Laboratory of Economics and Management
Sant'Anna School of Advanced Studies

Piazza Martiri della Libertà, 33 - 56127 PISA (Italy)
Tel. +39-050-883-343 Fax +39-050-883-344
Email: lem@sssup.it Web Page: http://www.lem.sssup.it/

LEM
Working Paper Series

Graph-Based Search Procedure for Vector Autoregressive Models

Alessio MONETA*
Peter SPIRTES†

* Sant'Anna School of Advanced Studies, Pisa, Italy
† Carnegie Mellon University, Pittsburgh, USA

2005/14 **June 2005**

ISSN (online) 2284-0400

Graph-based Search Procedure for Vector Autoregressive Models

Alessio Moneta
Laboratory of Economics and Management
Sant'Anna School of Advanced Studies
56127 Pisa, Italy
amoneta@sssup.it

Peter Spirtes
Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213
ps7z@andrew.cmu.edu

Abstract

In this paper we present a semi-automated search procedure to deal with the problem of the identification of the causal structure related to a vector autoregressive model. The structural form of the model is described by a directed graph and from the analysis of the partial correlations of the residuals the set of acceptable causal structures is derived.

1 INTRODUCTION

Vector autoregressive (VAR) models are a family of multiequation time-series models in which all variables are treated symmetrically, without dichotomizing the variables into "endogenous" and "exogenous". They were introduced into econometrics by Sims (1980), who demonstrated that VAR models provide a flexible and tractable framework for analyzing and summarizing the statistical properties of economic time-series. An identification problem is encountered when one wants to do structural inference or policy analysis. Basically, this problem corresponds to the problem of differentiating between causation and correlation (see Stock and Watson (2001)). In this paper we propose a graph-based search procedure to address this problem.

A zero-mean stationary VAR model can be written as:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t, \qquad (1)$$

where $Y_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$, $u_t = (u_{1t}, u_{2t}, \ldots, u_{kt})'$, and $A_1, \ldots A_p$ are $(k \times k)$ matrices. The components of u_t are white noise innovation terms, $E(u_t) = 0$ and u_t and u_s are independent for $s \neq t$. The matrix $\Sigma_u = E(u_t u_t')$ is in general nondiagonal. The relations among the contemporaneous components of Y_t, instead of appearing in the functional form (as in simultaneous equation models), are embedded in the covariance matrix of the innovations. From the estimation of equation (1), which is straightforwardly obtained by OLS, one does not get, in general, the structural relations among the variables, because numerous structures are compatible with a particular set of statistical associations. It is useful to assume (without loosing generality as to the family of linear models) that the data are generated by a structural equation of the form:

$$\Gamma Y_t = B_1 Y_{t-1} + \ldots B_p Y_{t-p} + C v_t, \qquad (2)$$

where v_t is a $(k \times 1)$ vector of serially uncorrelated structural disturbances with a mean of zero and a diagonal covariance matrix Σ_v.

The identification problem consists in finding a way to infer the unobserved parameters in (2) from the estimated form (1), where $A_i = \Gamma^{-1} B_i$ for $i = 1, \ldots, p$ and $u_t = \Gamma^{-1} C v_t$. The problem is that at most $k(k+1)/2$ unique, non-zero elements can be obtained from $\hat{\Sigma}_u$. On the other hand, there are $k(k+1)$ parameters in Γ and Σ_v and k^2 parameters to be identified in C. Even if it is assumed $C = I$ and the diagonal elements of Γ are normalized to 1, as it is typically done in the literature, at least $k(k-1)/2$ restrictions are required to satisfy the order condition for identification.

In order to address this problem, we associated a graph with the causal structure of the model and the properties of a given causal structure are obtained by analyzing the properties of the graph. Many ideas of this paper have been inspired by the method discussed in Swanson and Granger (1997). Related works are also Glymour and Spirtes (1988), Dahlhaus and Eichler (2000), Reale and Tunnicliffe Wilson (2001), Demiralp and Hoover (2003), and Haigh and Bessler (2004). This paper makes the following advances over the previous studies.

First, our work, like the mentioned papers, bases the search procedure upon tests of vanishing partial correlations among residuals. However, instead of treating the estimated residuals as if they were measured, our tests are based on the asymptotic distribution of the estimated residuals. This method raises the power of the tests, as some simulations have shown (see Moneta (2004)).

Second, we give a proof (Proposition 1) about the links between partial correlations among residuals and partial correlations among autoregressive variables in a VAR model. This proof, which we do not find in the previous studies, justifies rigorously the search procedure of deriving the contemporaneous causal structure from the vanishing partial correlation among residuals.

Third, we consider also the case of VAR models in which cycles and common shocks among contemporaneous variables are allowed.

The rest of the paper is organized as follows. The next section shows how the structural equation can be represented by a causal graph. The subsequent sections develop the search procedure. Two sections are preliminary: section 3 proofs a result about the link between partial correlations among residuals and partial correlations among contemporaneous variables, section 4 provides the testing procedure of vanishing partial correlation among residuals. Section 5 presents the search procedure for DAG models, which is extended to the feedback case in section 6. Section 7 extends the search to causal relations among lagged variables. Section 8 summarizes the search procedure.

2 CAUSAL GRAPH FOR THE STRUCTURAL MODEL

Let us suppose that a $(k \times 1)$ vector of time-series $Y_t = (y_{1t}, \ldots, y_{kt})'$ is governed by a structural model:

$$Y_t = \sum_{i=0}^{p} B_i Y_{t-i} + C v_t, \tag{3}$$

where the vector of the "structural disturbances" $v_t = (v_{1t}, \ldots, v_{kt})'$ is serially uncorrelated and $E(v_t v_t') = \Sigma_v$ is a diagonal matrix. The B_i's $(i = 0, \ldots, p)$ are $(k \times k)$ matrices. It is assumed that the equation (3) represents a causal structure which has generated the data. Such a causal structure can be represented by a directed graph, as can any linear structural equation model (see Spirtes *et al.* (2000) and Richardson and Spirtes (1999)). If we assume the presence only direct causal relationships (without feedbacks), the model can be represented by a directed acyclic graph (DAG). If we allow feedbacks (among contemporaneous variables: temporal backward causation is generally excluded) we deal with directed *cyclic* graphs. We illustrate our search procedure starting with the DAG model case. In section 6 we show how it can be extended allowing feedbacks and common shocks.

In Figure 1 an example of a DAG for the model of equation (3) is displayed. In this particular example $p = 1$, B_0 and B_1 are lower-triangular matrices and C is diagonal.

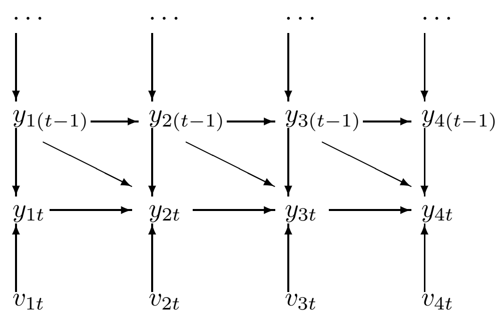


Figure 1: Example of causal DAG for equation (3).

There is a correspondence between each nonzero entry of the matrices $B_0, \ldots, B_p, C$ in equation (3) on the one hand, and each directed edge in the DAG representing it on the other hand. In this DAG there is a directed edge pointing from $y_{i,t-n}$ to $y_{j,t-m}$ (where $0 \le i, j, \le k$) if and only if either:

(i) $0 < n - m \le p$ and the entry corresponding to the j^{th} row and the i^{th} column of $B_{(n-m)}$ is different from zero; or

(ii) $n = m$, $i \neq j$ (we rule out self loops) and the entry corresponding to the j^{th} row and the i^{th} column of B_0 is different from zero.

There is a directed edge pointing from $v_{i,t}$ to $y_{j,t+h}$ if and only if $h = 0$ and the entry corresponding to the j^{th} row and the i^{th} column of C is different from zero. There is no directed edge pointing from $y_{i,t}$ to $v_{j,t+h}$ for any i, j and h. There is no directed edge pointing from $v_{i,t}$ to $v_{j,t+h}$ for any i, j and h.

Notice that since it is assumed that the causal structure is representable by means of a DAG, feedback loops are excluded. This is the same as assuming that if the (i, j) entry of B_0 is different from zero, then the (j, i) entry of B_0 must be equal to zero. The extension to graphs in which undirected edges between contemporaneous variables are allowed (while edges between lagged variables remain directed) is addressed in Section 6.

We assume that $C = I_k$, so that the relations among the contemporaneous components of Y_t are embedded only in the matrix B_0. In Section 6 we generalize by allowing $C \neq I_k$, and adapt the algorithm given in this section to a more complex pattern. However, assuming $C = I_k$ does not impede a structural shock $v_{i,t}$ to affect simultaneously components of Y_t besides $y_{i,t}$. This assumption means only that, for example, $v_{i,t}$ affects

$y_{j,t}$ through the effect of $y_{i,t}$ on $y_{j,t}$ and not directly. In many contexts the two situations are observationally equivalent.

The method proposed here is consistent with the Structural VAR approach (see e.g. Bernanke (1986)). First, we need to estimate the reduced form:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + u_t, \tag{4}$$

where $A_i = (I - B_0)^{-1} B_i$, for $i = 1, \ldots, p$. The vector $u_t = (I - B_0)^{-1} v_t$ is a serially uncorrelated vector of disturbances. It holds that:

$$u_t = B_0 u_t + v_t \tag{5}$$

Second, from the estimate of the covariance matrix of u_t ($\hat{\Sigma}_u$), we want to test all the possible vanishing partial correlations among the elements of u_t. Finally, such tests are used to constrain the possible causal relationships among the contemporaneous variables. The next sections illustrate this procedure. For convenience, we assume the vector of the error terms u_t to be normally distributed. However, the testing and search procedure can be extended to non-Gaussian processes (see footnote 2).

3 PARTIAL CORRELATIONS AMONG RESIDUALS

We want to show that partial correlations among the residuals u_t in (4) are tied to partial correlations among the contemporaneous components of Y_t.

Proposition 1. *Let $u_{1t}, \ldots, u_{kt}$ be the residuals of k OLS regressions of $y_{1t}, \ldots, y_{kt}$ on the same vector $J_{t-1} = (y_{1(t-1)}, \ldots, y_{k(t-1)}, \ldots, y_{1(t-p)}, \ldots, y_{k(t-p)})$. Let u_{it} and u_{jt} ($i \neq j$) be any two distinct elements of $\{u_{1t}, \ldots, u_{kt}\}$, $\mathcal{U}_t$ any subset of $\{u_{1t}, \ldots, u_{kt}\} \setminus \{u_{it}, u_{jt}\}$ and $\mathcal{Y}_t$ the corresponding subset of $\{y_{1t}, \ldots, y_{kt}\} \setminus \{y_{it}, y_{jt}\}$, so that u_{gt} is in $\mathcal{U}_t$ iff y_{gt} is in $\mathcal{Y}_t$, for $g = 1, \ldots, k$. Then it holds that:*

$$\rho(u_{it}, u_{jt} | \mathcal{U}_t) = \rho(y_{it}, y_{jt} | \mathcal{Y}_t, J_{t-1}).$$

Proof of Proposition 1. See Appendix.

To test vanishing partial correlations among residuals we apply a procedure illustrated in the next section.

If we consider only multivariate normal distributions, vanishing partial correlations and conditional independence relationships are equivalent. Therefore, if we consider a DAG with a set of vertices $\mathbf{X} = \{X_1, \ldots, X_n\}$ and a normal probability distribution $P(\mathbf{X})$, that jointly satisfy Markov and Faithfulness condition[1], it holds that: $\rho(X_i, X_j | X^{(h)}) = 0$ if and only if X_i *is independent from* X_j *given* $X^{(h)}$ if and only if X_i *and* X_j *are d-separated by* $X^{(h)}$, where $X^{(h)}$ is any subset of $\mathbf{X} \setminus \{X_i, X_j\}$ and $i \neq j$.[2]

4 TEST OF VANISHING PARTIAL CORRELATIONS AMONG RESIDUALS

In this section we provide a procedure to test the null hypotheses of vanishing correlations and vanishing partial correlations among the residuals. Tests are based on asymptotic results.

Let us write the VAR we are estimating in a more compact form, denoting $X'_t = [Y'_{t-1},\ Y'_{t-2}, \ldots,\ Y'_{t-p}$, which has dimension $(1 \times kp)$ and $\Pi' = [A_1,\ A_2, \ldots,\ A_p]$, which has dimension $(k \times kp)$. We can write: $Y_t = \Pi' X_t + u_t$. The maximum likelihood estimate of Π turns out to be given by

$$\hat{\Pi}' = \left[\sum_{t=1}^{T} Y_t X'_t\right] \left[\sum_{t=1}^{T} X_t X'_t\right]^{-1}.$$

Moreover, the ith row of $\hat{\Pi}'$ is

$$\hat{\pi}'_i = \left[\sum_{t=1}^{T} y_{it} X'_t\right] \left[\sum_{t=1}^{T} X_t X'_t\right]^{-1},$$

which coincides with the estimated coefficient vector from an OLS regression of y_{it} on X_t (Hamilton (1994), p. 293). The maximum likelihood estimate of the matrix of variance and covariance among the error terms Σ_u turns out to be $\hat{\Sigma}_u = (1/T) \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t$, where

[1]The Markov condition and the Faithfulness condition are conditions on the probability distribution P of a set of variables V and the DAG that associates a vertex to every element of V. The assumption that these two conditions are satisfied bases the causal search algorithms developed by Spirtes *et al.* (2000). The Markov condition says that any vertex in a DAG $\mathcal{G}$ is conditionally independent of its nondescendants (excluding its parents), given its parents, under P. The Faithfulness condition states the following. Let $\mathcal{G}$ be a causal graph with vertex set V and P be a probability distribution over the vertices in V such that $\mathcal{G}$ and P satisfy the Markov condition. $\mathcal{G}$ and P satisfy the Faithfulness condition if and only if every conditional independence relation true in P is entailed by the Markov condition applied to $\mathcal{G}$. (See Spirtes *et al.* (2000), pp. 29-31).

[2]However, some results of Spirtes *et al.* (2000, p. 47) show that assuming the Faithfulness condition for linear systems is equivalent to assume that in a graph G the vertices A and B are d-separated given a subset C of the vertices of G if and only if $corr(A, B|C) = 0$, *without* any normality assumption.

$\hat{u}_t = Y_t - \hat{\Pi}' X_t$. Therefore the maximum likelihood estimate of the covariance between u_{it} and u_{jt} is given by the (i,j) element of $\hat{\Sigma}_u$: $\hat{\sigma}_{ij} = (1/T)\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}$.

Proposition 2 (Hamilton 1994, p. 301).
Let $Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + u_t$ where $u_t \sim$ i.i.d. $N(\mathbf{0}, \Sigma_u)$ and where roots of $|I_k - A_1 z - A_2 z^2 - \ldots - A_p z^p| = 0$ lie outside the unit circle. Let $\hat{\Sigma}_u$ be the maximum likelihood estimate of Σ_u. Then

$$\sqrt{T}[vech(\hat{\Sigma}_u) - vech(\Sigma_u)] \xrightarrow{d} N(\mathbf{0},\, \Omega), \qquad (6)$$

where $\Omega = 2\boldsymbol{D}_k^{+}(\Sigma_u \otimes \Sigma_u)(\boldsymbol{D}_k^{+})'$, $\boldsymbol{D}_k^{+'} \equiv (\boldsymbol{D}_k'\boldsymbol{D}_k)^{-1}\boldsymbol{D}_k'$ and $\boldsymbol{D}_k$ is the duplication matrix.

Therefore, to test the null hypothesis that $\rho(u_{it}, u_{jt}) = 0$, we use the Wald statistic:

$$\frac{T\ (\hat{\sigma}_{ij})^2}{\hat{\sigma}_{ii}\hat{\sigma}_{jj}\ +\ \hat{\sigma}_{ij}^2}\ \approx\ \chi^2(1).$$

The Wald statistic for testing vanishing partial correlations is obtained by means of the delta method (see e.g. Lehmann and Casella (1998), p. 61).

For example, for $k = 4$, suppose we want to test $\rho(u_1, u_3|u_2) = 0$. First, notice that $\rho(u_1, u_3|u_2) = 0$ if and only if $\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23} = 0$, where σ_{ij} is the (i,j) element of Σ_u. Let us define a function $g : \mathbf{R}^{k(k+1)/2} \to \mathbf{R}$, such that $g(vech(\Sigma_u)) = \sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23}$. Thus,

$$\nabla g' = (0,\ -\sigma_{23},\ \sigma_{22},\ 0,\ \sigma_{13},\ -\sigma_{12},\ 0,\ 0,\ 0,\ 0).$$

The delta method implies that:

$$\sqrt{T}[(\hat{\sigma}_{22}\hat{\sigma}_{13} - \hat{\sigma}_{12}\hat{\sigma}_{23})\ -\ (\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23})] \xrightarrow{d} N(0, \nabla g'\Omega\nabla g).$$

The Wald test of the null hypothesis $\rho(u_1, u_3|u_2) = 0$ is given by:

$$\frac{T(\hat{\sigma}_{22}\hat{\sigma}_{13} - \hat{\sigma}_{12}\hat{\sigma}_{23})^2}{\nabla g'\Omega\nabla g} \approx \chi^2(1).$$

Tests for higher order correlations follow analogously.

5 SEARCHING FOR THE CAUSAL GRAPH AMONG CONTEMPORANEOUS VARIABLES

In this section an algorithm to identify the causal graph among the contemporaneous variables is presented. In real applications the output of the algorithm is an unique DAG only in very special cases (usually when some background knowledge is incorporated, besides the d-separation relations). In most of the cases, the algorithm just allows us to narrow significantly the set of possible DAGs and we obtain, as output, a pattern of DAGs. Therefore, further a priori knowledge is necessary to select the appropriate DAG from this pattern.

Proposition 1 implies that testing a vanishing partial correlation coefficient between u_{it} and u_{jt} given some other components $u_{qt}, \ldots, u_{pt}$ is equivalent to testing a vanishing partial correlation coefficient between y_{it} and y_{jt} given some other components $y_{qt}, \ldots, y_{pt}$ and J_{t-1}. Therefore, from tests on all the possible partial correlations among the components of u_t we can obtain d-separation relations for the graphical causal model representing the structural equation (3). The next proposition proves that the d-separation relations that we obtain correspond to all the possible d-separation relations among the contemporaneous variables for the graph induced on the contemporaneous variables $y_{1t}, \ldots, y_{kt}$ alone.

Proposition 3. *Let us call $\mathcal{G}$ the causal DAG representing equation (3) and $\mathcal{G}_{Y_t}$ the subgraph of $\mathcal{G}$ induced on $y_{1t}, \ldots, y_{kt}$. Let J_{t-1} and $\mathcal{Y}_t$ be the same as in Proposition 1. y_{it} and y_{jt} are d-separated by $\mathcal{Y}_t$ and J_{t-1} in $\mathcal{G}$, if and only if y_{it} and y_{jt} are d-separated by $\mathcal{Y}_t$ in $\mathcal{G}_{Y_t}$.*

Proof of Proposition 3. See Appendix.

The next proposition shows that d-connection (d-separation) relations entail some restrictions on the graph in terms of adjacencies among the vertices and directions of the edges. The goal is to justify the procedures given by the search algorithm below.

Proposition 4. *$\mathcal{G}_{Y_t}$ is defined as in Proposition 3. Let us assume $P(\boldsymbol{X})$ to be a probability distribution over the variables $\boldsymbol{X}$ that form $\mathcal{G}_{Y_t}$, such that $< \mathcal{G}_{Y_t}, P(\boldsymbol{X}) >$ satisfies the Markov and Faithfulness conditions. Then:*

- *(i) for all distinct vertices y_{it} and y_{jt} of $\mathcal{G}_{Y_t}$, y_{it} and y_{jt} are adjacent in $\mathcal{G}$ if and only if y_{it} and y_{jt} are d-connected in $\mathcal{G}_{Y_t}$ conditional on every set of vertices of $\mathcal{G}_{Y_t}$ that does not include y_{it} and y_{jt}; and*

- *(ii) for all vertices y_{ht}, y_{it} and y_{jt} such that y_{ht} is adjacent to y_{it} and y_{it} is adjacent to y_{jt}, but y_{ht} and y_{jt} are not adjacent, $y_{ht} \longrightarrow y_{it} \longleftarrow y_{jt}$ is a subgraph of $\mathcal{G}_{Y_t}$ if and only if y_{ht}, y_{jt} are d-connected in $\mathcal{G}_{Y_t}$ conditional on every set of vertices of $\mathcal{G}_{Y_t}$ containing y_{it} but not y_{ht} or y_{jt}.*

Proof of Proposition 4. This proposition is a particular case of a theorem proved in Spirtes *et al.* (2000, theorem 3.4, p. 47) and in Verma and Pearl (1990).

The goal of the algorithm described in Table 1 is to obtain a (possibly narrow) class of DAGs, which contains the causal structure among the contemporaneous variables $\mathcal{G}_{Y_t}$. The algorithm, which is an adaptation of the PC algorithm of Spirtes *et al.* (2000), starts from a complete undirected graph $\mathcal{C}$ among the k components of Y_t (in which each vertex is connected with every other vertex) and uses d-separation relations to eliminate and direct as many edges as it is possible. Notice that it would be equivalent to directly applying the algorithm to the components of the vector of residuals u_t and to consider vanishing partial correlations among them as d-separation relations. The graph induced on the components of u_t would be the same as the graph induced on the components of Y_t.

6 FEEDBACKS AND COMMON SHOCKS

The preceding sections complied with a sometimes useful simplification, namely that the statistical dependencies among the measured variables (which constitute the multivariate time series $\{Y_t\}$) are due only to directed causes, ruling out the possibility of feedbacks or of unmeasured common causes. In this section the search procedure is extended to consider the possibility that the data generating process is representable through a structure in which feedbacks (namely bi-directed causes) and particular latent variables (common shocks) are allowed.

Spirtes *et al.* (2000) develop an algorithm (FCI algorithm), which infers features of the DAGs from a probability distribution when there may be latent common causes, while Richardson and Spirtes (1999) develop an algorithm (CCD algorithm), which infers features of directed *cyclic* graphs from a probability distribution when there are *no* latent common causes. An open question is whether there are comparable algorithms for inferring features of directed graphs (cyclic or acyclic) even when there may be latent common causes. In fact distinguishing between feedbacks and latent variables is a difficult task, which the analysis of vanishing partial correlation alone seems not to solve.

We propose an automatic search procedure that produces as output an undirected graph. The undirected edges that form the undirected graph reflect an epistemological (more than ontological) reason: we do not know if the presence of an undirected edge denotes a feedback, a latent variable or a directed cause in the data generating process. In particular cases, an undirected edge may correspond to no direct connection at all in the data generating process, as we illustrate

Table 1: Search algorithm 1 (adapted from the PC Algorithm of Spirtes *et. al.* 2000).

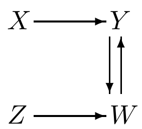
A.)
Form the complete undirected graph $\mathcal{C}$ on the vertex set $y_{1t}, \ldots, y_{kt}$. Let *Adjacencies*$(\mathcal{C}, y_{it})$ be the set of vertices adjacent to y_{it} in $\mathcal{C}$ and let *Sepset* (y_{ht}, y_{it}) be any set of vertices S so that y_{ht} and y_{it} are d-separated given S;
B.)
$n = 0$
repeat :
    repeat :
        select an ordered pairs of variables y_{ht} and y_{it} that are adjacent in $\mathcal{C}$ such that *Adjacencies*$(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ has cardinality greater than or equal to n, and a subset S of *Adjacencies*$(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ of cardinality n, and if y_{ht} and y_{it} are d-separated given S in $\mathcal{G}_{Y_t}$ delete edge y_{ht} — y_{it} from $\mathcal{C}$;
    until all ordered pairs of adjacent variables y_{ht} and y_{it} such that *Adjacencies*$(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ has cardinality greater than or equal to n and all subsets S of *Adjacencies*$(\mathcal{C}, y_{ht})$ $\backslash\{y_{it}\}$ of cardinality n have been tested for d-separation;
    $n = n + 1$;
until for each ordered pair of adjacent variables y_{ht}, y_{it}, *Adjacencies*$(\mathcal{C}, y_{ht})$ $\backslash\{y_{it}\}$ is of cardinality less than n;
C.)
for each triple of vertices y_{ht}, y_{it}, y_{jt} such that the pair y_{ht}, y_{it} and the pair y_{it}, y_{jt} are each adjacent in $\mathcal{C}$ but the pair y_{ht}, y_{jt} is not adjacent in $\mathcal{C}$, orient y_{ht} — y_{it} — y_{jt} as $y_{ht} \longrightarrow y_{it} \longleftarrow y_{jt}$ if and only if y_{it} never belongs to a *Sepset*(y_{ht}, y_{jt});
D.)
repeat :
    if $y_{at} \longrightarrow y_{bt}$, y_{bt} and y_{ct} are adjacent, y_{at} and y_{ct} are not adjacent and y_{bt} belongs to every *Sepset*(y_{at}, y_{ct}), then orient y_{bt} — y_{ct} as $y_{bt} \longrightarrow y_{ct}$;
    if there is a directed path from y_{at} to y_{bt}, and an edge between y_{at} and y_{bt}, then orient y_{at} — y_{bt} as $y_{at} \longrightarrow y_{bt}$;
until no more edges can be oriented.

below.

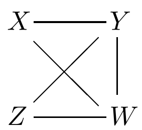
The search algorithm displayed in Table 2 is an adaptation of the common first and second part of the PC, FCI, and CCD algorithm and the PC algorithm. The algorithm starts from a complete undirected graph among the contemporaneous variables and just eliminates all the edges between two variables which are d-separated by any other variable.

We leave to background knowledge the criterion to decide whether the undirected edge represents a feedback, a latent variable, a direct cause, or actually no direct connection. However, there is another statistical check: if the restrictions on the contemporaneous variables of a VAR model are over-identifying, they can be tested according to a χ^2 statistics.

As we mentioned, the presence of an undirected edge in the output of the search algorithm proposed here does not correspond necessarily to the presence of an edge in the causal graph of the data generating process. For example, suppose we have four variables X, Y, Z, W, and the graph related to the data generating process is:



Then the output of the search algorithm of Table 2.2. will be:



Thus, for instance, the undirected edge between X and W in the output of the algorithm does not correspond to an edge in the data generating graph, although X is d-connected to W by any possible set of vertices. Analogous results are obtained in the presence of a latent variable, instead of a feedback, between Y and W in the data generating process. This means that when we obtain an output of this type, we have to be careful to consider the possibility of edges corresponding to no edges in the data generating graph.

Table 2: Search algorithm 2 (adapted from common steps of PC-FCI-CCD algorithms of Spirtes *et al.* (2000) and Richardson and Spirtes (1999)).

A.)
From the estimated covariance matrix of the VAR residuals test all the possible partial correlations among the residuals (using the Wald test procedure described in section 4).
B.)
Form the complete undirected graph $\mathcal{C}$ on the vertex set $y_{1t}, \ldots, y_{kt}$. Let *Adjacencies*$(\mathcal{C}, y_{it})$ be the set of vertices adjacent to y_{it} in $\mathcal{C}$ and let *Sepset* (y_{ht}, y_{it}) be any set of vertices S so that y_{ht} and y_{it} are d-separated given S;
C.)
$n = 0$
repeat :
    repeat :
        select an ordered pairs of variables y_{ht} and y_{it} that are adjacent in $\mathcal{C}$ such that *Adjacencies*$(\mathcal{C}, y_{ht}) \backslash \{y_{it}\}$ has cardinality greater than or equal to n, and a subset S of *Adjacencies*$(\mathcal{C}, y_{ht}) \backslash \{y_{it}\}$ of cardinality n, and if y_{ht} and y_{it} are d-separated given S in $\mathcal{G}_{Y_t}$ delete edge y_{ht} — y_{it} from $\mathcal{C}$;
    until all ordered pairs of adjacent variables y_{ht} and y_{it} such that *Adjacencies*$(\mathcal{C}, y_{ht}) \backslash \{y_{it}\}$ has cardinality greater than or equal to n and all subsets S of *Adjacencies*$(\mathcal{C}, y_{ht})$ $\backslash \{y_{it}\}$ of cardinality n have been tested for d-separation;
    $n = n + 1$;
until for each ordered pair of adjacent variables y_{ht}, y_{it}, *Adjacencies*$(\mathcal{C}, y_{ht})$ $\backslash \{y_{it}\}$ is of cardinality less than n;

7 STRUCTURAL RELATIONS AMONG LAGGED VARIABLES

Once structural relations among contemporaneous variables are obtained, we can use this information to impose some constraints on structural relations among different time point realizations of the components of Y_t. Indeed, the graphical model among the contemporaneous variables, which we obtain from the algorithms described above, implies zeros in the matrix $(I - B_0)$, derived from equations:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + u_t, \tag{7}$$

$$u_t = B_0 u_t + v_t. \quad (8)$$

It is also possible to test the zeros in the matrices $A_1, \ldots, A_p$ using asymptotic test procedures. Let us consider the matrices $B_0, B_1, \ldots, B_p$ of the structural form of the model:

$$Y_t = \sum_{i=0}^{p} B_i Y_{t-i} + v_t. \quad (9)$$

It turns out that $B_i = (I - B_0)\ A_i$, for $i = 1, \ldots, p$. From tests on the zeros of the matrices $A_1, \ldots, A_p$ and on the zeros of the matrix $(I - B_0)$ (entailed by the causal graph obtained from the search algorithms) one can derive the zeros of the matrices $B_1, \ldots, B_p$. Each zero in a matrix B_i implies a lack of edge in the causal graph of the model, according to the rules we set in the section 2.

8 SUMMARY OF THE PROCEDURE AND APPLICATIONS

The search procedure for identifying the graph of the structural model can be summarized as follows:
Step 1: Estimate a VAR via OLS.
Step 2: Estimate the covariance matrix of the residuals from the reduced form.
Step 3: Test all the possible vanishing partial correlations among the residuals (according to the procedure described in section 4) and list the consequent d-separation sets among the contemporaneous variables.[3]
Step 4: Apply the search algorithms (plus background knowledge) described in Tables 1 and 2 to such d-separation sets in order to determine the causal structure among the contemporaneous variables.
Step 5: Apply the procedure described in section 7 to determine the structure among the lagged variables.

The natural application of this method is to macroeconomic data, in which VAR models are mainly used to study the dynamic effect of structural shocks (impulse response functions analysis), which are given precise economic interpretation (shock to productivity, monetary policy shock, etc.). In order to get a reliable interpretation of the shocks, one has to find that transformation of the reduced form system of equations, which is consistent with the causal structure among the data. Since such transformation consistent in imposing a contemporaneous causal structure on the data (via matrix B_0 in equation (3)), this method permits to choose the most reliable transformations.

[3]Actually, to apply the algorithm one does not need to test all the vanishing partial correlations, but just those requested by the algorithm. However, when we deal with a VAR model with few variables (as it has to be the case), it may be useful to have a complete list of the vanishing partial correlations and the corresponding d-separation sets.

Moneta (2003) is an attempt to apply a simple version of this method, which only uses search algorithm 1 (ruling out cycles and latent variables), in order to identify the structural shocks associated with the following US macroeconomic variables: output, consumption, investment, money, interest rate, and inflation. The results point out that not only shocks associated to real macroeconomic variables (output, consumption and investment) but also shocks associated to nominal variables (money, inflation and interest rates) have a considerable effect on macroeconomic fluctuations (at all frequencies). This result shows how US data are not consistent with the Real Business Cycle hypothesis, which claims that a single productivity shock is driving output fluctuations.

The application considered by Moneta (2004) deals with the problem of finding the most appropriate measure of the exogenous monetary policy shock in US economy. The method allows cycles and common shocks among contemporaneous variables (using search algorithm 2). Background knowledge about the central bank operating procedures is used to further discriminate among the causal structures output of the algorithm. The results suggest that a good measure of monetary policy shock is that portion of shock to nonborrowed reserves orthogonal to shock to total reserve.

APPENDIX
Proofs

Proof of Proposition 1

Let $<\mathcal{U}_t>$ be the n-uple of the ordered elements of $\mathcal{U}_t$ and $<\mathcal{Y}_t>$ the corresponding n-uple of the ordered elements of $\mathcal{Y}_t$. We prove the proposition by induction on the length of $<\mathcal{U}_t>$.
(i) Base Case: $<\mathcal{U}_t> = \emptyset$. We want to prove that: $\rho(u_{it}, u_{jt}) = \rho(y_{it}, y_{jt}|J_{t-1})$. We know that:

$$\rho(y_{it}, y_{jt}|J_{t-1}) = \frac{cov(y_{it}, y_{jt}|J_{t-1})}{\sqrt{var(y_{it}|J_{t-1})var(y_{jt}|J_{t-1})}}$$

Since[4] $u_{it} = y_{it} - cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}$ and $u_{jt} = y_{jt} - cov(y_{jt}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}$, then
$cov(u_{it}, u_{jt}) =$
$cov(y_{it} - cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}, y_{jt} - cov(y_{jt}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}) =$
$= cov(y_{it}, y_{jt}) + cov(cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}, cov(y_{jt}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}) -$

[4]If an intercept is added the substance of the proof does not change.

$cov(cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}, y_{jt})-$
$cov(y_{it}, cov(y_{jt}, J_{t-1})var(J_{t-1})^{-1}J_{t-1}) =$
$= cov(y_{it}, y_{jt}) + cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}var(J_{t-1})$
$cov(J_{t-1}, y_{jt})var(J_{t-1})^{-1}-$
$cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}cov(J_{t-1}, y_{jt}) -$
$cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}cov(J_{t-1}, y_{jt})=$
$= cov(y_{it}, y_{jt}) - cov(y_{it}, J_{t-1})var(J_{t-1})^{-1}$
$cov(J_{t-1}, y_{jt}) = = cov(y_{it}, y_{jt}|J_{t-1})$.
In similar way, $var(u_{it}) = var(y_{it}|J_{t-1})$ and $var(u_{jt}) = var(y_{jt}|J_{t-1})$. Therefore, $\frac{cov(y_{it}, y_{jt}|J_{t-1})}{\sqrt{var(y_{it}|J_{t-1})var(y_{jt}|J_{t-1})}} = \frac{cov(u_{it}, u_{jt})}{\sqrt{var(u_{it})var(u_{jt})}} = \rho(u_{it}, u_{jt})$.
(ii) Induction Case: suppose the proposition holds for $< \mathcal{U}_t >= u_{1t}, \ldots, u_{nt}$. Let us prove it holds for $< \mathcal{U}_t >= u_{1t}, \ldots, u_{(n+1)t}$. $\rho(y_{it}, y_{jt}|y_{1t}, \ldots, y_{(n+1)t}, J_{t-1}) =$
$= [\rho(y_{it}, y_{jt}|y_{1t}, \ldots, y_{nt}, J_{t-1})-$
$\rho(y_{it}, y_{(n+1)t}|y_{1t}, \ldots, y_{nt}, J_{t-1})$
$\rho(y_{jt}, y_{(n+1)t}|y_{1t}, \ldots, y_{nt}, J_{t-1})]/$
$[\sqrt{1 - \rho^2(y_{it}, y_{(n+1)t}|y_{1t}, \ldots, y_{nt}, J_{t-1})}$
$\sqrt{1 - \rho^2(y_{jt}, y_{(n+1)t}|y_{1t}, \ldots, y_{nt}, J_{t-1})}] =$
$= [\rho(u_{it}, u_{jt}|u_{1t}, \ldots, u_{nt}) -$
$\rho(u_{it}, u_{(n+1)t}|u_{1t}, \ldots, u_{nt})$
$\rho(u_{jt}, u_{(n+1)t}|u_{1t}, \ldots, u_{nt})]/$
$[\sqrt{1 - \rho^2(u_{it}, u_{(n+1)t}|u_{1t}, \ldots, u_{nt})}$
$\sqrt{1 - \rho^2(u_{jt}, u_{(n+1)t}|u_{1t}, \ldots, u_{nt})}] =$
$= \rho(u_{it}, u_{jt}|u_{1t}, \ldots, u_{(n+1)t})$.

Proof of Proposition 3
(i) Suppose y_{it} and y_{jt} are d-separated by J_{t-1} and $\mathcal{Y}_t$ in $\mathcal{G}$. If there is a path in $\mathcal{G}$ between y_{it} and y_{jt} that contains only components of $\mathcal{Y}_t$ (and possibly of u_t), such path is not active. Then any path in $\mathcal{G}_{Y_t}$ between y_{it} and y_{jt} is not active. Then y_{it} and y_{jt} are d-separated by $\mathcal{Y}_t$ in $\mathcal{G}_{Y_t}$.
(ii) Suppose y_{it} and y_{jt} are d-separated by $\mathcal{Y}_t$ in $\mathcal{G}_{Y_t}$. Then, if there is an active path between y_{it} and y_{jt} in $\mathcal{G}$, such path must contain a component of J_{t-1} which is not a collider, since there are no directed edge from any component of Y_t pointing to any component of J_{t-1}. Therefore such path is not active relative to J_{t-1} in $\mathcal{G}$ and y_{it} and y_{jt} are d-separated by J_{t-1} and $\mathcal{Y}_t$ in $\mathcal{G}$.

References

Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Bernanke, B.S. (1986) "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-100.

Dahlhaus R., and M. Eichler (2000) "Causality and Graphical Models for Time Series," in P. Green, N. Hjort, and S. Richardson (eds.), *Highly Structured Stochastic Systems*, Oxford: Oxford University Press.

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

Glymour, C., and P. Spirtes (1988) "Latent Variables, Causal Models and Overidentifying Constraints," *Journal of Econometrics* 39(1), 175-198.

Haigh, M., and D.A. Bessler (2004) "Causality and Price Discovery: An Application of Directed Acyclic Graphs", *Journal of Business* (forthcoming).

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

Lehmann, E. L., and G. Casella (1998) *Theory of Point Estimation*, New York: Springer Verlag.

Moneta, A. (2003) "Graphical Models and Structural Vector Autoregressions" LEM Working Paper, Sant'Anna School of Advanced Studies, Pisa.

Moneta, A. (2004) "Graphical Causal Models and VAR-based Macroeconometrics," unpublished PhD thesis, Sant'Anna School of Advanced Studies, Pisa.

Reale, M., and G. Tunnicliffe Wilson (2001) "Identification of Vector AR Models with Recursive Structural Errors Using Conditional Independence Graphs," *Statistical Methods and Applications* 10(1), 49-65.

Richardson, T., and P. Spirtes (1999) "Automated Discovery of Linear Feedback Models," in C. Glymour, and G.F. Cooper (eds.), *Computation, Causation and Discovery.* Menlo Park, CA: AAAI Press and The MIT Press.

Sims, C.A. (1980) "Macroeconomics and Reality," *Econometrica*, 48(1), 1-47.

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

Stock, J.H., and M.W. Watson (2001) "Vector Autoregressions," *Journal of Economic Perspectives*, 15(4), 101-115.

Swanson, N.R., and C.W.J. Granger (1997) "Impulse Response Function Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association*, 92(437), 357-367.

Verma, T., and J. Pearl (1990) "Equivalence and Synthesis of Causal Models," *Proceedings of the Sixth Conference on Uncertainty in AI*, Mountain View CA: Association for Uncertainty in AI.