be to use graph-theoretic algorithms to select the contemporaneous causal ordering of the SVAR and then to apply algorithms such as PcGets to the individual equations.

NOTES

1. I concentrate here on zero restrictions, although SVARs are sometimes identified in other ways.

2. The early history of the VAR program is recounted in Hoover (1988, Ch. 8).

3. These approaches are not well known to econometricians. Glymour and Sprites (1988) contributed to a special issue of the *Journal of Econometrics* on causality, and Pearl's book has been reviewed by Swanson (2002), LeRoy (2002), and Hoover (2003). The most extensive review, which raises searching questions about the methods, is Neuberg (2003), to which Pearl (2003) replies.

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

5. The name "PC algorithm" derives from the names of its authors, Peter and Clark (Pearl 2000, p. 50).

6. Of course, in the cross-section applications that Glymour et al. have mind, the covariance matrix is estimated from raw data. In this case, it is estimated from residuals that are themselves

estimates of the true residuals. This introduces another source of uncertainty. What is more, as Swanson and Granger (1997, p. 361) note, estimates of the conditional correlation may be inconsistent under the null hypothesis of a particular graph. Some evidence of whether the additional uncertainty or the inconsistency of the estimates is practically important is provided by Monte Carlo studies such as those of Demiralp and Hoover (2004).

7. See Hoover (1990; 1991; 2001, Chs. 8–10), Hoover and Sheffrin (1992), and Hoover and Siegler (2000).

8. Breitung and Swanson (2002) provide a careful analysis of how different sorts of temporal aggregation may render the apparent Granger-causal order of the aggregated system a misleading guide to the true Granger-causal order of the underlying, disaggregated system.

REFERENCES

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

Breitung, J. & N.R. Swanson (2002) Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis* 23, 651–665.

Cooley, T.F. & S.F. LeRoy (1985) Atheoretical macroeconomics: A critique, *Journal of Monetary Economics* 16, 283–308.

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

Demiralp, S. (2000) *The Structure of Monetary Policy and Transmission Mechanism*. Ph.D diss., Department of Economics, University of California, Davis.

Demiralp, S. & K.D. Hoover (2004) Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and Statistics* 65, supplement, 745–767.

Glymour, C. & P. Spirtes (1988) Latent variables, causal models and overidentifying constraints. *Journal of Econometrics* 39, 175–198.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

Granger, C.W.J. (1988) Some recent developments in a concept of causality. *Journal of Econometrics* 39, 199–211.

Hoover, K.D. (1988) *The New Classical Macroeconomics*. Blackwell.

Hoover, K.D. (1990) The logic of causal inference: Econometrics and the conditional analysis of causality. *Economics and Philosophy* 6, 207–234.

Hoover, K.D. (1991) The causal direction between money and prices: An alternative approach. *Journal of Monetary Economics* 27, 381–423.

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

Hoover, K.D. (2003) Review of Judea Pearl's *Causality: Models, Reasoning, and Inference*. *Economic Journal* 113, F411–F413.

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

Hoover, K.D. & S.J. Perez (1999b) Reply to our discussants. *Econometrics Journal* 2, 244–247.

Hoover, K.D. & S.M. Sheffrin (1992) Causation, spending and taxes: Sand in the sandbox or tax collector for the welfare state? *American Economic Review* 82, 225–248.

Hoover, K.D. & M. Siegler (2000) Taxing and spending in the long view: The causal structure of U.S. fiscal policy after 1791. *Oxford Economic Papers* 52, 745–773.

Leamer, E.E. (1985) Vector autoregressions for causal inference. In K. Brunner & A.H. Meltzer (eds.), *Understanding Monetary Regimes*, Carnegie-Rochester Conference Series on Public Policy, vol. 22, pp. 225–304. North-Holland.

LeRoy, S.F. (2002) *Causality: Models, Reasoning, and Inference:* A review of Judea Pearl, *Causality*. *Journal of Economic Methodology*. 9, 100–103.

Moneta, A. (2003) Graphical Models for Structural Vector Autoregressions. Typescript, S. Anna School of Advanced Studies, Pisa, Italy.

Neuberg, L.G. (2003) Review of *Causality: Models, Reasoning, and Inference* by Judea Pearl. *Econometric Theory* 19, 675–685.

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

Pearl, J. (2003) Comments on Neuberg's review of *Causality*. *Econometric Theory* 19, 686–689.

Richardson, T. (1996) A discovery algorithm for directed cyclical graphs. In F. Jensen & E. Horwitz (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Twelfth Congress*, pp. 462–469. Morgan Kaufman.

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

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

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

Swanson, N.R. (2002) Review of Judea Pearl's *Causality: Models, reasoning and inference. Journal of Economic Literature* 40(3), 925–926.

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

Wold, H.O.A. (1949) Statistical estimation of economic relationships. *Econometrica* 17, supplement, 1–21.