Graphical Models for the Identification of Causal Structures in Multivariate Time Series 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, USA
ps7z@andrew.cmu.edu

1 INTRODUCTION

In this paper we present a semi-automated search procedure to deal with the problem of the identification of the contemporaneous causal structure connected to a large class of multivariate time series models. We refer in particular to multivariate models, such as vector autoregressive (VAR) and dynamic factor (DF) model, in which the background or theoretical knowledge is not sufficient or enough reliable to build a structural equations model. VAR models deal with a small number of time series models (the maximum number is typically between 6 and 8), while DF models deal with a large number of time series, possibly larger than the number of observation (T) over time. Both VAR and DF models have proven to be very efficient in the macroeconomic and financial literature to address different empirical issues, such as forecasting, summarizing the statistical properties of the data, and building economics indicators (of business cycles, for instance). Moreover, DF models can be used in the financial literature to estimate insurable risk and in the macroeconomic literature to learn about aggregate behavior on the basis of microeconomic data (sectors, regions). (Forni and Lippi (1997), Forni et al. (2000), and Stock and Watson (2001) are useful references).

However, there is another important task that macroeconometricians usually try to address. This is the task of identifying and measuring the structure of the economy, namely the complex set of causal processes that have generated the data. Such task is called *structural analysis* and is intimately connected with the possibility of giving policy advice and answering counterfactual questions (asking what would happen to the economy if the policymaker intervened on a particular variable). However, it is fraught with the problem that many different structural relations are consistent with the same data, usually called the *problem of identification.*

In this paper we propose to use graphical causal models for constructing partial information about the contemporaneous causal structure of the data generating process starting from statistical properties (partial correlations) of the data. In other words, our method will permit the exclusion of a large set of causal structures which are not consistent with some statistical properties, under the assumption that any causal structure among random variables is tied to a particular configuration of partial correlations over the same random variables. We will recover only causal structures concerning contemporaneous variables, but, as shown below, this is sufficient to identify the model.

Our work makes a step forward in the literature concerned with the applications of techniques based on graph-theory to the problem of residuals orthogonalization in VAR models, developed by Bessler and Lee (2002), Demiralp and Hoover (2003), Moneta (2004), Reale and Tunnicliffe Wilson (2001), and Swanson and Granger (1997). The main innovations are the extension to a more general framework which includes DF models, the possibility of dealing with feedbacks or common shocks and the adaptation of the algorithms developed by Spirtes *et al.* (2000) to the multivariate time series framework.

2 THE PROBLEM OF IDENTIFICATION

Let us briefly illustrate the problem of identification in VAR and DF models.

2.1 VAR Model

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 $Y_t = (y_{1t}, \ldots, y_{kt})'$, $u_t = (u_{1t}, \ldots, u_{kt})'$, 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:

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

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 $u_t = \Gamma^{-1} v_t$. 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. Thus, at least $k(k-1)/2$ restrictions are required to satisfy the order condition for identification.

2.2 Dynamic Factor Model

A DF model is a vector of n time series represented as the sum of two unobservable orthogonal components, a *common component* driven by few (fewer than n) *common factors*, and an *idiosyncratic* component driven by n idiosyncratic factors. The formal representation of a DFM is the following:

$$x_{it} = A_{i1}(L) u_{1t} + \ldots + A_{iq}(L) u_{qt} + \xi_{it}, \qquad (3)$$

for $i = 1, \ldots, n$, where $U_t = (u_{1t}, \ldots, u_{qt})'$ is an orthonormal white-noise vector (i.e. u_{jt} has unit variance and is orthogonal to u_{st} for any $j \neq s$), and $A_{ij}(L) = \sum_{k=0}^{\infty} a_{ij,k} L^k$ is a polynomial (finite or infinite) in the lag operator L. The polynomial $\chi_{it} = A_{i1}(L) u_{1t} + \ldots + A_{iq}(L) u_{qt}$ is the common component, while the elements $u_{1t}, \ldots, u_{qt}$ are the common factors (or *common shocks*). The idiosyncratic component is represented by ξ_{it}.

Let us suppose we have estimated (with the method proposed by Forni et al. (2000) the DFM represented in equation (3). The estimation of the common components χ_{it} does not imply identification of the common shocks. Indeed, there exists an infinite number of representations which are observationally equivalent to (3), like

$$\chi_{it} = B_{i1}(L) v_{1t} + \ldots + B_{iq}(L) v_{qt}, \qquad (4)$$

where $V_t = (v_{1t}, \ldots, v_{qt})'$ is an orthonormal white-noise vector such that $V_t = S' U_t$, S being a unitary matrix (i.e. $SS' = I$, see Forni et al. (2004)). Suppose a specific equation (4) is the structural model and we have estimated (3). In the typical economic and financial applications, it is crucial to give a precise economic interpretation to the common shocks, while the idiosyncratic shocks can be left in the reduced form. This is sufficient, for instance, to study the dynamic effect of common macroeconomic shocks to the economic system.

The identification problem of the common shocks consists in justifying the imposition of enough (at least $q(q-1)/2$) restrictions on the matrix S', in order to recover V_t. Notice the analogy with the VAR model case, in which the restrictions are to be imposed on Γ. The main difference between VAR and DF model case is that in the latter the number of shocks does not increase with the number of variables, while in the former the restrictions required to achieve identification increases as the square of the number of variables included in the model.

3 RECOVERING THE STRUCTURAL FORM WITH GRAPHICAL MODELS

The last section has shown that to solve the problem of identification, once we have estimated the reduced form shocks $U_t = (u_{1t}, \ldots, u_{rt})'$ (where $r = k$ in the VAR case and $r = q$ in the DF case), we have to recover the structural shocks $V_t = (v_{1t}, \ldots, v_{rt})'$, imposing at least $r(r+1)/2$ zero-restrictions on the matrix H entering the following equation:

$$V_t = H U_t, \qquad (5)$$

where $H = \Gamma$ in the VAR case, and $H = S'$ in the DF case. The application of graphical models aims to infer the causal structures among the elements of U_t starting from the partial correlations among the same elements. What is is usually recovered is not a unique causal structure, but a set of possible causal structures, which in almost all of the cases is sufficient to identify the model. In the VAR case, as shown in Swanson and Granger (1997) and Moneta (2003), the causal structure among the elements of U_t corresponds to the causal structure among the contemporaneous elements of Y_t. We consider two different algorithms for causal inference.

3.1 Directed Acyclic Graphs

The first algorithm is applied in the case we assume that the causal relations among the elements of U_t are direct and they do not form ant cycle (feedback loops are not allowed). In this case the causal structures can be described by Directed Acyclic Graphs (DAG), as the one shown in Figure 1.

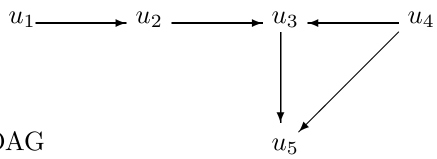

Figure 1: DAG

The algorithm (see extended version of the paper) is based on two general conditions about the connections between causation and partial correlations (Spirtes *et al.* (2000)):

1. *Causal Markov Condition*: Any random variable (with probability distribution P) corresponding to the vertices of the DAG $\mathcal{G}$ is conditionally independent of its graphical nondescendants (excluded its graphical parents), given its graphical parents, under P.

2. *Faithfulness Condition*: $\mathcal{G}$ and P satisfy the Faithfulness Condition if and only if every conditional independence relation true in P is entailed by the Causal Markov Condition applied to $\mathcal{G}$.

The algorithm, which is an adaptation of the PC algorithm of Spirtes *et al.* (2000) to the time-series framework, 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 vanishing partial correlations to eliminate and direct as many edges as it is possible.

3.2 Feedbacks and Latent Variables

The preceding sections complied with a sometimes useful simplification, namely that the statistical dependencies among the components of U_t are due only to direct 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 bidirected causes) and particular latent variables 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: in many cases, we cannot tell whether 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 (see extended version of the paper).

The search algorithm, that is presented in the extended version of the paper, 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 eliminates some edges using information on vanishing partial correlations. 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 are over-identifying, they can be tested according to a χ^2 statistics.

4 EMPIRICAL APPLICATIONS

In this section we discuss three empirical examples that utilize our method. The first and second example use macroeconomic US data 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 transformations consist in imposing a contemporaneous causal structure on the data, this method permits the choice of the most reliable transformations. In the first example we build a VAR model using the data of Moneta (2003) (which are an extension of the King et al. (1991) data). In the second example we build a DF model using the data of Forni et al. (2004) (which are also an extension of the King et al. (1991) data). Both examples use the search algorithm described in section 3.1 to identify the contemporaneous causal structure and 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.

A third example deals with the problem of finding the most appropriate measure of the exogenous monetary policy shock in US economy. We build a VAR model using the Bernanke and Mihov (1998) and Moneta (2004) data. The method allows cycles and common shocks among contemporaneous variables (using the search algorithm described in section 3.2). Figure 2 displays the output of the search algorithm, that is the causal structure among the contemporaneous variables: GDP, PGDP (GDP deflator), PSCCOM (Dow-Jones index of spot commodity prices), FFR (federal funds rate), TR (total bank reserves), NBR (nonborrowed reserves). 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. Figure 3 shows the effects of an exogenous monetary policy shocks on GDP for the different specifications of the causal structure.

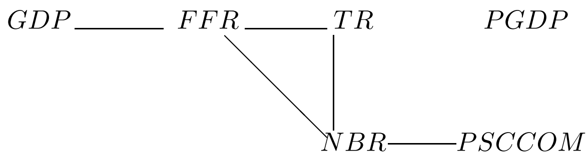

Figure 2: contemporaneous causal structure.

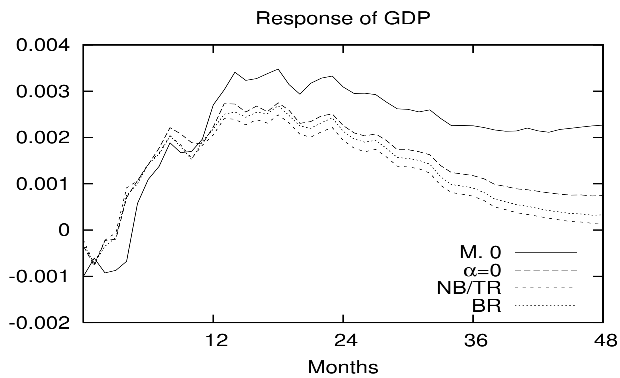

Figure 3: responses of GDP to one-standard-deviation monetary shock for the sample 1965:1-1996:12

References

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

Bessler, D.A., and S. Lee, 2002, Money and prices: US data 1869-1914 (a study with directed graphs), Empirical Economics, 27 (3), 427-446.

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.

Forni, M., and M. Lippi, 1997, *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Oxford: Oxford University Press.

Forni, M., M. Hallin, M. Lippi and L. Reichlin, 2000, The generalized factor model: identification and estimation. The Review of Economics and Statistics 80, 540-554.

Forni, M., D. Giannone, M. Lippi and L. Reichlin, 2004, Opening the Black Box: Structural Factor Models versus Structural VARs, CEPR Discussion Paper 4133.

King, R.G., C.I. Plosser, J.H. Stock, and M.W. Watson, 1991, Stochastic trends and economic fluctuations, *American Economic Review*, 81(4), 819-840.

Moneta, A., 2003, Graphical Models and Structural Vector Autoregressions, LEM Working Paper 07, Sant'Anna School of Advanced Studies, Pisa.

Moneta, A., 2004, Identification of Monetary Policy Shocks: A Graphical Causal Approach, *Notas Económicas*, 20.

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.

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.